

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN  
KHOA HỆ THÔNG THÔNG TIN



BÁO CÁO ĐỒ ÁN MÔN HỌC  
KHO DỮ LIỆU VÀ OLAP

Đề tài:

PHÂN TÍCH HOẠT ĐỘNG ĐẶT PHÒNG KHÁCH SẠN TỪ  
NĂM 2015 ĐẾN 2017

Giảng viên hướng dẫn: ThS. Đỗ Thị Minh Phụng

Lớp: IS217.M22

Sinh viên thực hiện - Nhóm 3:

Đặng Vũ Phương Uyên - 19520345

Nguyễn Thị Thu Phương - 19522066

TP.HCM, ngày 20 tháng 05 năm 2022

## MỤC LỤC

<b>CHƯƠNG I. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU .....</b>	<b>1</b>
<b>1. Phát biểu về dữ liệu .....</b>	<b>1</b>
1.1. Lý do chọn đề tài .....	1
1.2. Giới thiệu nguồn dữ liệu .....	1
1.3. Thuộc tính của tập dữ liệu .....	2
1.4. Tiền xử lý dữ liệu bằng Python .....	7
<b>2. Xây dựng kho dữ liệu .....</b>	<b>14</b>
2.1. Khái niệm Dimensional Modeling .....	14
2.2. Xây dựng sơ đồ bông tuyết .....	14
2.3. Dim_Reservation_Status .....	15
2.4. Dim_Hotel_Type .....	15
2.5. Dim_ReservationDate .....	16
2.6. Dim_Customer .....	16
2.7. Dim_Customer_Type .....	16
2.8. Dim_Deposit_Type .....	16
2.9. Dim_Distribution_Channel .....	17
2.10. Dim_Market_Segment .....	17
2.11. Dim_Arrival_Time .....	17
2.12. Fact_Hotel_Booking .....	18
<b>3. Nội dung 15 câu truy vấn.....</b>	<b>20</b>
3.1. Thống kê tổng số lượng khách hàng đặt phòng, tổng doanh thu của từng loại khách sạn theo từng năm. ....	20
3.2. Thống kê có bao nhiêu khách hàng là người lớn, bao nhiêu khách hàng là trẻ em và bao nhiêu khách hàng là em bé trong từng tháng của từng năm.....	20
3.3. Liệt kê khách hàng của từng loại khách hàng có doanh số cao nhất trong năm 2017. ....	20
3.4. Cho biết top 10 quốc gia có lượng khách du lịch cao nhất trong năm 2016, sắp xếp các giá trị theo thứ tự tăng dần.....	20
3.5. Liệt kê số lượng đơn đặt phòng ở ‘China’ và ‘Việt Nam’ theo từng năm và loại khách hàng là ‘Transient’. ....	20

3.6. Tìm các quốc gia có số lượng đơn đặt phòng lớn hơn lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn. ....	20
3.7. Liệt kê 5 khách hàng có tên bắt đầu bằng chữ D ở Bồ Đào Nha (PRT), nếu khách hàng có tổng chi tiêu > 500 và là khách hàng quay lại thì xuất "Khách hàng thân thiết". Nếu tổng chi tiêu > 500 và là khách hàng lần đầu tiên tới thì xuất "Khách hàng tiềm năng". Ngược lại xuất "Khách hàng vãng lai.".....	20
3.8. Liệt kê các quốc gia thu được lợi nhuận năm 2017 nhiều hơn năm 2016 và sắp xếp theo lợi nhuận giảm dần.....	20
3.9. Liệt kê các phân khúc thị trường mà tổng doanh thu của của kênh phân phối "Corporate" đó cộng lại tối thiểu 500 (lấy theo thứ tự doanh thu tăng dần) hoặc những phân khúc thị trường có số lượng lượt đặt phòng của kênh phân phối "Direct" tối thiểu 500... ..	20
3.10. Liệt kê tổng số đêm thuê của các quốc gia trong năm 2015 trừ (ngày nhận phòng vào thứ 7 và không đặt tiền cọc trước của trong năm 2015). .....	20
3.11. Với mỗi tháng đưa ra 3 email của khách hàng có chi tiêu cao nhất.....	20
3.12. Liệt kê các quý theo từng năm của quốc gia Bồ Đào Nha, thuộc top 50% quý có tổng yêu cầu đặc biệt trong tổng số yêu cầu đặc biệt thu được. ....	20
3.13. Liệt kê những phân khúc thị trường và kênh phân phối có yêu cầu bãi đỗ cho từ 5 đến 10 xe.....	21
3.14. Thông kê tổng thời gian chờ của từng hình thức thanh toán đặt cọc mà khách hàng đặt phòng theo từng tháng, quý, năm.....	21
3.15. Tính trung bình ngày chờ trên mỗi loại khách hàng của loại hình phân khúc thị trường của mỗi quốc gia và được thực hiện với năm thay đổi trạng thái bằng năm nhận phòng trừ loại hình phân khúc là Complementary. ....	21
<b>CHƯƠNG II : TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)</b> .....	<b>22</b>
1. Khái niệm .....	22
2. Chuẩn bị công cụ và Data Warehouse .....	22
3. Tạo project SSIS trong Visual Studio 2017. ....	30
4. Lọc và làm sạch dữ liệu.....	35
5. Tạo bảng Dim và bảng Fact.....	43
5.1. Bảng DimHotelType .....	44
5.2. Bảng DimDistributionChannel.....	54
5.3. Bảng DimMarketSegment.....	62
5.4. Bảng DimDepositType.....	71

5.5. Bảng DimReservationStatus .....	79
5.6. Bảng DimCustomerType.....	88
5.7. Bảng DimCustomer.....	96
5.8. Bảng DimArrivalTime .....	107
5.9. Bảng DimReservationDate.....	114
5.10. Bảng FactHotelBooking.....	123
<b>6. Tạo Execute SQL Task .....</b>	<b>146</b>
<b>CHƯƠNG III : PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS) .....</b>	<b>154</b>
1. Tạo project SSAS trong Visual Studio 2019 .....	154
2. Xác định dữ liệu nguồn (DEFINE DATA SOURCE) .....	155
3. Xác định khung nhìn dữ liệu nguồn ( DEFINE DATA SOURCE VIEW) ....	159
4. Xác định Cube và tạo mới các Measures .....	163
4.1. Xác định Cube và các bảng chiều.....	163
4.2. Xác định các Measures .....	167
5. Tính chất của các bảng Dimensions và Attribution nâng cao.....	169
5.1. Định nghĩa các tính chất của bảng Dimension:.....	169
5.2. Gom nhóm các thuộc tính .....	177
5.3. Xóa dòng dữ liệu Unknown .....	182
6. Phân cấp trong các bảng chiều.....	186
6.1. Arrival Time .....	186
6.2. Reservation Date.....	193
7. Định nghĩa Calculate, Named Set và Key Performance Indicators (KPIs)...	198
7.1. Calculate .....	198
7.2. Named Set.....	200
7.3. Key Performance Indicators (KPIs).....	202
8. Chạy Project SSAS .....	205
9. Quá trình phân tích dữ liệu bảng công cụ SSAS trên các khối Cube, công cụ Pivot Excel và ngôn ngữ MDX .....	206
9.1. Thống kê tổng số lượng khách hàng đặt phòng, tổng doanh thu của từng loại khách sạn theo từng năm . .....	206

9.2. Thống kê có bao nhiêu khách hàng là người lớn, bao nhiêu khách hàng là trẻ em và bao nhiêu khách hàng là em bé trong từng tháng của từng năm.....	208
9.3. Liệt kê khách hàng của từng loại khách hàng có doanh số cao nhất trong năm 2017.. ..	210
9.4. Cho biết top 10 quốc gia có lượng khách du lịch cao nhất trong năm 2016, sắp xếp các giá trị theo thứ tự tăng dần. ....	215
9.5. Liệt kê số lượng đơn đặt phòng ở ‘China’ và ‘Việt Nam’ theo từng năm và loại khách hàng là ‘Transient’ .....	219
9.6. Tìm các quốc gia có số lượng đơn đặt phòng lớn hơn lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn.....	221
9.7. Liệt kê 5 khách hàng có tên bắt đầu bằng chữ D ở Bồ Đào Nha (PRT), nếu khách hàng có tổng chi tiêu > 500 và là khách hàng quay lại thì xuất "Khách hàng thân thiết". Nếu tổng chi tiêu > 500 và là khách hàng lần đầu tiên tới thì xuất "Khách hàng tiềm năng". Ngược lại xuất "Khách hàng vãng lai." .....	223
9.8. Liệt kê các quốc gia thu được lợi nhuận năm 2017 nhiều hơn năm 2016 và sắp xếp theo lợi nhuận giảm dần.....	229
9.9. Liệt kê các phân khúc thị trường mà tổng doanh thu của kênh phân phối "Corporate" đó cộng lại tối thiểu 500 (lấy theo thứ tự doanh thu tăng dần) hoặc những phân khúc thị trường có số lượng lượt đặt phòng của kênh phân phối "Direct" tối thiểu 500... ..	234
9.10. Liệt kê tổng số đêm thuê của các quốc gia trong năm 2015 trừ (ngày nhận phòng vào thứ 7 và không đặt tiền cọc trước của trong năm 2015) .....	239
9.11. Với mỗi tháng đưa ra 3 email của khách hàng có chi tiêu cao nhất. ....	242
9.12. Liệt kê các quý theo từng năm của quốc gia Bồ Đào Nha, thuộc top 50% quý có tổng yêu cầu đặc biệt trong tổng số nhu cầu đặc biệt thu được.....	246
9.13. Liệt kê những phân khúc thị trường và kênh phân phối có yêu cầu bãi đỗ cho từ 5 đến 10 xe.....	251
9.14. Thống kê tổng thời gian chờ của từng hình thức thanh toán đặt cọc mà khách hàng đặt phòng theo từng tháng, quý, năm.....	255
9.15. Tính trung bình ngày chờ trên mỗi loại khách hàng của loại hình phân khúc thị trường của mỗi quốc gia và được thực hiện với năm thay đổi trạng thái bằng năm nhận phòng trừ loại hình phân khúc là Complementary .....	257
<b>CHƯƠNG IV. QUÁ TRÌNH LẬP BÁO BIỂU (SSRS).....</b>	<b>262</b>
1. Cài đặt Microsoft Sql Server 2019 Reporting Services .....	262
2. Quá trình lập báo biểu bằng công cụ Report Builder.....	272

2.1. Tạo kết nối với kho dữ liệu.....	272
2.2. Thực hiện báo biểu bảng công cụ Report Builder.....	275
<b>3. Quá trình lập báo biểu bằng công cụ Google Data Studio .....</b>	<b>306</b>
3.1. Tạo file CSV từ dữ liệu đã lọc từ quá trình SSIS.....	306
3.2. Đưa dữ liệu và lọc dữ liệu từ công cụ Google Data Studio. ....	307
3.3. Quá trình thực hiện báo biểu bằng công cụ Google Data Studio.....	309
<b>CHƯƠNG V. QUÁ TRÌNH DATA MINING.....</b>	<b>325</b>
1. Tạo project Data Mining trong Visual Studio 2019 .....	325
2. Xác định dữ liệu nguồn (DEFINE DATA SOURCE) .....	327
3. Xác định khung nhìn dữ liệu nguồn (DEFINE DATA SOURCE VIEW) .....	330
4. Thiết lập cấu trúc Data Mining.....	335
5. Deploy project Data Mining .....	343
7. Xây dựng thêm Model.....	346
8. So sánh các thuật toán khác .....	350
9. Thực hiện lại khi xóa thuộc tính Id_Reservation_Status .....	350
10. Mô hình các thuật toán khi loại bỏ Id_Reservation_Status .....	351
11. So sánh kết quả giữa hai lần chạy.....	362
11.1. So sánh và đánh giá các thuật toán bằng đồ thị Lift .....	362
11.2. Dự đoán kết quả .....	365
<b>BẢNG PHÂN CÔNG CÔNG VIỆC .....</b>	<b>367</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>368</b>

## CHƯƠNG I. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU

### 1. Phát biểu về dữ liệu

#### 1.1. Lý do chọn đề tài

Trong những năm qua với sự phát triển của ngành du lịch thì sự đóng góp không nhỏ vào thành công chung của ngành phải kể đến trước hết là lĩnh vực kinh doanh khách sạn. Trong đó bộ phận lễ tân khách sạn nói chung và hoạt động đặt phòng nói riêng là nhân tố chủ chốt giúp thúc đẩy hoạt động kinh doanh của khách sạn.

Đối với khách hàng, quá trình đặt phòng làm tăng cơ hội về một thỏa thuận cho chỗ ở, khách lưu trú sẽ yên tâm khi đến nơi. Đối một khách sạn, dữ liệu đặt phòng có thể giúp quản lý tìm xu hướng và đưa ra dự báo về kinh doanh và quản lý tốt hơn trải nghiệm của khách trong suốt mùa kinh doanh bình thường cũng như mùa cao điểm.

Nhận thấy được tầm quan trọng của quá trình đặt phòng trong lĩnh vực khách sạn, nhóm đã quyết định chọn đề tài phân tích hoạt động đặt phòng của các khách sạn từ 2015-2017.

#### 1.2. Giới thiệu nguồn dữ liệu

##### 1.2.1. Khái niệm tập dữ liệu.

Tập dữ liệu tương ứng với nội dung của một bảng cơ sở dữ liệu hoặc một ma trận dữ liệu thống kê, trong đó mỗi cột của bảng đại diện cho một biến cụ thể và mỗi hàng tương ứng với một thành viên nhất định của tập dữ liệu được đề cập.

##### 1.2.2. Tên nguồn dữ liệu

- Tên dataset : Hotel Booking Demand.
- Ngày cập nhật gần nhất : 06/2021.

##### 1.2.3. Đơn vị cung cấp

- Đây là dữ liệu được lấy từ bài báo *Hotel Booking Demand Datasets* trong tạp chí *Data in Brief* số 22 xuất bản vào tháng 02/2019 được viết bởi Nuno Antonio, Ana Almeida và Luis Nunes.

#### 1.2.4. Mô tả số dòng, cột, thời gian thu thập

- Bộ dữ liệu gồm 119390 dòng dữ liệu và 36 cột thuộc tính chứa quan sát các hoạt động đặt phòng từ các city hotel và resort hotel.
- Dữ liệu được thu nhập vào khoảng thời gian từ 01/07/2015 – 31/08/2017

#### 1.2.5. Hướng chủ đề

- Kho dữ liệu được xây dựng với hướng chủ đề kinh doanh khách sạn.
- Nguồn tải dataset : <https://www.kaggle.com/mojtaba142/hotel-booking>

#### 1.3. Thuộc tính của tập dữ liệu

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu
1	Hotel	Kiểu khách sạn. Có hai kiểu: 1. Resort Hotel. 2. City Hotel.	Text
2	Is_canceled	Hủy đặt phòng. Có hai giá trị: 1. Giá trị 0: không hủy. 2. Giá trị 1: hủy.	Number
3	Lead_time	Số ngày từ ngày khách hàng đặt chỗ và được ghi nhận vào hệ thống quản lý của khách sạn (PSM) đến ngày khách hàng đến nhận phòng.	Number
4	Arrival_date_time	Năm của ngày khách hàng đến.	Number
5	Arrvial_date_month	Tháng của ngày khách hàng đến với 12 giá trị từ tháng 1 đến tháng 12.	Text
6	Arrvial_date_week_number	Số thứ tự tuần trong năm của ngày khách hàng đến với giá trị từ 1 đến 53.	Number
7	Arrvial_date_of_month	Số ngày trong tháng của ngày khách hàng đến với giá trị từ 1 đến 31.	Number

*Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

8	Stays_in_weekend_nights	Số đêm cuối tuần (thứ bảy hoặc chủ nhật) mà khách có lưu trú hoặc đặt phòng tại khách sạn	Number
9	Stays_in_week_nights	Số đêm trong tuần (từ Thứ Hai đến Thứ Sáu) mà khách đã lưu trú hoặc đặt phòng tại khách sạn (được tính bằng cách đếm số đêm trong tuần).	Number
10	Adults	Số lượng khách hàng là người lớn	Number
11	Children	Số lượng khách hàng là trẻ em	Number
12	Babies	Số lượng khách hàng là em bé	Number
13	Meal	Cho biết các bữa ăn được bao gồm trong giá phòng mà khách hàng đặt, có 3 giá trị chính : + BB ( Bed & Breakfast) : Chỉ bao gồm bữa sáng. + HB (Half Board) : Bữa sáng và bữa tối được bao gồm trong giá phòng. Trong một số trường hợp, bạn có thể chọn nhận bữa trưa thay vì bữa sáng - khách sạn sẽ xác nhận điều này khi khách hàng đến. + FB (Full Board) : Được bao gồm bữa sáng, bữa trưa và buổi tối. + SC (Self-care) : Khách hàng tự phục vụ.	Text
14	Country	Quốc tịch của khách hàng.	Text
15	Market_segment	Thể hiện các phân khúc khách hàng của khách sạn : + Offline TA/TO (Travel Agency/ Tour Operator) : khách hàng từ các đại lý du lịch và công ty điều hành tour. + Online TA : khách hàng từ các đại lý du lịch nhưng bằng hình thức online trực tuyến. + Aviation : khách hàng đến từ các hãng hàng không.	Text

		<ul style="list-style-type: none"> <li>+ Complementary : khách hàng được hưởng ưu đãi miễn phí ( vd : khi một chuyến bay bị hủy, hãng hàng không có thể sẽ sắp xếp cho khách hàng chỗ nghỉ trong khách sạn và khi check-out khách hàng không phải trả tiền)</li> <li>+ Corporate: khách hàng đến từ các công ty, doanh nghiệp</li> <li>+ Groups: khách hàng đến từ các hội nhóm, tổ chức</li> <li>+ Direct: khách hàng đặt phòng trực tiếp tại khách sạn</li> <li>+ Undefined: không xác định</li> </ul>	
16	Distribution_channel	<p>Để tiếp cận với lượng lớn khách thuê phòng tiềm năng, các khách sạn đều liên kết bán phòng qua nhiều kênh phân phối phòng khách sạn. Có ba loại:</p> <ul style="list-style-type: none"> <li>+ TA/TO ( Travel Agency/Tour Operator): là kênh phân phối thông qua các đại lý du lịch và công ty điều hành tour.</li> <li>+ Direct : là kênh phân phối trực tiếp ( khách hàng đặt phòng trực tiếp tại khách sạn)</li> <li>+ Corporate : là kênh phân phối thông qua các công ty, doanh nghiệp</li> <li>+ Undefined : không xác định.</li> </ul>	Text
17	Is_repeated_guest	<p>Cho biết lượt đặt chỗ có phải từ khách hàng đã lưu trú nhiều lần tại khách sạn hay là lần đầu đến</p> <p>Có hai giá trị:</p> <ol style="list-style-type: none"> <li>1. 0: khách hàng tới lần đầu.</li> <li>2. 1: khách hàng cũ.</li> </ol>	Number

18	Previous_cancellations	Số lượng lượt đặt phòng đã bị khách hàng hủy trước lượt đặt phòng hiện tại.	Number
19	Previous_booking_not_canceled	Số lượng lượt nhận phòng thành công của khách hàng trước lượt đặt phòng hiện tại.	Number
20	Reserved_room_type	Mã loại phòng được đại diện bằng các chữ cái.	Text
21	Assigned_room_type	Mã loại phòng được chỉ định cho lượt đặt phòng. Đôi khi loại phòng được chỉ định khác với loại phòng đã đặt vì lý do hoạt động của khách sạn(khách sạn nhận số lượng đặt phòng nhiều hơn số lượng phòng thực tế ) hoặc do yêu cầu của khách hàng.	Int
22	Booking_changes	Số lượng sửa đổi đối với lượt đặt phòng kể từ thời điểm đặt phòng được nhập trên PMS cho đến thời điểm nhận phòng hoặc lượt đặt phòng bị hủy bỏ	Number
23	Deposit_type	Thể hiện hình thức đặt cọc trước: + No Deposit : không có tiền đặt cọc trước + Non Refund : một khoản đặt cọc đã được thực hiện bằng giá trị của tổng chi phí lưu trú. + Refundable : một khoản đặt cọc đã được thực hiện với giá trị dưới tổng chi phí lưu trú.	Text
24	Agent	Thể hiện ID của đại lý du lịch đã đặt chỗ	Number
25	Company	ID của công ty / tổ chức đã thực hiện đặt phòng hoặc chịu trách nhiệm thanh toán đặt phòng	Number
26	Days_in_waiting_list	Số ngày lượt đặt chỗ ở trong danh sách chờ trước khi nó được xác nhận với khách hàng	Number
27	Customer_type	Thể hiện loại khách hàng : + Group : Khi khách hàng đặt phòng theo một nhóm. + Transient : Khách tạm trú là những khách lưu trú qua đêm không có ý định ở lại lâu dài.	Text

		+ Transient party: Khách tạm trú theo nhóm lưu trú qua đêm không có ý định ở lại lâu dài.	
28	Adr	Doanh thu bình quân hằng ngày ( Avarage Daily Rate) được tính bằng cách chia tổng của tất cả các giao dịch lưu trú cho tổng số đêm lưu trú.	Number
29	Required_car_parking_spaces	Số chỗ đậu xe ô tô theo yêu cầu của khách hàng.	Number
30	Total_of_special_requests	Số lượng yêu cầu đặc biệt của khách hàng (ví dụ: có giường đôi hoặc ở tầng cao)	Number
31	Reservation_status	Thể hiện tình trạng đặt phòng : + Check-Out : Khách hàng đã check-in nhưng đã rời đi. + No-Show: Khách hàng không check-in và đã thông báo lý do cho khách sạn. + Cancel : Khách hàng hủy lượt đặt phòng.	Text
32	Reservation_status_date	Ngày mà trạng thái cuối cùng của lượt đặt phòng được cập nhật. Thuộc tính này có thể được sử dụng cùng với Reservation_status để biết khi nào đặt phòng bị hủy hoặc khi nào khách hàng đã trả phòng khách sạn.	Date
33	Name	Họ tên của khách hàng	Text
34	Email	Email của khách hàng	Text
35	Phone_number	Số điện thoại khách hàng	Number
36	Credit_card	Số thẻ tín dụng của khách hàng	Number

#### 1.4. Tiết xử lý dữ liệu bằng Python

- Import các thư viện cần thiết cho quá trình xử lý.

```
# Linear algebra
import numpy as np
# Data processing
import pandas as pd
# Data visualization
import seaborn as sns
%matplotlib inline
from matplotlib import pyplot as plt
from matplotlib import style
# Datetime
import datetime
from datetime import datetime
```

Hình 1.2: Các thư viện cần thiết cho quá trình xử lý.

- Đọc tập dữ liệu.

```
with open("E:\\IS217\\TienXuLyDuLieu\\hotel_booking.csv") as f:
    df = pd.read_csv(f)
```

Hình 1.3: Đọc dữ liệu đầu vào.

- Kiểm tra 5 dòng dữ liệu đầu tiên.

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	

5 rows x 36 columns

Hình 1.4: Năm dòng đầu trong tập dữ liệu

- Thống kê mô tả trước khi tiền xử lý dữ liệu.
  - Thống kê số lượng nhận phòng và hủy đặt phòng.

```

fig, ax = plt.subplots(figsize=(8, 6))

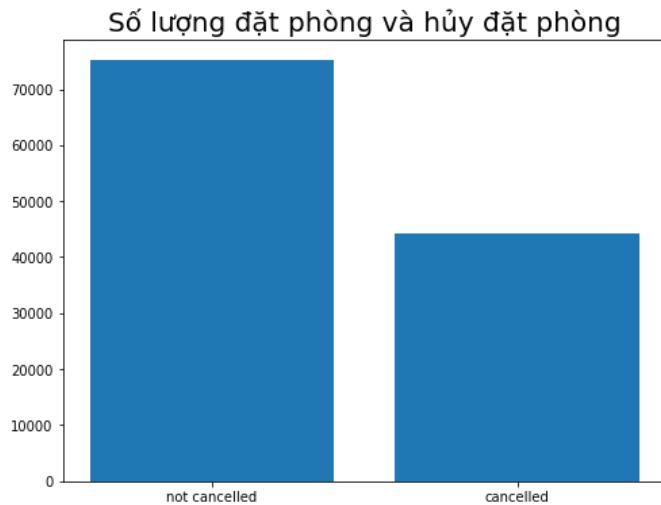
vcs = df['is_canceled'].value_counts()

ax.bar(vcs.index, vcs)
ax.set_xticks([0, 1])
ax.set_xticklabels(['not cancelled', 'cancelled'])
ax.set_title('Số lượng đặt phòng và hủy đặt phòng', fontsize=20)

```

Python

Hình 1.5: Code thống kê số lượng nhận phòng và hủy đặt phòng.



Hình 1.4: Kết quả thống kê.

- Kiểm định mối quan hệ giữa dữ liệu bị hủy đặt phòng trong hai loại hình khách sạn trong khoảng thời gian từ khi đặt phòng tới khi nhận được phòng.

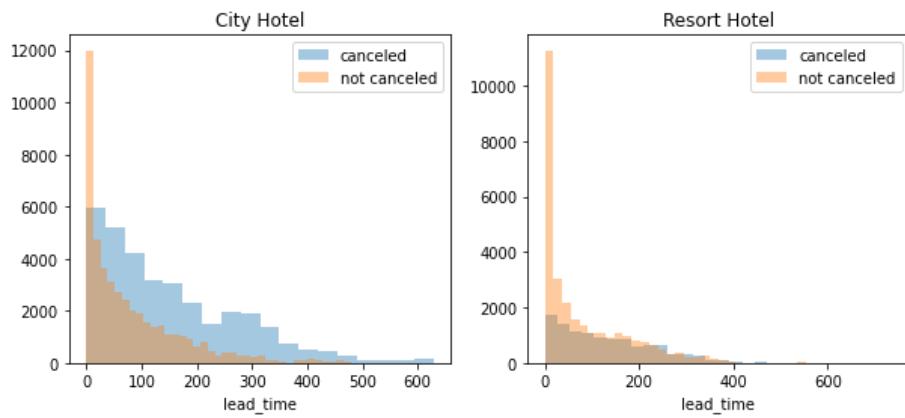
```

canceled = 'canceled'
not_canceled = 'not canceled'
fig, axes = plt.subplots(nrows = 1, ncols = 2, figsize = (10,4))
city = df[df['hotel']=='City Hotel']
resort = df[df['hotel']=='Resort Hotel']
ax = sns.distplot(city[city['is_canceled']==1].lead_time.dropna(), bins = 18,label = canceled, ax =axes[0],kde = False)
ax = sns.distplot(city[city['is_canceled']==0].lead_time.dropna(), bins = 40,label =not_canceled, ax =axes[0],kde = False)
ax.legend()
ax.set_title('City Hotel')
ax = sns.distplot(resort[resort['is_canceled']==1].lead_time.dropna(), bins = 18,label = canceled, ax =axes[1],kde = False)
ax = sns.distplot(resort[resort['is_canceled']==0].lead_time.dropna(), bins = 40,label =not_canceled, ax =axes[1],kde = False)
ax.legend()
_=ax.set_title('Resort Hotel')

```

Hình 1.6: Code vẽ đồ thị thể hiện mối quan hệ giữa dữ liệu bị hủy đặt phòng trong hai loại hình khách sạn trong khoảng thời gian từ khi đặt phòng tới khi nhận được phòng.

## Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 1.7: Kết quả thống kê.

- Thống kê sự tương quan giữa việc thuê phòng vào cuối tuần với hủy đặt phòng tại khách sạn.

```

total = df.groupby('stays_in_weekend_nights')['is_canceled'].count()[:5]
cancels = df.groupby('stays_in_weekend_nights')['is_canceled'].sum()[:5]

cutoff = 3
x = list(total.index[:cutoff])
x.append(cutoff)

y_total = list(total.values[:cutoff])
y_total.append(np.sum(total.values[cutoff:]))

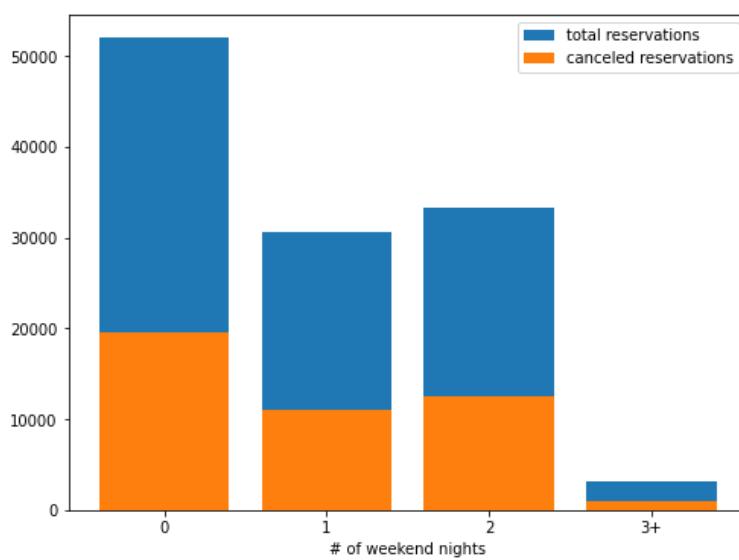
y_cancel = list(cancels.values[:cutoff])
y_cancel.append(np.sum(cancels.values[cutoff:]))
```

```

fig, ax = plt.subplots(figsize=(8, 6))

ax.bar(x, y_total, label='total reservations')
ax.bar(x, y_cancel, label='canceled reservations')
ax.legend()
ax.set_xticks([0, 1, 2, 3])
ax.set_xticklabels(['0', '1', '2', '3+'])
ax.set_xlabel('# of weekend nights')
```

Hình 1.8: Code thống kê sự tương quan giữa việc thuê phòng vào cuối tuần với hủy đặt phòng tại khách sạn



Hình 1.9: Kết quả thống kê.

- Tiên xử lý dữ liệu.
  - Xóa thuộc tính agent và company vì có quá nhiều dữ liệu Null.

```
# Xóa cột agent  
df = df.drop(["agent"], axis = 1)  
# Xóa cột company  
df = df.drop(["company"], axis = 1)
```

Python

Hình 1.10: Xóa thuộc tính agent và company.

- Xóa thuộc tính credit\_card vì ẩn nhiều thông tin.

```
# Xóa cột credit_card  
df = df.drop(["credit_card"], axis = 1)
```

Python

Hình 1.11: Xóa thuộc tính credit\_card.

- Xóa các dòng dữ liệu children có giá trị Null vì quá ít sẽ không ảnh hưởng nhiều.

```
# Xóa dòng dữ liệu mà thuộc tính children có giá trị null  
df.dropna(axis=0,inplace=True,subset=['children'])
```

Python

Hình 1.12: Xóa các dòng dữ liệu trong children có giá trị Null.

- Thay thế các dòng dữ liệu Null trong country thành các giá trị có tần suất xuất hiện cao nhất.

```
#Xem xét thông tin thuộc tính country  
df['country'].describe()
```

Python

```
count      118902  
unique       177  
top        PRT  
freq      48590  
Name: country, dtype: object
```

```
# Ta nhận thấy giá trị "PRT" phổ biến nhất nên ta chọn giá trị đó là  
common_value='PRT'
```

```
for dataset in [df]:  
    dataset['country'] = dataset['country'].fillna(common_value)
```

Python

Hình 1.13: Thay thế các giá trị trong dòng dữ liệu trong country có giá trị Null thành PRT.

- Xóa các dòng dữ liệu có giá trị Undefined vì thiếu thông tin.

```
df.replace('Undefined', np.nan, inplace=True)
```

Python

Hình 1.14: Thay thế giá trị Undefined thành NaN.

```
# Xóa tất cả các giá trị Undefined
df = df.dropna()
```

Python

Hình 1.15: Xóa các giá trị NaN trong tập dữ liệu.

- Chuyển đổi thuộc tính bữa ăn (meal) từ chữ sang số.

```
meals={"BB":1,"FB":2,"HB":3,"SC":4}

for dataset in [df]:
    dataset['meal']=dataset['meal'].map(meals)
```

Python

Hình 1.16: Chuyển đổi thuộc tính từ dạng chữ sang dạng số.

- Chuyển đổi thuộc tính phòng (reserved\_room\_type và assigned\_room\_type) từ chữ sang số.

```
rooms={"A":1, "B":2, "C":3, "D":4, "E":5, "F":6, "G":7, "H":8, "I":9, "K":10, "L":11, "P":12}

for dataset in [df]:
    dataset['reserved_room_type']=dataset['reserved_room_type'].map(rooms)
    dataset['assigned_room_type']=dataset['assigned_room_type'].map(rooms)
```

Hình 1.17: Chuyển đổi thuộc tính từ dạng chữ sang dạng số.

- Chuyển đổi thuộc tính tháng (arrival\_date\_month) từ chữ sang số.

```
for dataset in [df]:
    dataset.loc[dataset['arrival_date_month']=='January','arrival_date_month_number']=1
    dataset.loc[dataset['arrival_date_month']=='February','arrival_date_month_number']=2
    dataset.loc[dataset['arrival_date_month']=='March','arrival_date_month_number']=3
    dataset.loc[dataset['arrival_date_month']=='April','arrival_date_month_number']=4
    dataset.loc[dataset['arrival_date_month']=='May','arrival_date_month_number']=5
    dataset.loc[dataset['arrival_date_month']=='June','arrival_date_month_number']=6
    dataset.loc[dataset['arrival_date_month']=='July','arrival_date_month_number']=7
    dataset.loc[dataset['arrival_date_month']=='August','arrival_date_month_number']=8
    dataset.loc[dataset['arrival_date_month']=='September','arrival_date_month_number']=9
    dataset.loc[dataset['arrival_date_month']=='October','arrival_date_month_number']=10
    dataset.loc[dataset['arrival_date_month']=='November','arrival_date_month_number']=11
    dataset.loc[dataset['arrival_date_month']=='December','arrival_date_month_number']=12
    dataset['arrival_date_month_number']=dataset['arrival_date_month_number'].astype(int)
```

Hình 1.18: Chuyển đổi thuộc tính từ dạng chữ sang dạng số.

- Sửa tên thuộc tính arrival\_date\_name thành arrival\_date\_name\_abbrev.

```
df.rename(index=str, columns={"arrival_date_month": "arrival_date_month_name"}, inplace=True)
```

*Hình 1.19: Đổi tên thuộc tính.*

- Thêm thuộc tính id\_reservation\_status\_date lấy ý tưởng từ reservation\_status\_date theo cấu trúc YYYYMMDD.

```
df['reservation_status_date']=pd.to_datetime(df['reservation_status_date'])
df["id_reservation_status_date"] = df["reservation_status_date"].dt.strftime('%Y%m%d')
```

Python

*Hình 1.20: Tạo thuộc tính id\_reservation\_status\_date.*

- Tạo thuộc tính arrival\_date\_full dựa vào các thuộc tính khác như arrival\_date\_year(YYYY), arrival\_date\_month\_number (MM) và arrival\_date\_day\_of\_month(DD).

```
cols=["arrival_date_year","arrival_date_month_number","arrival_date_day_of_month"]
df['arrival_date_full'] = df[cols].apply(lambda x: '-'.join(x.values.astype(str)), axis="columns")
```

Python

```
df['arrival_date_full']=pd.to_datetime(df['arrival_date_full'])
for dataset in [df]:
    print(dataset["arrival_date_full"])
```

Python

*Hình 1.21: Tạo thuộc tính arrival\_date\_full*

- Tạo thuộc tính chi tiết chi arrival\_date\_full.

```
df["id_arrival_date"] = df["arrival_date_full"].dt.strftime('%Y%m%d')
```

*Hình 1.22: Tạo thuộc tính id\_arrival\_date.*

```
df["arrival_date_quarter"] = df["arrival_date_full"].dt.quarter
for dataset in [df]:
    print(dataset["arrival_date_quarter"])
```

*Hình 1.23: Tạo thuộc tính arrival\_date\_quarter.*

```
df["arrival_date_day_of_week"] = df["arrival_date_full"].dt.dayofweek
for dataset in [df]:
    print(dataset["arrival_date_day_of_week"])
```

*Hình 1.24: Tạo thuộc tính arrival\_date\_day\_of\_week.*

```
df["arrival_date_day_name"] = df["arrival_date_full"].dt.day_name()
for dataset in [df]:
    print(dataset["arrival_date_day_name"])
```

*Hình 1.25: Tạo thuộc tính arrival\_date\_day\_name.*

```
df['arrival_date_month_name_abbrev'] = df['arrival_date_full'].dt.month_name().str[:3]
for dataset in [df]:
    print(dataset["arrival_date_month_name_abbrev"])
```

Hình 1.26: Tạo thuộc tính arrival\_date\_month\_name\_abbrev.

```
df["arrival_date_weekday_flag"] = df["arrival_date_day_of_week"]
for dataset in [df]:
    dataset['arrival_date_weekday_flag']=dataset['arrival_date_weekday_flag'].astype(int)
    dataset.loc[dataset['arrival_date_weekday_flag']<=5,'arrival_date_weekday_flag']=0
    dataset.loc[dataset['arrival_date_weekday_flag']>5,'arrival_date_weekday_flag']=1
```

Hình 1.28: Tạo thuộc tính arrival\_date\_weekday\_flag.

- Kết quả của quá trình tiền xử lý:

- Dữ liệu gồm có 118 216 dòng dữ liệu và 43 cột.

Index: 118216 entries, 0 to 119389

Data columns (total 43 columns):

#	Column	Non-Null Count	Dtype
0	hotel	118216	non-null object
1	is_canceled	118216	non-null int64
2	lead_time	118216	non-null int64
3	arrival_date_year	118216	non-null int64
4	arrival_date_month_name	118216	non-null object

A	B	C	D	E	F	G	H	I	J	K
1 [hotel]	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_week	stays_in_weekend_nights	stays_in_week_nights	adults	children
2 Resort Hotel	0	342	2015	July	27	1	0	0	0	2
3 Resort Hotel	0	737	2015	July	27	1	0	0	0	2
4 Resort Hotel	0	7	2015	July	27	1	0	1	1	1
5 Resort Hotel	0	13	2015	July	27	1	0	1	1	1
6 Resort Hotel	0	14	2015	July	27	1	0	2	2	2
7 Resort Hotel	0	14	2015	July	27	1	0	2	2	2
8 Resort Hotel	0	0	2015	July	27	1	0	0	2	2
9 Resort Hotel	0	9	2015	July	27	1	0	2	2	2
10 Resort Hotel	1	85	2015	July	27	1	0	3	2	2
11 Resort Hotel	1	75	2015	July	27	1	0	3	2	2
12 Resort Hotel	1	23	2015	July	27	1	0	4	2	2
13 Resort Hotel	0	35	2015	July	27	1	0	4	2	2
14 Resort Hotel	0	68	2015	July	27	1	0	4	2	2
15 Resort Hotel	0	18	2015	July	27	1	0	4	2	2
16 Resort Hotel	0	37	2015	July	27	1	0	4	2	2
17 Resort Hotel	0	68	2015	July	27	1	0	4	2	2
18 Resort Hotel	0	37	2015	July	27	1	0	4	2	2
19 Resort Hotel	0	12	2015	July	27	1	0	1	2	2
20 Resort Hotel	0	0	2015	July	27	1	0	1	2	2
21 Resort Hotel	0	7	2015	July	27	1	0	4	2	2
22 Resort Hotel	0	37	2015	July	27	1	1	4	1	2
23 Resort Hotel	0	72	2015	July	27	1	2	4	2	2

Hình 1.16.1,2: Dữ liệu sau khi xử lý

## **2. Xây dựng kho dữ liệu**

### **2.1. Khái niệm Dimensional Modeling**

- Mô hình chiều dữ liệu (Dimensional modeling) là một kỹ thuật thiết kế logic thường được sử dụng trong phần thiết kế data warehouse nhằm tìm cách trình bày dữ liệu trong một khuôn khổ tiêu chuẩn trực quan cho phép truy cập hiệu suất cao.

+ Bảng Fact : Chứa các dữ liệu thuộc nhóm định lượng hoặc sự kiện, có tính chất đo lường mà người dùng cần phân tích và được thiết kế dựa trên business process.

+ Bảng Dimension: Chứa các thuộc tính mô tả của bảng fact.

+ Bảng Dim là bảng dữ liệu tĩnh, và bảng Fact là dữ liệu động được nạp bằng các thao tác. Khoá ngoại của Fact được tạo bởi khoá của các bảng Dim. Nghĩa là khoá chính của các bảng Dim chính là khoá ngoại của bảng Fact.

### **2.2. Xây dựng sơ đồ bông tuyết**

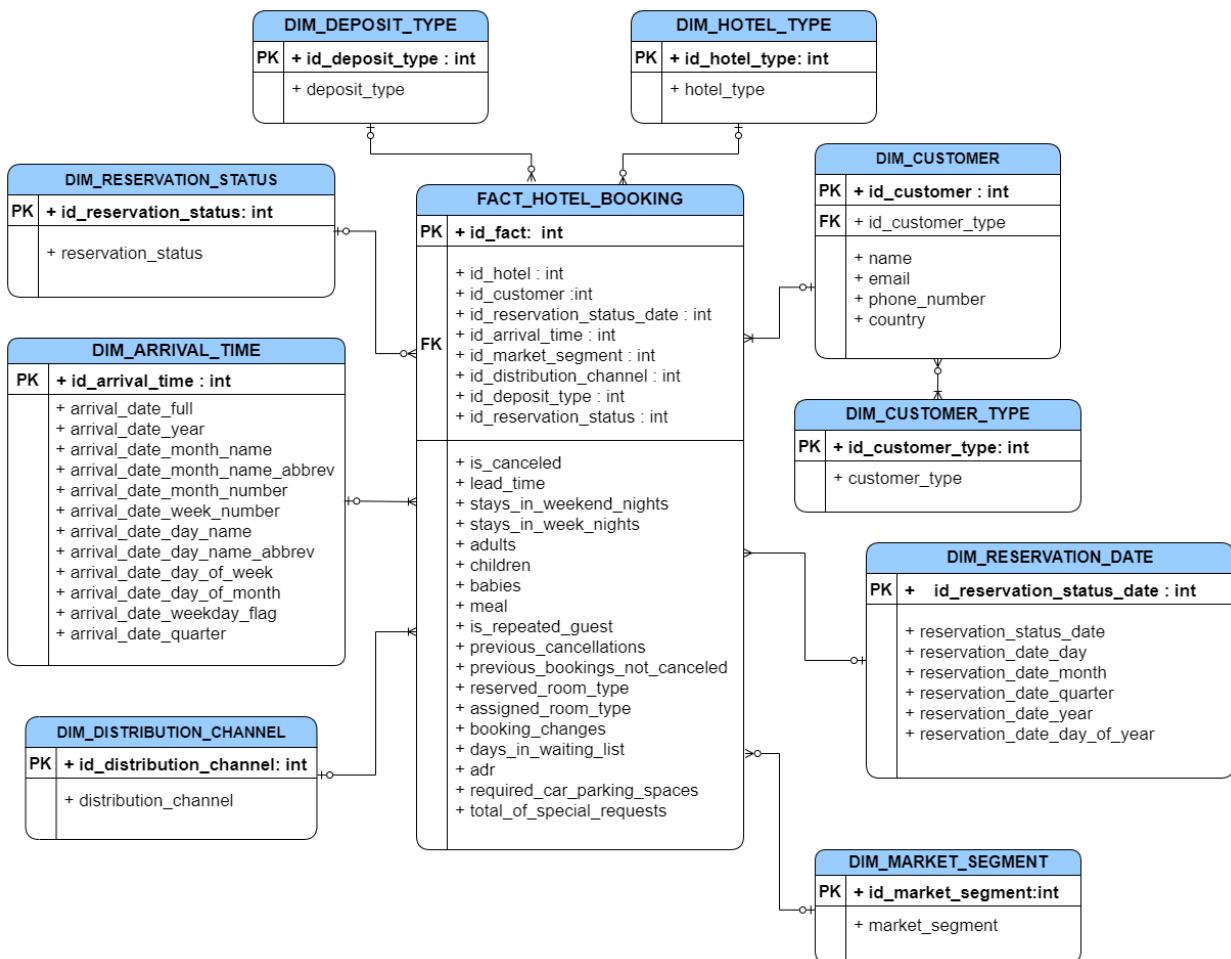
#### **a) Khái niệm sơ đồ bông tuyết**

- Là dạng mở rộng của lược đồ hình sao làm giảm dư thừa dữ liệu bằng cách chuẩn hoá các bảng Dimension.

- Một lược đồ được gọi là lược đồ bông tuyết nếu một hoặc nhiều bảng Dimension không kết nối trực tiếp với bảng dữ kiện mà phải kết nối thông qua các bảng dimension khác.

- Các bảng trong sơ đồ bông tuyết thường được chuẩn hóa thành dạng chuẩn thứ ba. Mỗi bảng Dimension thực hiện chính xác một cấp trong hệ thống phân cấp.

### b) Sơ đồ bông tuyết minh họa



Hình 2.1: Sơ đồ bông tuyết của tập dữ liệu nhu cầu đặt phòng khách sạn.

#### 2.3. Dim\_Reservation\_Status

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_reservation_status	PK	int	Mã lượt đặt phòng	
reservation_status		varchar	Trạng thái của lượt đặt phòng	

#### 2.4. Dim\_Hotel\_Type

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_hotel_type	PK	int	Mã của khách sạn	
hotel_type		varchar	Tên của khách sạn	

### 2.5. Dim\_ReservationDate

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_reservation_status_date	PK	int	Mã ngày trạng thái đặt phòng	
reservation_status_date		date	Ngày của trạng thái đặt phòng	
reservation_date_day		int	Số ngày của trạng thái đặt phòng	
reservation_date_month		varchar	Tháng của ngày trạng thái đặt phòng	
reservation_date_quarter		int	Quý của ngày trạng thái đặt phòng	
reservation_date_year		int	Năm của ngày trạng thái đặt phòng	
reservation_date_day_of_year		int	Số ngày trong năm của ngày trạng thái đặt phòng	

### 2.6. Dim\_Customer

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_customer	PK	int	Mã khách hàng	
id_customer_type	FK	int	Mã loại khách hàng	
name		varchar	Tên khách hàng	
email		varchar	Email của khách hàng	
phone_number		varchar	Số điện thoại của khách hàng	
credit_card		varchar	Số thẻ tín dụng của khách hàng	
country		varchar	Quốc tịch của khách hàng	

### 2.7. Dim\_Customer\_Type

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_customer_type	PK	int	Mã loại khách hàng	
customer_type		varchar	Tên loại khách hàng	

### 2.8. Dim\_Deposit\_Type

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_deposit_type	PK	int	Mã hình thức đặt cọc	
deposit_type		varchar	Tên hình thức đặt cọc	

### 2.9. Dim\_Distribution\_Channel

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_distribution_channel	PK	int	Mã kênh bán hàng	
distribution_channel		varchar	Tên kênh bán hàng	

### 2.10. Dim\_Market\_Segment

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_market_segment	PK	int	Mã phân khúc khách hàng	
market_segment		varchar	Tên của các phân khúc khách hàng	

### 2.11. Dim\_Arrival\_Time

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_arrival_time	PK	int	Mã ngày khách đến.	
arrival_date_full		date	Ngày khách hàng đến.	
arrival_date_year		int	Năm của ngày khách hàng đến.	
arrival_date_month_name		int	Tháng của ngày khách hàng đến với 12 giá trị từ tháng 1 đến tháng 12. (bằng số)	
arrival_date_month_name_abbrev		varchar	Tháng của ngày khách hàng đến với 12 giá trị từ tháng 1 đến tháng 12. (bằng chữ)	
arrival_date_month_number		varchar	Tháng của ngày khách hàng đến với 12 giá trị từ tháng 1 đến tháng 12. (bằng chữ viết tắt)	
arrival_date_week_number		int	Số thứ tự tuần trong năm của ngày khách hàng đến với giá trị từ 1 đến 53.	
arrival_date_day_name		varchar	Ngày trong tuần với giá trị từ thứ hai đến chủ nhật (bằng chữ).	
arrival_date_day_name_abbrev		varchar	Ngày trong tuần với giá trị từ thứ hai đến chủ nhật (bằng chữ viết tắt).	

arrival_date_day_of_week		int	Ngày trong tuần với giá trị từ thứ hai đến chủ nhật (bằng số).	
arrival_date_day_of_month		varchar	Số ngày trong tháng của ngày khách hàng đến với giá trị từ 1 đến 31	
arrival_date_weekday_flag		int	Ngày trong tuần.	
arrival_date_quarter		int	Quý của ngày khách hàng đến có giá trị từ 1 đến 4.	

## 2.12. Fact\_Hotel\_Booking

Tên thuộc tính	Ràng buộc	Kiểu dữ liệu	Mô tả thuộc tính	NULL
id_fact	PK	int	Mã của sự kiện đặt phòng.	
id_hotel	FK	int	Mã của khách sạn.	
id_customer	FK	int	Mã khách hàng.	
id_arrival_time	FK	int	Mã ngày đến của khách hàng.	
id_market_segment	FK	int	Mã phân khúc khách hàng	
id_distribution_channel	FK	int	Mã kênh phân phối bán hàng,	
id_deposit_type	FK	int	Mã hình thức đặt cọc	
id_reservation_status	FK	int	Mã trạng thái đặt phòng	
id_reservation_date	FK	int	Mã ngày trang thái đặt phòng	
is_canceled		int	Hủy đặt phòng. Có hai giá trị: 1.0: không hủy. 2. 1: hủy.	
lead_time		int	Số ngày từ ngày khách hàng đặt chỗ và được ghi nhận vào hệ thống quản lý của khách sạn (Property Management System) đến ngày khách hàng đến nhận phòng.	
stays_in_weekend_nights		int	Số đêm cuối tuần (thứ bảy hoặc chủ nhật) mà khách có lưu trú hoặc đặt phòng tại khách sạn	

stays_in_week_nights		int	Số đêm trong tuần (từ Thứ Hai đến Thứ Sáu) mà khách đã lưu trú hoặc đặt phòng tại khách sạn (được tính bằng cách đếm số đêm trong tuần).	
adults		int	Số lượng khách hàng là người lớn	
children		int	Số lượng khách hàng là trẻ em	
babies		int	Số lượng khách hàng là em bé	
meal		int	Cho biết các bữa ăn được bao gồm trong giá phòng mà khách hàng đặt.	
is_repeated_guest		int	Khách hàng cũ.	
reserved_room_type		varchar	Mã loại phòng được đại diện bằng các chữ cái.	
assigned_room_type		int	Mã loại phòng được chỉ định cho lượt đặt phòng.	
booking_changes		int	Số lượng sửa đổi đối với lượt đặt phòng kể từ thời điểm đặt phòng được nhập trên PMS cho đến thời điểm nhận phòng hoặc lượt đặt phòng bị hủy bỏ	
days_in_waiting_list		int	Số ngày lượt đặt chờ ở trong danh sách chờ trước khi nó được xác nhận với khách hàng	
adr		float	Doanh thu bình quân hằng ngày được tính bằng cách chia tổng của tất cả các giao dịch lưu trú cho tổng số đêm lưu trú.	
required_car_parking_spaces		int	Số chỗ đậu xe ô tô theo yêu cầu của khách hàng.	
total_of_special_requests		int	Số lượng yêu cầu đặc biệt của khách hàng	

### 3. Nội dung 15 câu truy vấn

- 3.1. Thống kê tổng số lượng khách hàng đặt phòng, tổng doanh thu của từng loại khách sạn theo từng năm.
- 3.2. Thống kê có bao nhiêu khách hàng là người lớn, bao nhiêu khách hàng là trẻ em và bao nhiêu khách hàng là em bé trong từng tháng của từng năm.
- 3.3. Liệt kê khách hàng của từng loại khách hàng có doanh số cao nhất trong năm 2017.
- 3.4. Cho biết top 10 quốc gia có lượng khách du lịch cao nhất trong năm 2016, sắp xếp các giá trị theo thứ tự tăng dần.
- 3.5. Liệt kê số lượng đơn đặt phòng ở 'China' và 'Việt Nam' theo từng năm và loại khách hàng là 'Transient'.
- 3.6. Tìm các quốc gia có số lượng đơn đặt phòng lớn hơn lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn.
- 3.7. Liệt kê 5 khách hàng có tên bắt đầu bằng chữ D ở Bồ Đào Nha (PRT), nếu khách hàng có tổng chi tiêu > 500 và là khách hàng quay lại thì xuất "Khách hàng thân thiết". Nếu tổng chi tiêu > 500 và là khách hàng lần đầu tiên tới thì xuất "Khách hàng tiềm năng". Ngược lại xuất "Khách hàng vãng lai.".
- 3.8. Liệt kê các quốc gia thu được lợi nhuận năm 2017 nhiều hơn năm 2016 và sắp xếp theo lợi nhuận giảm dần.
- 3.9. Liệt kê các phân khúc thị trường mà tổng doanh thu của kênh phân phối "Corporate" đó cộng lại tối thiểu 500 (lấy theo thứ tự doanh thu tăng dần) hoặc những phân khúc thị trường có số lượng lượt đặt phòng của kênh phân phối "Direct" tối thiểu 500.
- 3.10. Liệt kê tổng số đêm thuê của các quốc gia trong năm 2015 trừ (ngày nhận phòng vào thứ 7 và không đặt tiền cọc trước của trong năm 2015).
- 3.11. Với mỗi tháng đưa ra 3 email của khách hàng có chi tiêu cao nhất.
- 3.12. Liệt kê các quý theo từng năm của quốc gia Bồ Đào Nha, thuộc top 50% quý có tổng yêu cầu đặc biệt trong tổng số yêu cầu đặc biệt thu được.

3.13. Liệt kê những phân khúc thị trường và kênh phân phối có yêu cầu bão tố cho từ 5 đến 10 xe.

3.14. Thống kê tổng thời gian chờ của từng hình thức thanh toán đặt cọc mà khách hàng đặt phòng theo từng tháng, quý, năm.

3.15. Tính trung bình ngày chờ trên mỗi loại khách hàng của loại hình phân khúc thị trường của mỗi quốc gia và được thực hiện với năm thay đổi trạng thái bằng năm nhận phòng trừ loại hình phân khúc là Complementary.

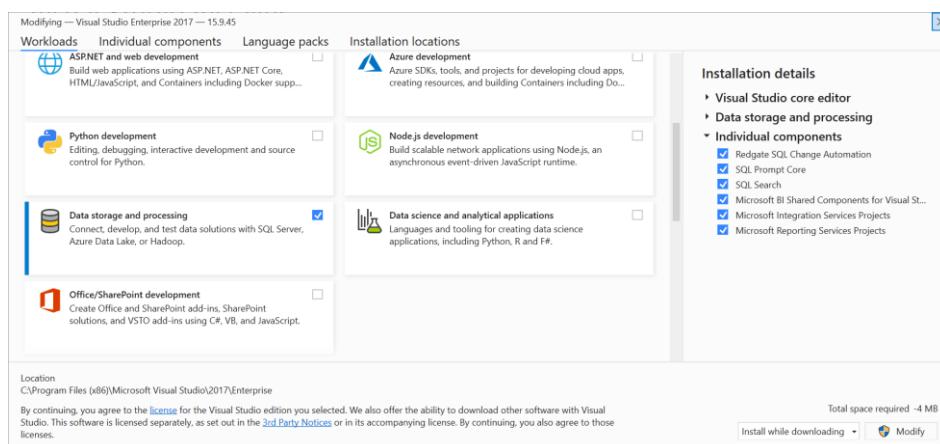
## **CHƯƠNG II : TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)**

### **1. Khái niệm**

- SSIS là tên viết tắt của SQL Server Integration Services, là một trong các thành phần chính của Microsoft SQL Server Enterprise Edition.
- SSIS là công cụ dùng để thực hiện các tác vụ quản lý tích hợp dữ liệu (Data Integration) và là thành phần chính trong các ứng dụng của data warehouse như lưu trữ dữ liệu, trích xuất và tải dữ liệu, quản lý dữ liệu, ...

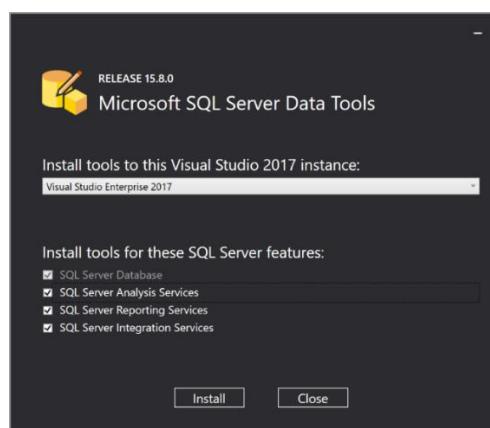
### **2. Chuẩn bị công cụ và Data Warehouse**

- Download công cụ **SQL Server Data Tools**.
- Chọn Data storage and processing và chọn những mục như hình minh họa.



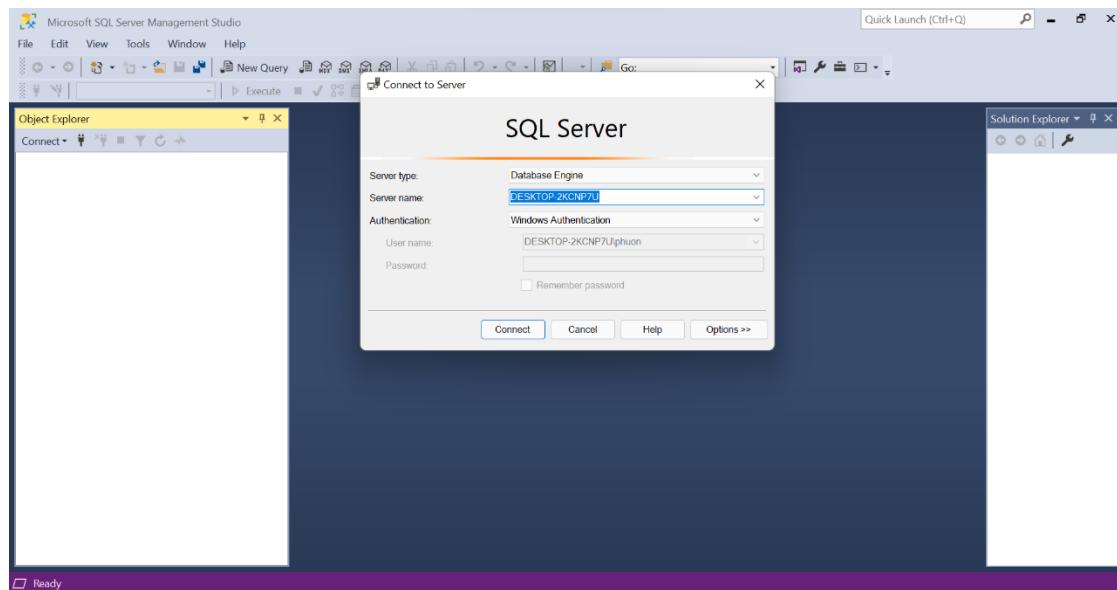
*Hình 2. 1. Giao diện tải SQL Server Data Tools trong Visual Studio 2017*

- Tải các công cụ SSDT cho phiên bản Visual Studio Enterprise 2017.



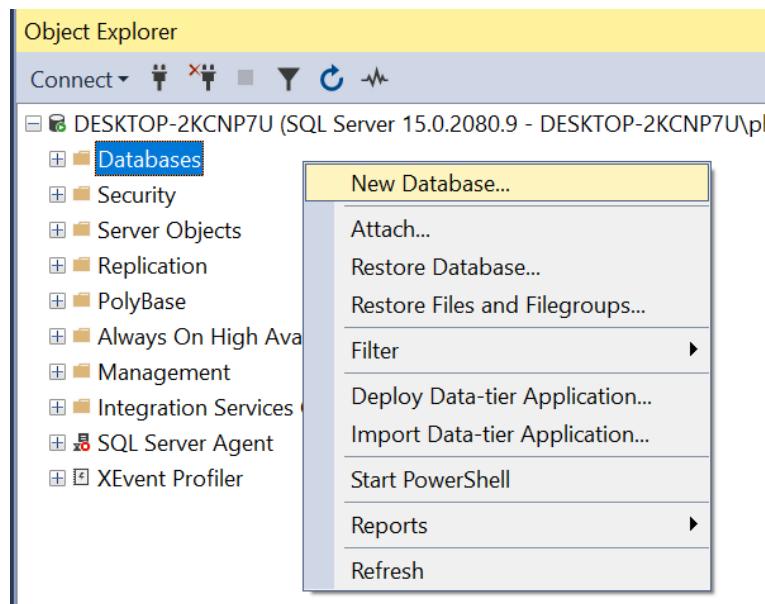
*Hình 2. 2. Các công cụ SSDT*

- Các bước tạo kho dữ liệu bằng Microsoft SQL Server Management Studio.
  - Bước 1: Mở Microsoft SQL Server Management Studio và kết nối với Server.



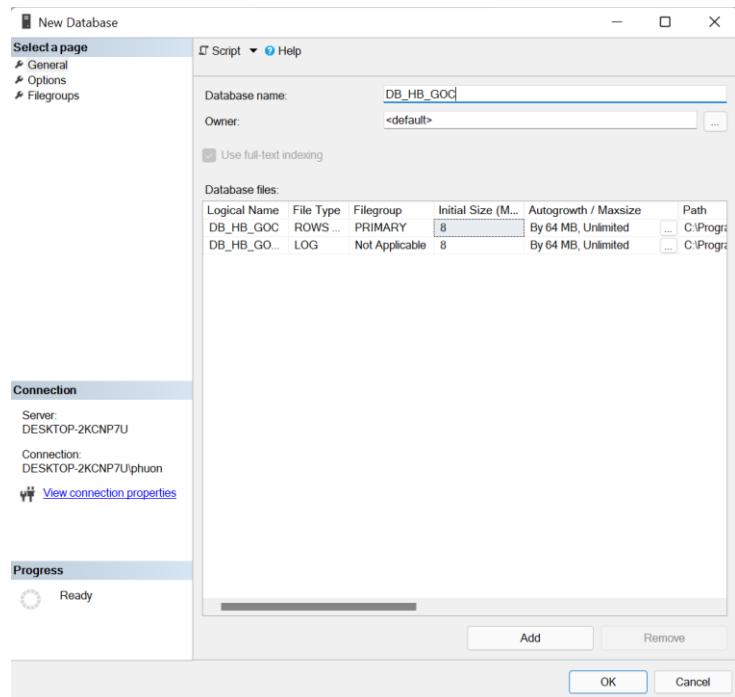
*Hình 2. 3. Mở SQL Server và kết nối với Server*

- Bước 2: Tạo kho dữ liệu DB\_HB\_GOC, chọn New Database để mở giao diện tạo cơ sở dữ liệu mới.



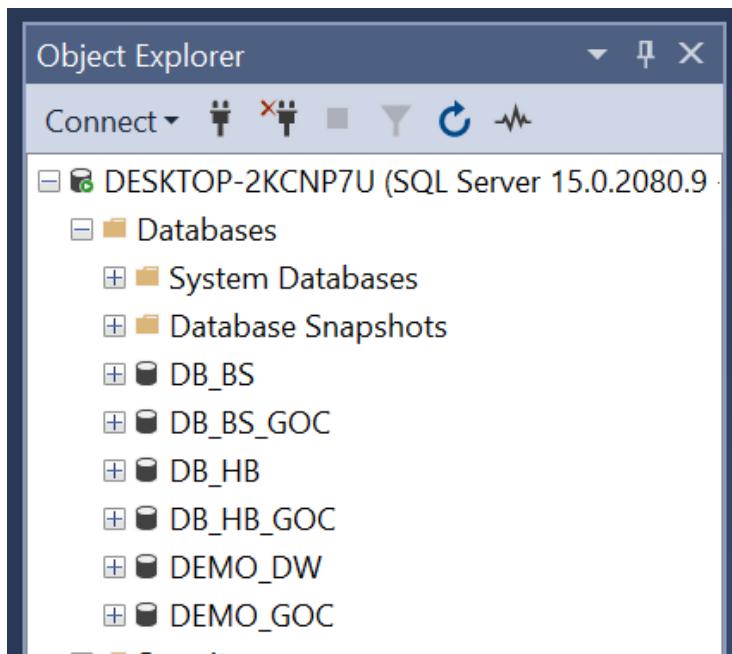
*Hình 2. 4. Tạo kho dữ liệu*

- Bước 3: Nhập tên database mới và nhấn OK để kết thúc quá trình tạo database.



Hình 2. 5. Đặt tên kho dữ liệu

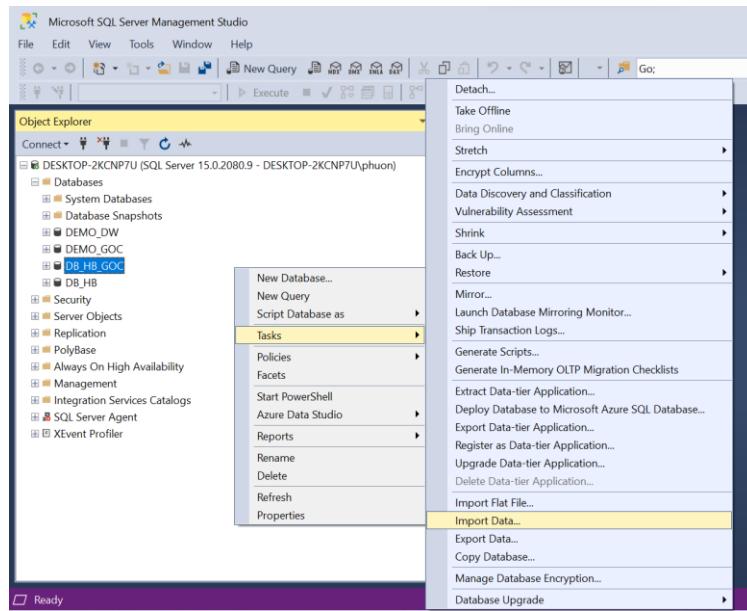
- Bước 4: Làm tương tự với DB\_HB, ta được:



Hình 2. 6. Kho dữ liệu đã tạo được

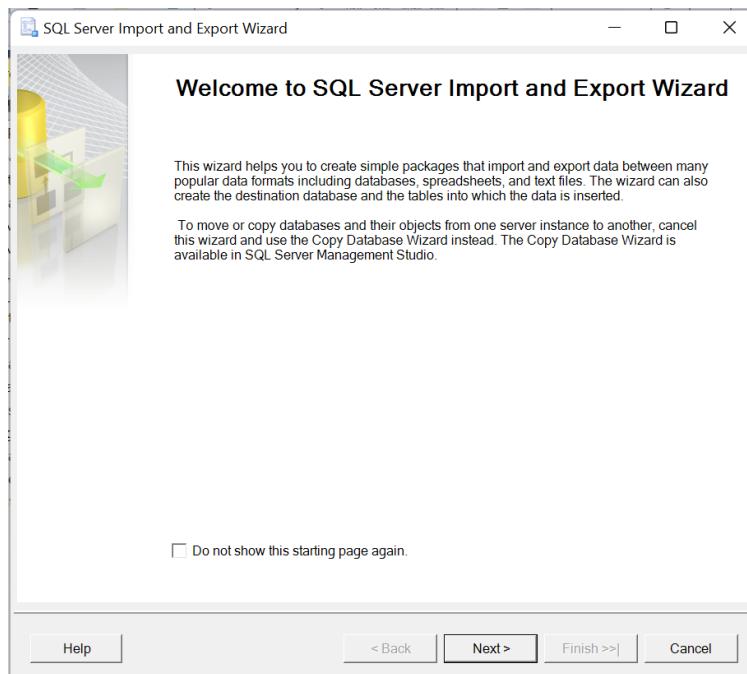
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 5: Import dữ liệu hotel\_booking (đã lọc và làm sạch ở chương 1) để tiến hành tách bảng ở phần tiếp theo:
  - + Chọn database sẽ nhập dữ liệu → Tasks → Import Data ...



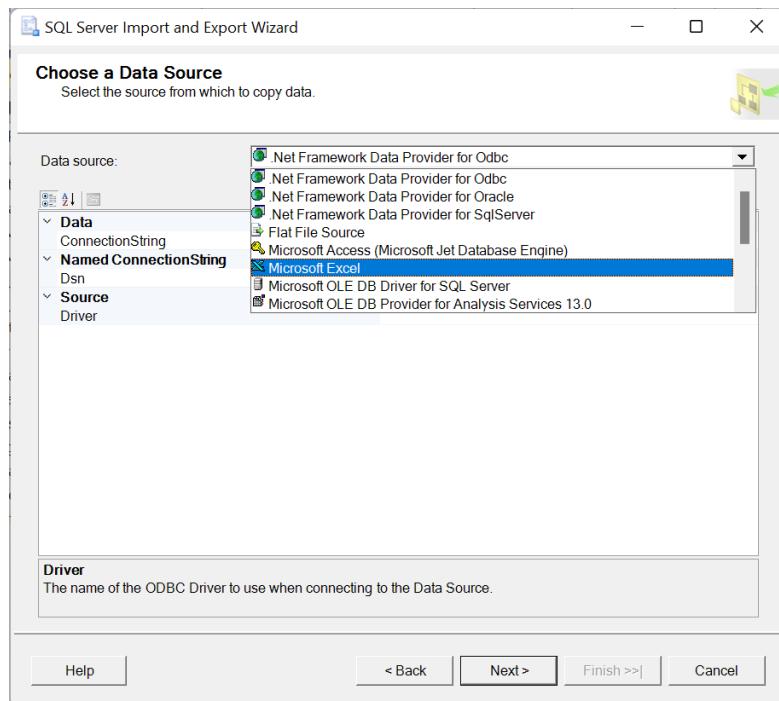
Hình 2. 7. Import dữ liệu vào DB\_HB\_GOC

- + Chọn Next để tiếp tục quá trình.



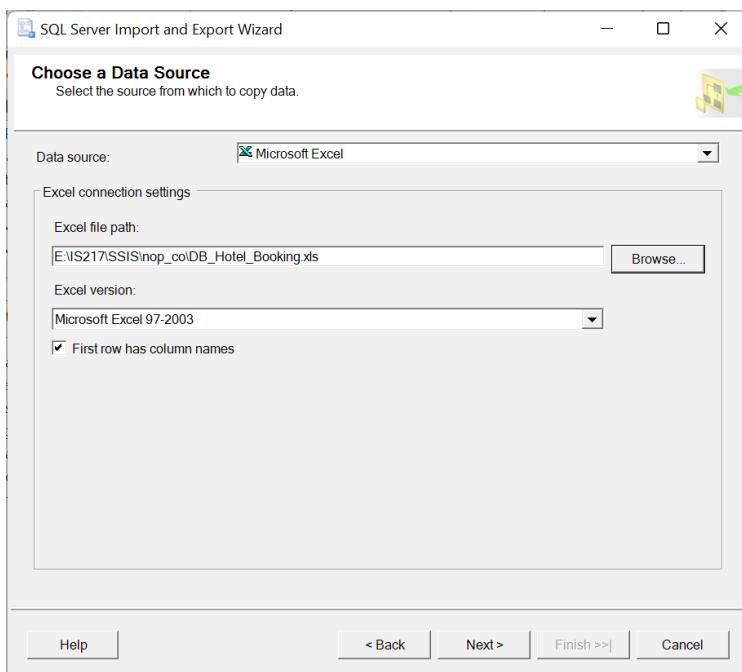
Hình 2. 8. Cửa sổ import dữ liệu

- + Trong khung data source, chọn Microsoft Excel.



*Hình 2. 9. Chọn data source*

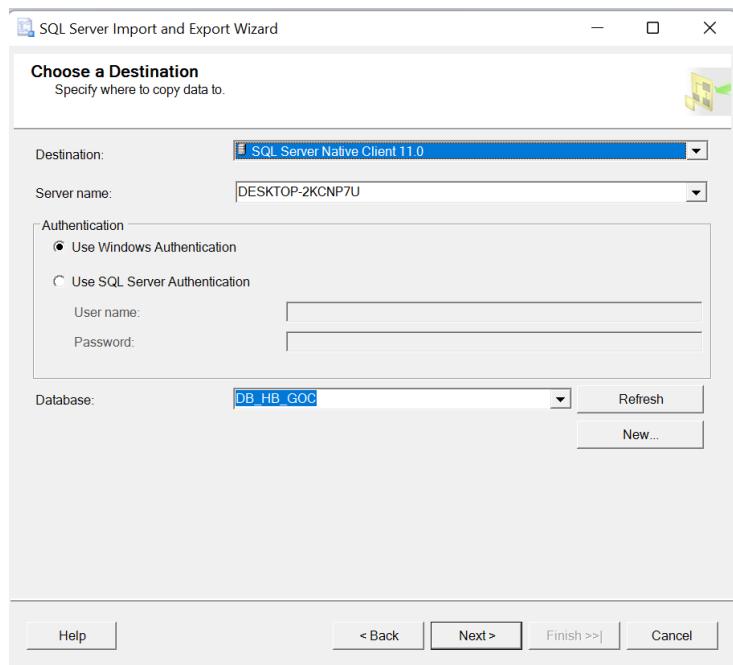
- + Nhấn Browse để chọn file excel cần nạp vào rồi chọn Next.



*Hình 2. 10. Thiết lập đường dẫn đến data source*

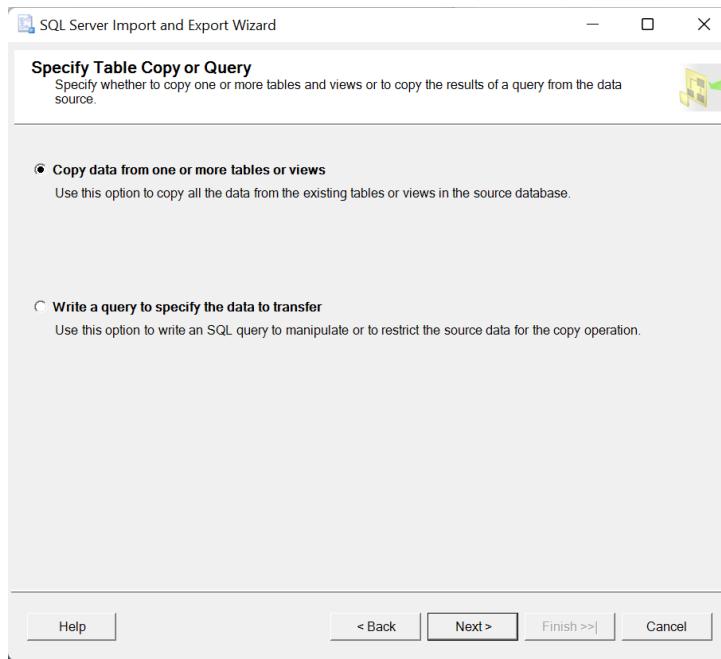
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- + Trong Destination, chọn SQL Server Native Client 11.0 → Next.



*Hình 2. 11.Chọn Destination*

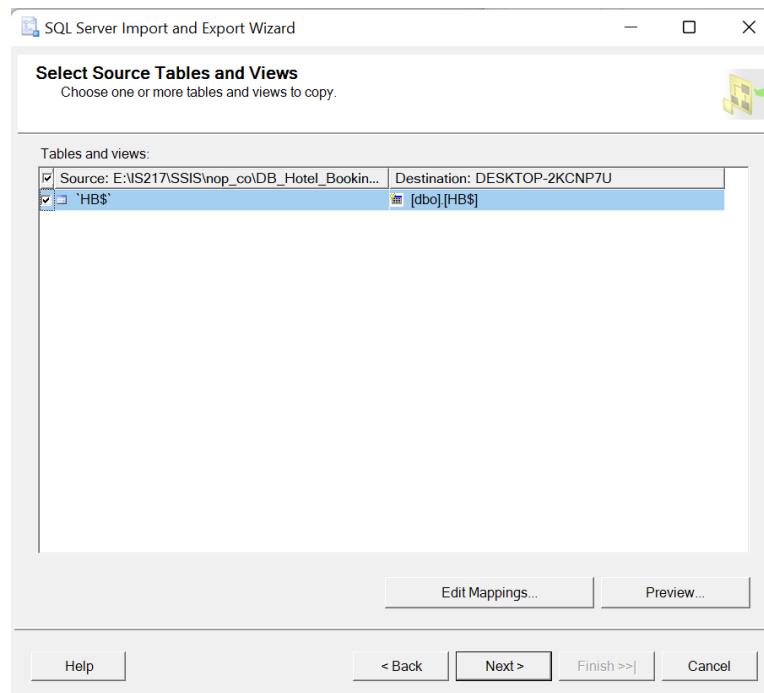
- + Chọn **Copy data from one or more tables or views** → Next.



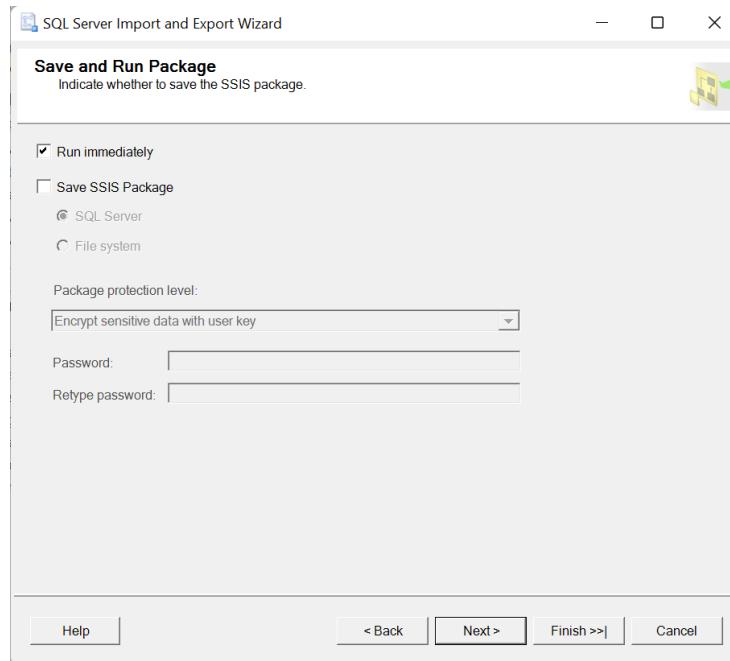
*Hình 2. 12.Chọn hình thức lấy dữ liệu*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

+ Chọn HB\$ tương ứng với sheet HB trong file excel đã dùng → Next.



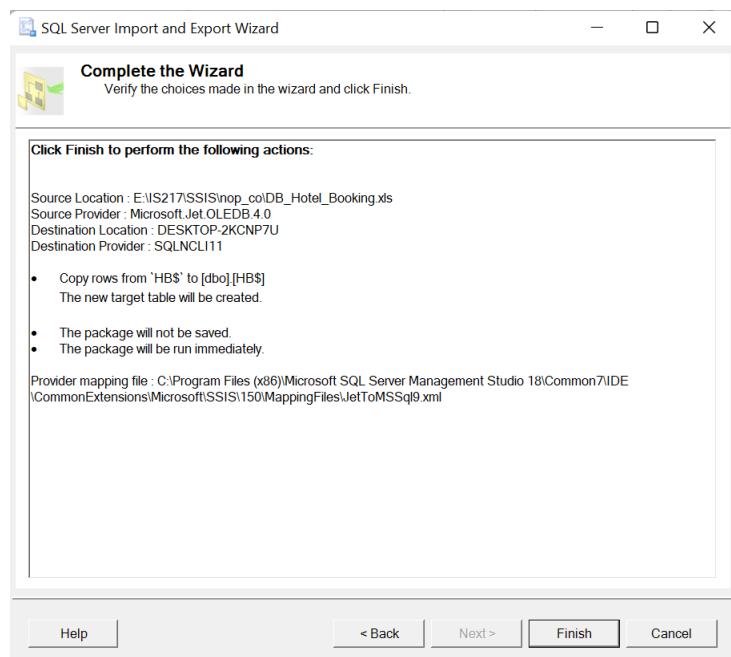
*Hình 2. 13. Chọn Source Tables và Views*



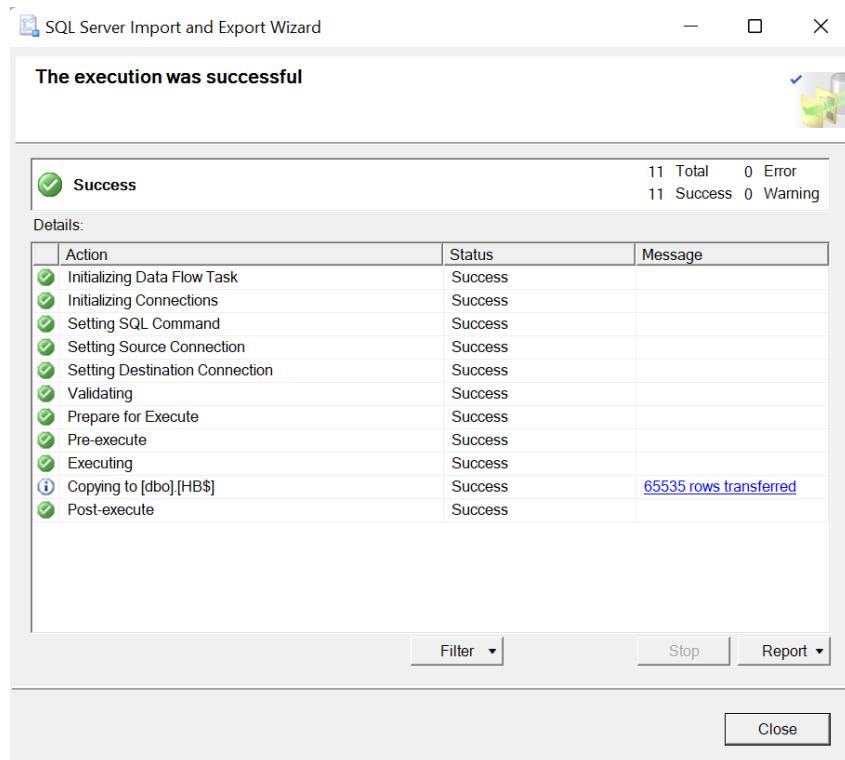
*Hình 2. 14. Lưu các thiết lập*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

+ Sau đó, nhấn Finish để tiến hành đổ dữ liệu vào database DB\_HB\_GOC.

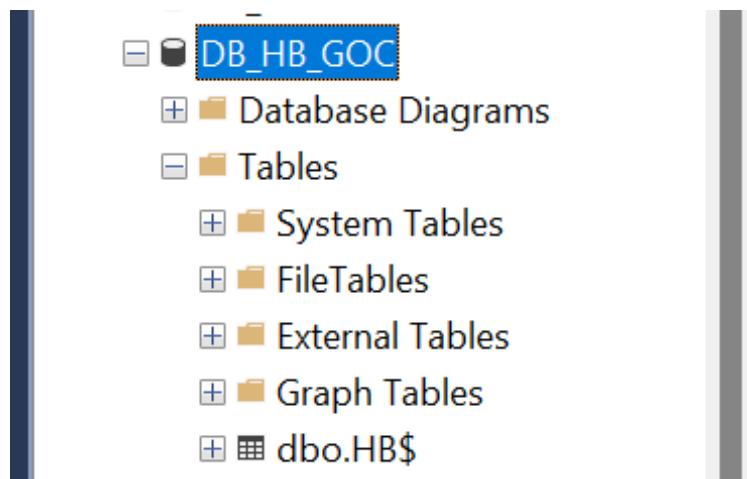


Hình 2. 15. Hoàn tất quá trình thiết lập



Hình 2. 16. Cửa sổ thông báo đổ dữ liệu thành công

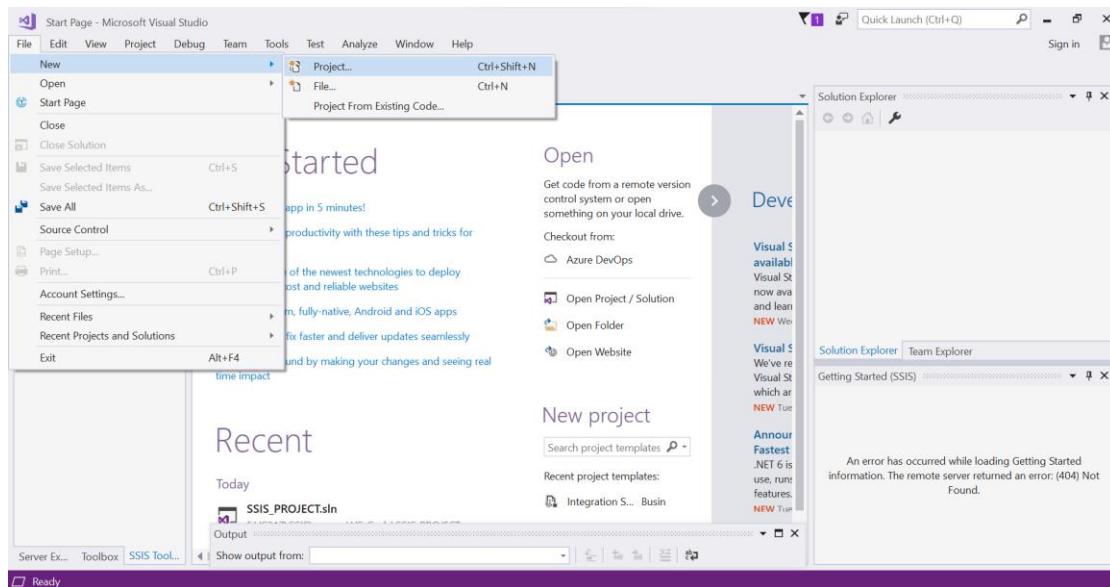
- + Sau khi import dữ liệu thành công, ta sẽ có được bảng [dbo].HB\$.



*Hình 2. 17. Bảng [dbo].HB\$ được tạo trong DB\_HB\_GOC khi đổ dữ liệu thành công*

### **3. Tạo project SSIS trong Visual Studio 2017.**

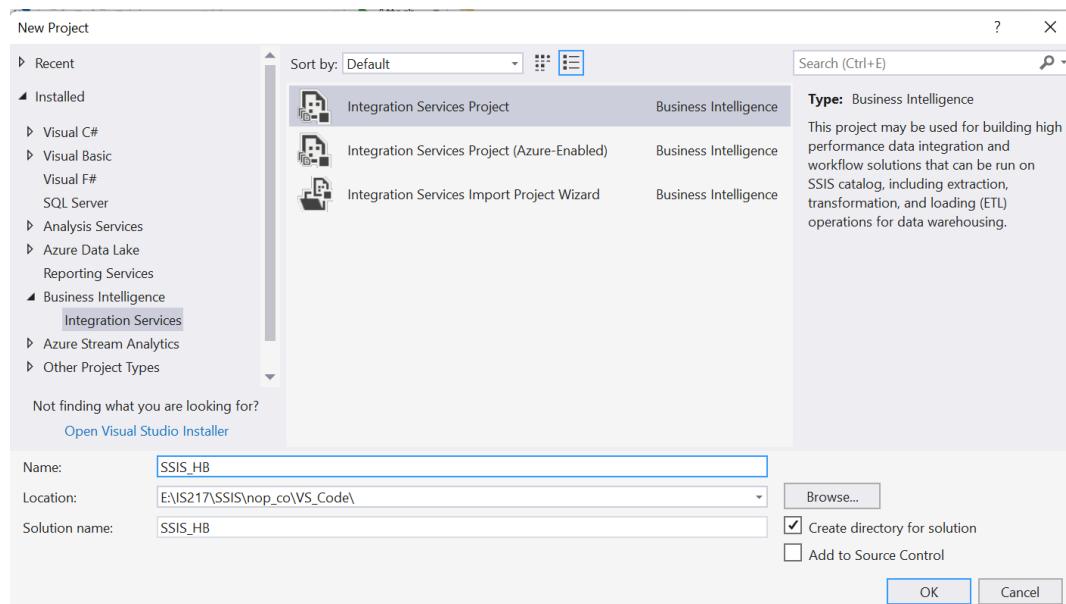
- Bước 1: Chọn File → New → Project để mở giao diện sau:



*Hình 2. 18. Tạo Project bằng công cụ BI trong Visual Studio 2017*

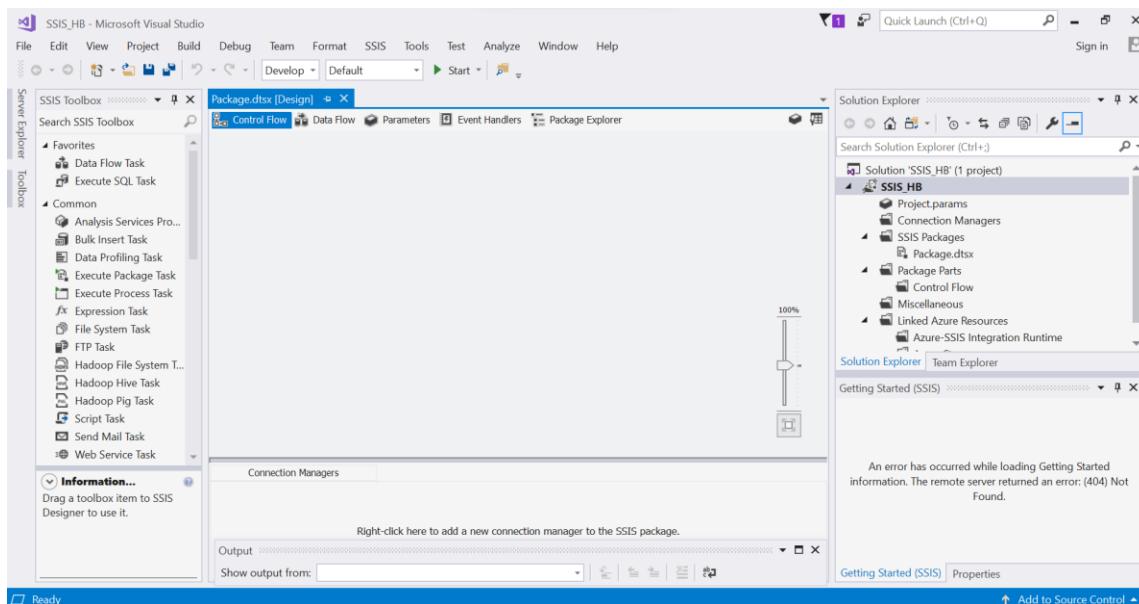
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 2: Chọn Business Intelligence. Sau đó, chọn công cụ integration Services project chọn đường dẫn để lưu project và đặt tên là SSIS\_HB.



Hình 2. 19.. Tạo Project bằng công cụ BI trong Visual Studio 2017

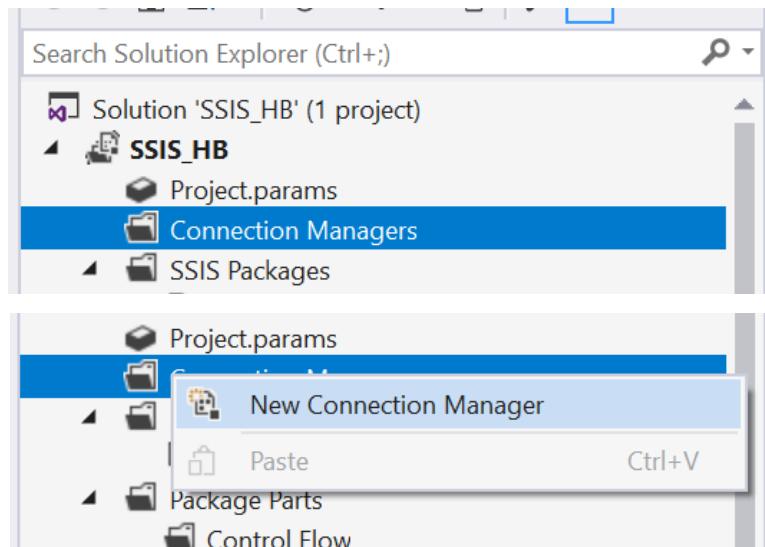
- Bước 3: Sau đó, nhấn OK và giao diện quá trình SSIS của công cụ BI sẽ hiện ra.



Hình 2. 20.Giao diện của project SISS sau khi tạo thành công

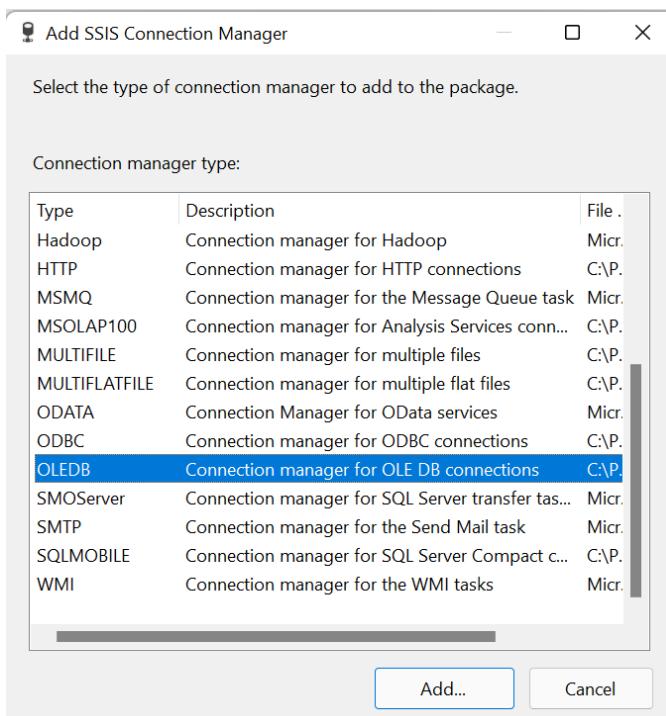
- + Các chức năng chính của SSIS nằm ở cột bên trái SSIS ToolBox.

- Bước 4: Tạo kết nối với cơ sở dữ liệu trong SQL Server.
- + Tại cửa sổ Solution Explorer, chọn Connection Managers → New Connection Manager



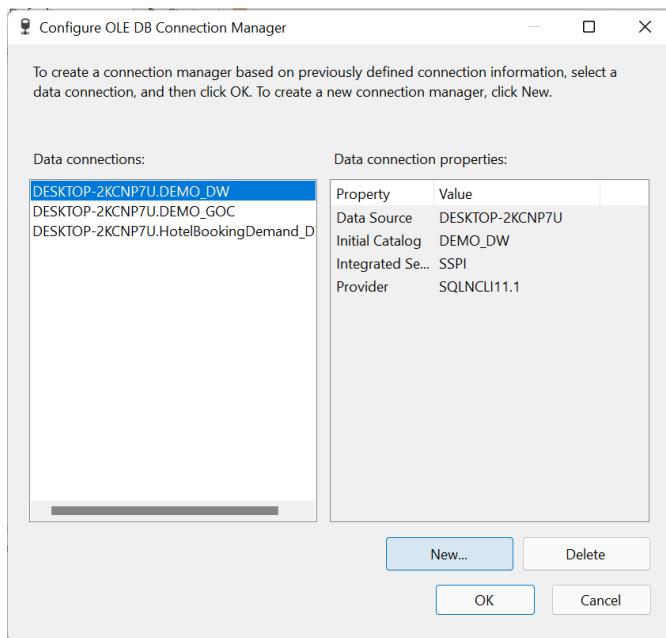
*Hình 2. 21. Tạo kết nối với cơ sở dữ liệu*

- + Chọn OLEDB → Add.



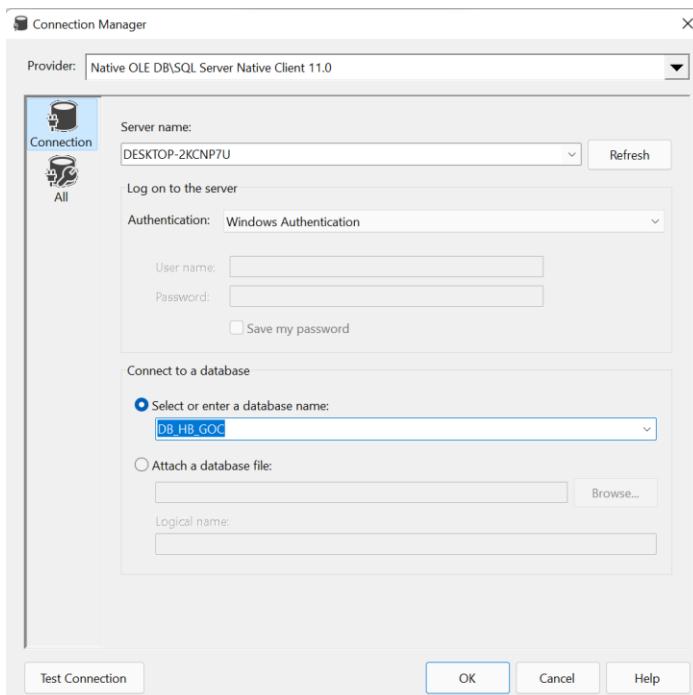
*Hình 2. 22. Thêm kết nối OLEDB*

- + Sau khi xuất hiện giao diện Configure OLE Connection Manager, chọn New:



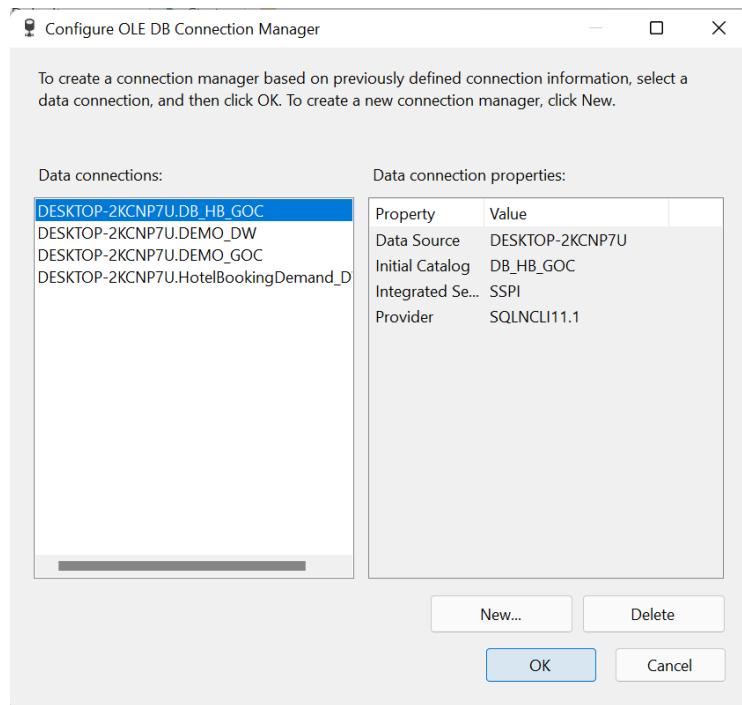
*Hình 2. 23. Thiết lập kết nối*

- + Nhập Server name và chọn Database name được dùng để tách dữ liệu và click OK.



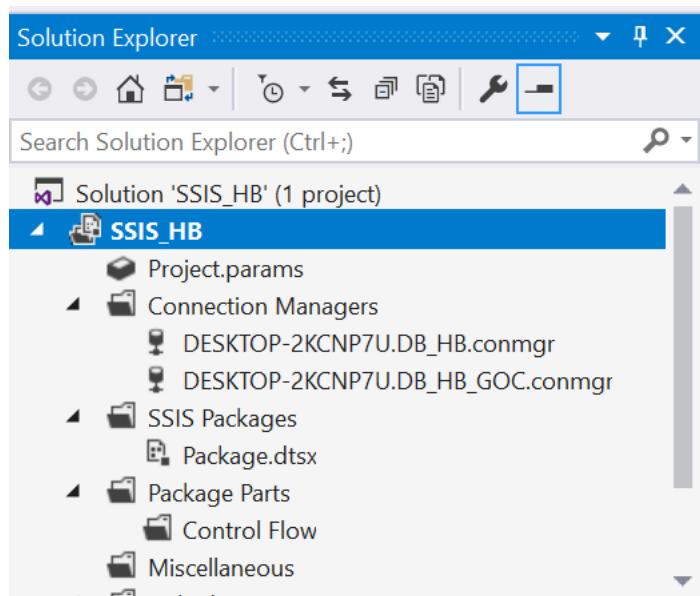
*Hình 2. 24.Nhập Server name và Database name*

- + Tiếp tục nhấn Ok để hoàn thành kết nối với cơ sở dữ liệu.



*Hình 2. 25. Hoàn thành kết nối với cơ sở dữ liệu*

- + Tiếp tục tạo kết nối với cơ sở dữ liệu DB\_HB, ta được:



*Hình 2. 26. Kết nối thành công với cơ sở dữ liệu*

#### 4. Lọc và làm sạch dữ liệu

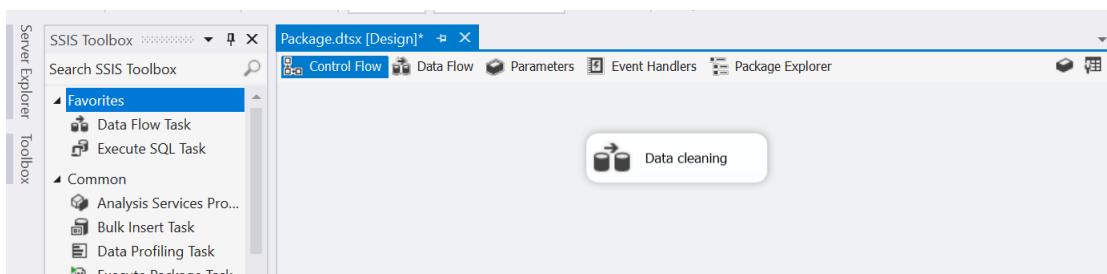
##### ❖ Các khái niệm:

- Trong SSIS có hai thành phần chính:
  - + *Data Transformation Run-time engine* : quản lý các control flow và package.
  - + *DataFlow engine*: Data Source, Transformations, Destination.
- **Data Flow Task**: là một task chính trong các task của control flow. Nhiệm vụ chính của DataFlow là biến đổi dữ liệu. Có ba thành phần data flow trong SSIS:
  - + *Sources* : trích xuất dữ liệu từ kho dữ liệu như bảng và view trong cơ sở dữ liệu quan hệ, tệp và Analysis Services databases và là nguồn cung cấp dữ liệu để biến đổi.
  - + *Transformations* : thực hiện các tác vụ như cập nhật, tóm tắt, làm sạch, hợp nhất và phân phối dữ liệu, sửa đổi giá trị trong cột, tra cứu giá trị trong bảng, làm sạch dữ liệu và tổng hợp giá trị cột.
  - + *Destinations* : ghi dữ liệu từ data flow vào một kho dữ liệu cụ thể hoặc tạo tập dữ liệu trên bộ nhớ. Một data flow có thể có nhiều Destinations tải dữ liệu vào nhiều kho khác nhau.
- **OLE DB Source**: Được sử dụng để trích xuất dữ liệu từ các loại cơ sở dữ liệu quan hệ theo chuẩn OLE-DB bằng cách sử dụng bảng, view hoặc là câu lệnh SQL.
  - + Trình chỉnh sửa của OLE DB Source cho phép chúng ta tùy chỉnh hình thức truy cập dữ liệu khác nhau nhằm tải dữ liệu như lấy từ bảng, view đã tồn tại hoặc từ một bảng mới hoặc là kết quả của câu lệnh SQL.
  - + Sử dụng trình quản lý kết nối OLE DB để kết nối với nguồn dữ liệu mà từ đó nó trích xuất dữ liệu.
- **OLE DB Destination**:
  - + Được sử dụng để tải dữ liệu vào các cơ sở dữ liệu quan hệ theo chuẩn OLE-DB bằng cách sử dụng bảng, view hoặc là câu lệnh SQL.

- + Trình chỉnh sửa của OLE DB Destination cho phép chúng ta tùy chỉnh hình thức truy cập dữ liệu khác nhau nhằm tải dữ liệu như lấy từ bảng, view đã tồn tại hoặc từ một bảng mới hoặc là kết quả của câu lệnh SQL.
- + Bao gồm các ánh xạ giữa các cột dữ liệu đầu vào và các cột trong nguồn dữ liệu đích.
  - + Các dữ liệu của cột các ánh xạ phải tương thích với nhau.
  - + Nếu cột dữ liệu đích không cho phép giá trị null thì bắt buộc phải có một cột ở nguồn dữ liệu đầu vào ánh xạ đến cột dữ liệu đích này nếu không sẽ xảy ra lỗi.
- **Data conversion** : Chuyển đổi kiểu dữ liệu trong cột dữ liệu ban đầu thành một kiểu dữ liệu khác và sao chép dữ liệu đó sang cột dữ liệu đích.
- **Sort** : Sắp xếp dữ liệu đầu vào theo thứ tự tăng dần hoặc giảm dần và sao chép dữ liệu đã sắp xếp vào dữ liệu đầu ra. Ngoài ra, sort cũng loại bỏ các hàng trùng lặp như một phần của cách sắp xếp của nó.

❖ **Thực hiện lọc và làm sạch dữ liệu**

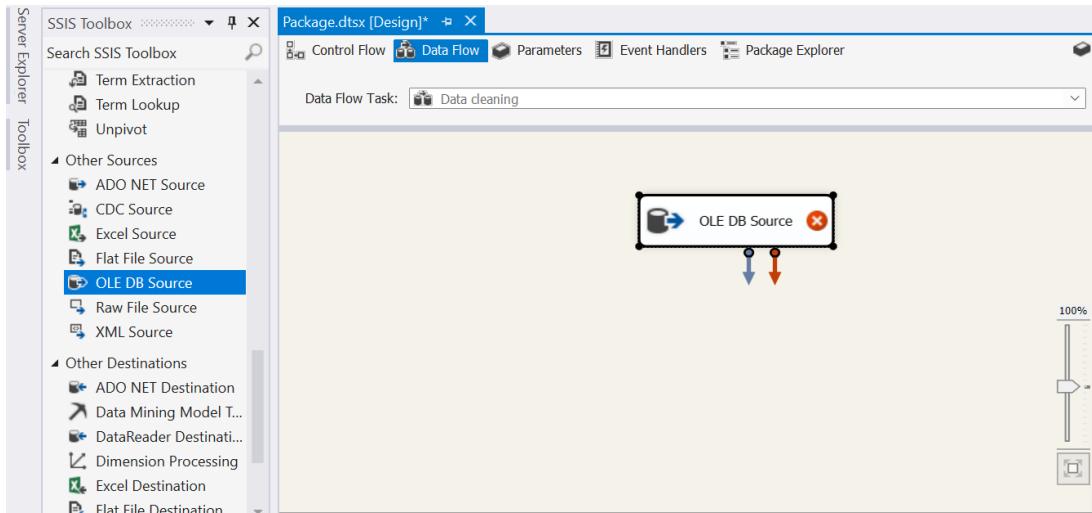
- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành Data Cleaning.



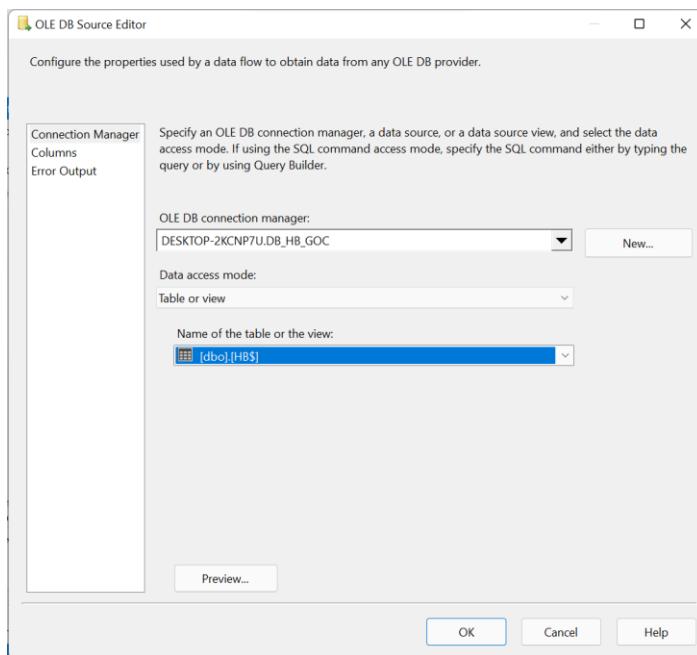
Hình 2. 27.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành Data Cleaning

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

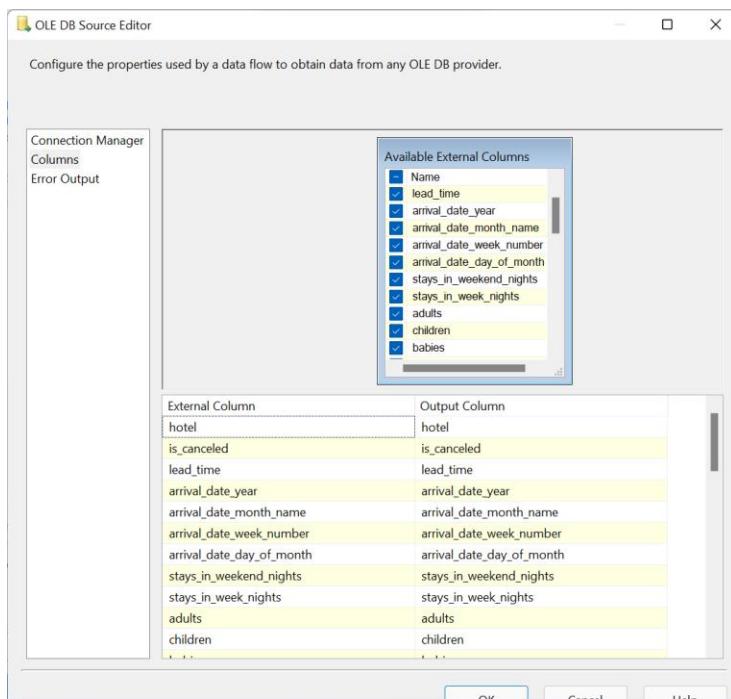
- Bước 2: Double click vào Data Cleaning để chuyển từ Control Flow sang Data Flow và chọn chức năng OLE DB Source và kết nối với DB\_HB\_GOC để nạp dữ liệu.



*Hình 2. 28. Nạp dữ liệu*

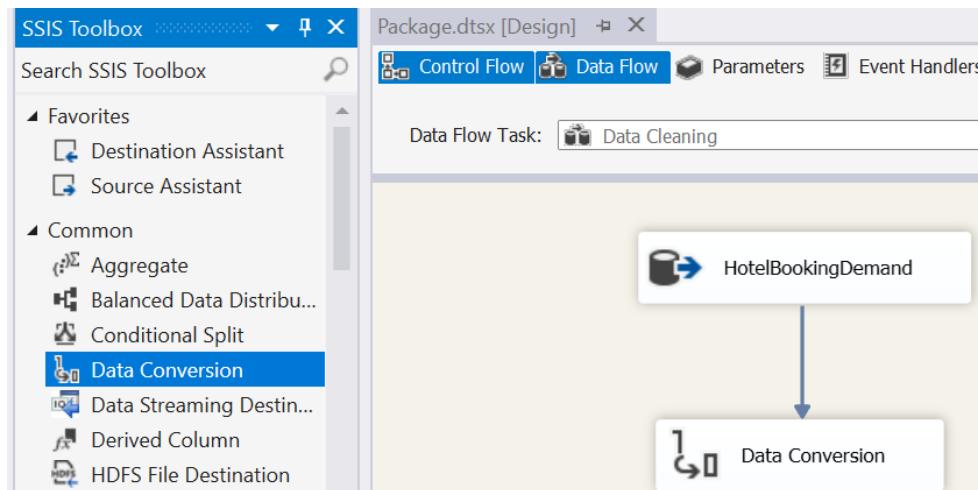


*Hình 2. 29. Chọn bảng*

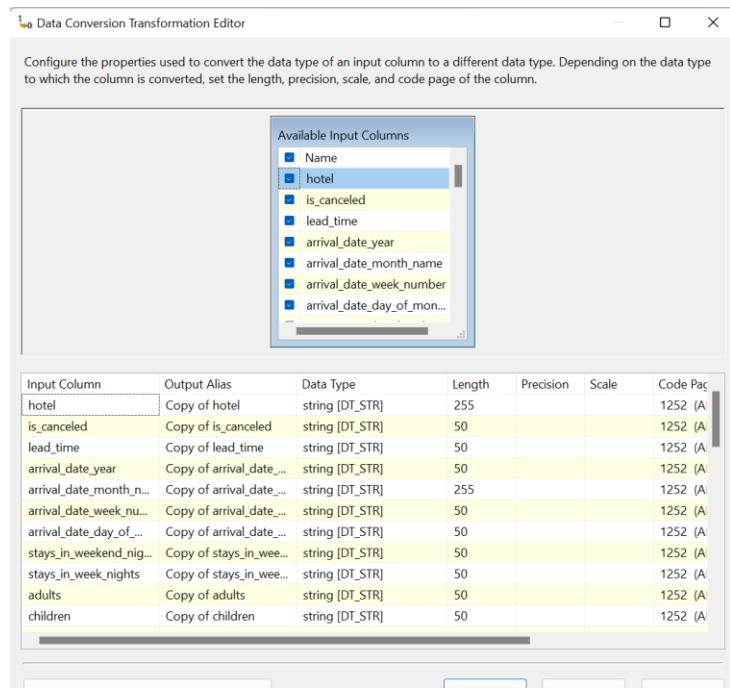


*Hình 2. 30. Các thuộc tính trong bảng dữ liệu được chọn*

- Bước 3: Tìm kiếm chức năng Data Conversion để chuyển tất cả các kiểu dữ liệu thành kiểu string để tiến hành phân tích ở bước sau.

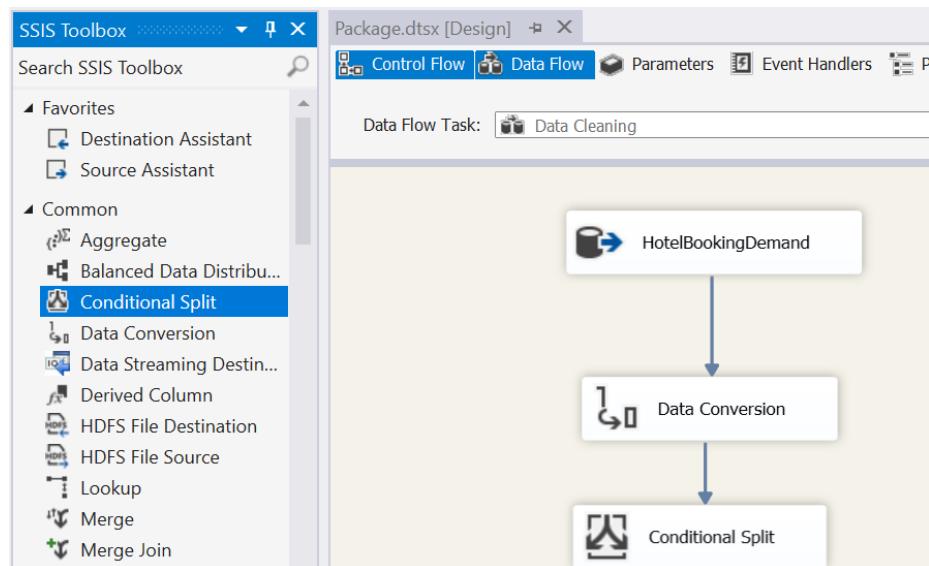


*Hình 2. 31. Chuyển kiểu dữ liệu bằng Data Conversion*



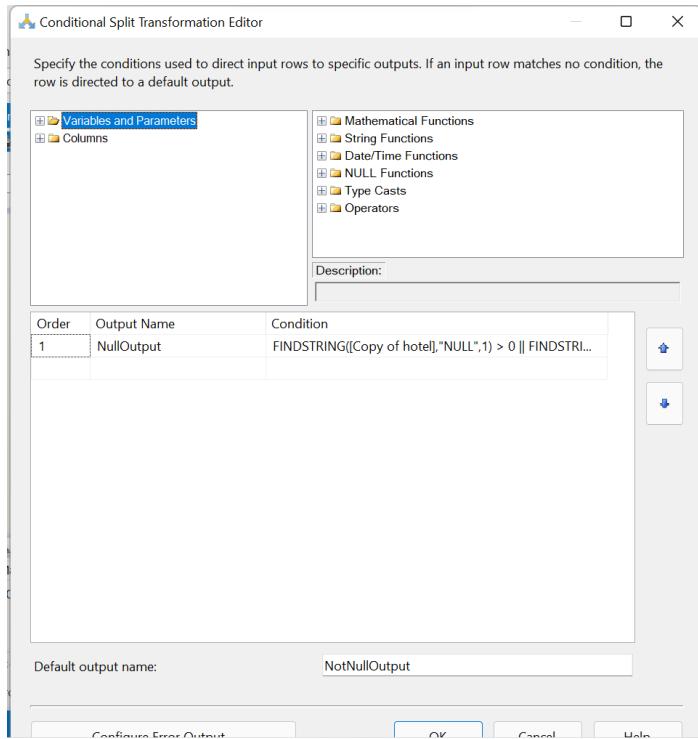
*Hình 2. 32. Thuộc tính của bảng sau khi chuyển đổi kiểu dữ liệu sang string*

- Bước 4: Dùng công cụ Conditional Split bắt đầu cắt dữ liệu có điều kiện, và lọc những dòng NULL ra khỏi database trước khi đưa vào kho dữ liệu.



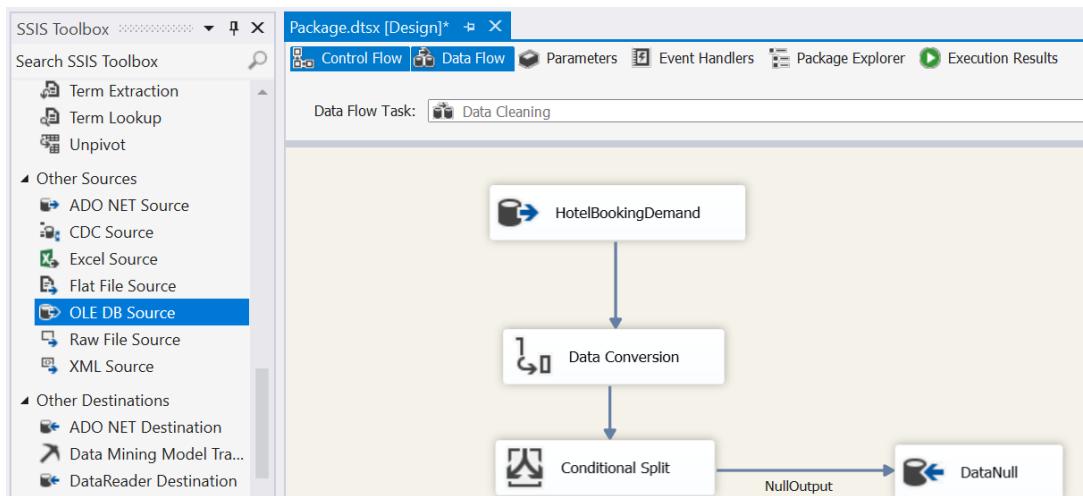
*Hình 2. 33. Cắt dữ liệu bằng Conditional Split*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



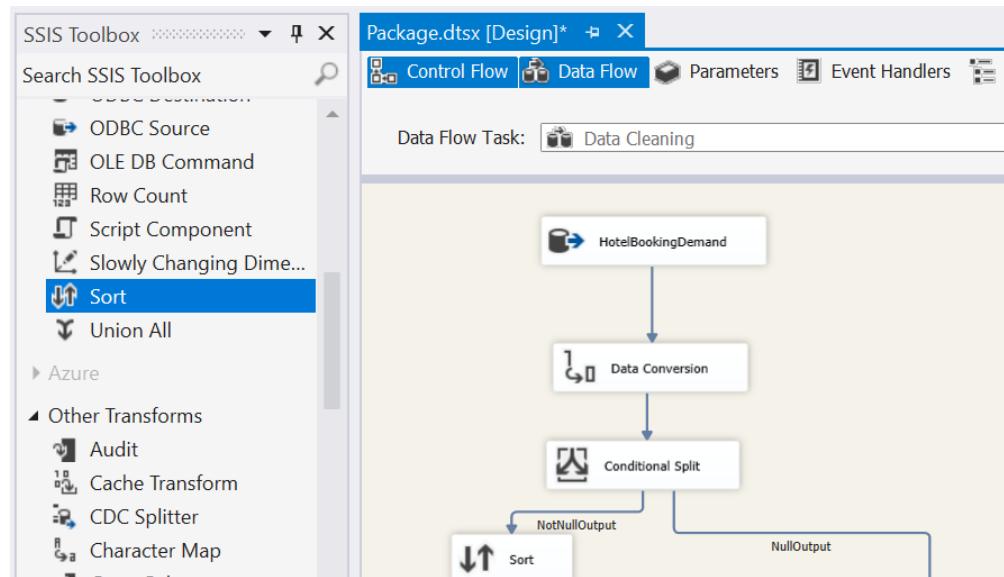
*Hình 2. 34. Lọc những dòng NULL trong dữ liệu*

- Bước 5: Kéo thả công cụ OLE DB Destination và đổi tên thành DataNull, kết nối với Condition Split bằng NullOutput.

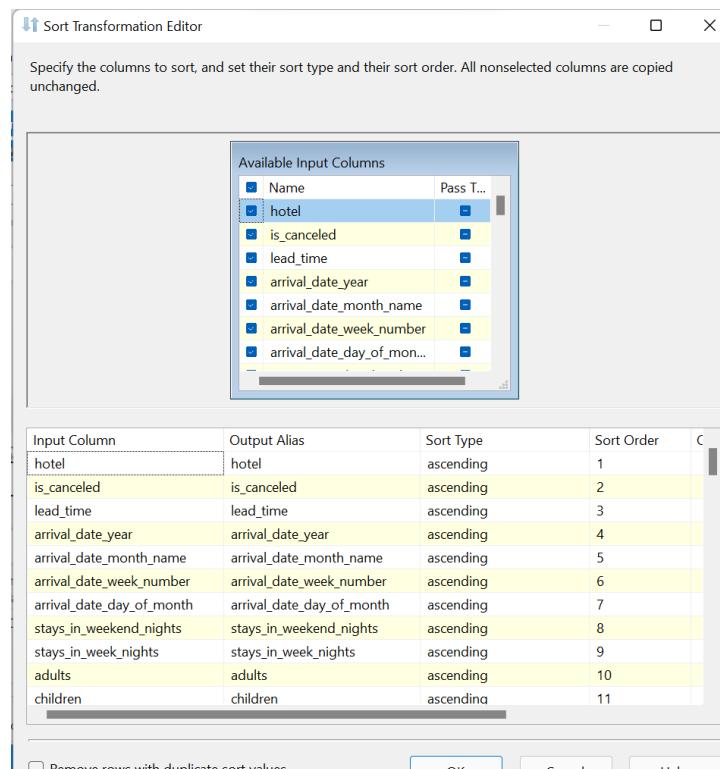


*Hình 2. 35. Kéo thả OLE DB Destiantion vào màn hình làm việc và đổi tên thành DataNull*

- Bước 6: Chọn công cụ Sort để tiến hành sắp xếp dữ liệu sẽ dùng để tách thành các bảng Dimensions và bảng Fact.

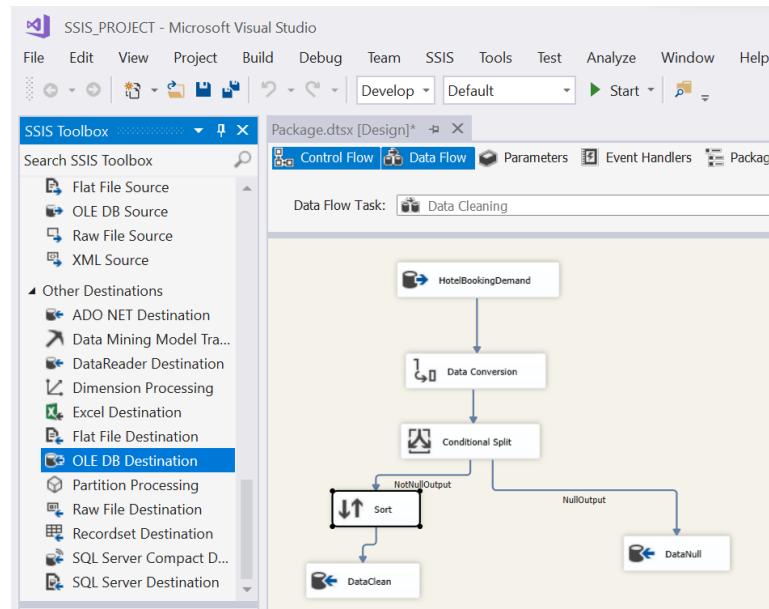


*Hình 2. 36. Dùng công cụ Sort để sắp xếp dữ liệu*



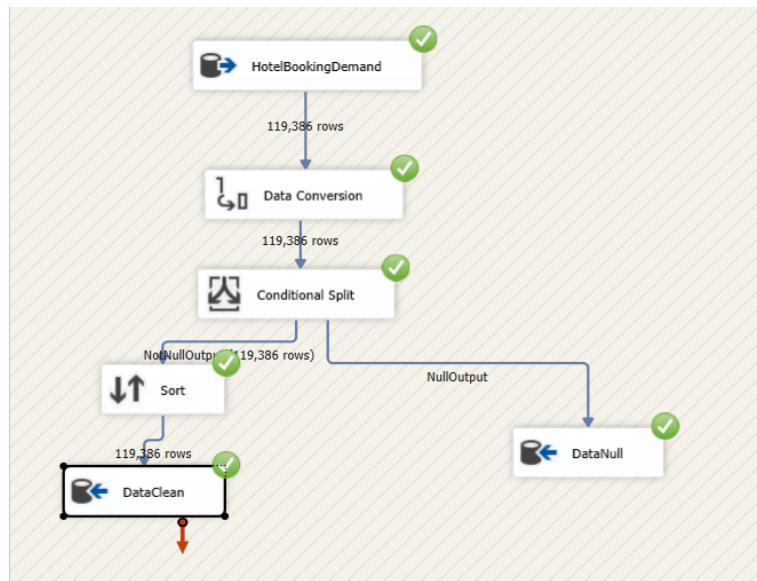
*Hình 2. 37. Thiết lập điều kiện sắp xếp dữ liệu*

- Bước 7: Nối Sort và OLE DB Source lại với nhau, để tạo bảng DataClean cho cơ sở dữ liệu DB\_HB\_GOC.



*Hình 2. 38. Tạo kết nối giữa Sort và OLE DB Source*

- Bước 8: Chọn Start để bắt đầu quá trình đồ dữ liệu. Khi đồ thành công, ta có kết quả như sau:



*Hình 2. 39. Đồ dữ liệu thành công*

- Bước 9: Kiểm tra bảng DataNull và DataClean trên SQL Server.

hotel	is_cancelled	lead_time	arrival_date_year	arrival_date_month_name	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights
1 City Hotel	0	56	2017	March	13	31	0
2 City Hotel	0	56	2017	March	13	31	0
3 City Hotel	0	56	2017	March	13	31	0
4 City Hotel	0	56	2017	March	13	31	0
5 City Hotel	0	56	2017	March	13	31	0
6 City Hotel	0	56	2017	May	18	4	2
7 City Hotel	0	56	2017	May	18	5	0
8 City Hotel	0	56	2017	May	19	11	0
9 City Hotel	0	56	2017	May	20	17	0
10 City Hotel	0	57	2015	December	52	25	2
11 City Hotel	0	57	2015	December	52	26	2
12 City Hotel	0	57	2015	December	52	26	4
13 City Hotel	0	57	2015	December	53	29	1
14 City Hotel	0	57	2015	December	53	30	0

Hình 2. 40.Kiểm tra bảng DataNull và DataClean trên SQL Server

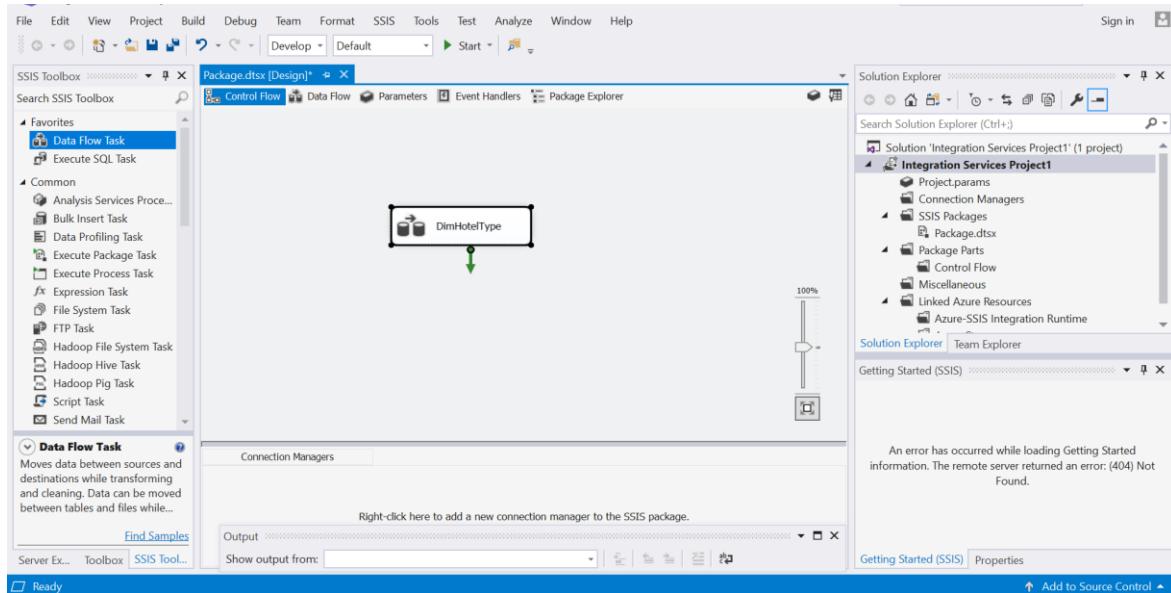
## 5. Tạo bảng Dim và bảng Fact.

### ❖ Các khái niệm :

- **Script component** : Script có thể được sử dụng như một source, một transformation, hoặc một destination. Thành phần này hỗ trợ một đầu vào và nhiều đầu ra :
  - + Nếu được sử dụng như một source thì Script component hỗ trợ nhiều đầu ra.
  - + Nếu được sử dụng như tranformation thì Script component hỗ trợ một đầu vào và nhiều đầu ra.
  - + Nếu được sử dụng như destination thì Script component hỗ trợ một đầu vào
- **Aggregate**: Được sử dụng Để thực hiện các phép tính tổng hợp trên các nhóm dữ liệu như Group by, Sum ,Average, Count, Count Distinct, Minimum, Maximum,...
- **Lookup** : Thực hiện tra cứu bằng cách nối dữ liệu trong các cột dữ liệu đầu vào với các cột trong tập dữ liệu tham chiếu thường để truy cập thông tin bổ sung trong bảng có liên quan dựa trên các giá trị trong các cột chung.

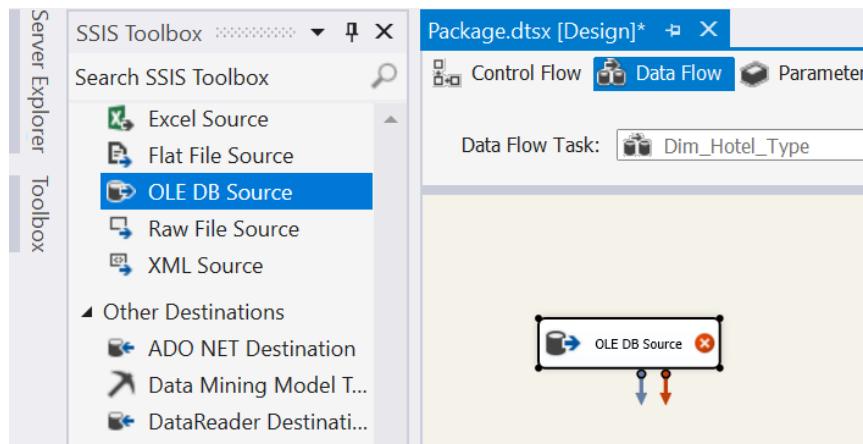
### 5.1. Bảng DimHotelType

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimHotelType.



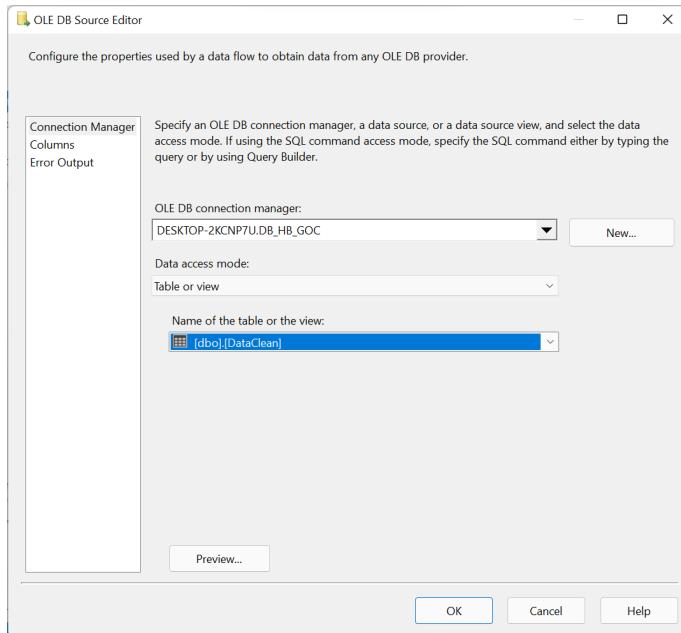
Hình 2. 41.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimHotelType

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



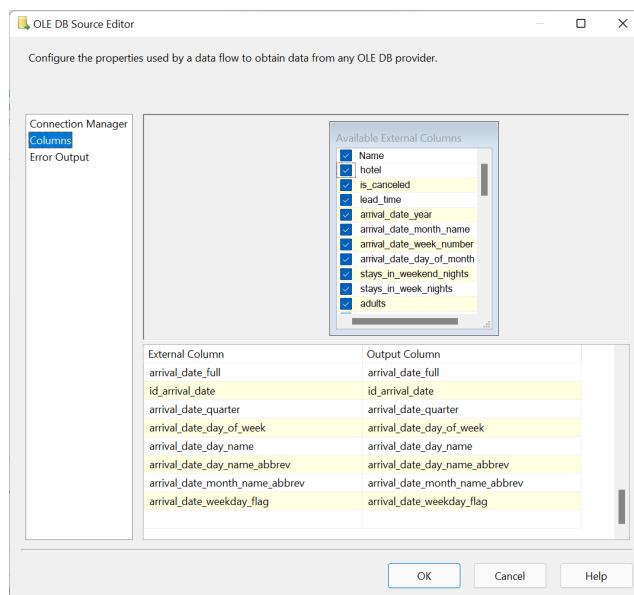
Hình 2. 42. Kéo chức năng OLE DB Source vào màn hình làm việc

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



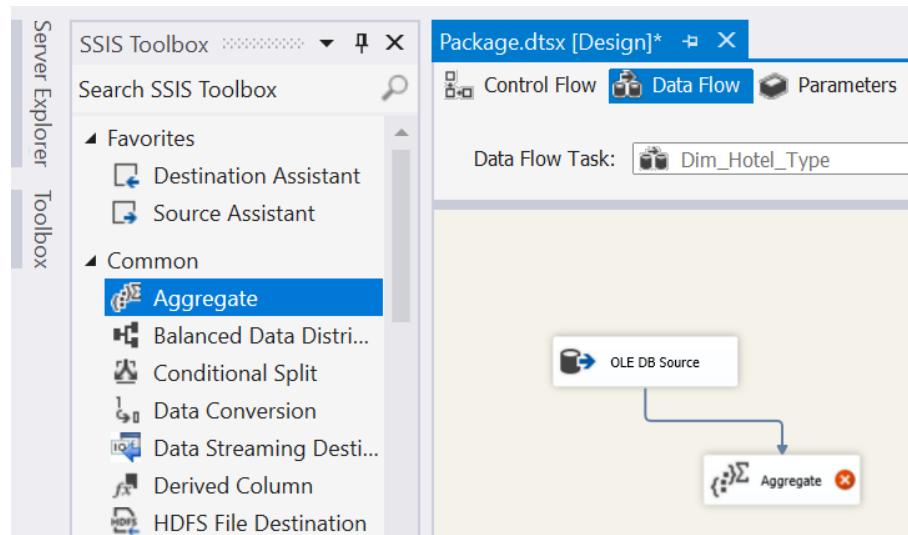
*Hình 2. 43. Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.

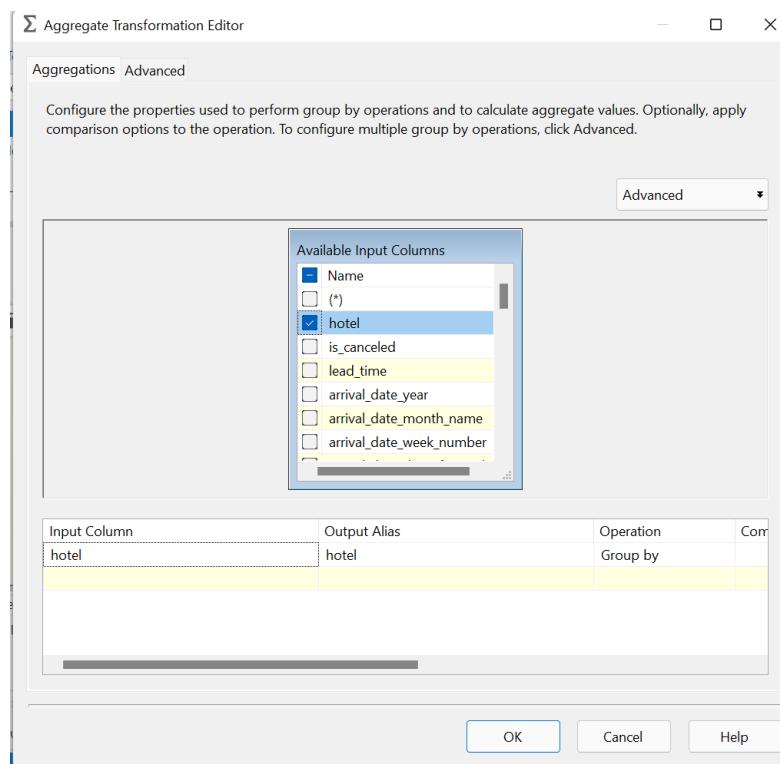


*Hình 2. 44. Chọn các columns cần sử dụng*

- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là hotel.

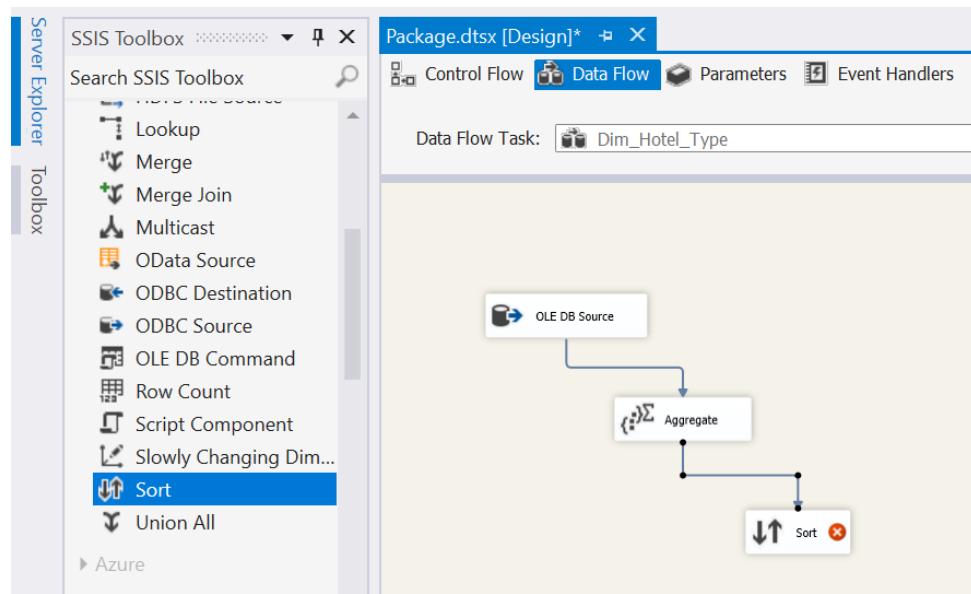


*Hình 2. 45. Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau*

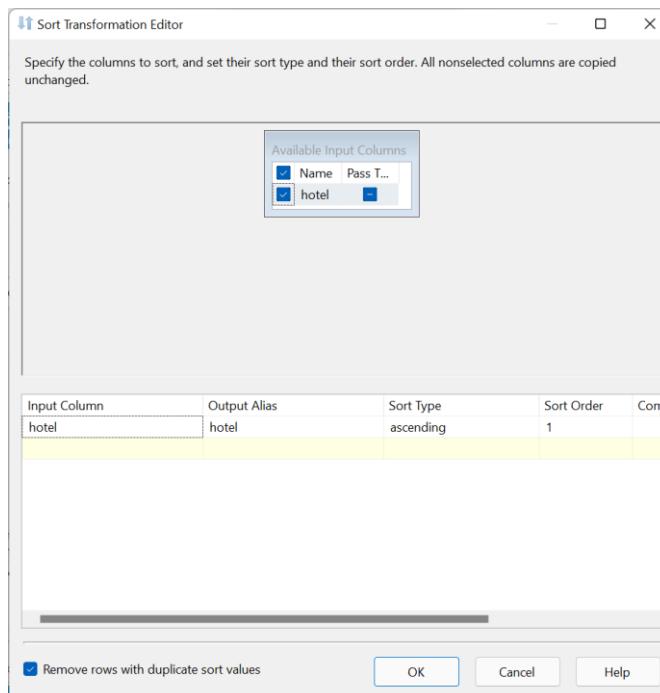


*Hình 2. 46. Chọn các thuộc tính cần gom nhóm*

- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.



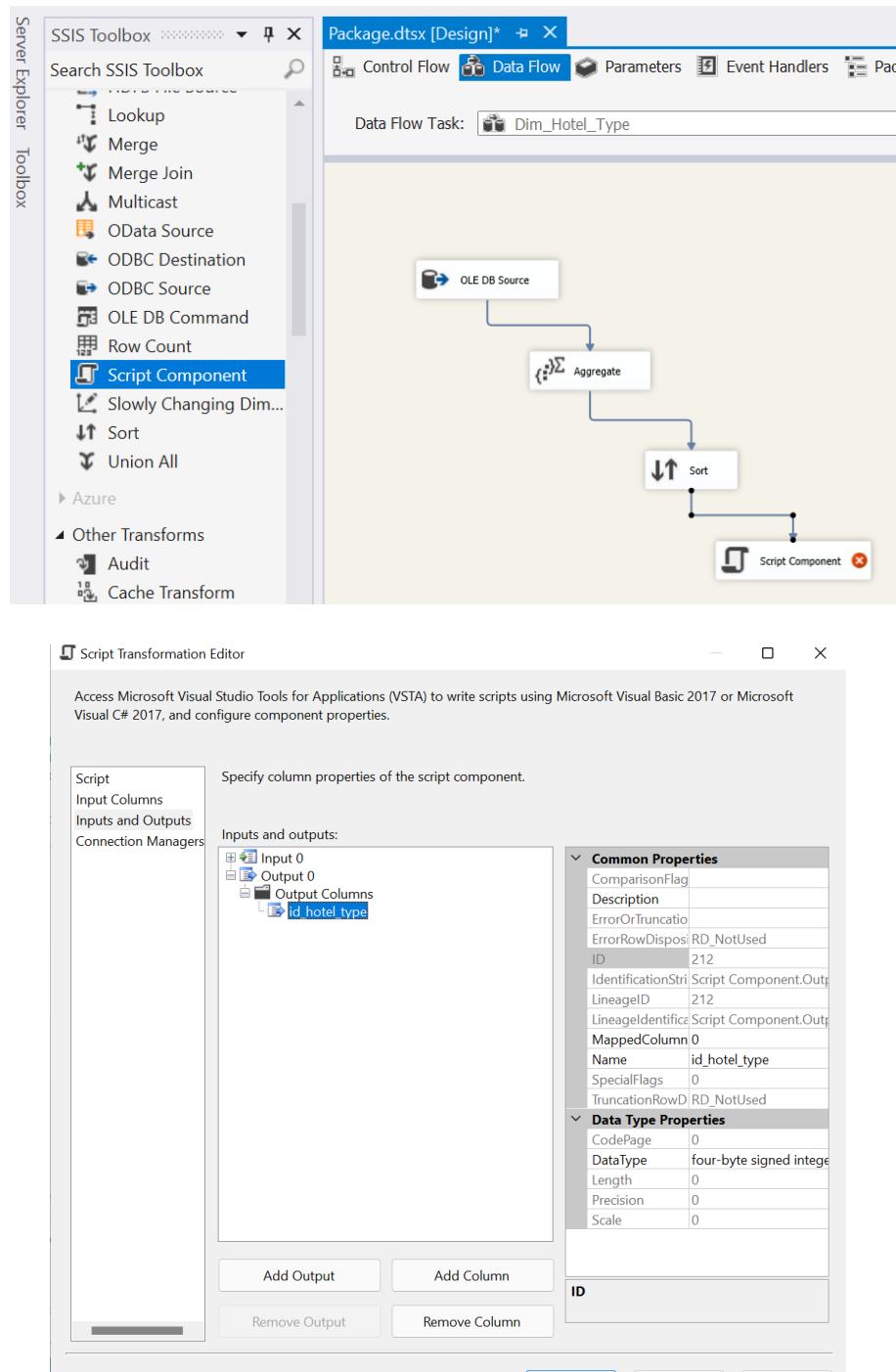
*Hình 2. 47. Sắp xếp các giá trị theo chiều tăng dần*



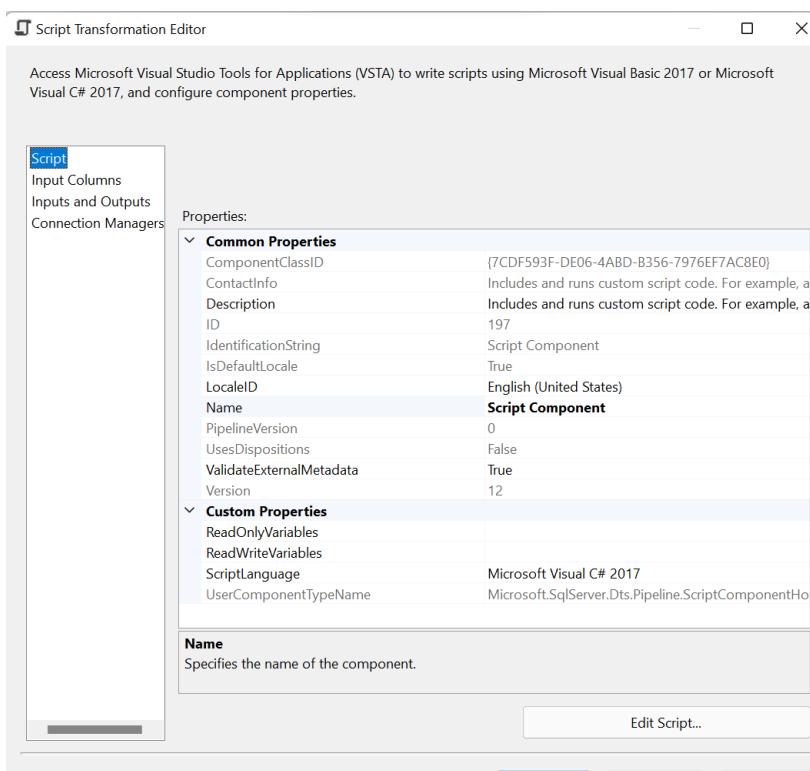
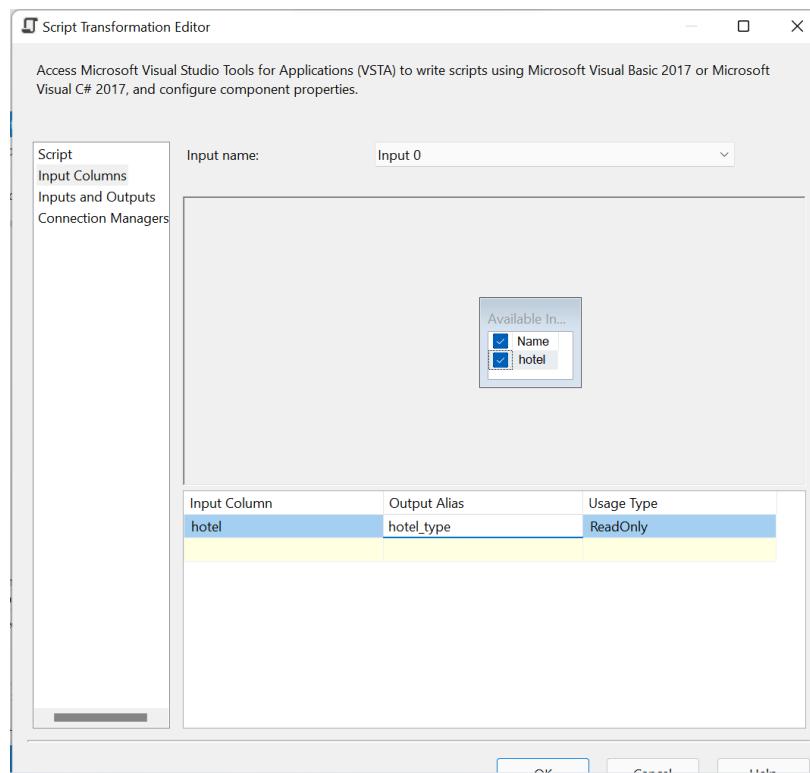
*Hình 2. 48. Sắp xếp các giá trị theo chiều tăng dần*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_hotel\_type và đổi tên thuộc tính từ hotel sang hotel\_type.



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

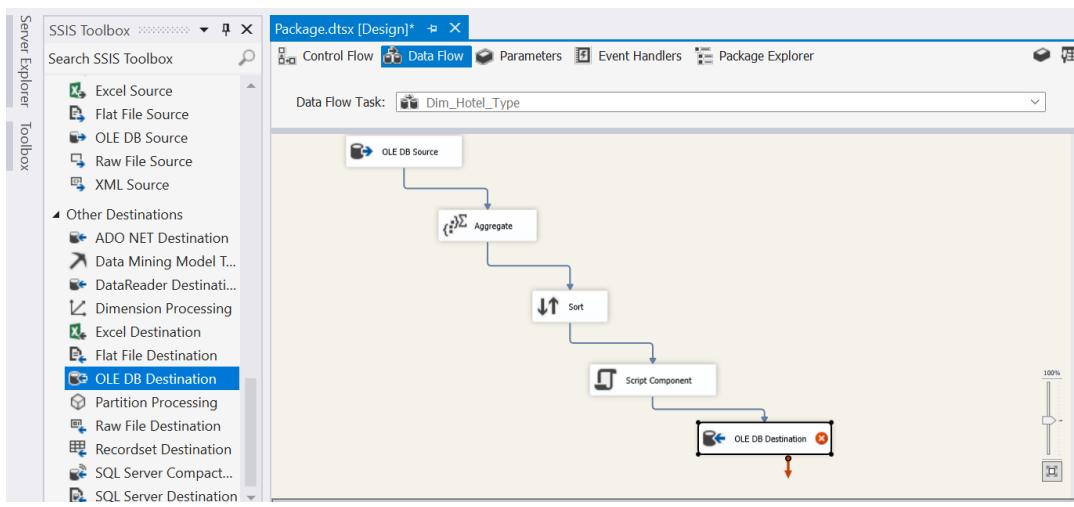
The screenshot shows the SSIS Script Component Editor with two code snippets:

```
public class ScriptMain : UserComponent
{
    Help: Using Integration Services variables and parameters
    Help: Using Integration Services Connection Managers
    Help: Firing Integration Services Events
    int count = 1;
    /// <summary>
    /// This method is called once, before rows begin to be processed in the component.
    ///
    /// You can remove this method if you don't need to do anything here.
    /// </summary>
    public override void PreExecute()
    {
        base.PreExecute();
        /*
         * Add your code here
         */
    }
}

2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    Row.idhoteltypes = count;
    count++;
}
```

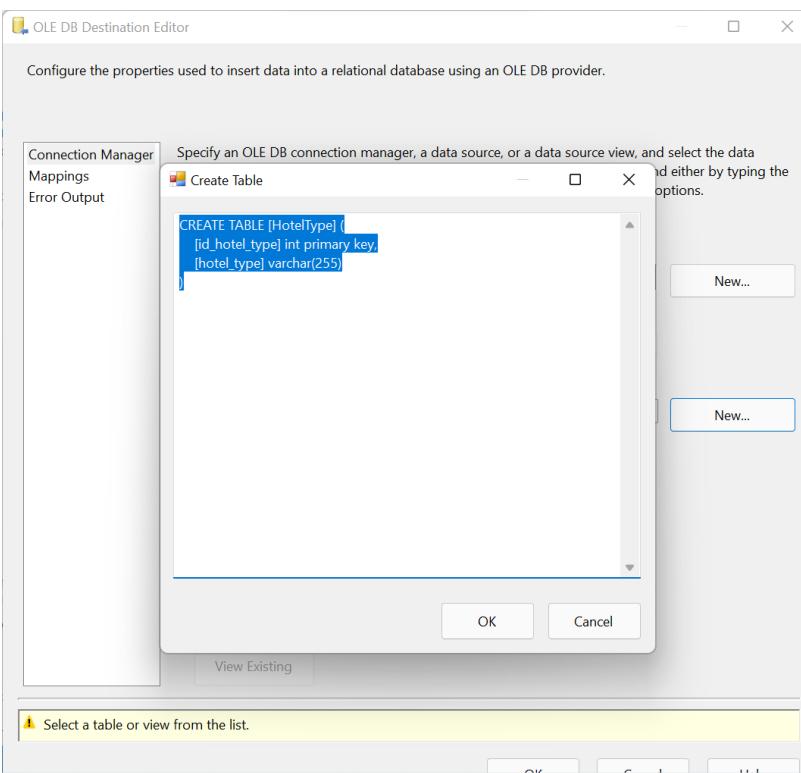
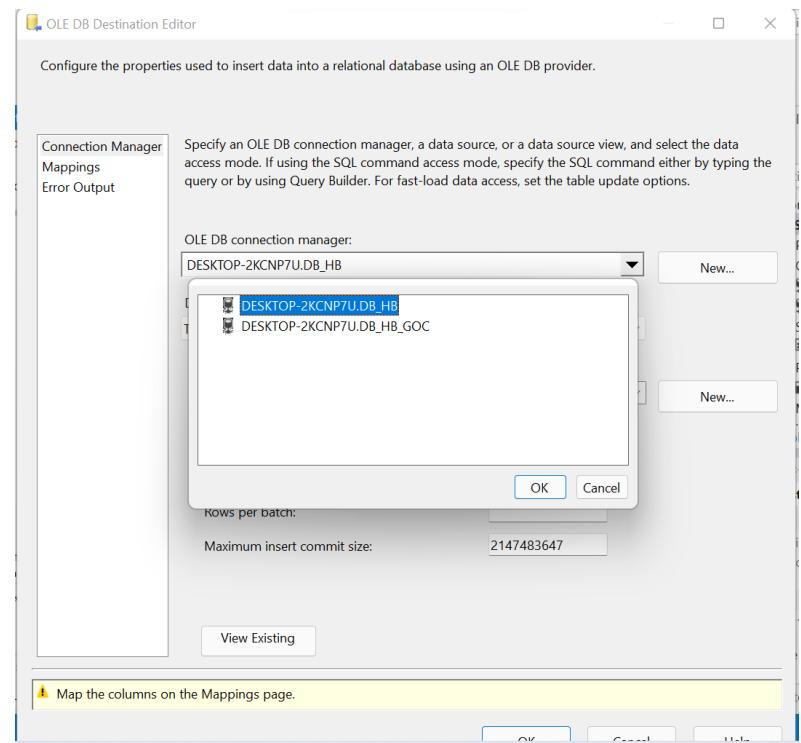
Hình 2. 49.1.,2.,3.,4.,5.,6. Sử dụng Script Component để tạo khóa chính id\_hotel\_type và đổi tên thuộc tính từ hotel sang hotel\_type

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là HotelType.

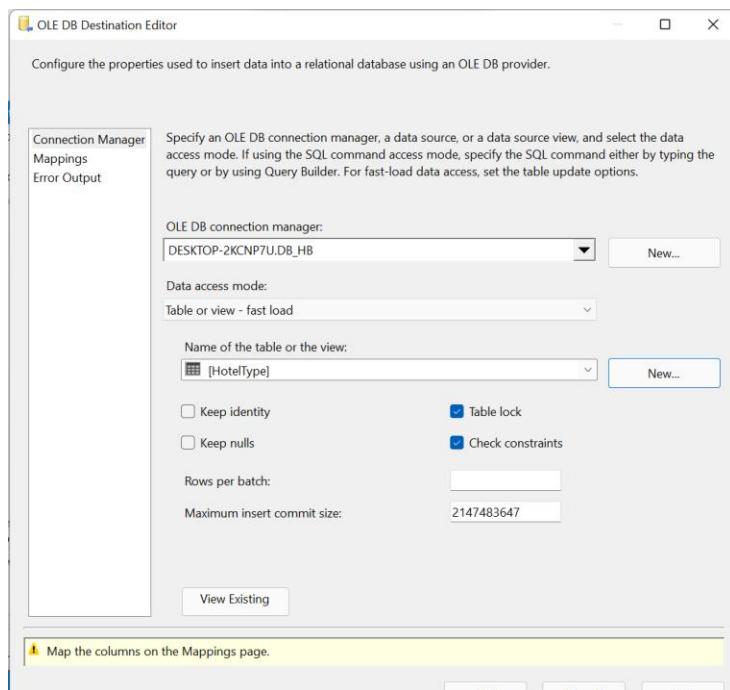


Hình 2. 50. Sử dụng OLE DB Destination để tạo bảng

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

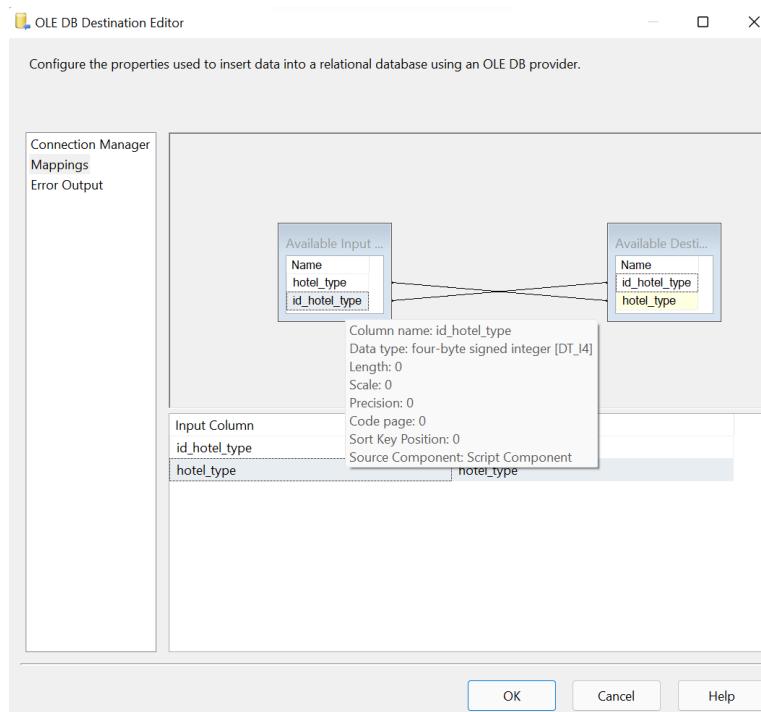


## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



*Hình 2. 51.1.,2.,3. Tạo bảng Hotel Type*

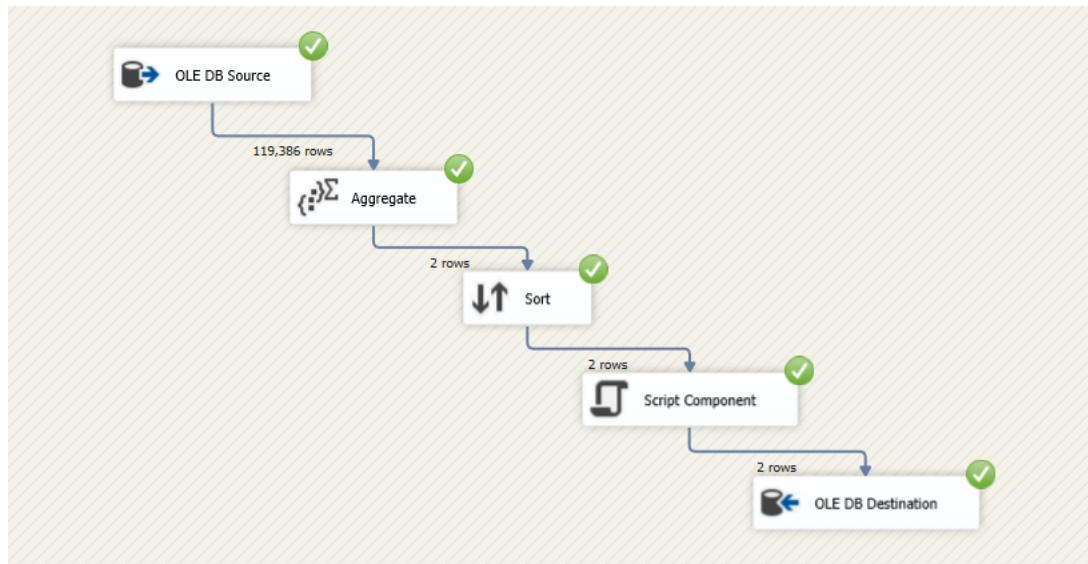
- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



*Hình 2. 52. Quá trình Mappings dữ liệu bảng HotelType*

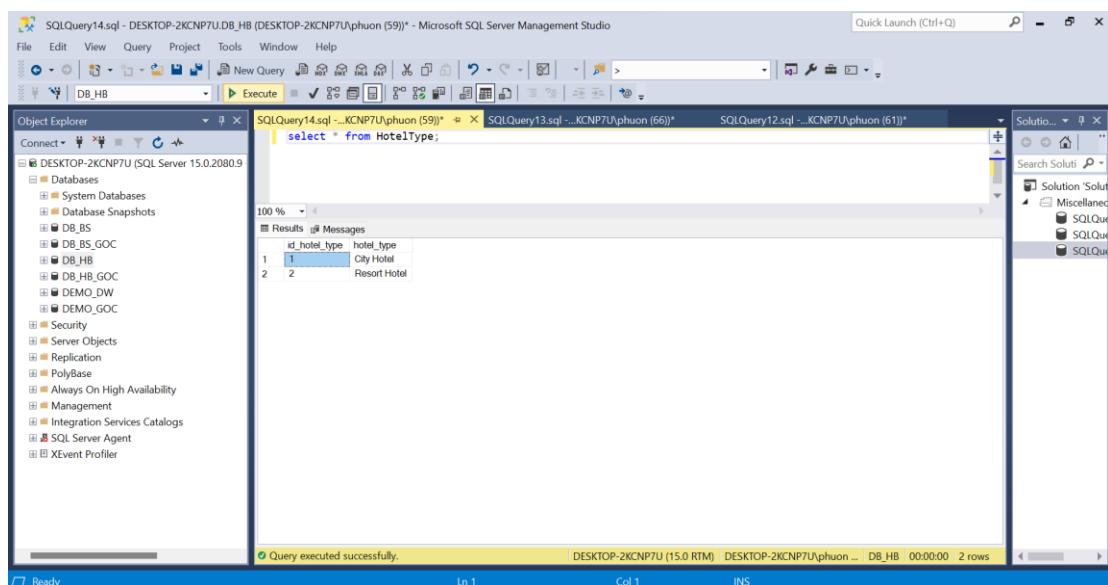
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 53.Hoàn thành đổ dữ liệu vào HotelType trong kho dữ liệu*

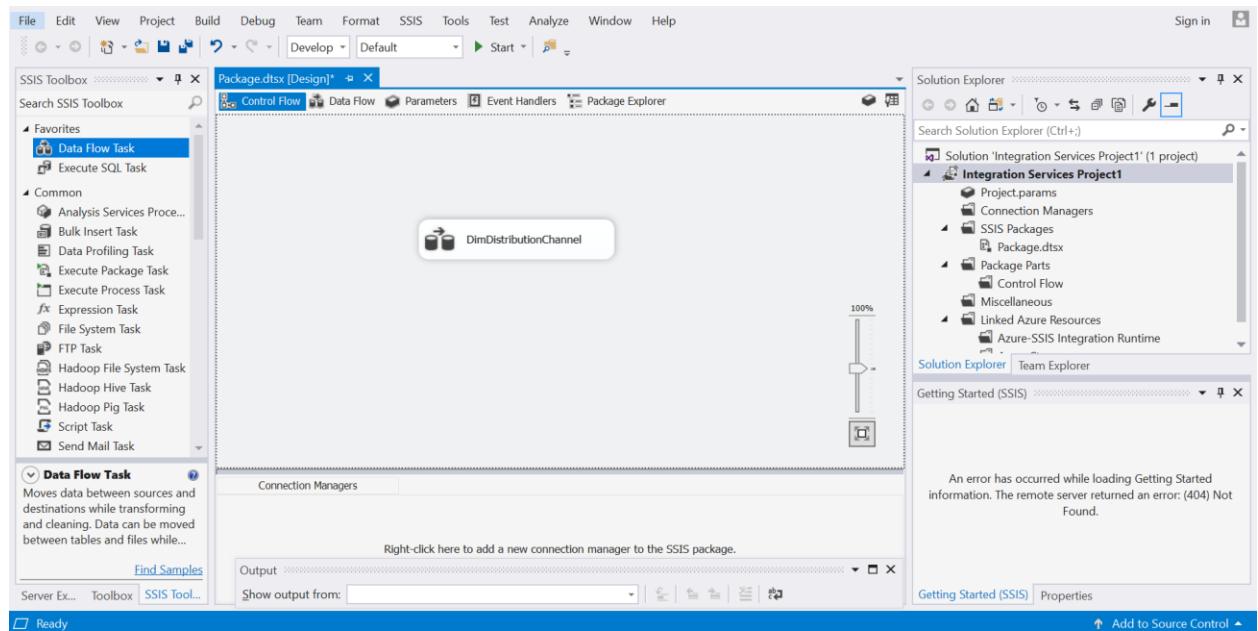
- Bước 11: Kiểm tra bảng HotelType trên SQL Server.



*Hình 2. 54.Kiểm tra bảng HotelType trên SQL Server*

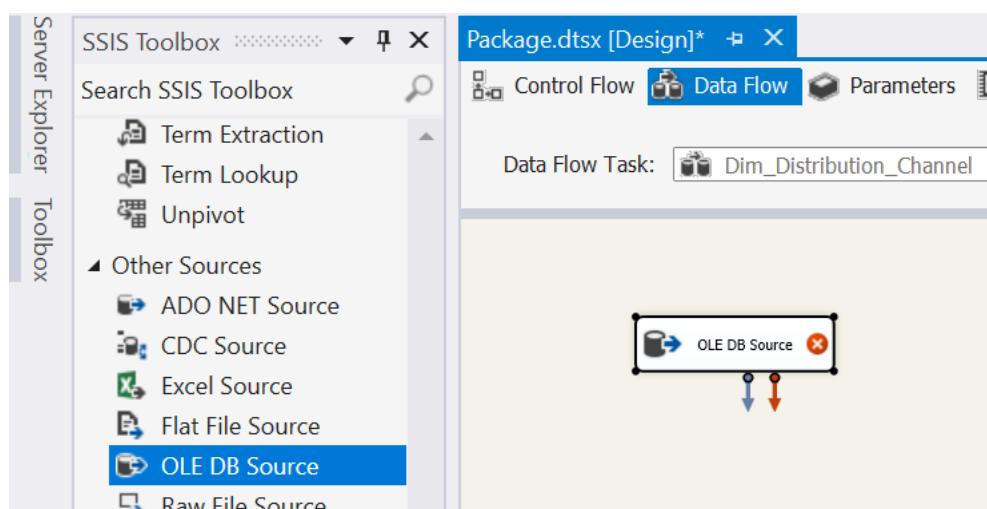
## 5.2. Bảng DimDistributionChannel

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimDistributionChannel.



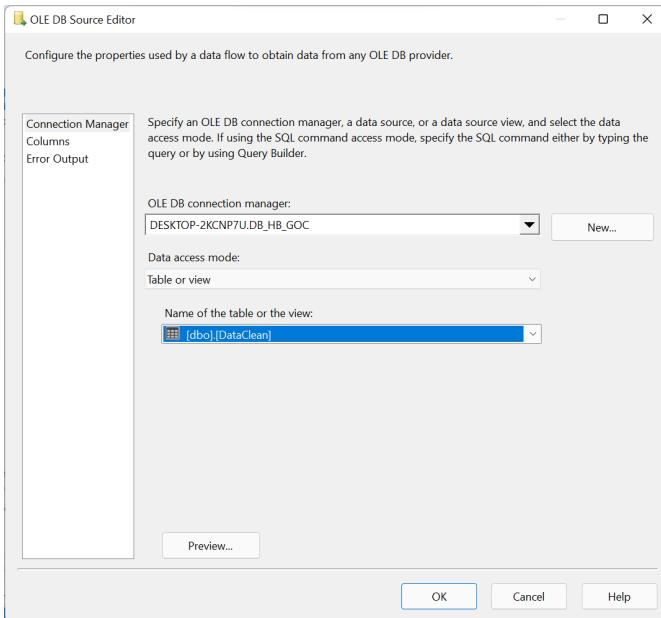
Hình 2. 55.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimDistributionChannel

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



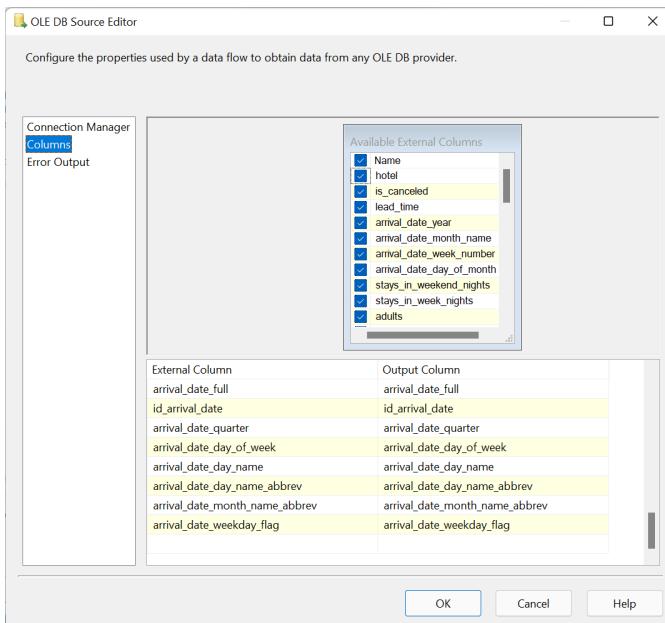
Hình 2. 56.Kéo chức năng OLE DB Source vào màn hình làm việc

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



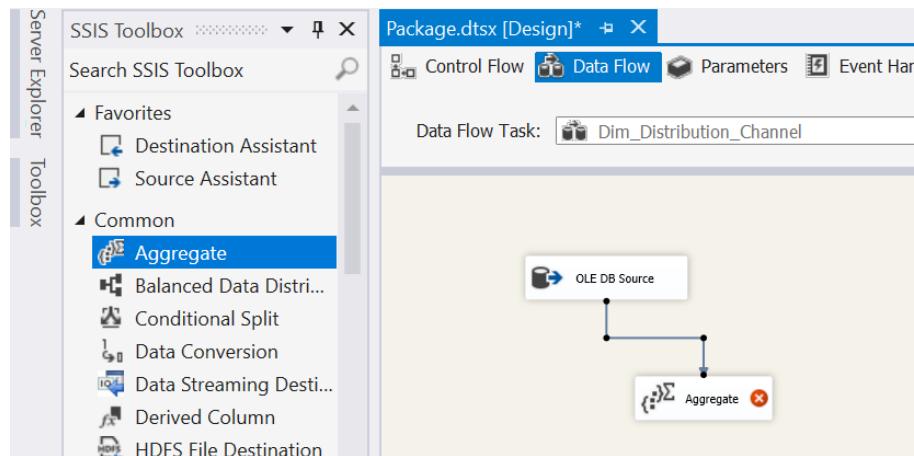
*Hình 2. 57. Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.

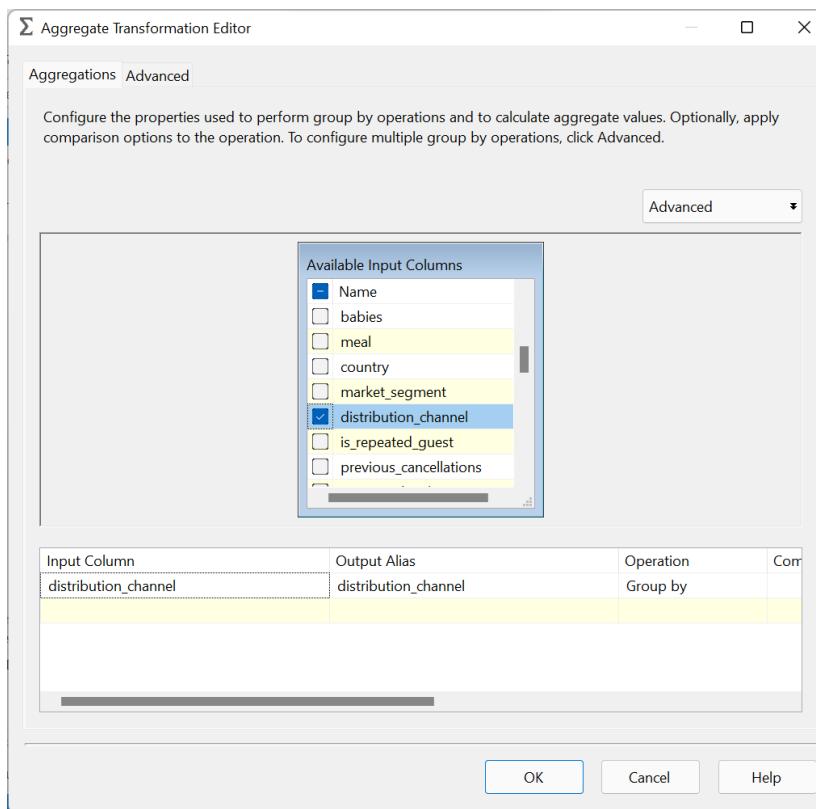


*Hình 2. 58. Chọn các columns cần sử dụng*

- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là distribution\_channel.

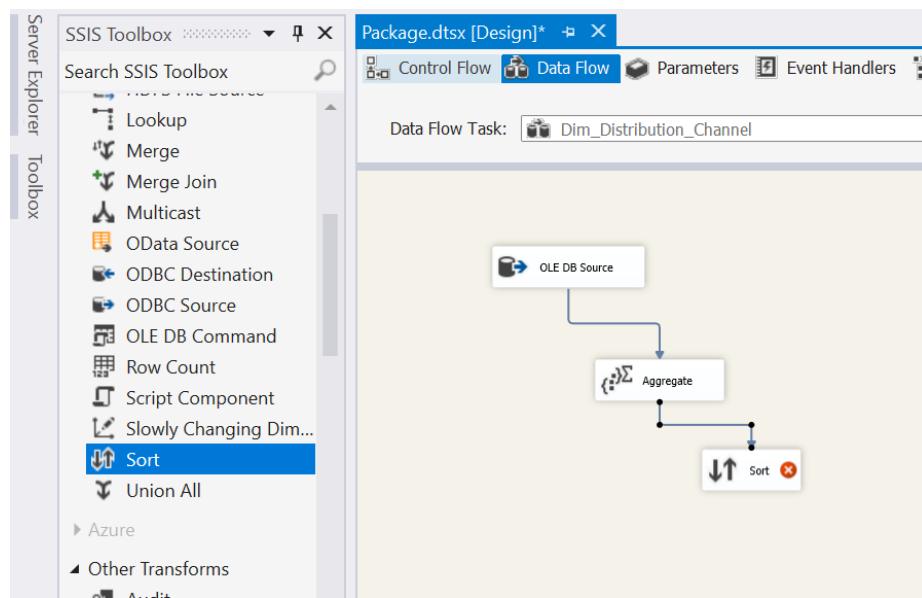


*Hình 2. 59.Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau*

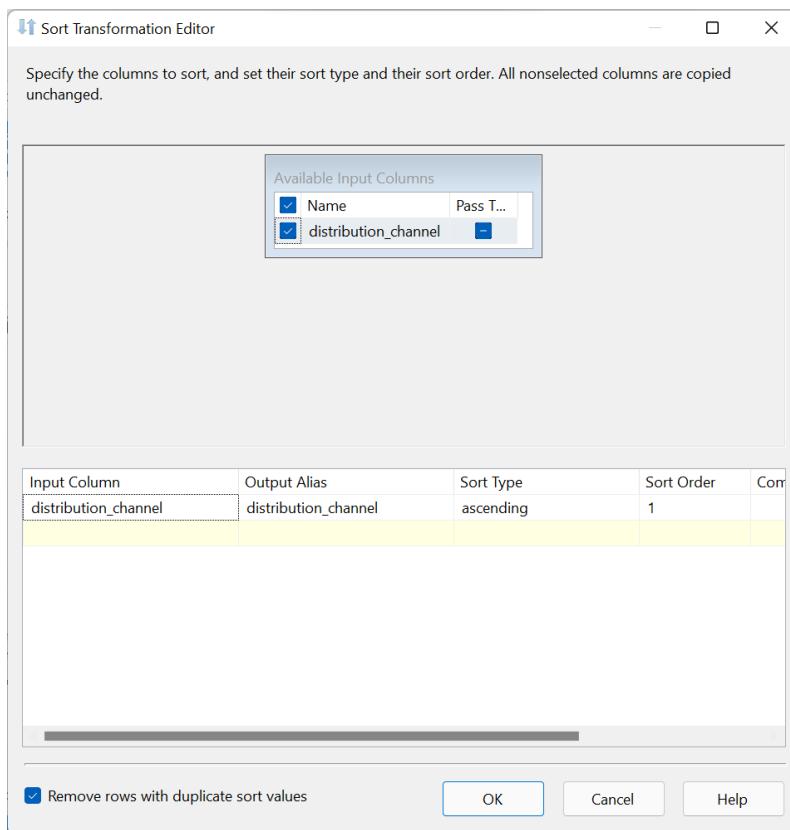


*Hình 2. 60.Chọn các thuộc tính cần gom nhóm*

- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.

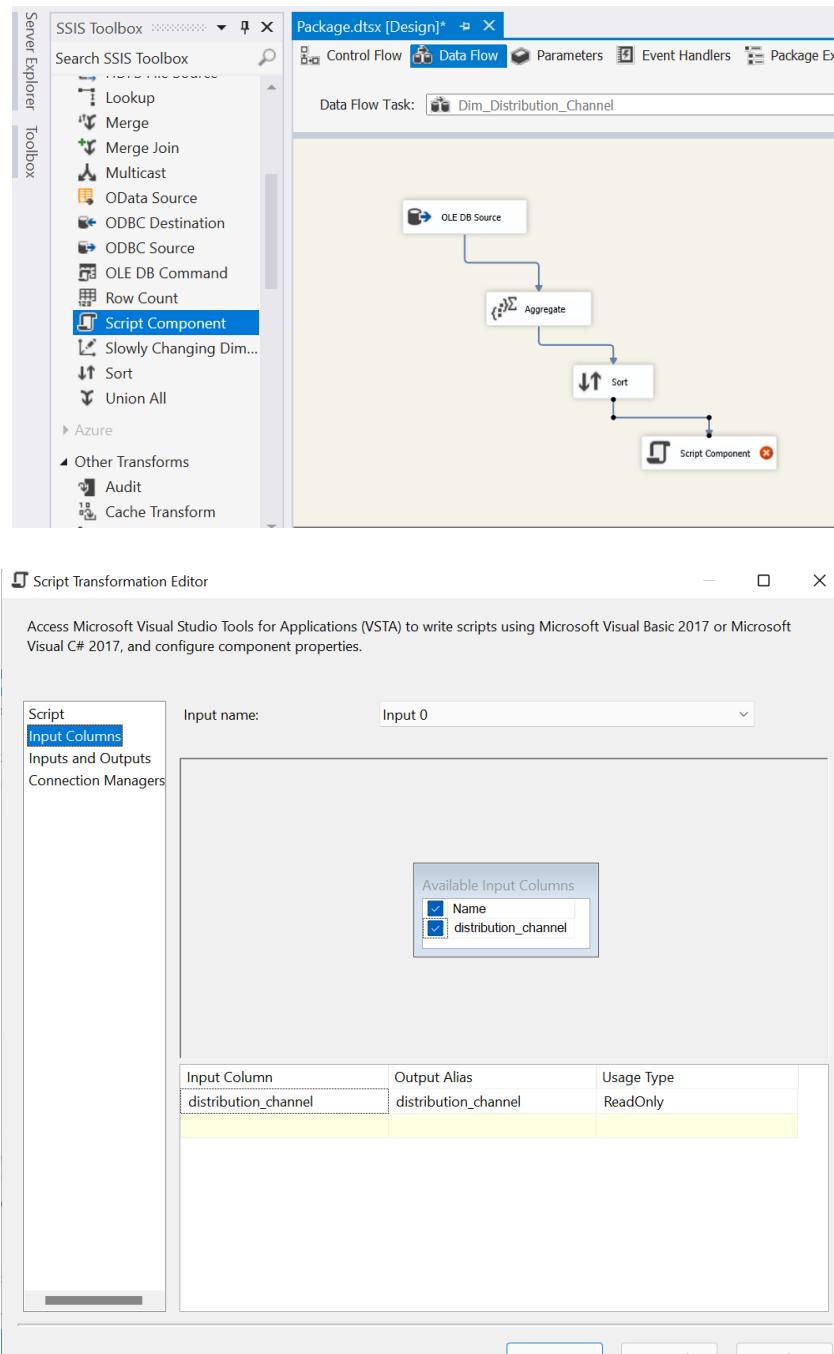


Hình 2. 61. Sắp xếp các giá trị theo chiều tăng dần

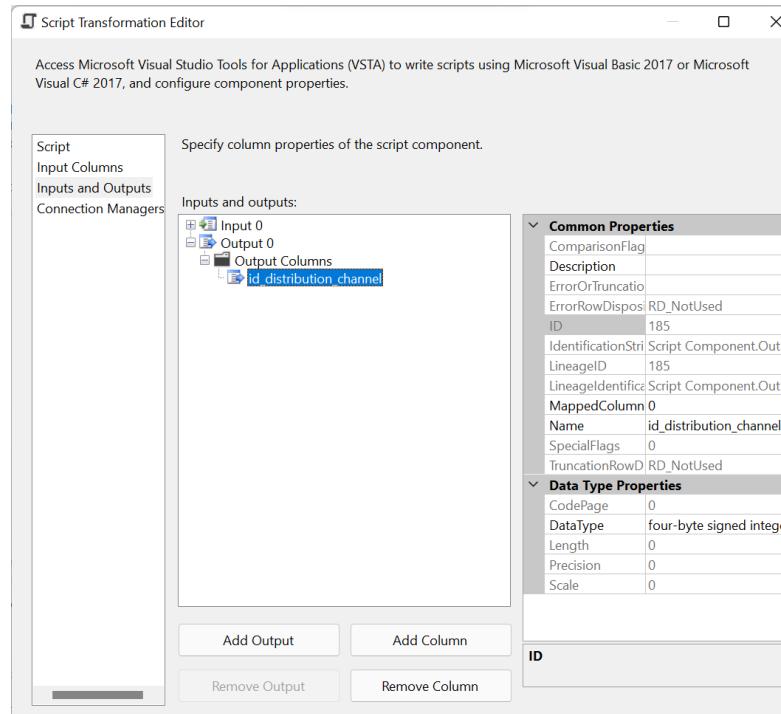


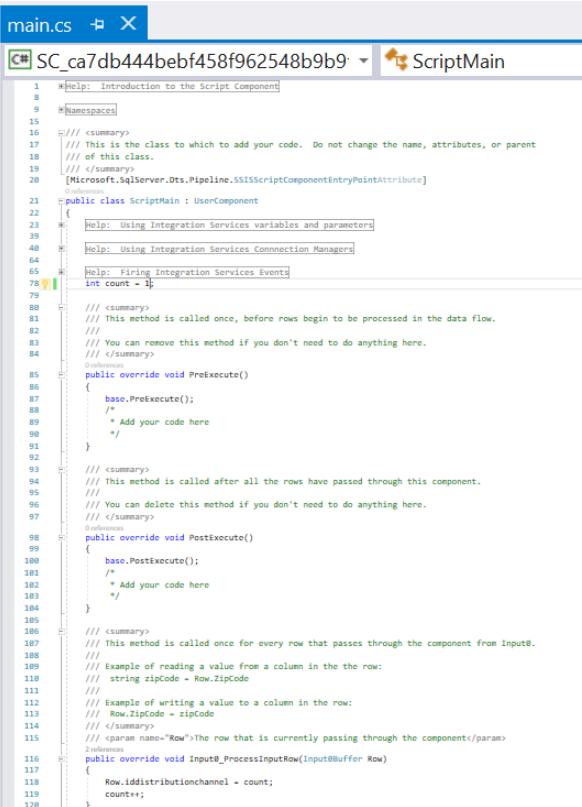
Hình 2. 62. Sắp xếp các giá trị theo chiều tăng dần

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_distribution\_channel.



Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22





The screenshot shows the Visual Studio IDE with the following details:

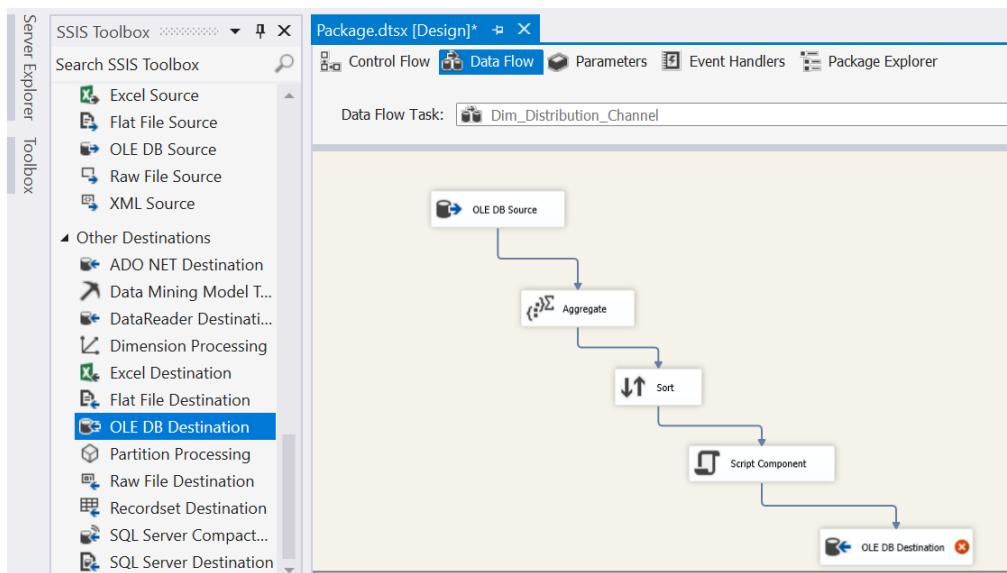
- Title Bar:** The title bar displays "main.cs" and "SC\_ca7db444bebfb458f962548b9b9".
- Toolbox:** The toolbox is visible on the left side of the interface.
- Code Editor:** The main area contains the C# code for the "ScriptMain" component. The code includes various XML comments (///) and annotations (Help:). It defines a public class "ScriptMain" with methods "PreExecute" and "PostExecute", and a private method "Input0\_ProcessInputRow".
- Status Bar:** The status bar at the bottom shows "21 lines" and "12000 bytes".

```
1  Help: Introduction to the Script Component
2
3  Namespaces
4
5  Summary
6  This is the class to which to add your code. Do not change the name, attributes, or parent
7  of this class.
8  Summary
9  [Microsoft.SqlServer.Dts.Pipeline.SSIScriptComponentEntryPointAttribute]
10
11  public class ScriptMain : UserComponent
12  {
13      Help: Using Integration Services variables and parameters
14
15      Help: Using Integration Services Connection Managers
16
17      Help: Firing Integration Services Events
18      int count = 0;
19
20      Summary
21      This method is called once, before rows begin to be processed in the data flow.
22
23      You can remove this method if you don't need to do anything here.
24      Summary
25      Indicates
26      public override void PreExecute()
27      {
28          base.PreExecute();
29
30          * Add your code here
31          */
32      }
33
34      Summary
35      This method is called after all the rows have passed through this component.
36
37      You can delete this method if you don't need to do anything here.
38      Summary
39      Indicates
40      public override voidPostExecute()
41      {
42          base.PostExecute();
43
44          * Add your code here
45          */
46      }
47
48      Summary
49      This method is called once for every row that passes through the component from Input0.
50
51      Example of reading a value from a column in the the row:
52      string zipCode = Row.ZipCode
53
54      Example of writing a value to a column in the row:
55      Row.ZipCode = zipCode
56      Summary
57      (param name="Row">The row that is currently passing through the component</param>
58      Indicates
59      public override void Input0_ProcessInputRow(Input0Buffer Row)
60      {
61          Row.Iddistributionchannel + count;
62          count++;
63      }
64
65      2 lines
66
```

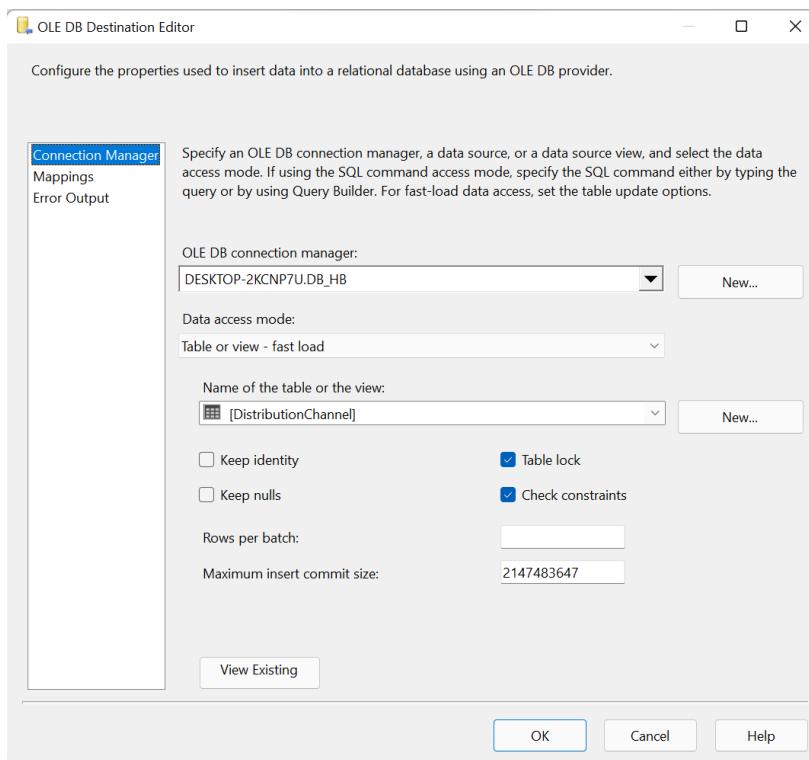
Hình 2. 63.1.,2.,3.4. Sử dụng Script Component để tạo khóa chính  
id distribution channel

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là DistributionChannel.

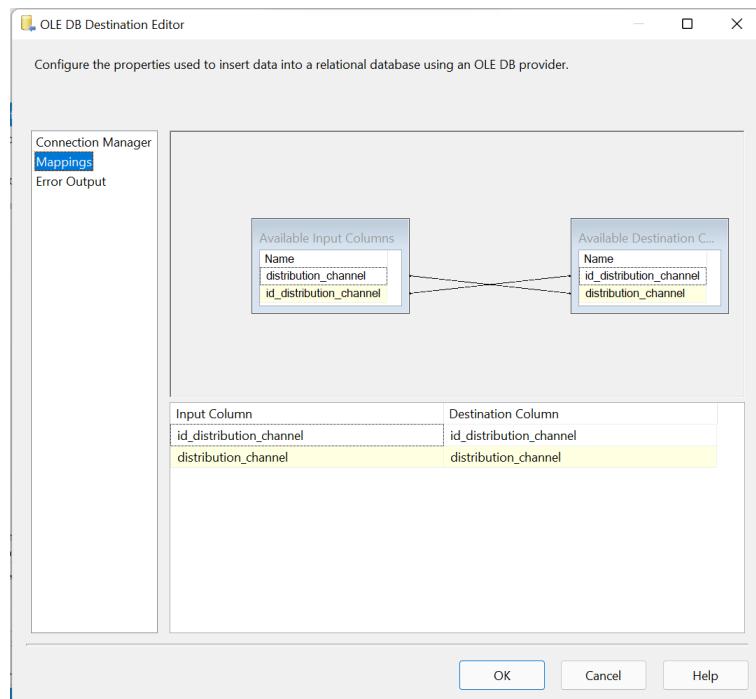


Hình 2. 64.Sử dụng OLE DB Destination để tạo bảng



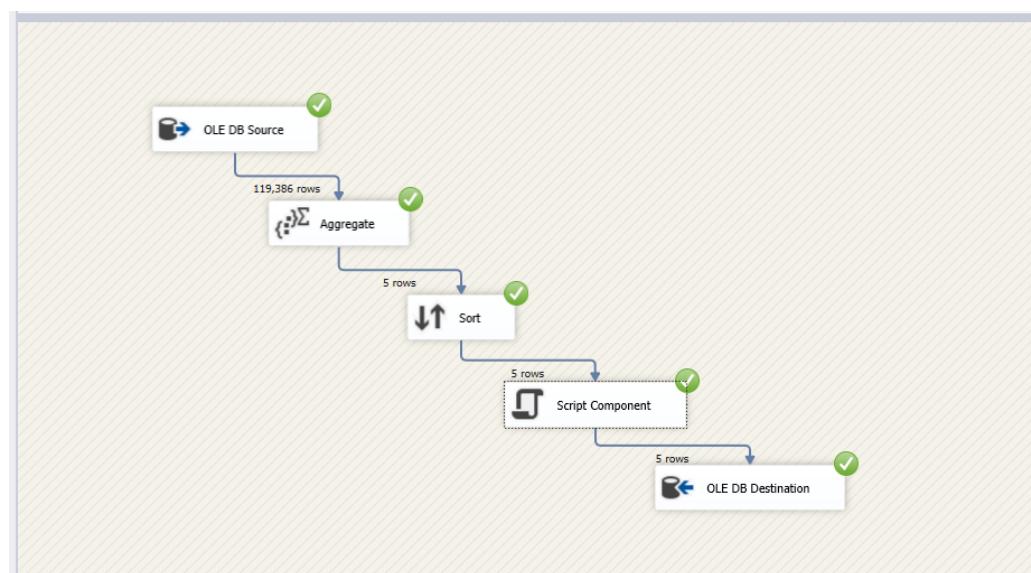
Hình 2. 65.Tạo bảng DistributionChannel

- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



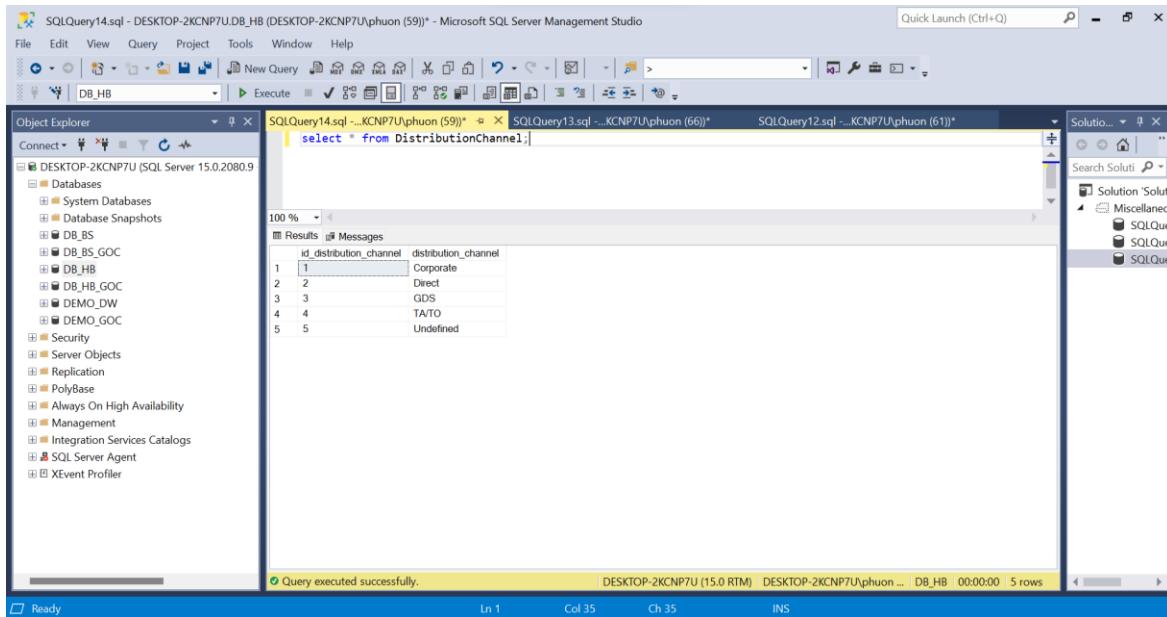
Hình 2. 66. Quá trình Mappings dữ liệu bảng DistributionChannel

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



Hình 2. 67. Hoàn thành đổ dữ liệu vào Distribution Channel trong kho dữ liệu

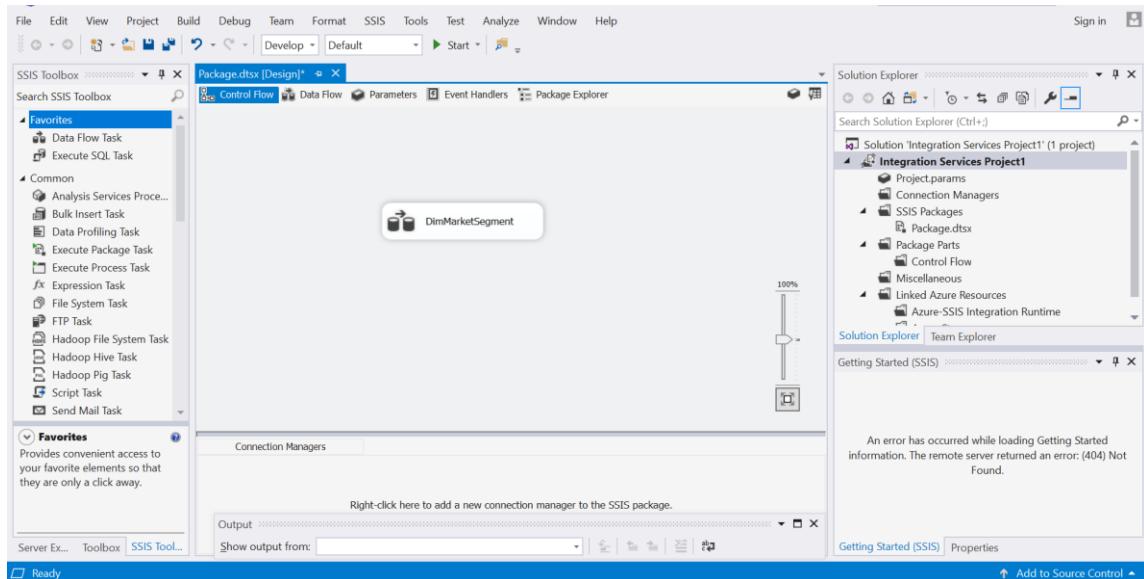
- Bước 11: Kiểm tra bảng DistributionChannel trên SQL Server.



Hình 2. 68.Kiểm tra bảng Distribution Channel trên SQL Server

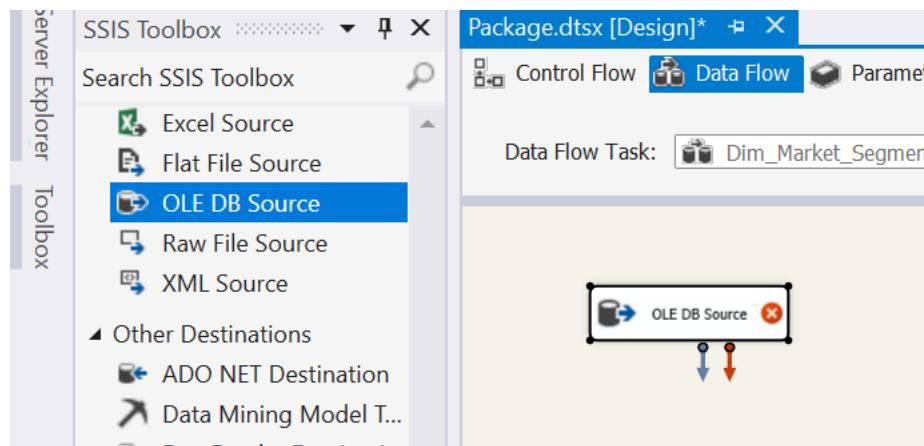
### 5.3. Bảng DimMarketSegment

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimMarketSegment.



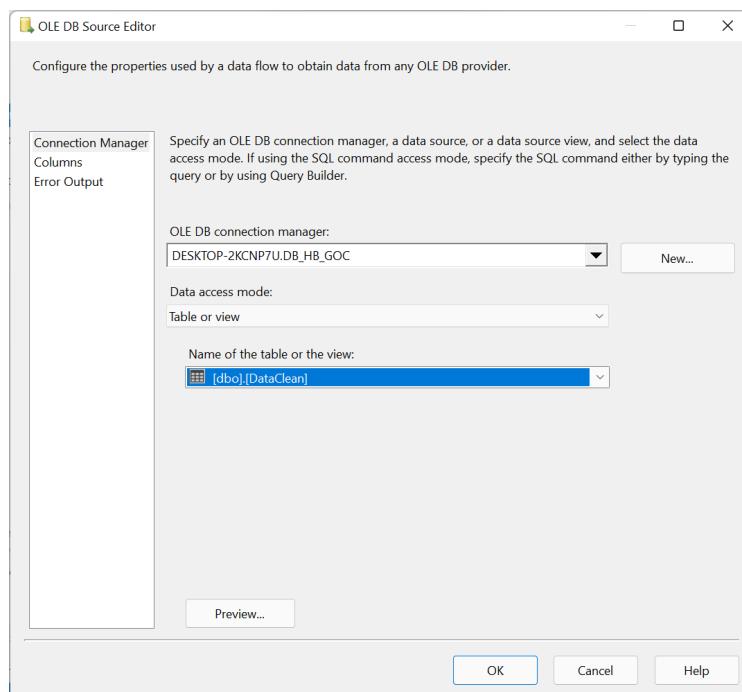
Hình 2. 69.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimMarketSegment

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



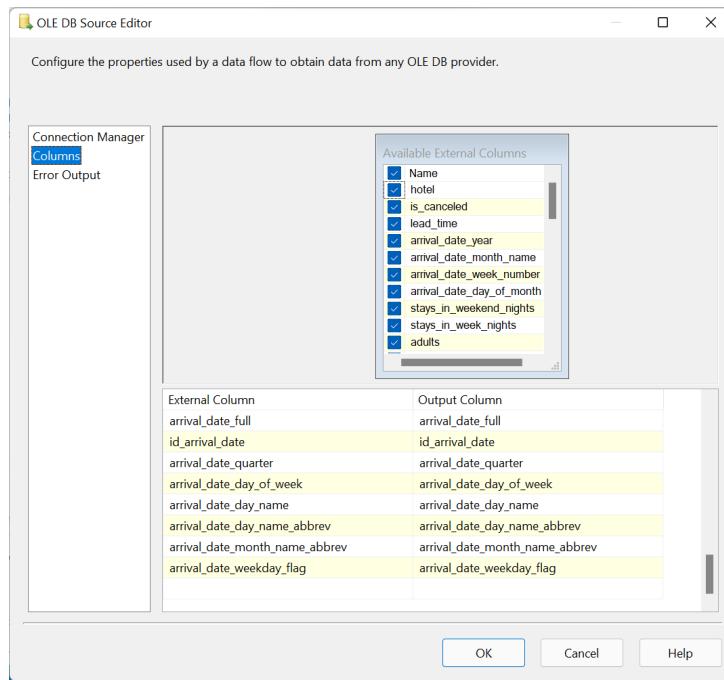
*Hình 2. 70.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



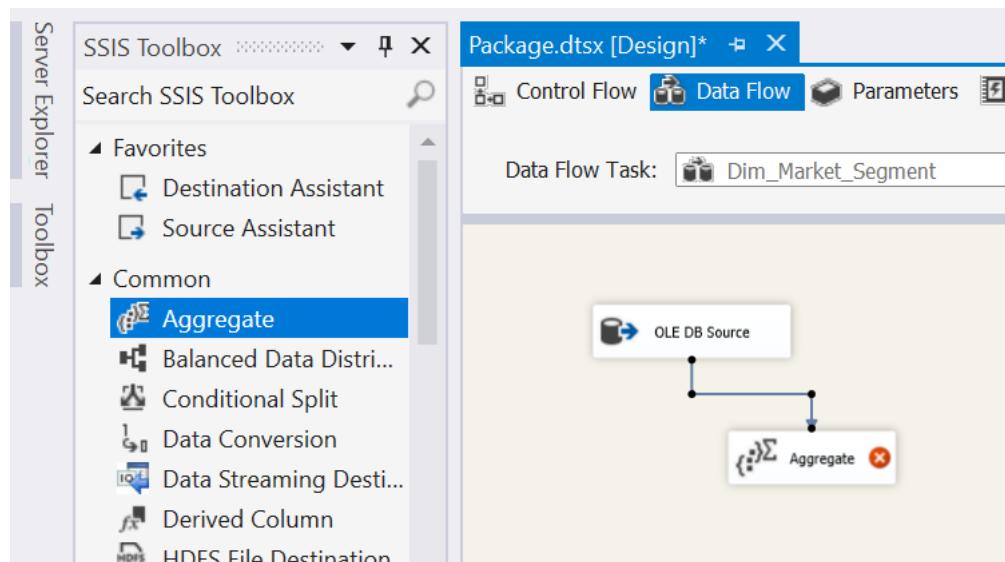
*Hình 2. 71.Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.

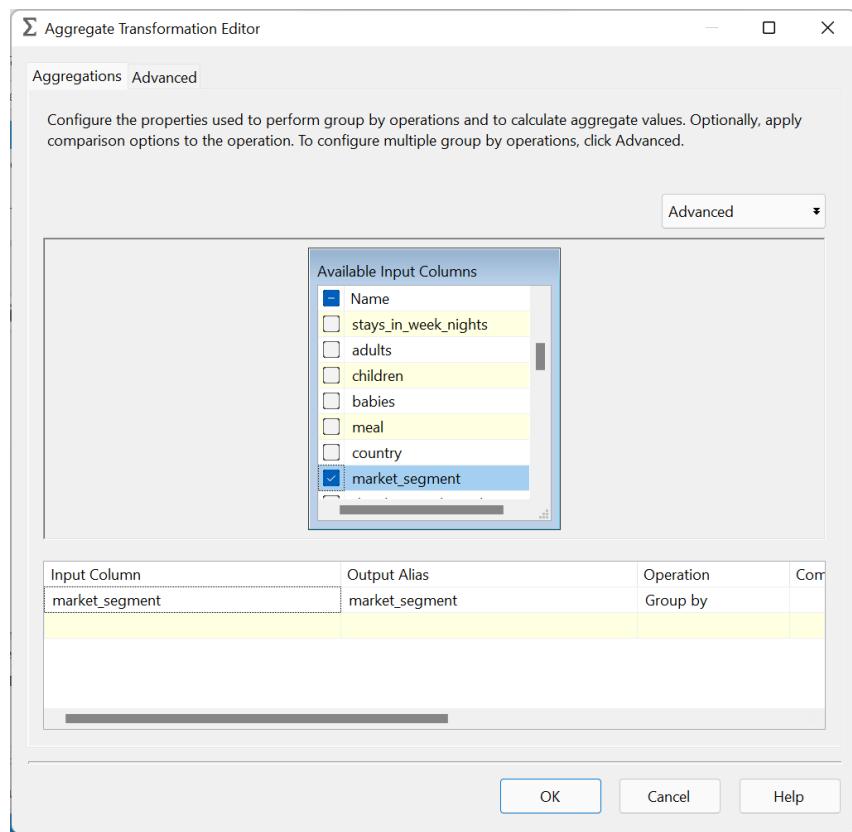


Hình 2. 72.Chọn các columns cần sử dụng

- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là market\_segment.

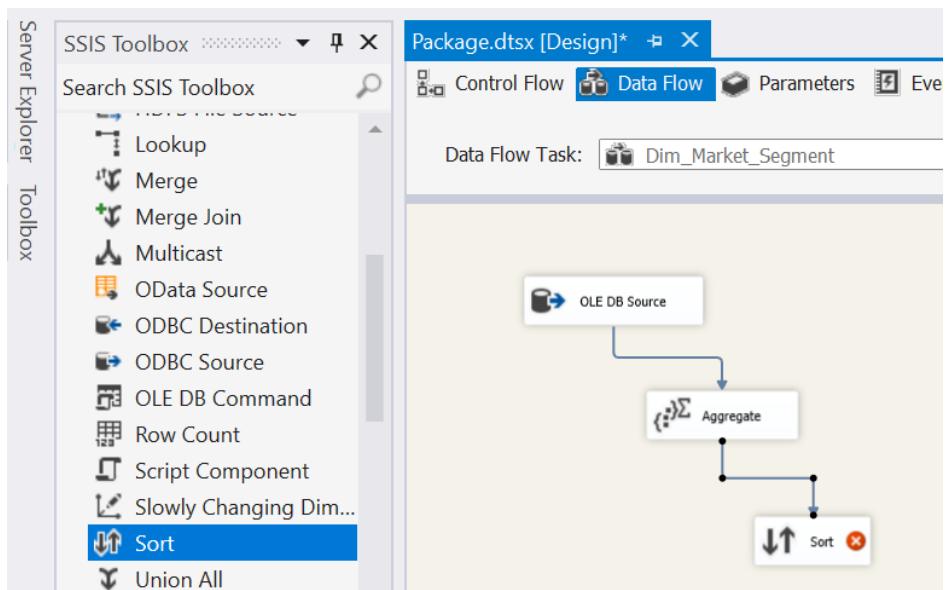


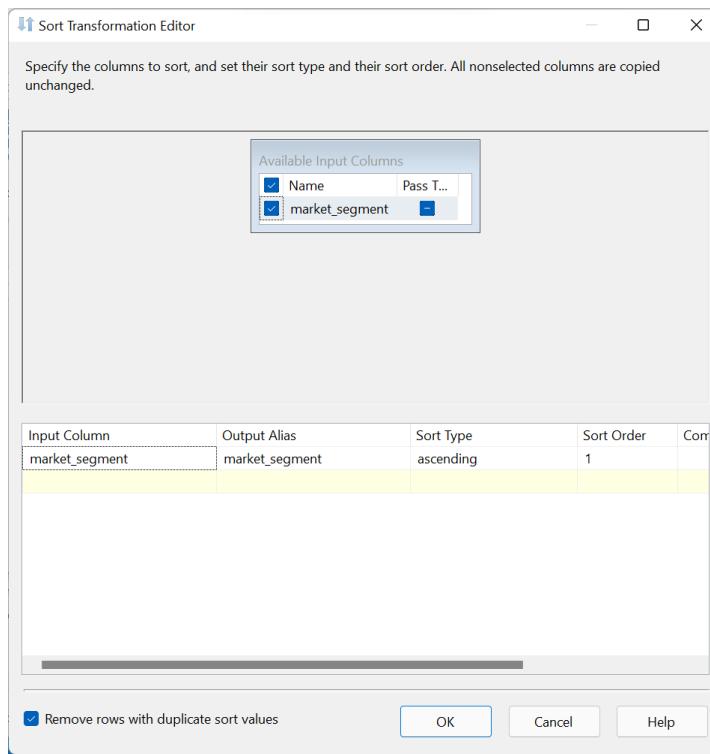
Hình 2. 73.Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau



Hình 2. 74. Chọn các thuộc tính cần gom nhóm

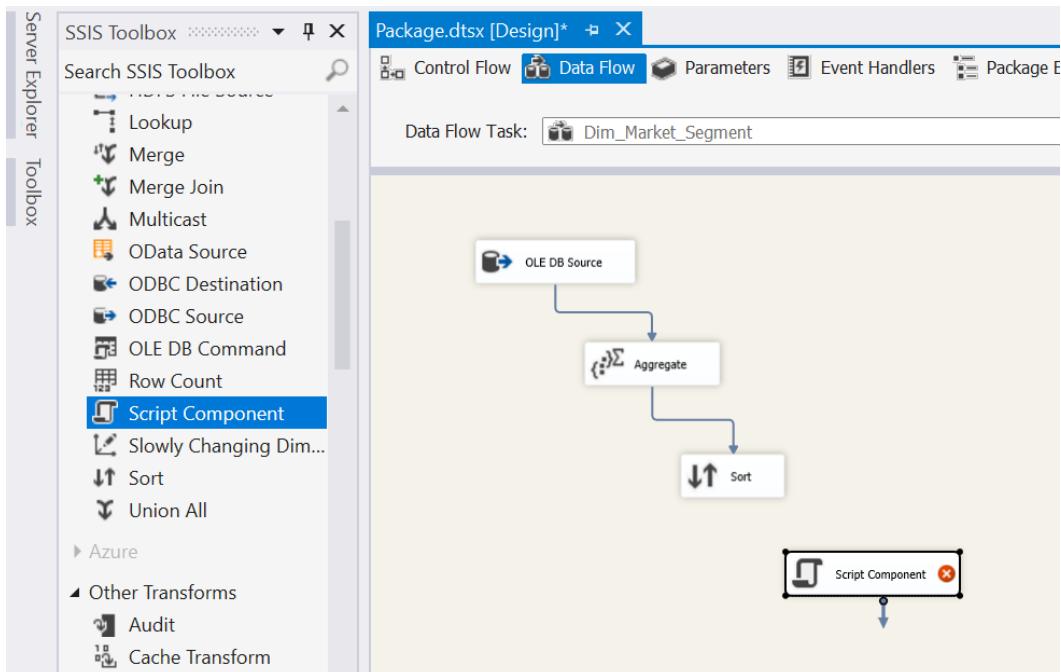
- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.



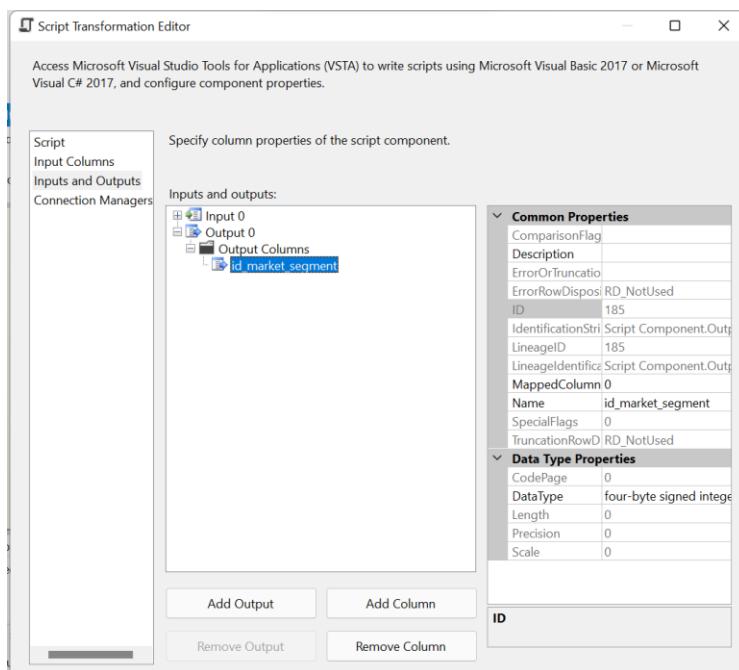
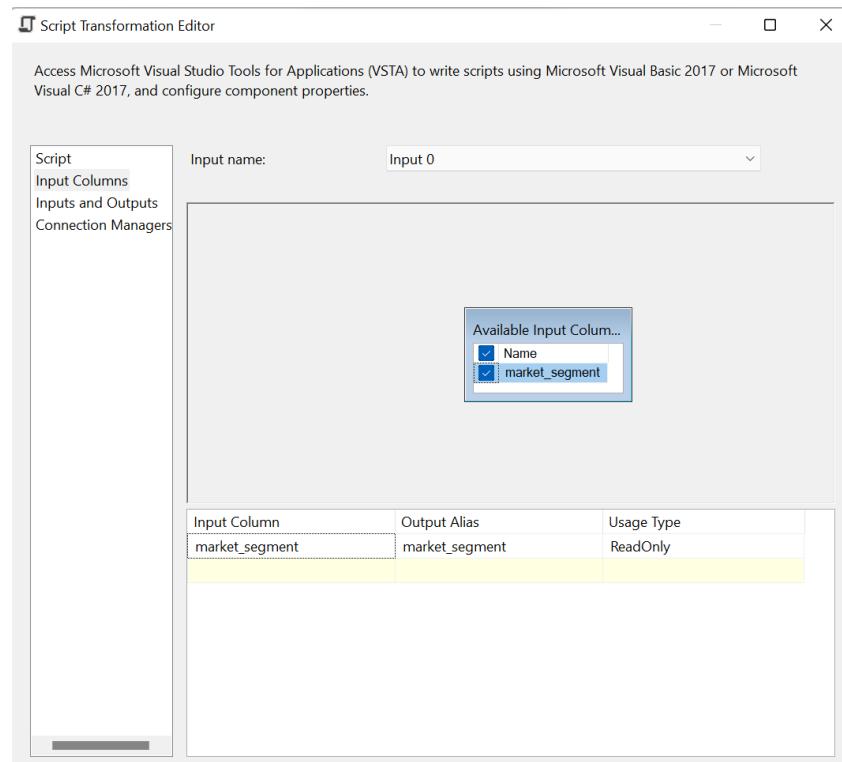


*Hình 2. 75.,76.Sắp xếp các giá trị theo chiều tăng dần*

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_market\_segment.



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



```

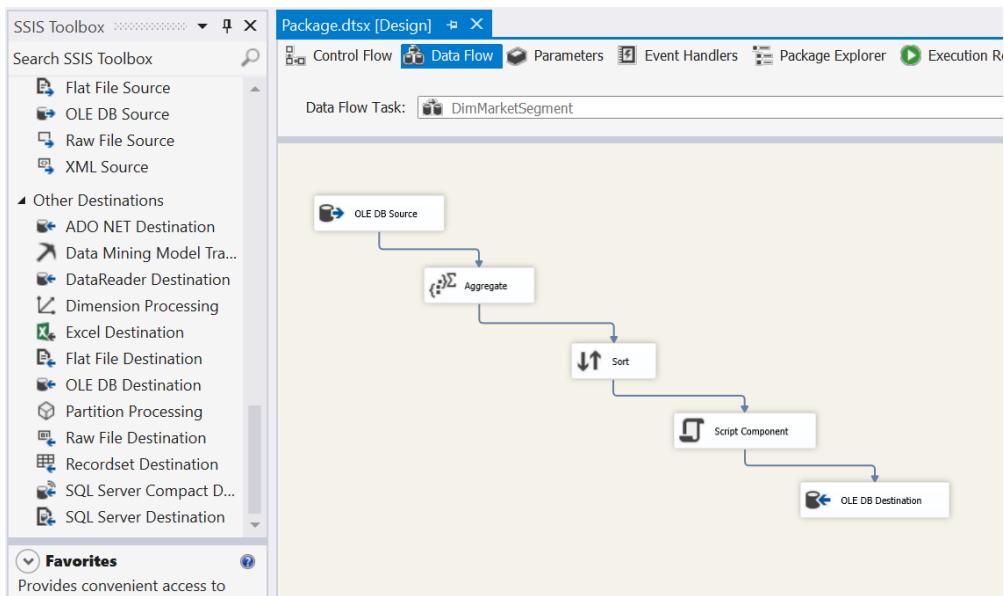
main.cs*  X
SC_e50646ed71ef417b83aea38672c  ScriptMain

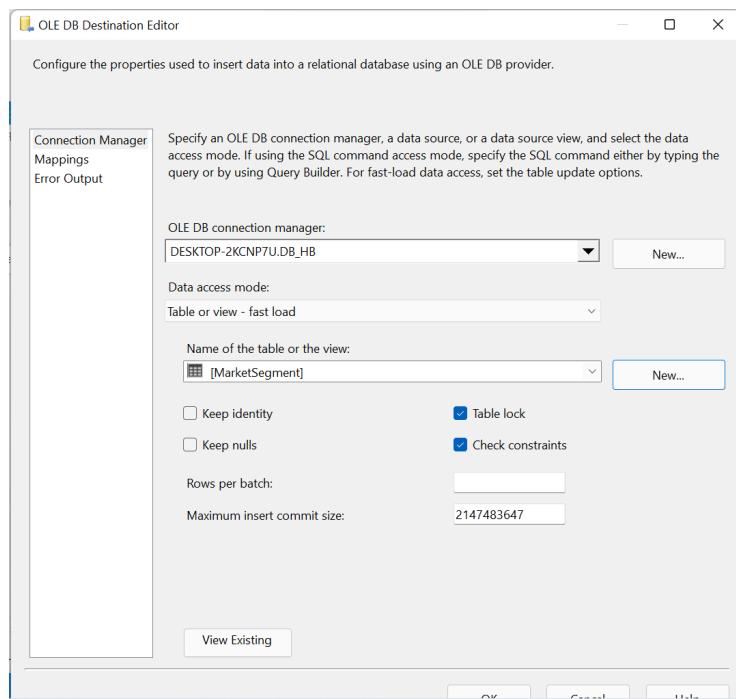
1  <!-- Help: Introduction to the Script Component-->
2  <!-- Namespaces-->
3  <!-- Summary-->
4  <!-- This is the class to which to add your code. Do not change the name, attributes, or parent-->
5  <!-- of this class.-->
6  <!-- />summary-->
7  [Microsoft.SqlServer.Dts.Pipeline.SSIScriptComponentEntryPointAttribute]
8
9  public class ScriptMain : UserComponent
10 {
11     <!-- Help: Using Integration Services variables and parameters-->
12     <!-- Help: Using Integration Services Connection Managers-->
13     <!-- Help: Firing Integration Services Events-->
14
15     int count = 1;
16
17     <!-- summary-->
18     <!-- This method is called once, before rows begin to be processed in the data flow.-->
19     <!-- You can remove this method if you don't need to do anything here.-->
20     <!-- />summary-->
21
22     public override void PreExecute()
23     {
24         base.PreExecute();
25         <!-- Add your code here-->
26     }
27
28     <!-- summary-->
29     <!-- This method is called after all the rows have passed through this component.-->
30     <!-- You can delete this method if you don't need to do anything here.-->
31     <!-- />summary-->
32
33     public override void PostExecute()
34     {
35         base.PostExecute();
36         <!-- Add your code here-->
37     }
38
39     <!-- summary-->
40     <!-- This method is called once for every row that passes through the component from Input0.-->
41     <!-- Example of reading a value from a column in the row:
42     <!-- string zipCode = Row.ZipCode-->
43     <!-- Example of writing a value to a column in the row:
44     <!-- Row.ZipCode = zipCode-->
45     <!-- />summary-->
46     <!-- <param name="Row">The row that is currently passing through the component</param>
47
48     public override void Input0_ProcessInputRow(Input0Buffer Row)
49     {
50         Row.idmarketsegment = count;
51         count++;
52     }
53
54 }

```

Hình 2. 76.1.,2.,3.,4. Sử dụng Script Component để tạo khóa chính id\_market\_segment

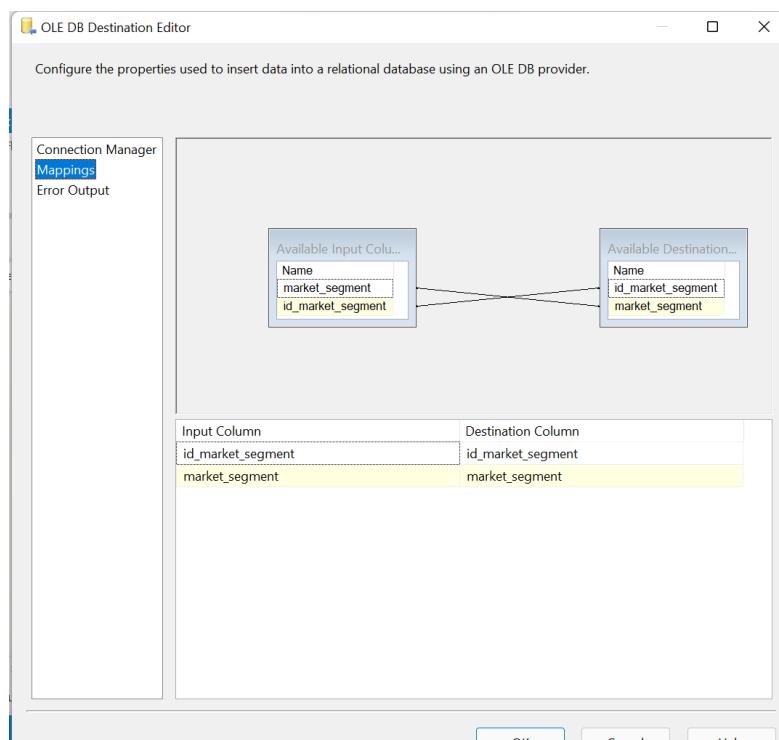
- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là MarketSegment.





*Hình 2. 77.1.,2. Tạo bảng Market Segment*

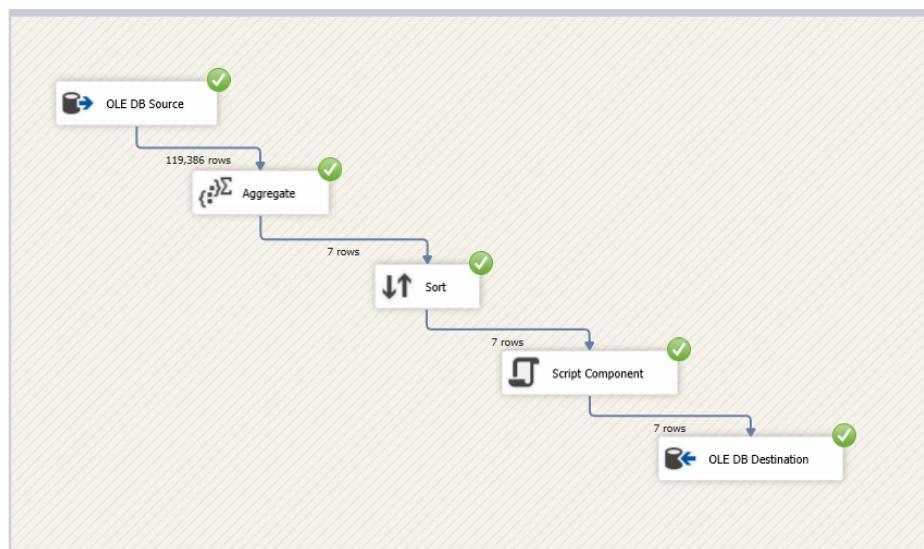
- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



*Hình 2. 78. Quá trình Mappings dữ liệu bảng MarketSegment*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 79.Hoàn thành đổ dữ liệu vào Market Segment trong kho dữ liệu*

- Bước 11: Kiểm tra bảng MarketSegment trên SQL Server.

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the database 'DB\_HB' is selected. In the center pane, a query window displays the following SQL code and results:

```
select * from MarketSegment;
```

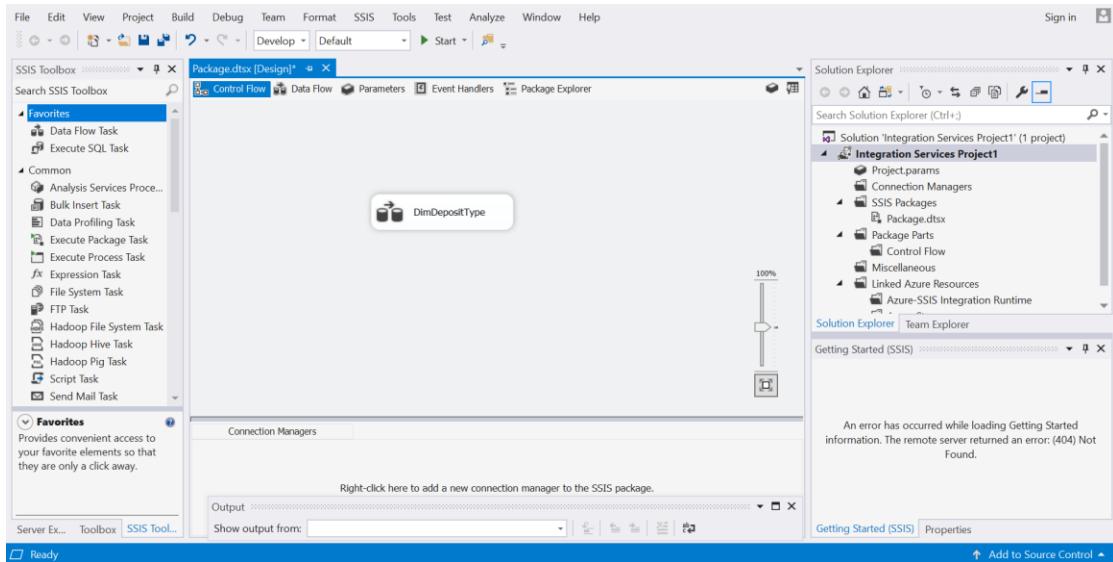
	id_market_segment	market_segment
1	1	Aviation
2	2	Complementary
3	3	Corporate
4	4	Direct
5	5	Groups
6	6	Offline TA/TO
7	7	Online TA

Below the results, a message states: "Query executed successfully." The status bar at the bottom indicates: DESKTOP-2KCNP7U (15.0 RTM) | DESKTOP-2KCNP7U\phuong ... | DB\_HB | 00:00:00 | 7 rows

*Hình 2. 80.Kiểm tra bảng MarketSegment trên SQL Server*

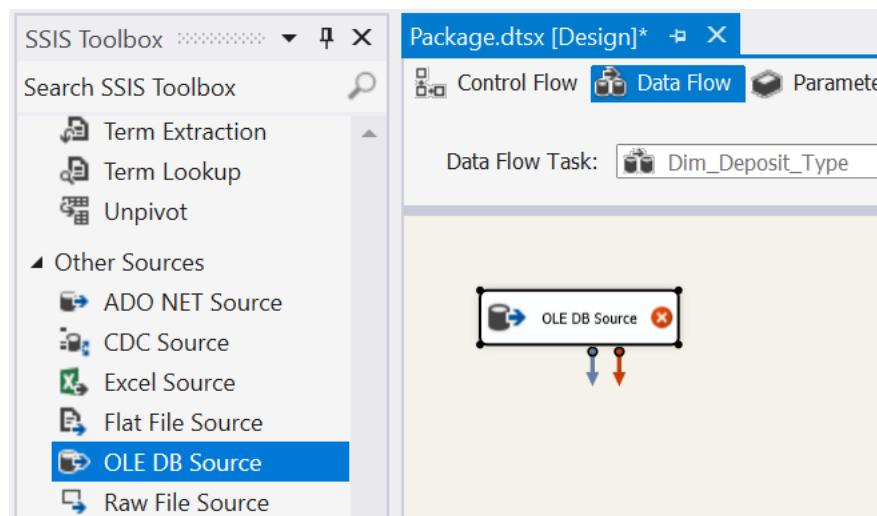
#### 5.4. Bảng DimDepositType

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimDepositType.



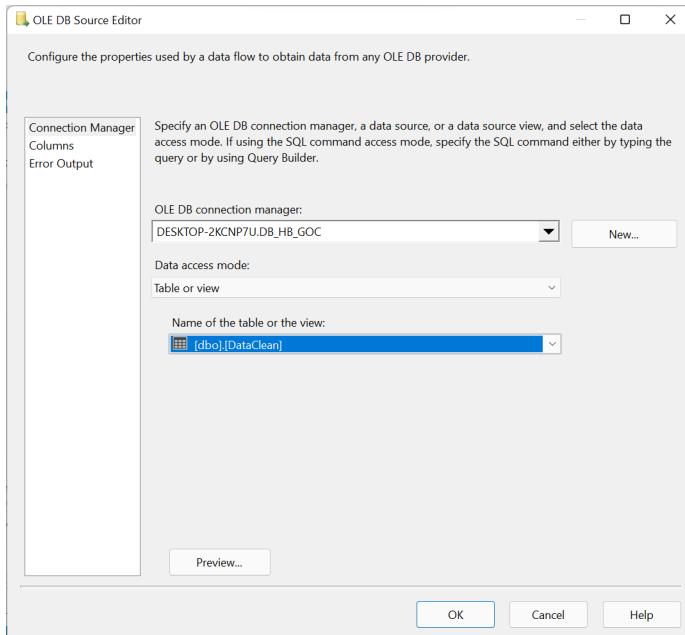
*Hình 2. 81.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimDepositType*

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



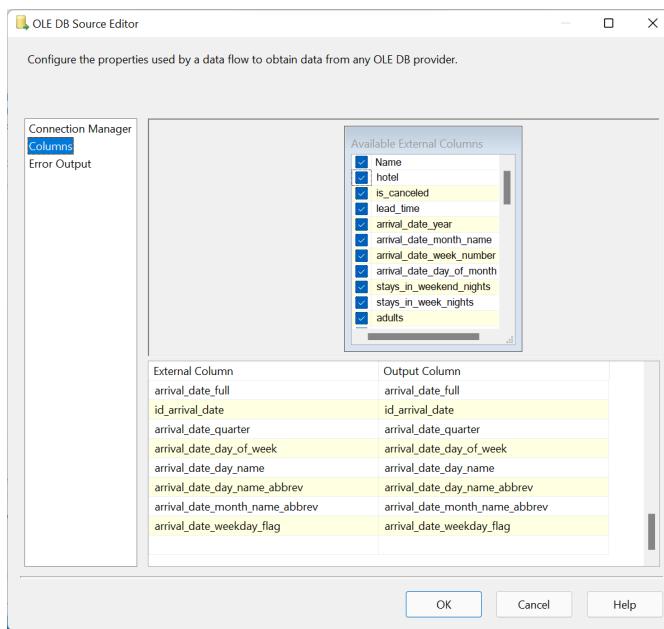
*Hình 2. 82.1.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



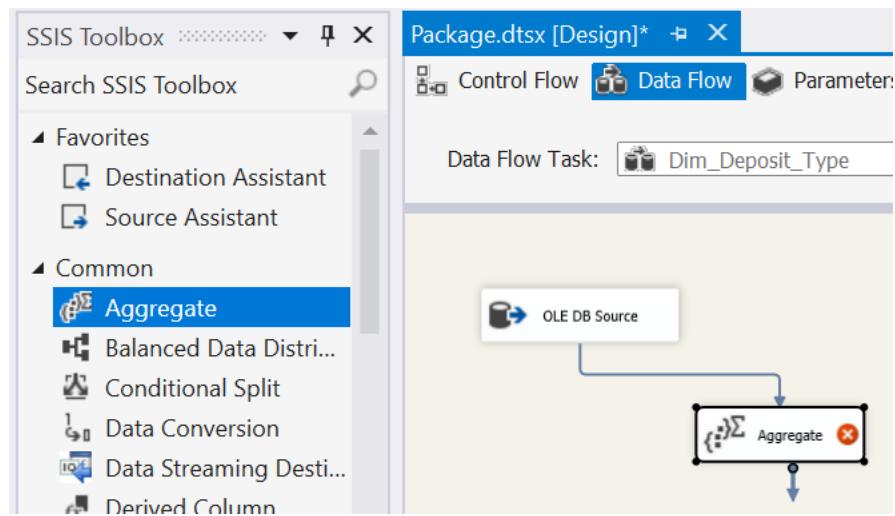
*Hình 2. 82.2. Chọn table sẽ lấy dữ liệu sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.

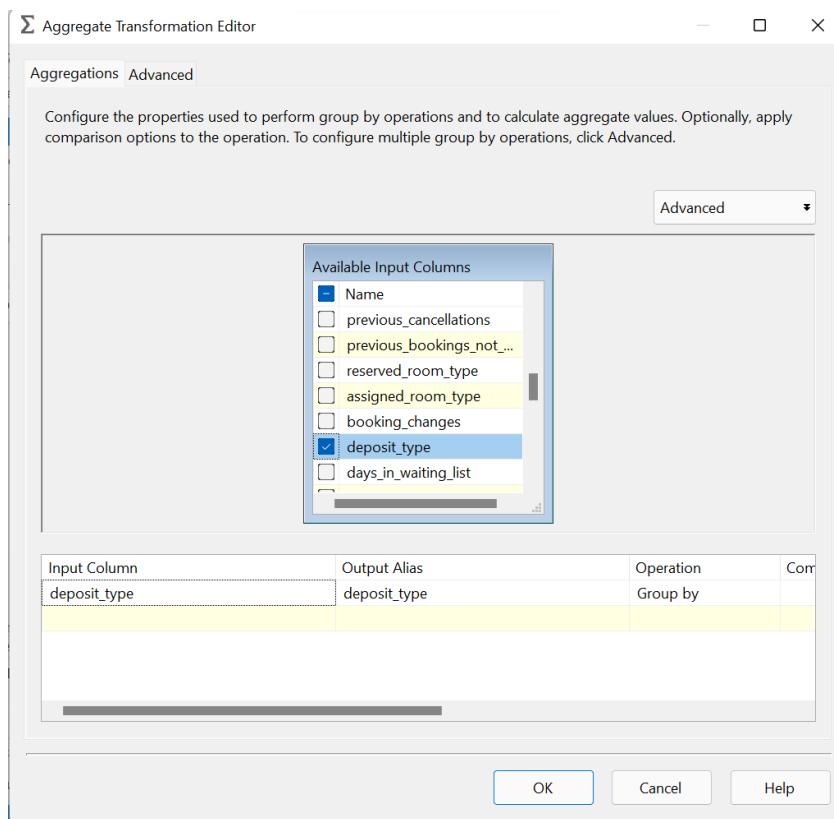


*Hình 2. 83. Chọn các columns cần sử dụng*

- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là deposit\_type.

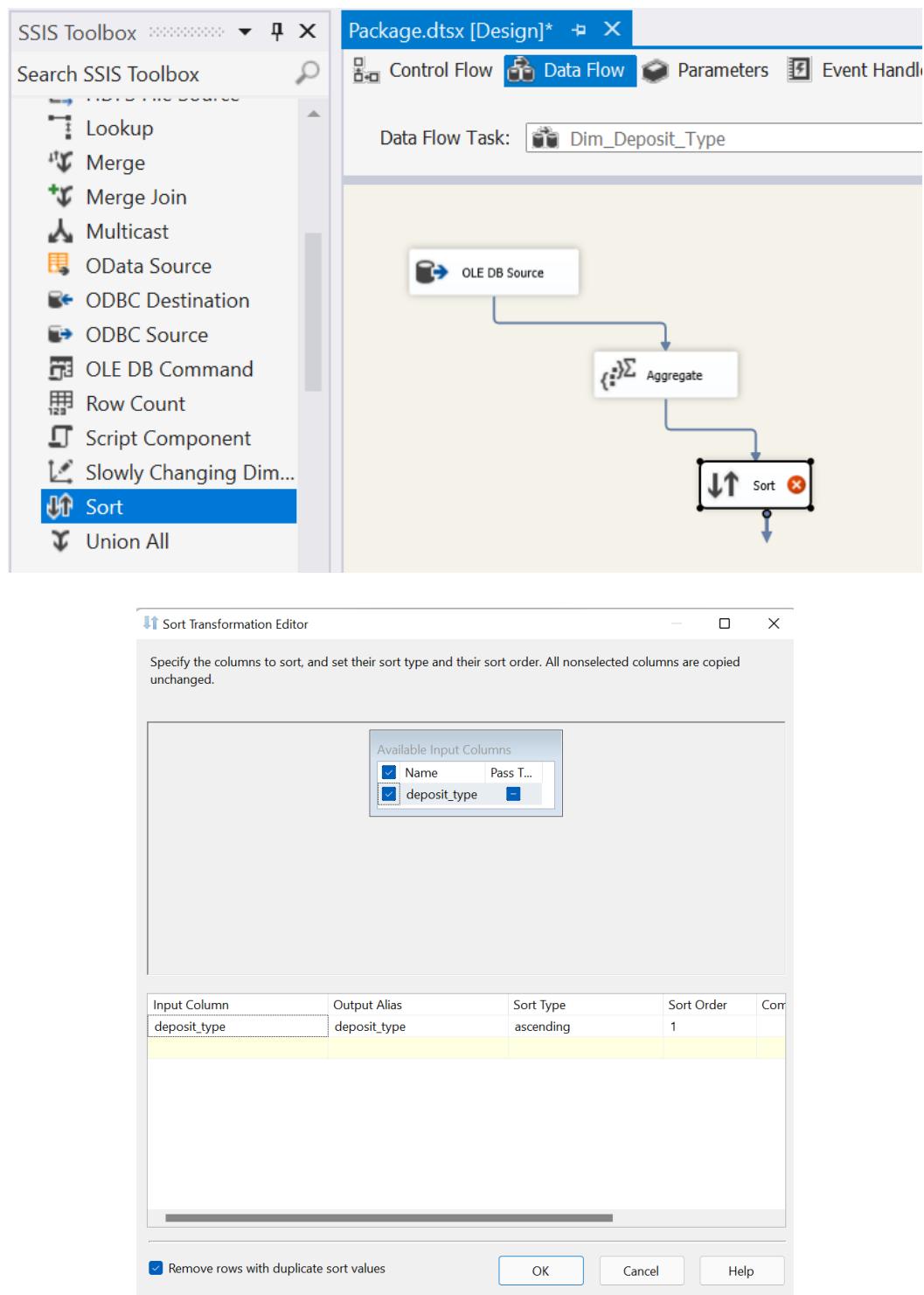


*Hình 2. 84. Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau*



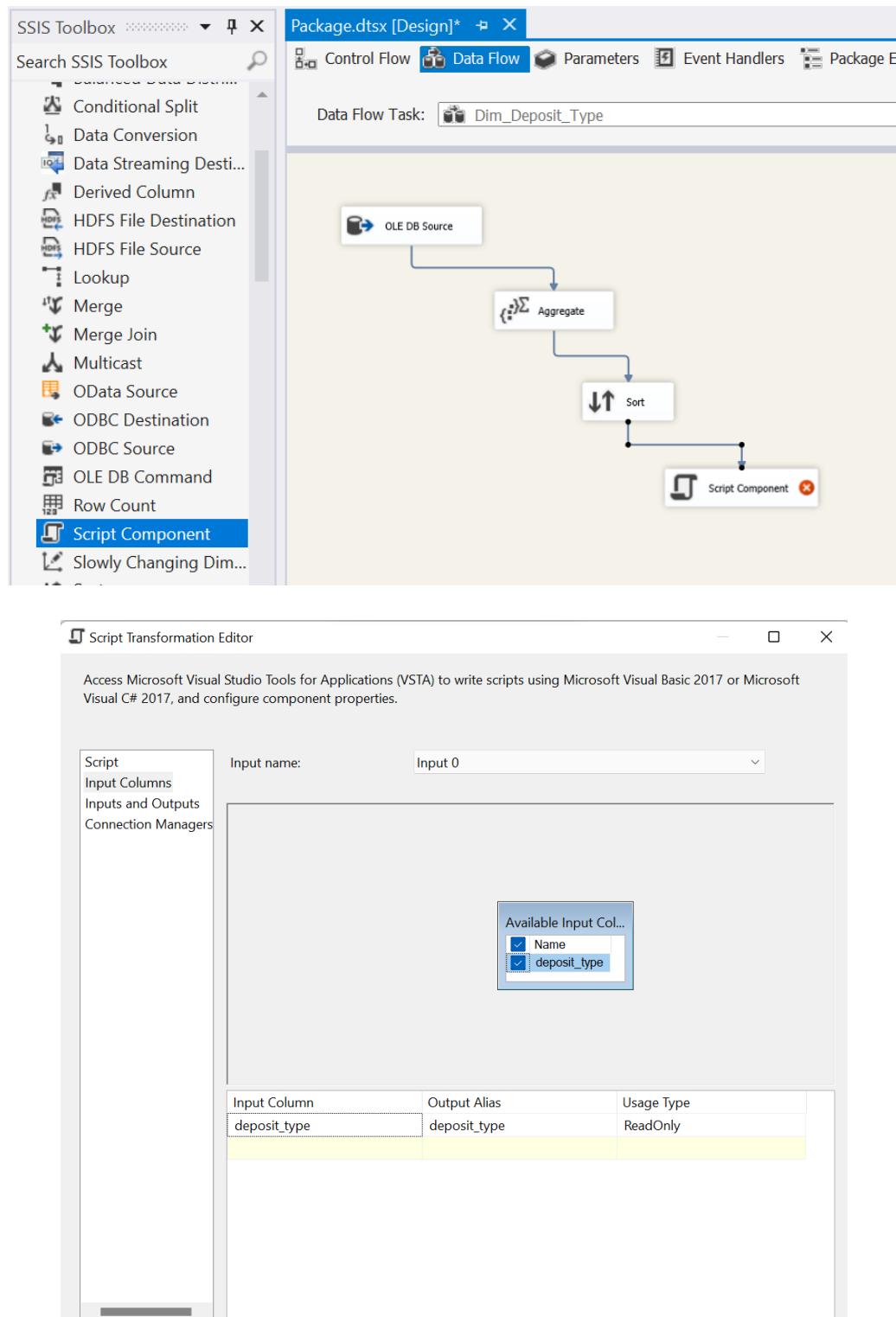
*Hình 2. 85. Chọn các thuộc tính cần gom nhóm*

- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.

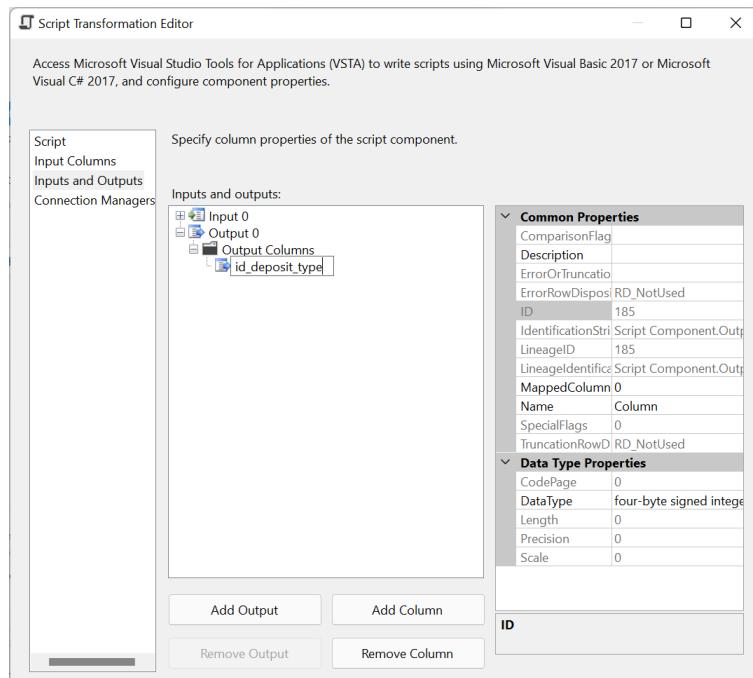


*Hình 2. 86.1,.2.Sắp xếp các giá trị theo chiều tăng dần*

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_deposit\_type.



## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



main.cs\*

```

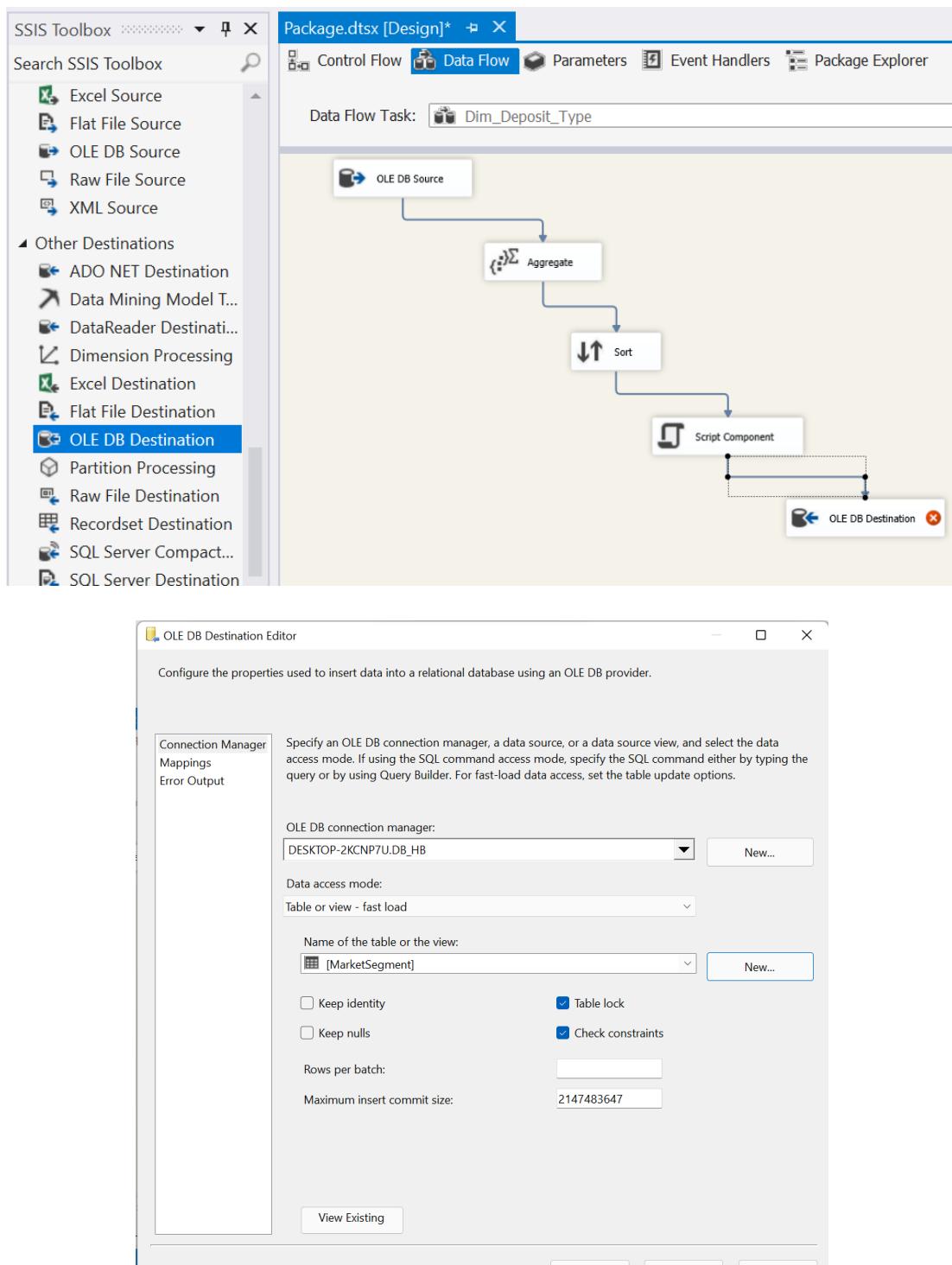
1  /*Help: Introduction to the Script Component*/
2
3  #Namespaces
4
5
6  /// <summary>
7  /// This is the class to which to add your code. Do not change the name, attributes, or parent
8  /// of this class.
9  /// </summary>
10 [Microsoft.SqlServer.Dts.Pipeline.SSISScriptComponentEntryPointAttribute]
11
12 public class ScriptMain : UserComponent
13 {
14
15     /* Help: Using Integration Services variables and parameters
16
17     /* Help: Using Integration Services Connection Managers
18
19     /* Help: Firing Integration Services Events
20
21     int count = 1;
22
23     /// <summary>
24     /// This method is called once, before rows begin to be processed in the data flow.
25
26     /// You can remove this method if you don't need to do anything here.
27
28     /// </summary>
29
30     public override void PreExecute()
31     {
32         base.PreExecute();
33
34         /*
35          * Add your code here
36          */
37     }
38
39     /// <summary>
40     /// This method is called after all the rows have passed through this component.
41
42     /// You can delete this method if you don't need to do anything here.
43
44     /// </summary>
45
46     public override void PostExecute()
47     {
48         base.PostExecute();
49
50         /*
51          * Add your code here
52          */
53     }
54
55     /// <summary>
56     /// This method is called once for every row that passes through the component from Input0.
57
58     /// Example of reading a value from a column in the the row:
59     string zipCode = Row.ZipCode;
60
61     /// Example of writing a value to a column in the row:
62     Row.ZipCode = zipCode;
63
64     /// <param name="Row">The row that is currently passing through the component</param>
65
66     public override void Input0_ProcessInputRow(Input0Buffer Row)
67     {
68         Row.iddeposittype = count;
69         count++;
70     }
71
72
73 }
74
75
76
77
78
79
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122

```

Hình 2. 87.,1.,2.,3.,4.Sử dụng Script Component để tạo khóa chính id\_deposit\_type

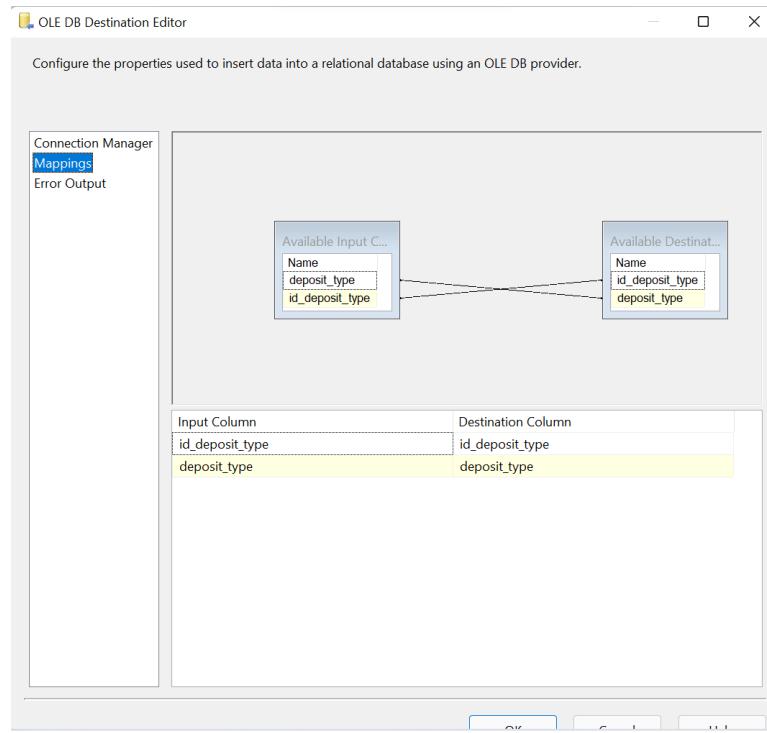
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là DepositType.



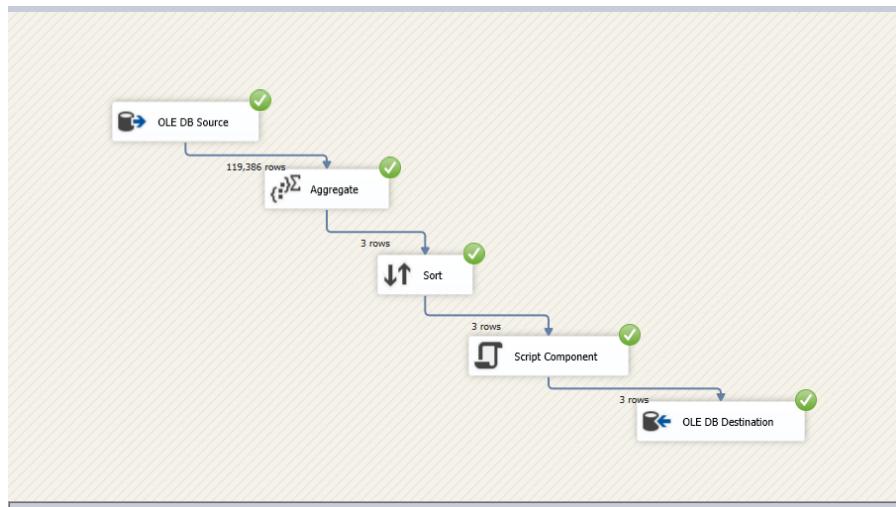
*Hình 2. 88.1.,2. Tạo bảng DepositType*

- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



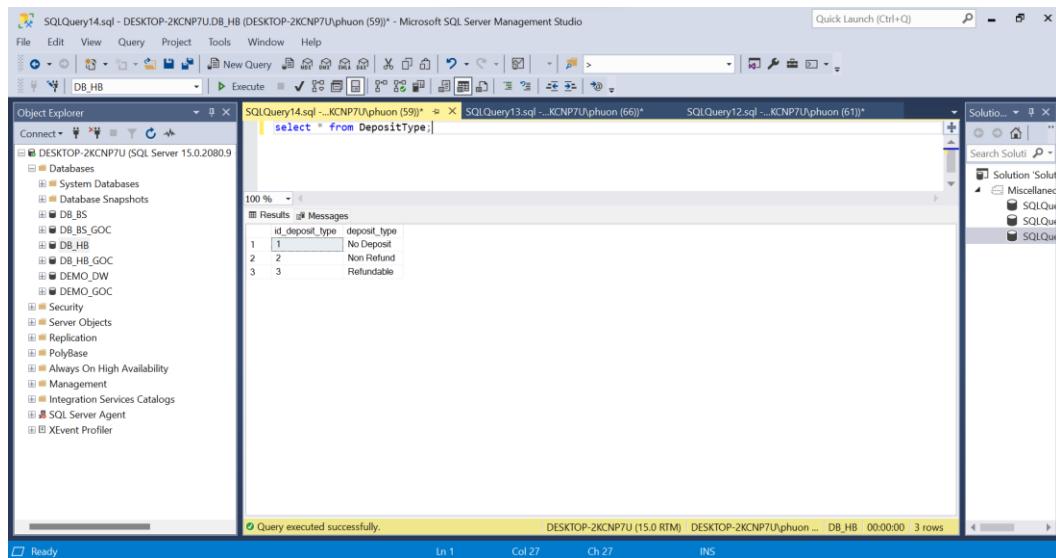
*Hình 2. 89..Quá trình Mappings dữ liệu bảng DepositType*

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 90.Hoàn thành đổ dữ liệu vào DepositType trong kho dữ liệu*

- Bước 11: Kiểm tra bảng DepositType trên SQL Server.



The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the database 'DB\_HB' is selected. In the center pane, a query window displays the following results:

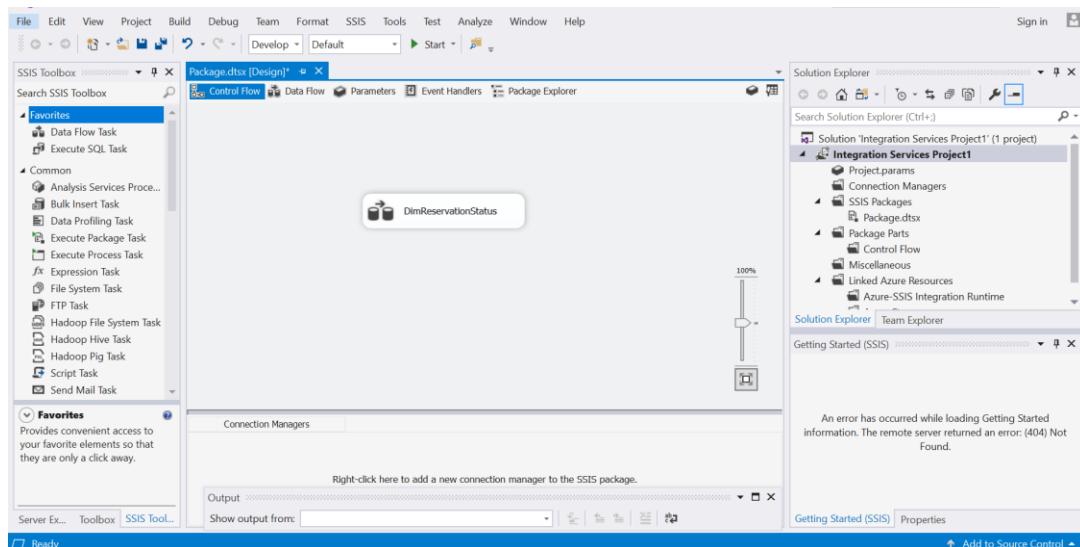
	id_deposit_type	deposit_type
1	1	No Deposit
2	2	Non Refund
3	3	Refundable

Below the results, a message states: "Query executed successfully." The status bar at the bottom indicates: DESKTOP-2KCNP7U (15.0 RTM) DESKTOP-2KCNP7U\phuong .. DB\_HB | 00:00:00 | 3 rows.

*Hình 2. 91.Kiểm tra bảng Deposit Type trên SQL Server*

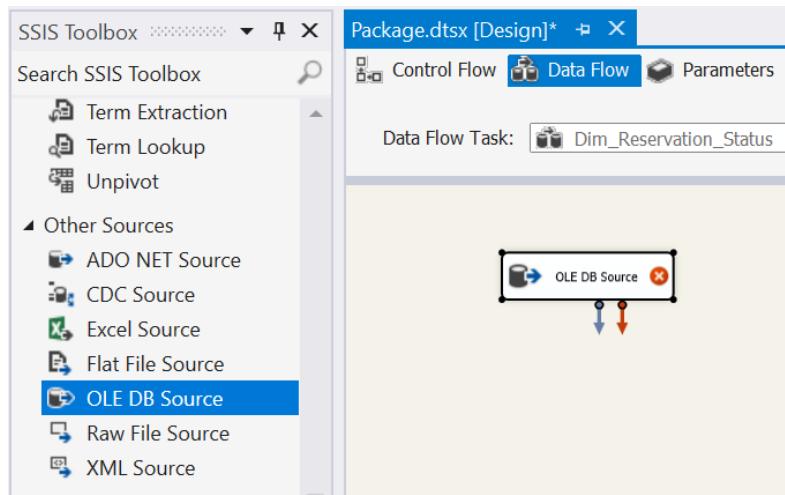
### 5.5. Bảng DimReservationStatus

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimReservationStatus.



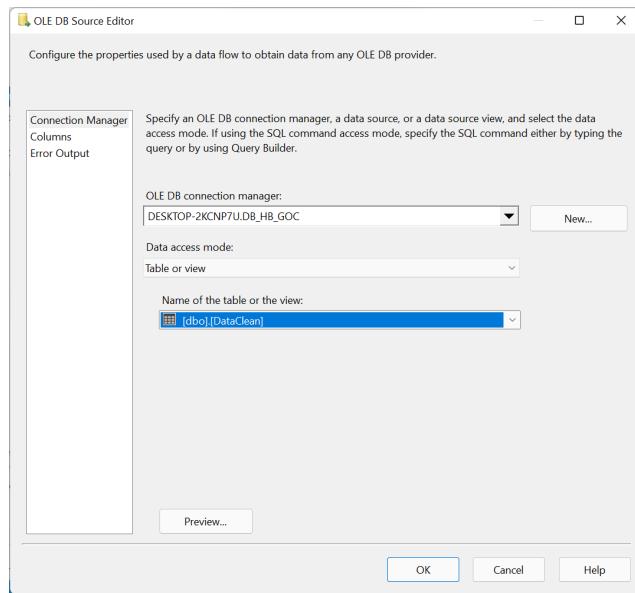
*Hình 2. 92.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimReservationStatus*

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



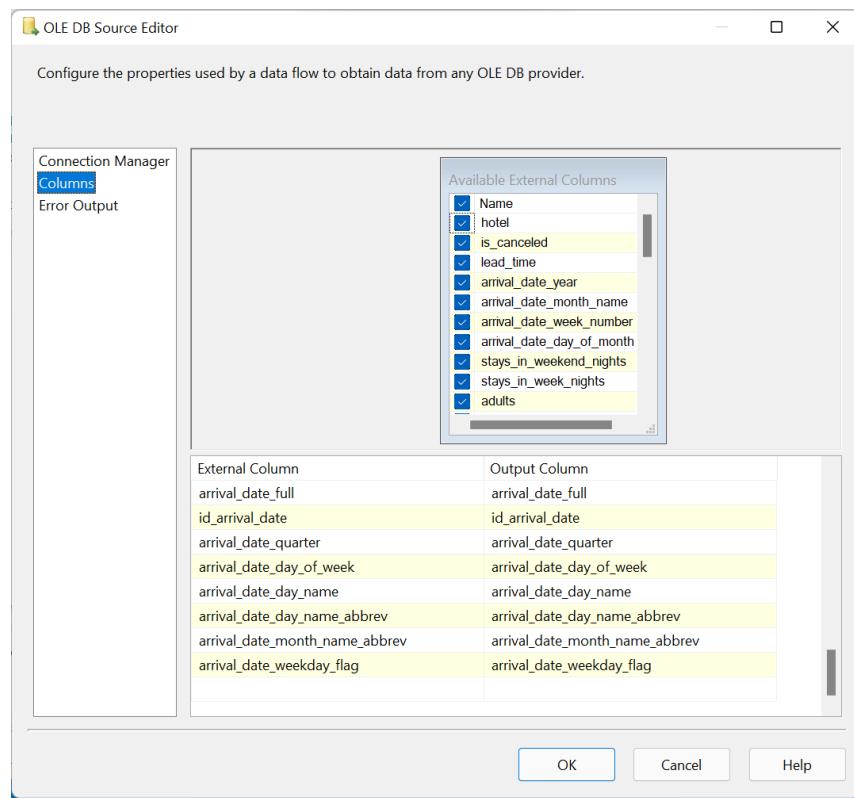
*Hình 2. 93.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



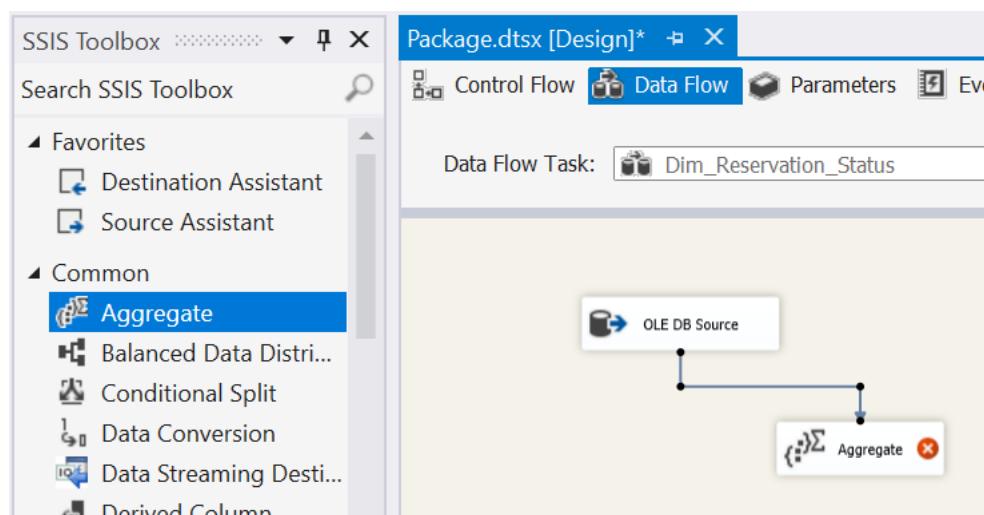
*Hình 2. 94.Chọn table sẽ lấy dữ liệu*

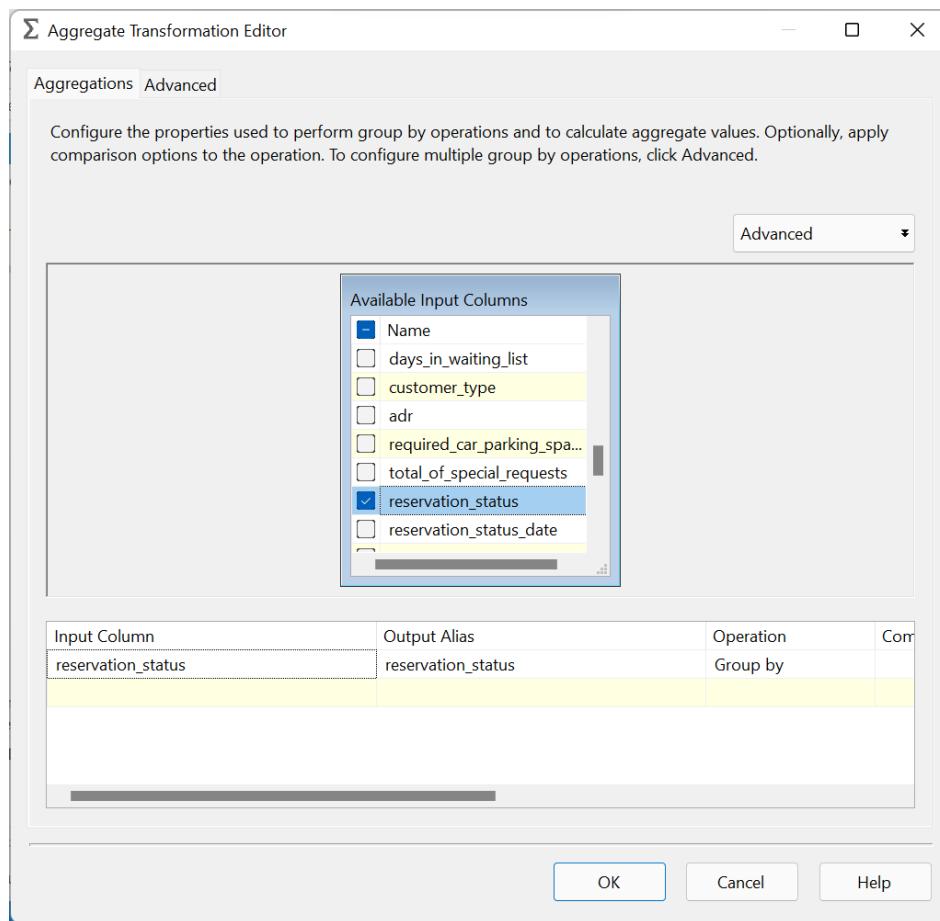
- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.



Hình 2. 95. Chọn các columns cần sử dụng

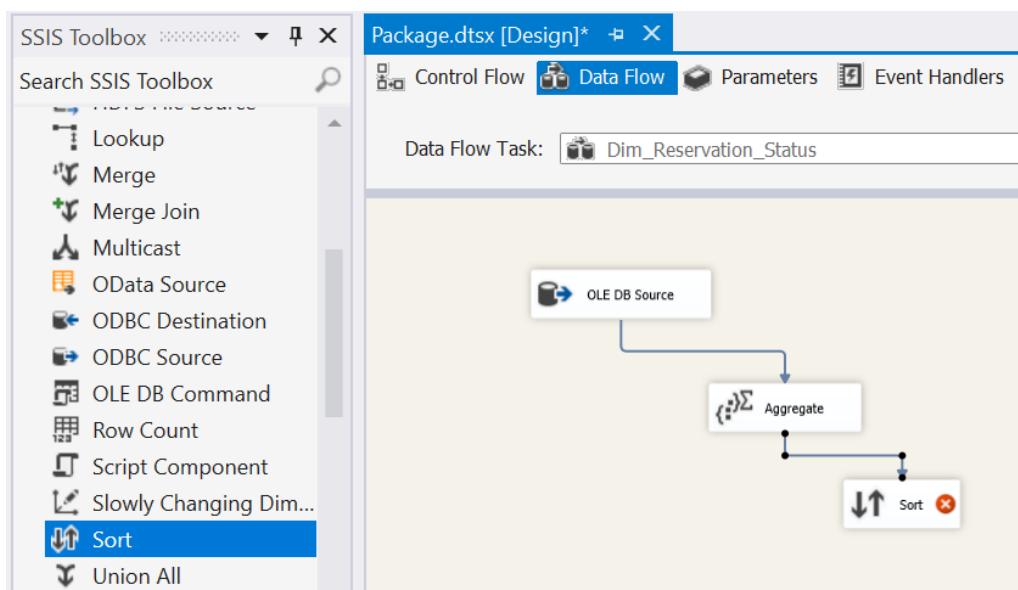
- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là reservation\_status.

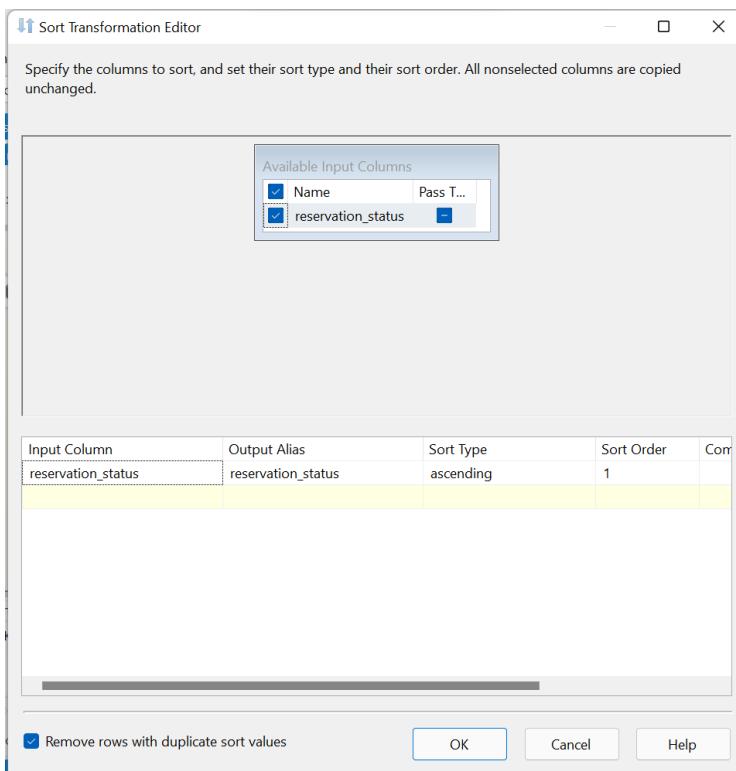




Hình 2. 96.1.,2.Dùng công cụ Aggregate để thực hiện lọc các thuộc tính trùng nhau

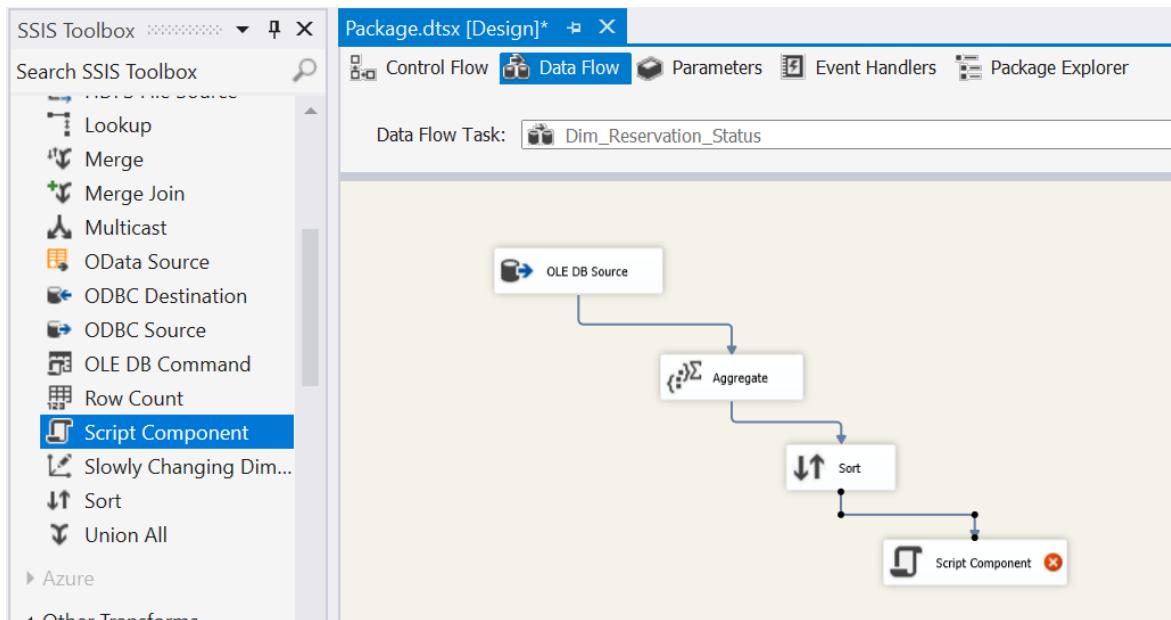
- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.



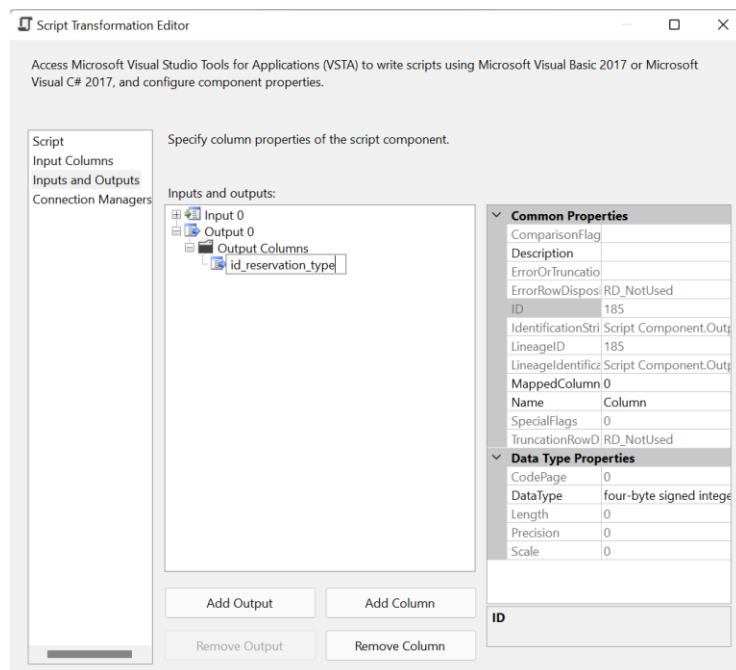
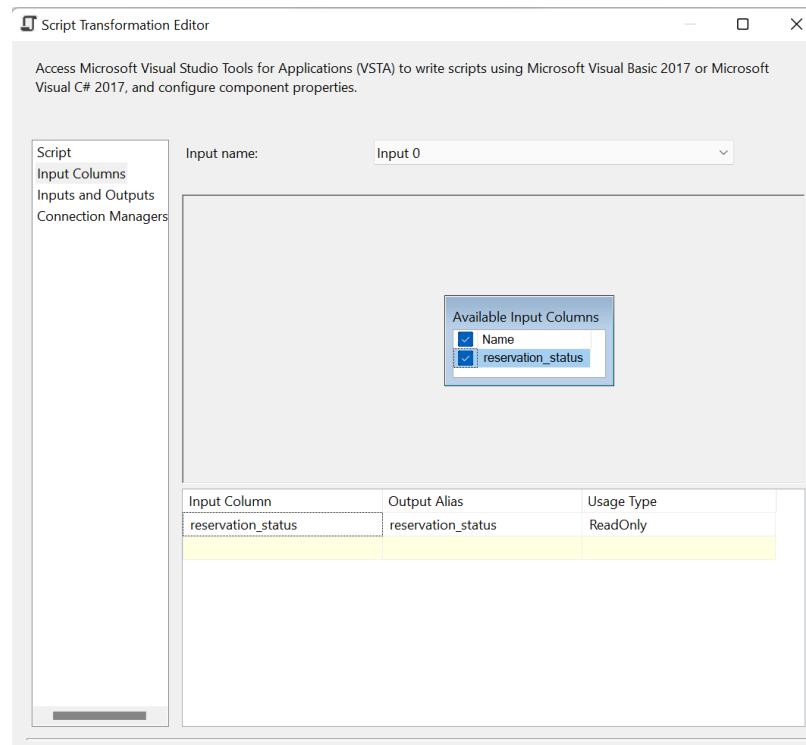


*Hình 2. 97. Sắp xếp các giá trị theo chiều tăng dần*

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_reservation\_status.



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

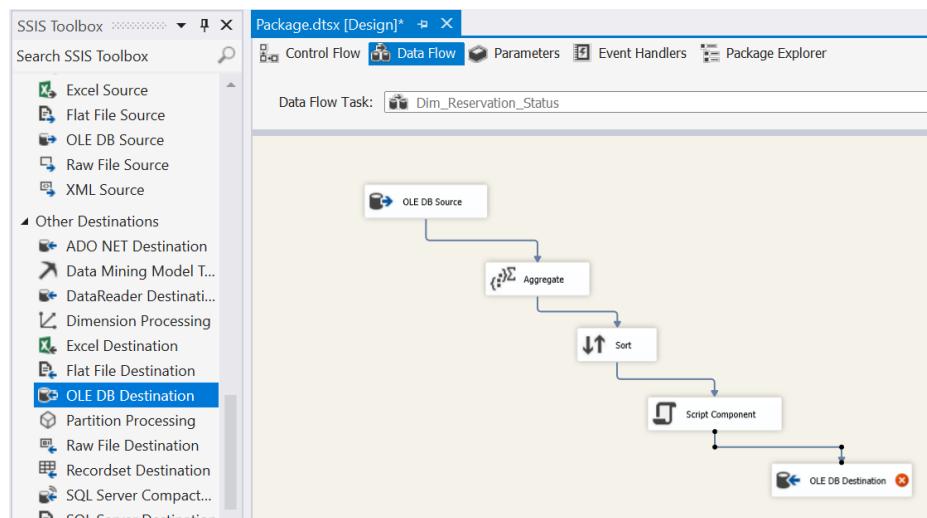
```

main.cs  ✎ X
C# SC_96f4a712e6ac4a0eb98c5834b1e  ScriptMain
18  /// of this class.
19  /// <summary>
20  [Microsoft.SqlServer.Dts.Pipeline.SSIScriptComponentEntryPointAttribute]
21  public class ScriptMain : UserComponent
22  {
23      Help: Using Integration Services variables and parameters
24
25      Help: Using Integration Services Connection Managers
26
27      Help: Firing Integration Services Events
28      int count = 1;
29
30      /// <summary>
31      /// This method is called once, before rows begin to be processed in the data flow.
32      /// You can remove this method if you don't need to do anything here.
33      /// </summary>
34
35      public override void PreExecute()
36      {
37          base.Preexecute();
38
39          /* Add your code here
40          */
41
42      }
43
44      /// <summary>
45      /// This method is called after all the rows have passed through this component.
46      /// You can delete this method if you don't need to do anything here.
47      /// </summary>
48
49      public override void PostExecute()
50      {
51          base.PostExecute();
52
53          /* Add your code here
54          */
55
56      }
57
58      /// <summary>
59      /// This method is called once for every row that passes through the component from Input0.
60      ///
61      /// Example of reading a value from a column in the row:
62      /// string zipCode = Row.ZipCode
63
64      /// Example of writing a value to a column in the row:
65      /// Row.ZipCode = zipCode
66      /// <summary>
67      /// <param name="Row">The row that is currently passing through the component</param>
68      /// </summary>
69
70      public override void Input0_ProcessInputRow(Input0Buffer Row)
71      {
72
73          /*
74          * Add your code here
75          */
76
77          Row.idreservationtype = count;
78
79          count++;
80
81      }
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
}

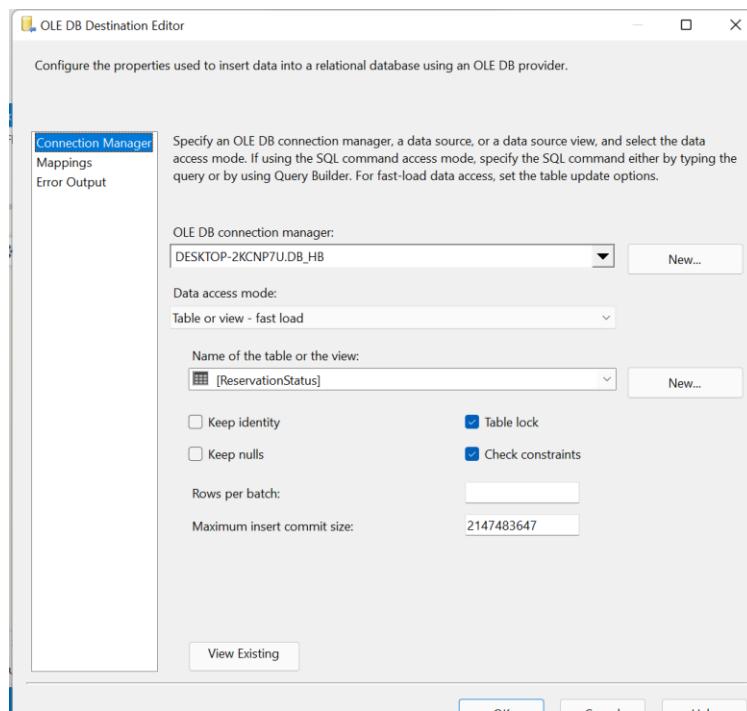
```

Hình 2. 98.1,.2.,3, 4.Sử dụng Script Component để tạo khóa chính id\_reservation\_status

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là ReservationStatus.

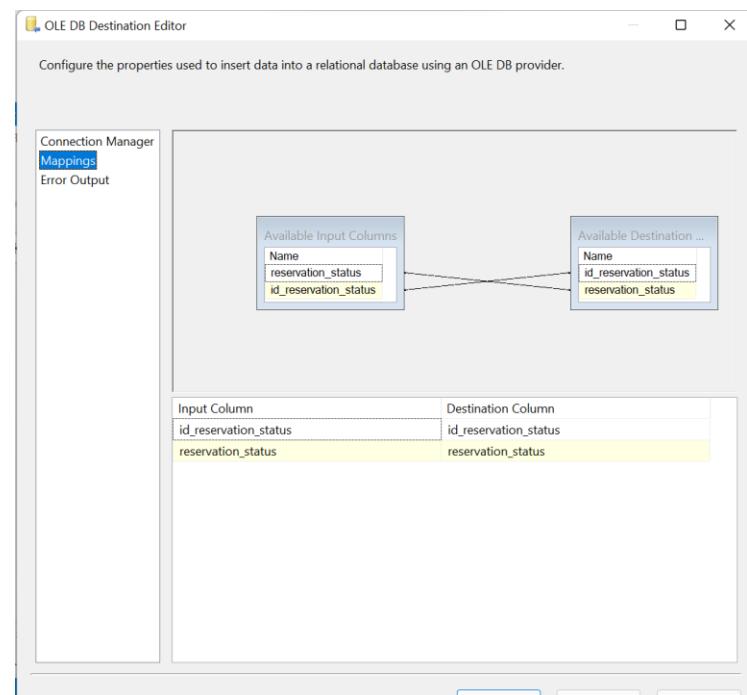


## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



*Hình 2. 99.Tạo bảng ReservationStatus*

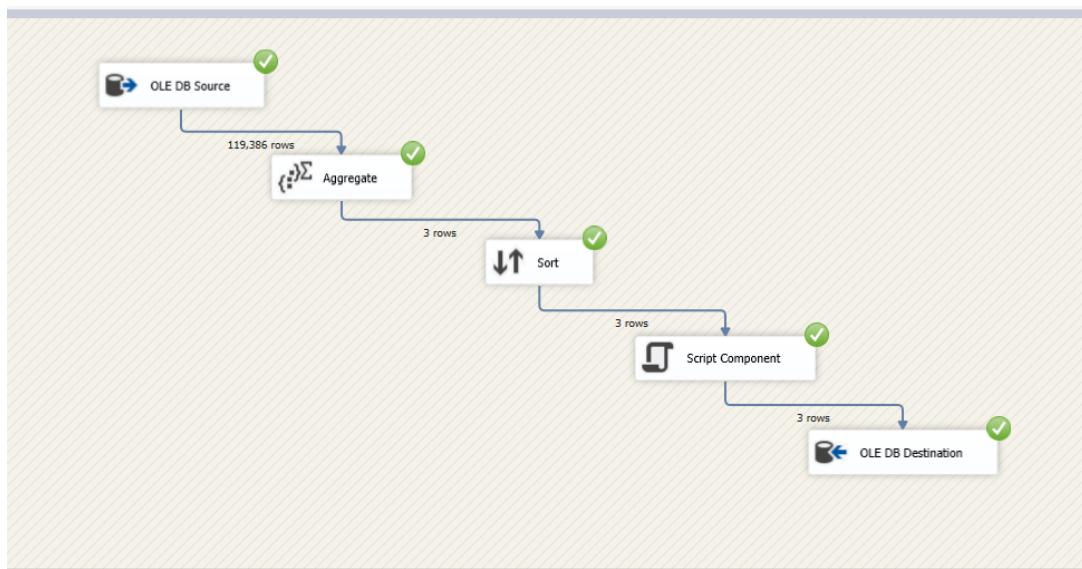
- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



*Hình 2. 100.Quá trình Mappings dữ liệu bảng ReservationStatus.*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 101.Hoàn thành đổ dữ liệu vào ReservationStatus trong kho dữ liệu*

- Bước 11: Kiểm tra bảng ReservationStatus trên SQL Server.

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the database "DB\_HB" is selected. In the center pane, a query window displays the following SQL code and results:

```
select * from ReservationStatus;
```

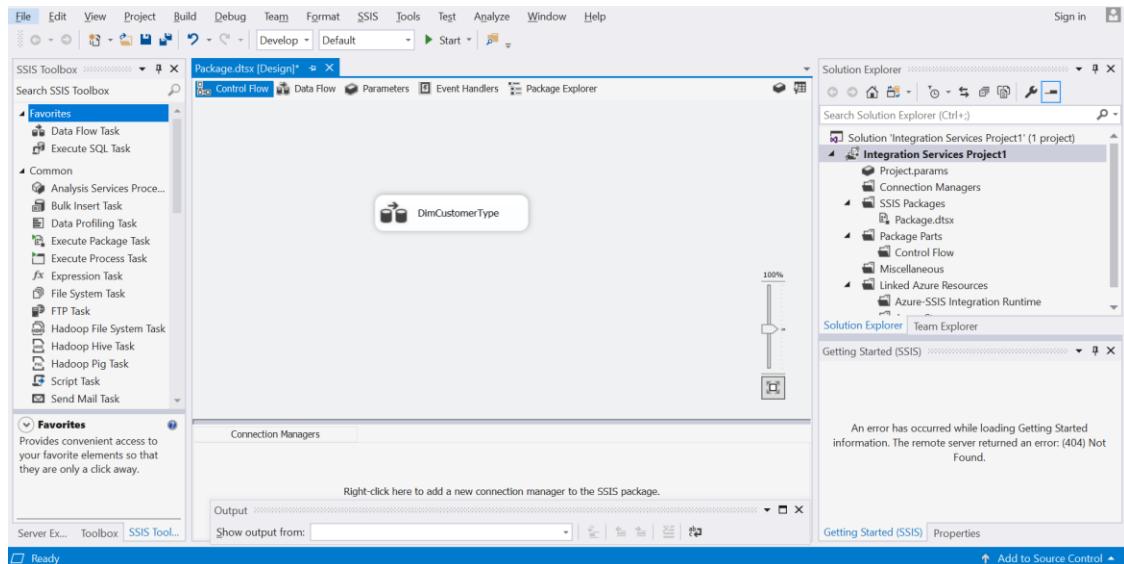
	id_reservation_status	reservation_status
1	1	Canceled
2	2	Check-Out
3	3	No-Show

The status bar at the bottom indicates "Query executed successfully." and "3 rows".

*Hình 2. 102.Kiểm tra bảng ReservationStatus trên SQL Server*

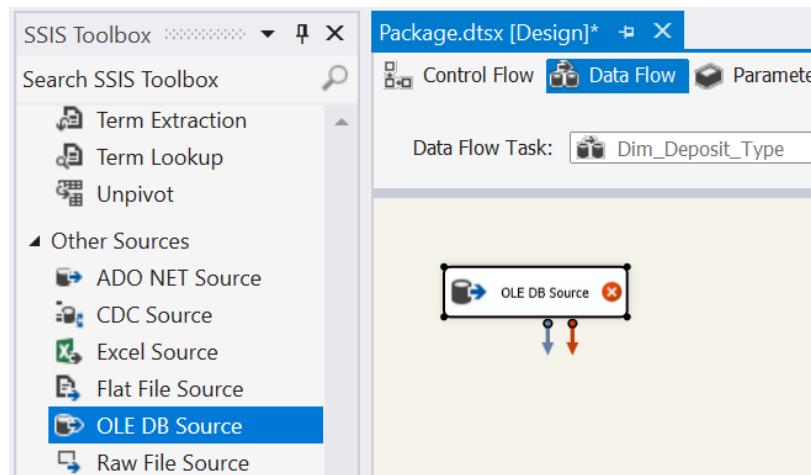
## 5.6. Bảng DimCustomerType

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimCustomerType.



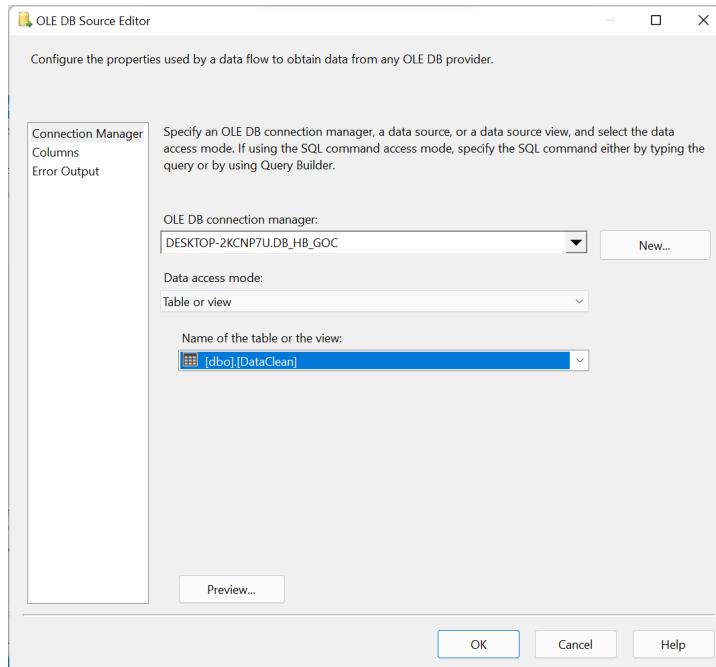
*Hình 2. 103.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimCustomerType*

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



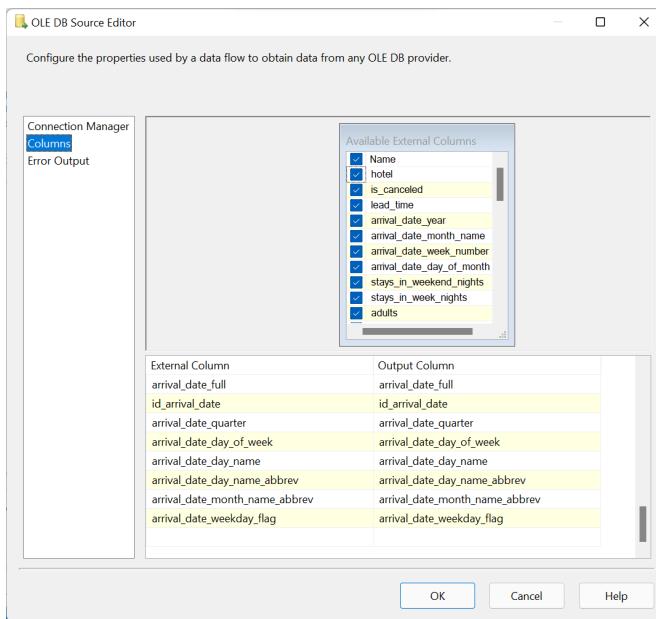
*Hình 2. 104 .Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



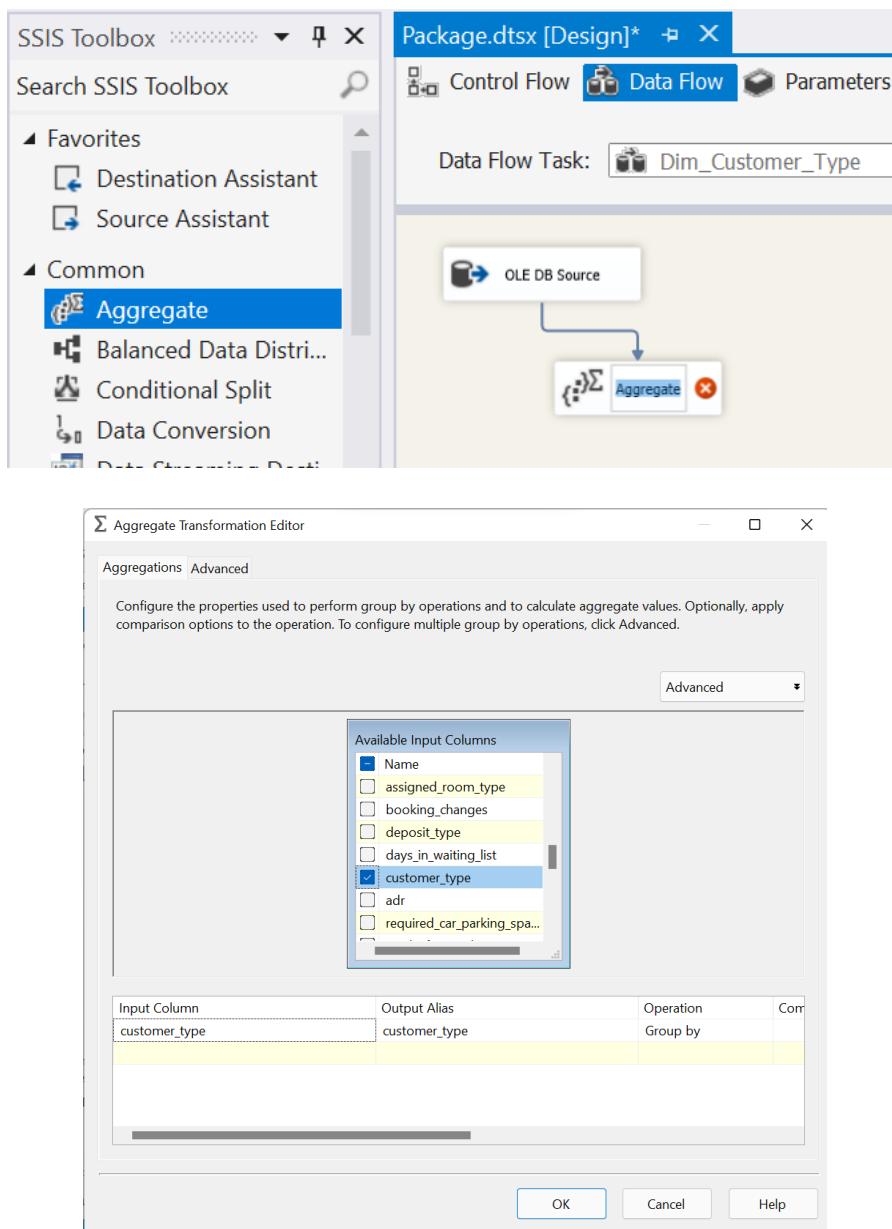
*Hình 2. 105 .Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.



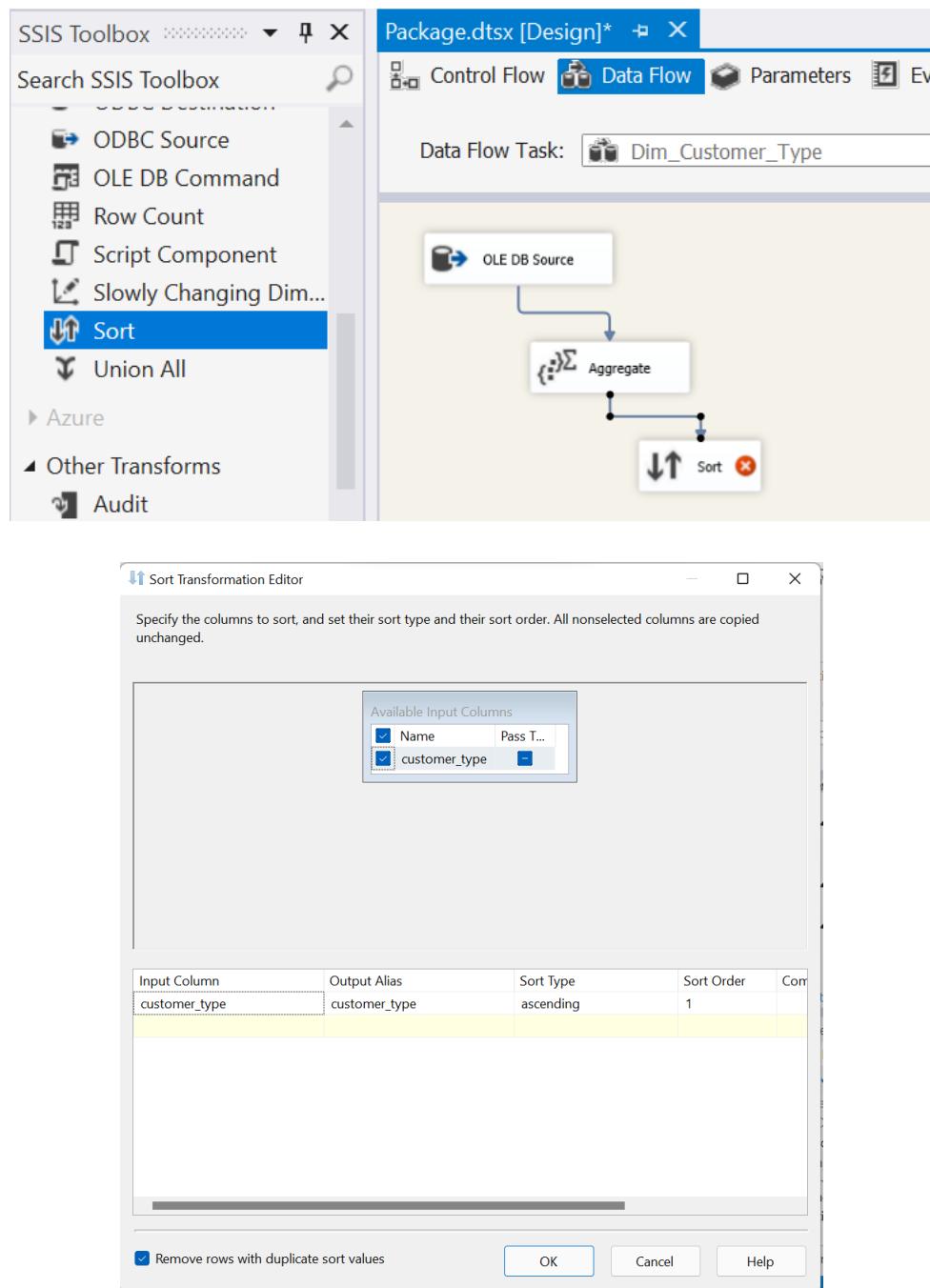
*Hình 2. 106.Chọn các columns cần sử dụng*

- Bước 5: Chọn các thuộc tính cần gom nhóm lên thuộc tính sử dụng là `customer_type`.



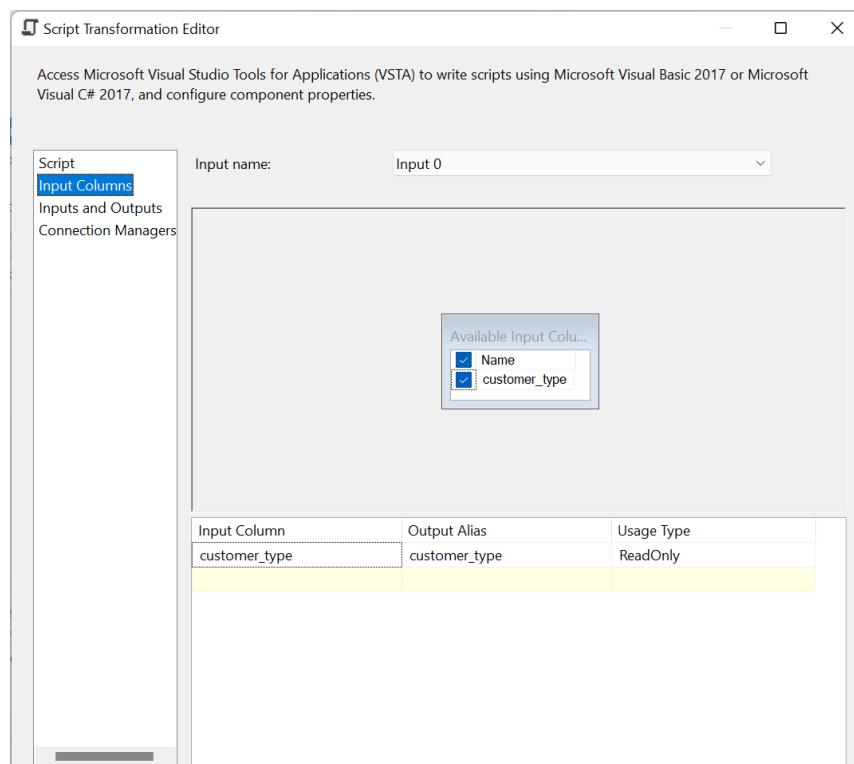
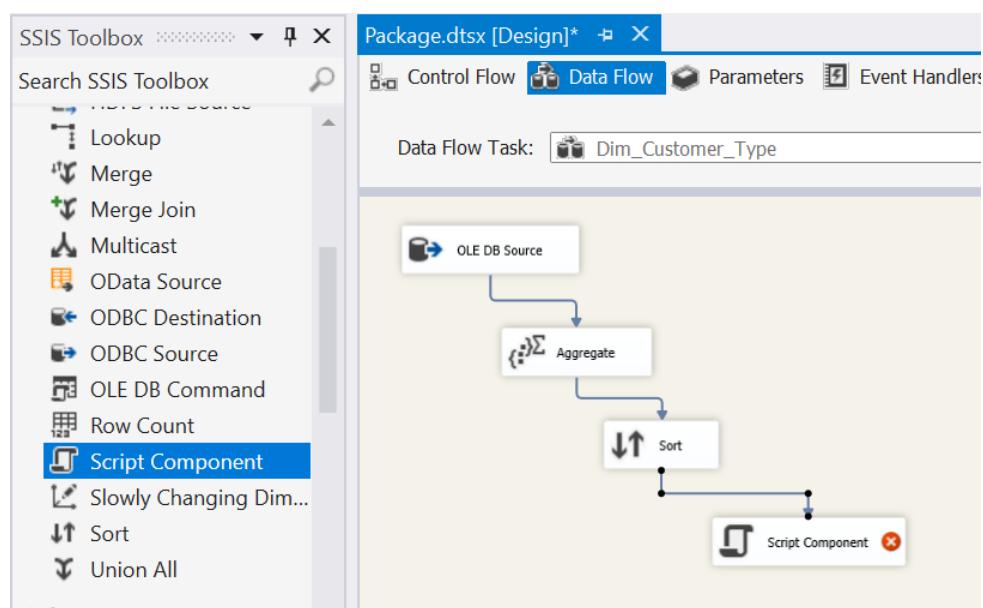
*Hình 2. 107.1,.2.Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau*

- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.

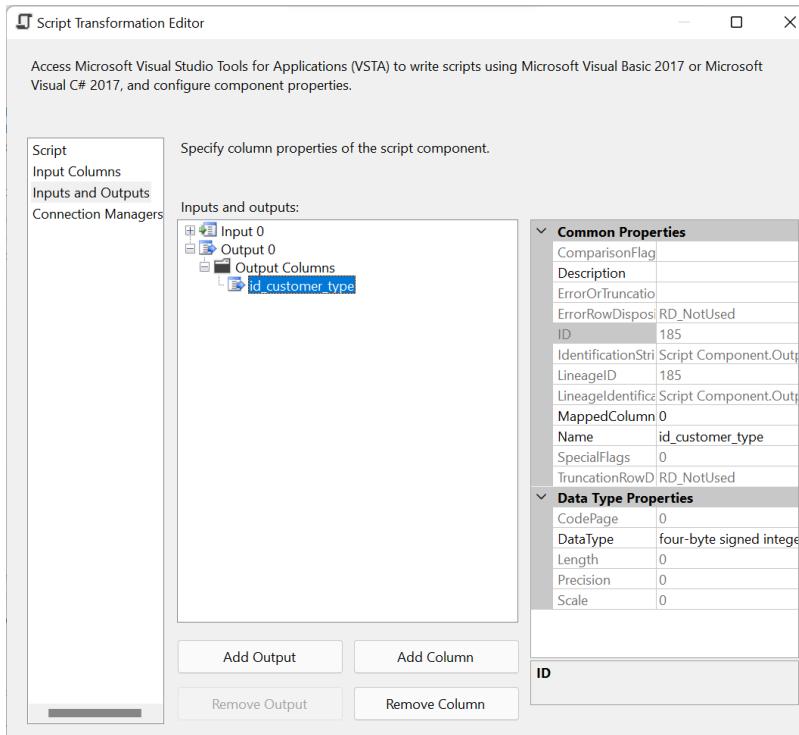


Hình 2. 108. Sắp xếp các giá trị theo chiều tăng dần

- Bước 7: Kéo thả công cụ Script Component để tạo khóa chính id\_customer\_type.



## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



main.cs\*

**SC\_13db605322ee4651a09e383781** ▾ **ScriptMain**

```

public class ScriptMain : UserComponent
{
    Help: Using Integration Services variables and parameters
    Help: Using Integration Services Connection Managers
    Help: Firing Integration Services Events
    int count = 1;
    /// <summary>
    /// This method is called once, before rows begin to be processed in the data flow.
    ///
    /// You can remove this method if you don't need to do anything here.
    /// </summary>
    public override void PreExecute()
    {
        base.PreExecute();
        /*
         * Add your code here
         */
    }

    /// <summary>
    /// This method is called after all the rows have passed through this component.
    ///
    /// You can delete this method if you don't need to do anything here.
    /// </summary>
    public override void PostExecute()
    {
        base.PostExecute();
        /*
         * Add your code here
         */
    }

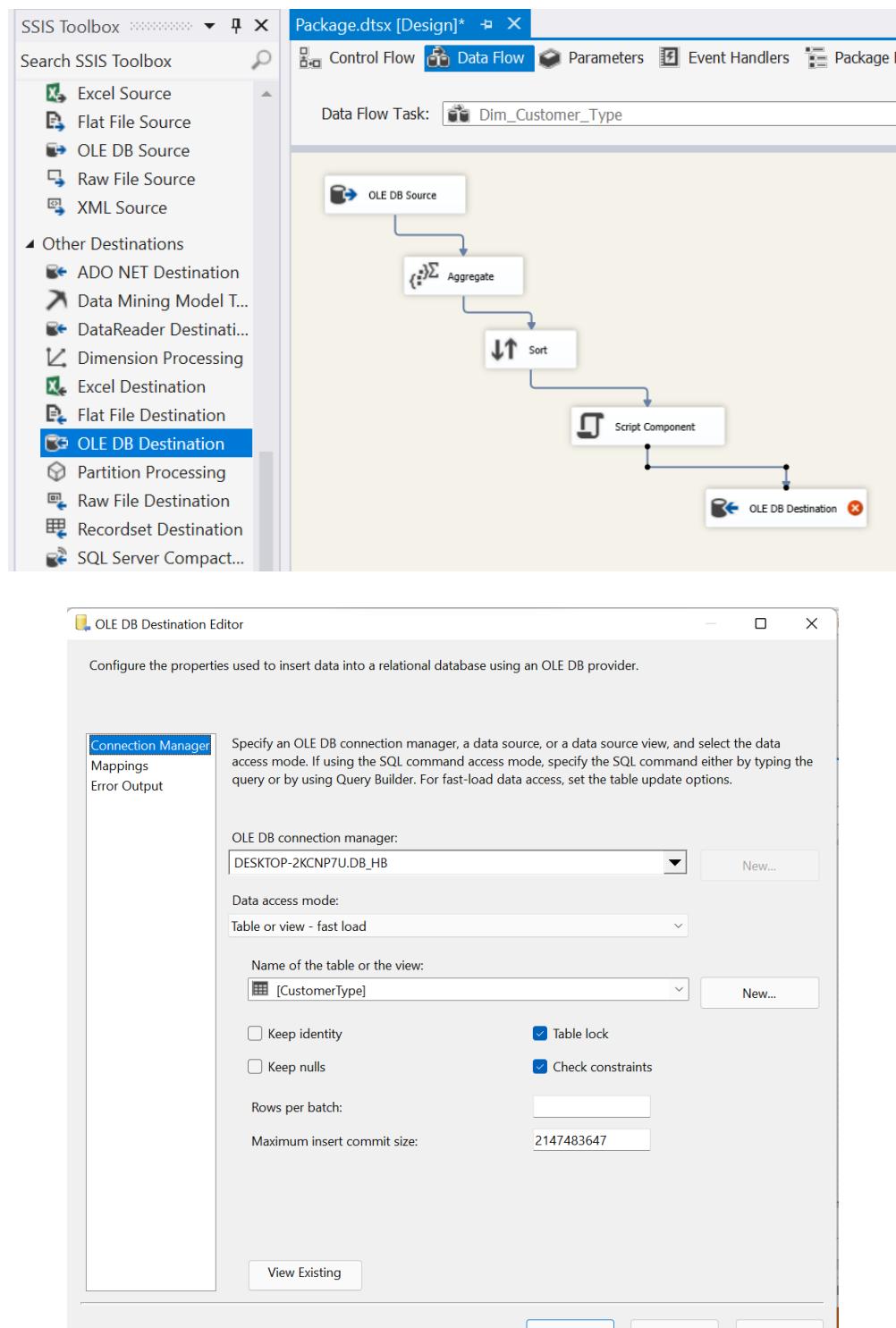
    /// <summary>
    /// This method is called once for every row that passes through the component from Input0.
    ///
    /// Example of reading a value from a column in the the row:
    /// string zipCode = Row.ZipCode
    ///
    /// Example of writing a value to a column in the row:
    /// Row.ZipCode = zipCode
    /// </summary>
    /// <param name="Row">The row that is currently passing through the component</param>
    public override void Input0_ProcessInputRow(Input0Buffer Row)
    {
        Row.idcustomer_type = count;
        count++;
    }
}

```

Hình 2. 109.1.,2.,3.,4.Sử dụng Script Component để tạo khóa chính id\_customer\_type

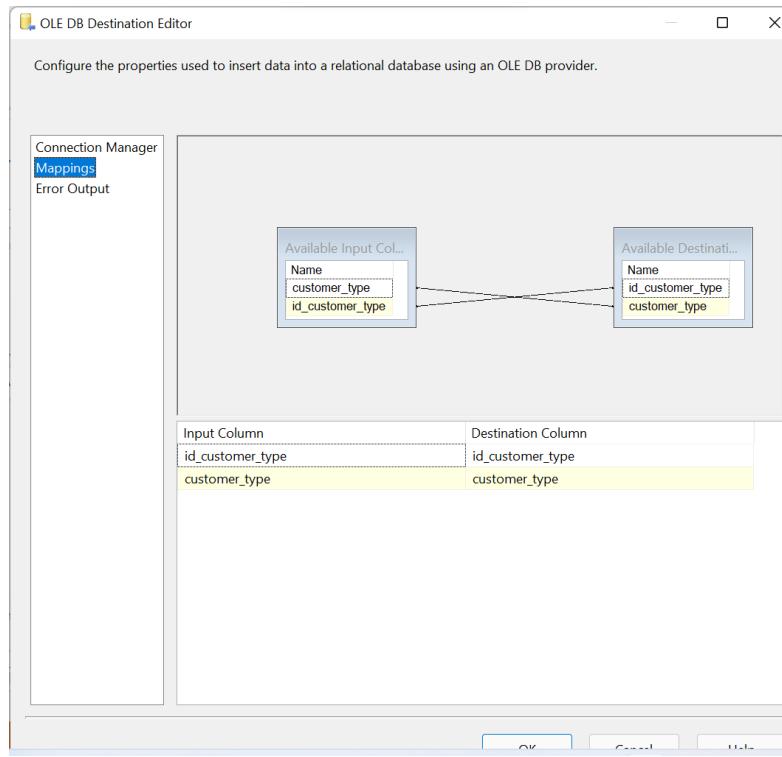
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là CustomerType.



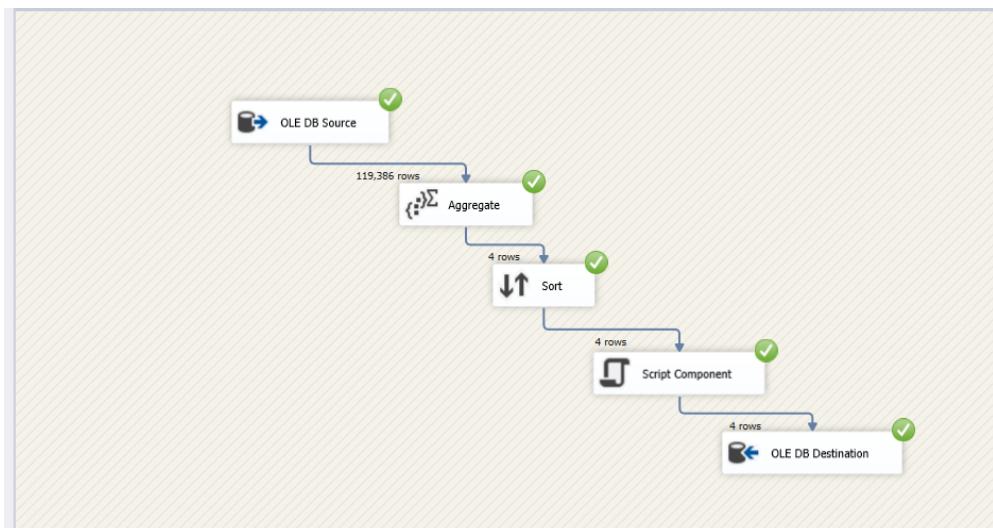
Hình 2. 110.1,.2.Tạo bảng CustomerType

- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



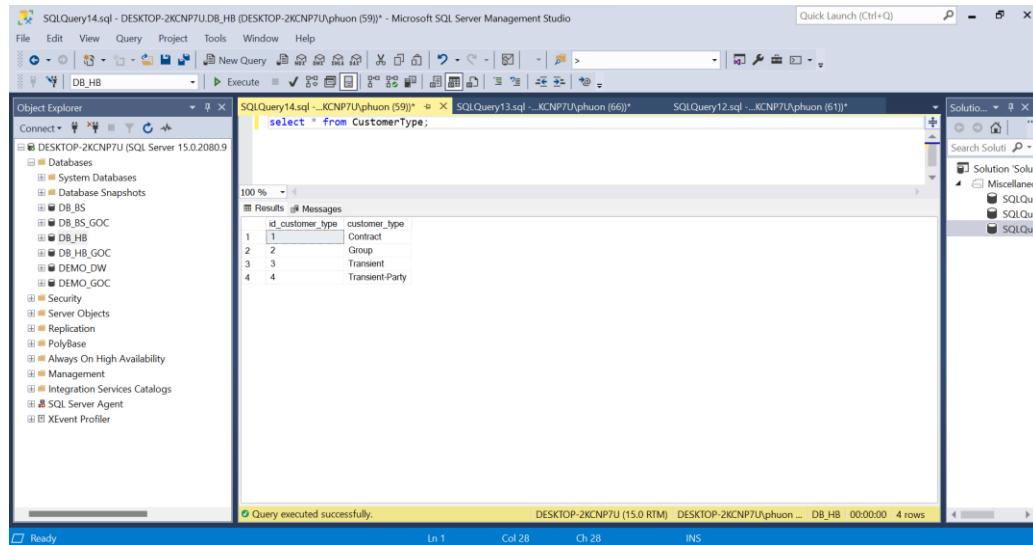
*Hình 2. 111.Quá trình Mappings dữ liệu bảng CustomerType*

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 112..Hoàn thành đổ dữ liệu vào CustomerType trong kho dữ liệu*

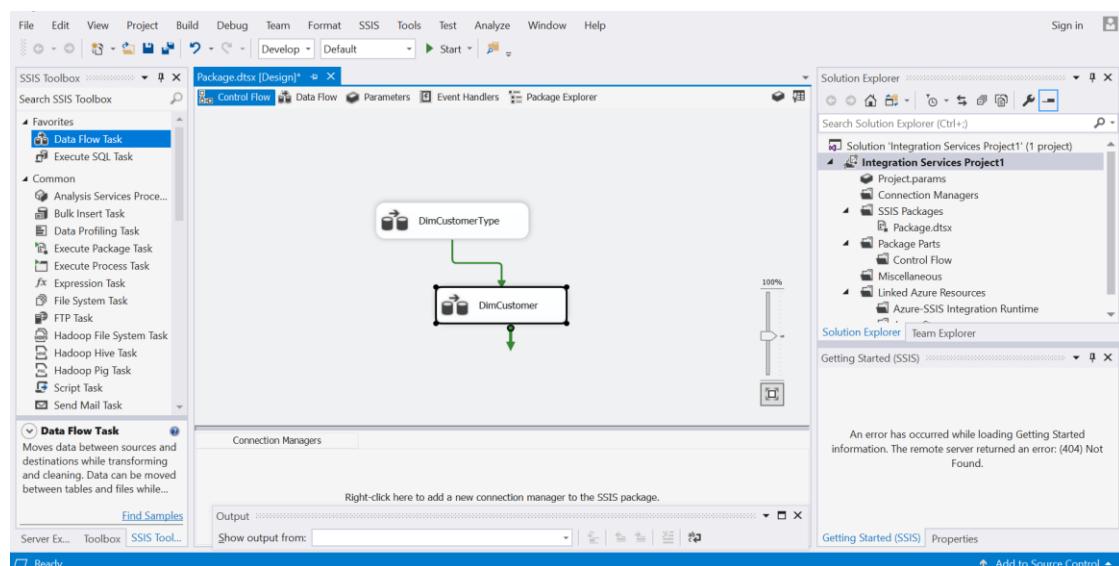
- Bước 11: Kiểm tra bảng CustomerType trên SQL Server.



Hình 2. 113.Kiểm tra bảng CustomerType trên SQL Server

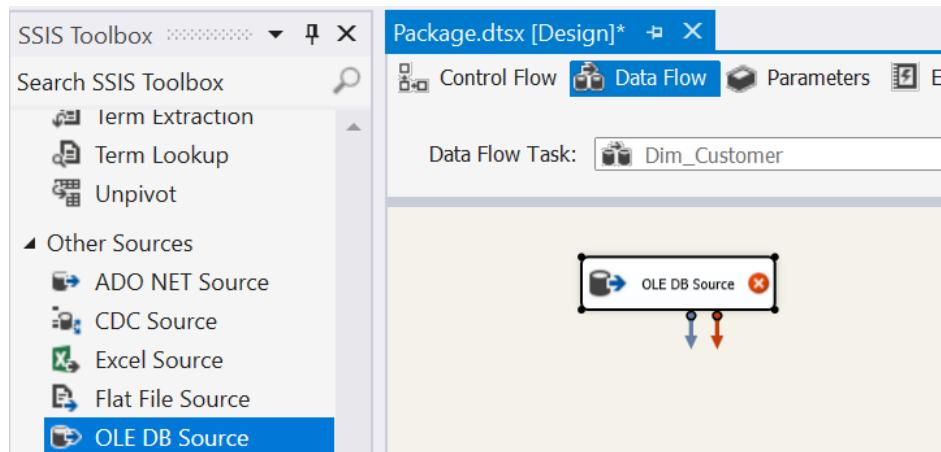
### 5.7. Bảng DimCustomer

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimCustomer và tạo đường liên kết từ DimCustomerType đến DimCustomer để thực hiện theo thứ tự lần lượt là DimCustomerType và DimCustomer.



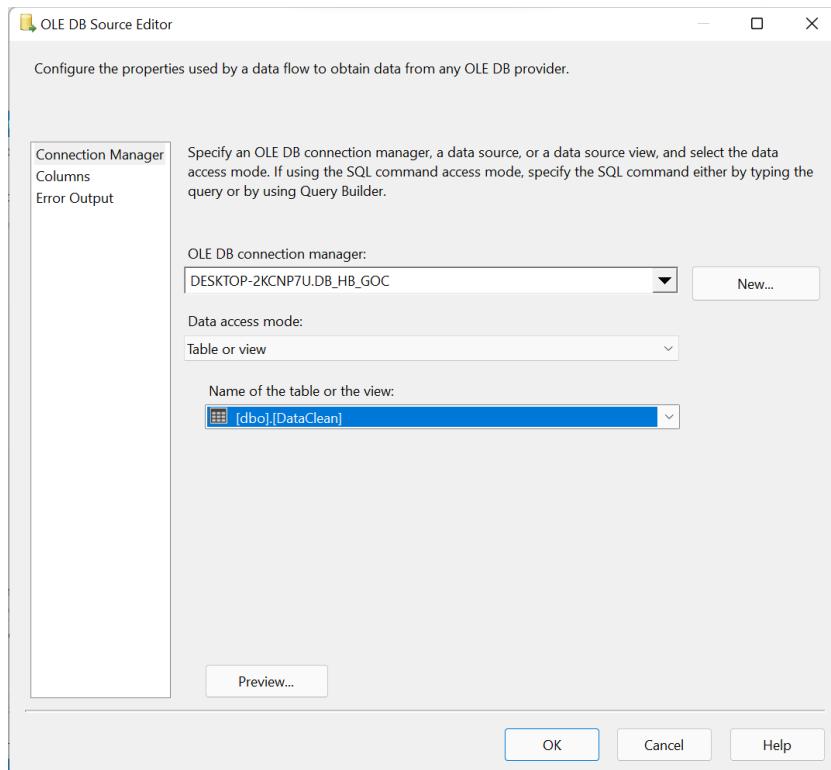
Hình 2. 114.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimCustomer

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



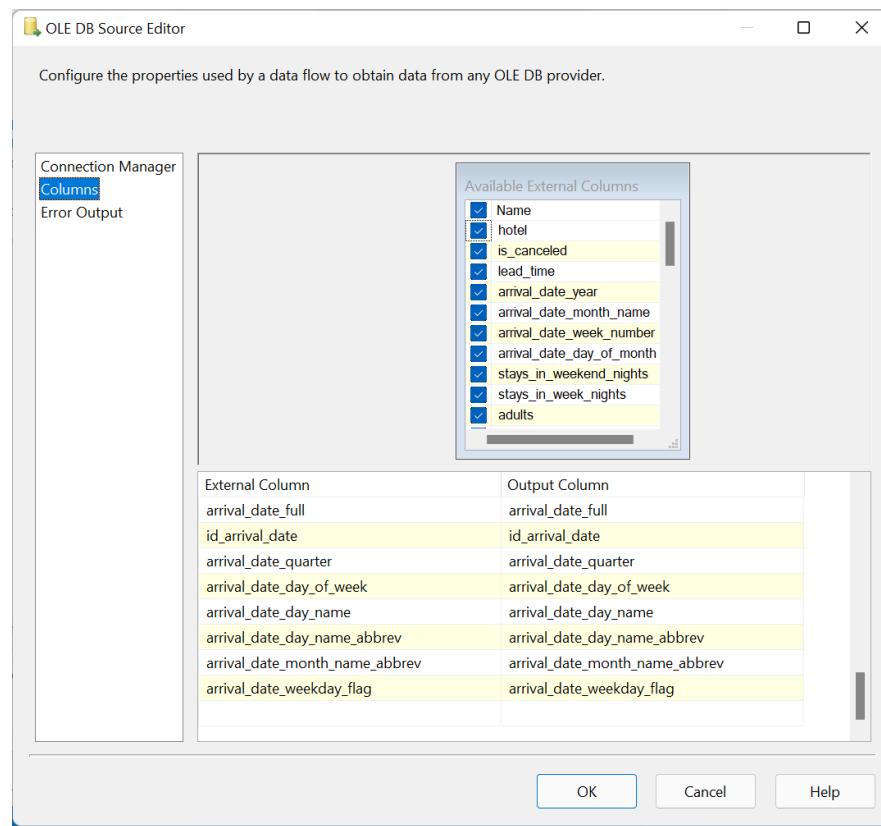
*Hình 2. 115.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



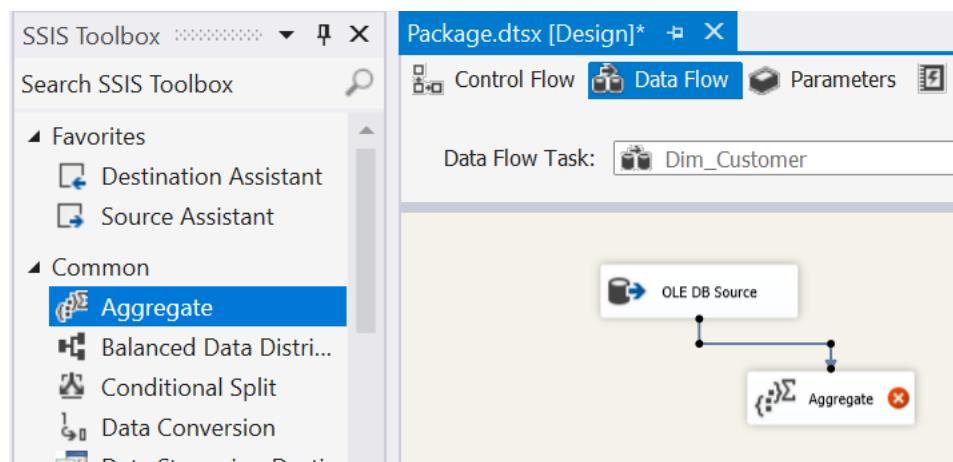
*Hình 2. 116.Chọn table sẽ lấy dữ liệu*

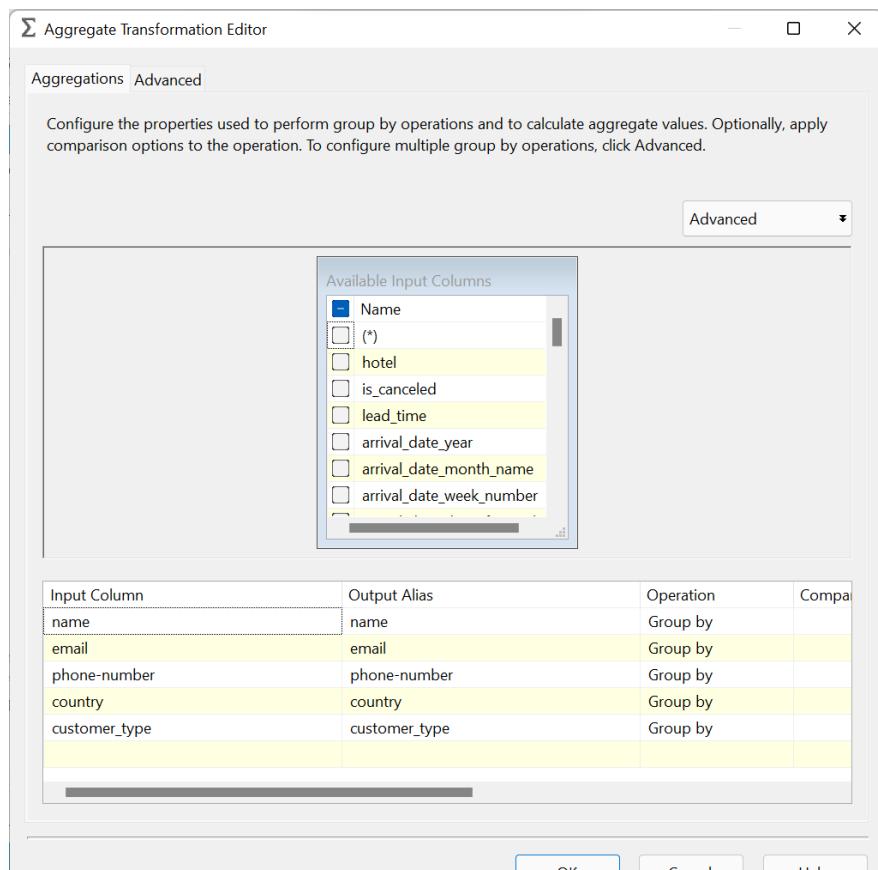
- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.



*Hình 2. 117.Chọn các columns cần sử dụng*

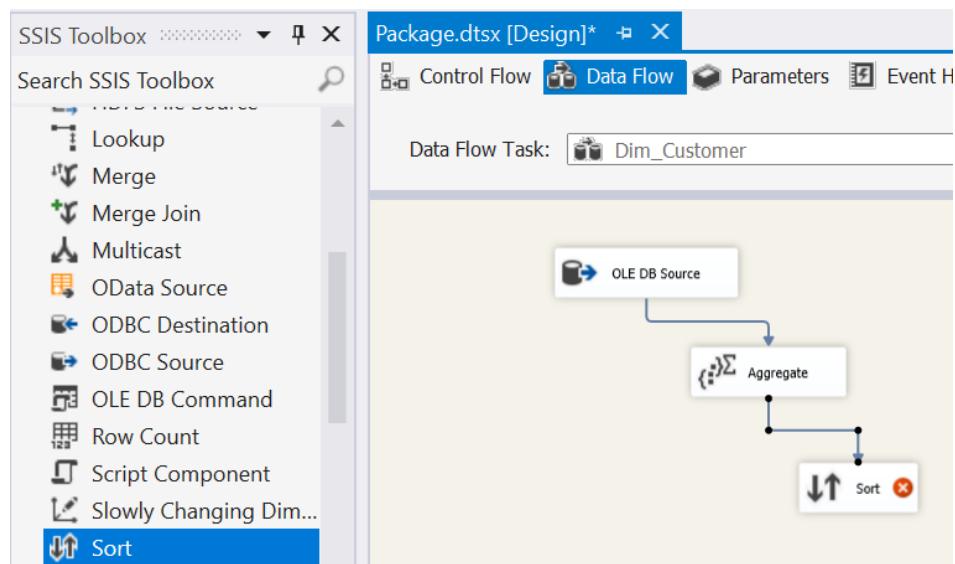
- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là name, email, phone-number, country, customer\_type.

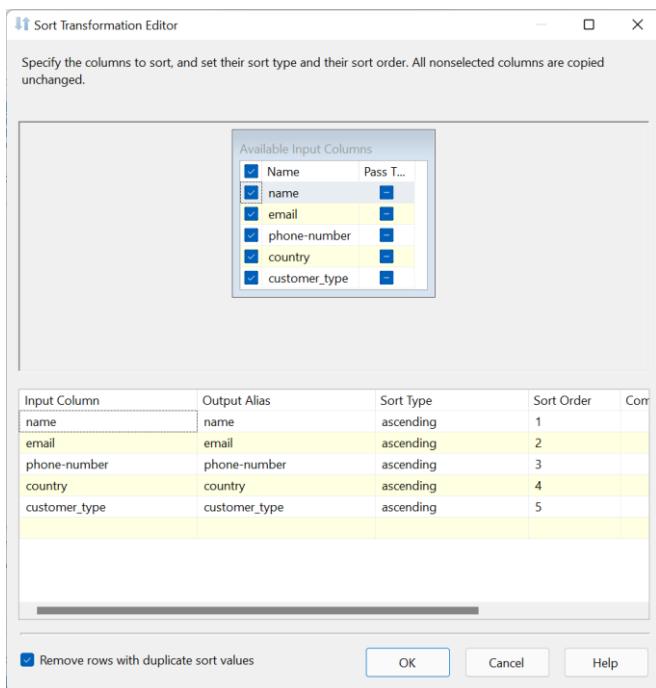




*Hình 2. 118.1.,2.Chọn các thuộc tính cần gom nhóm*

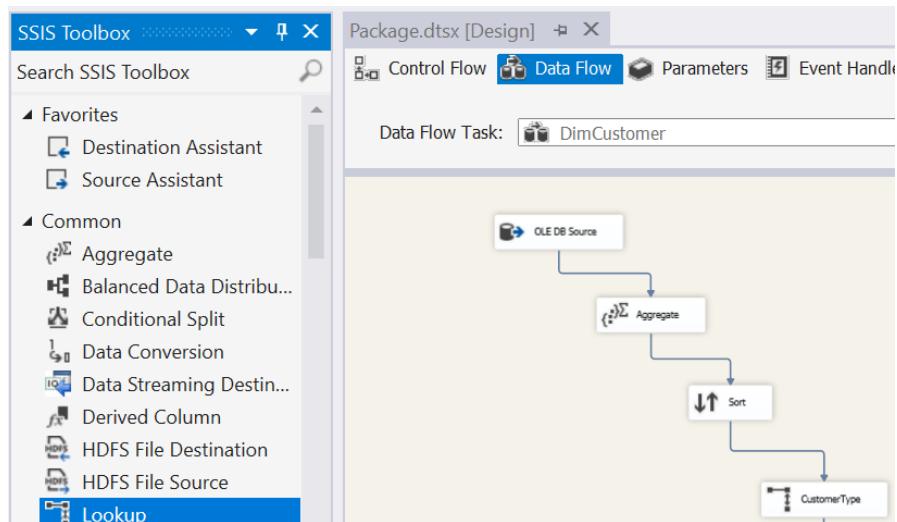
- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.



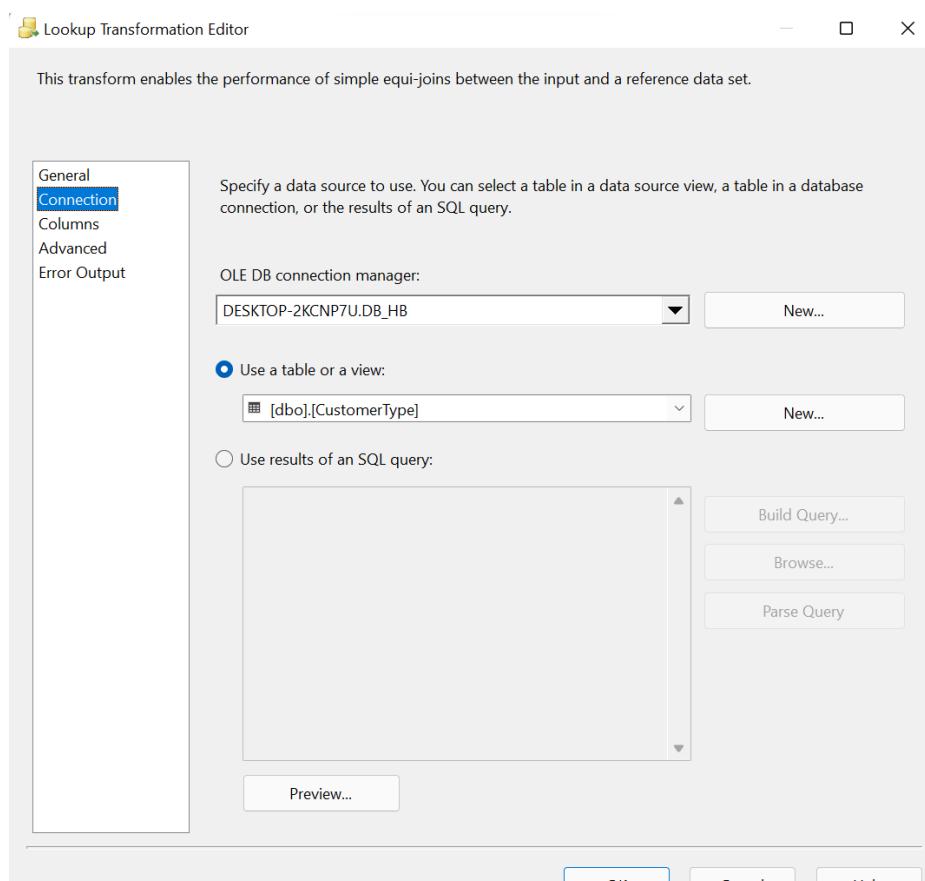
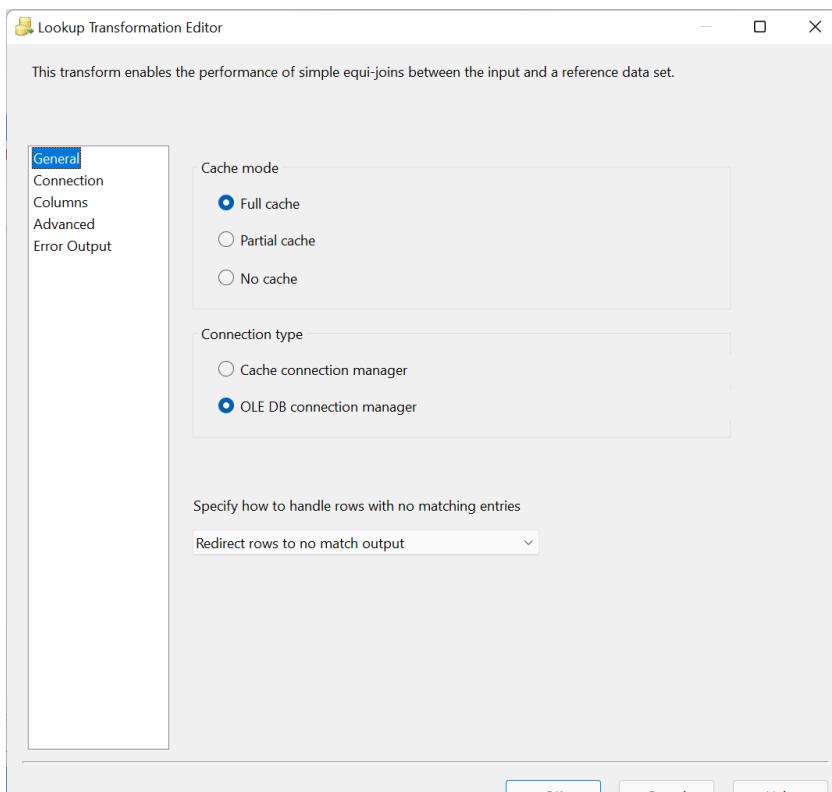


*Hình 2. 119.1.,2.Sắp xếp các giá trị theo chiều tăng dần*

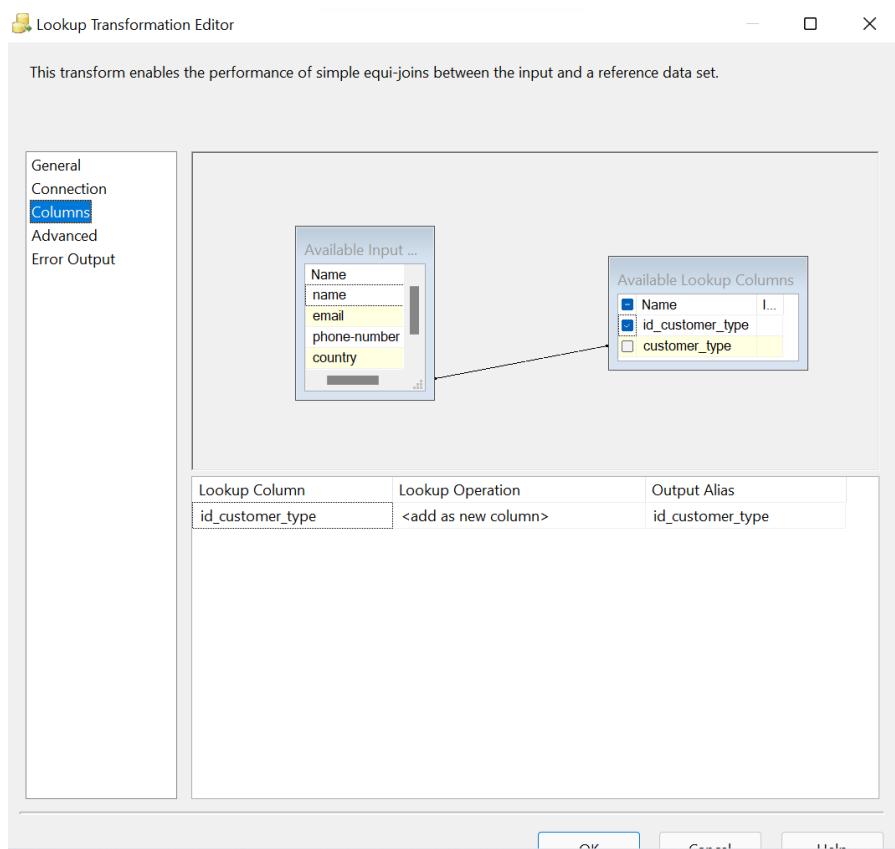
- Bước 7: Tìm kiếm công cụ Lookup trên SSIS Toolbox và đổi tên thành CustomerType để tạo khóa ngoại id\_customer\_type đến bảng Customer.



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

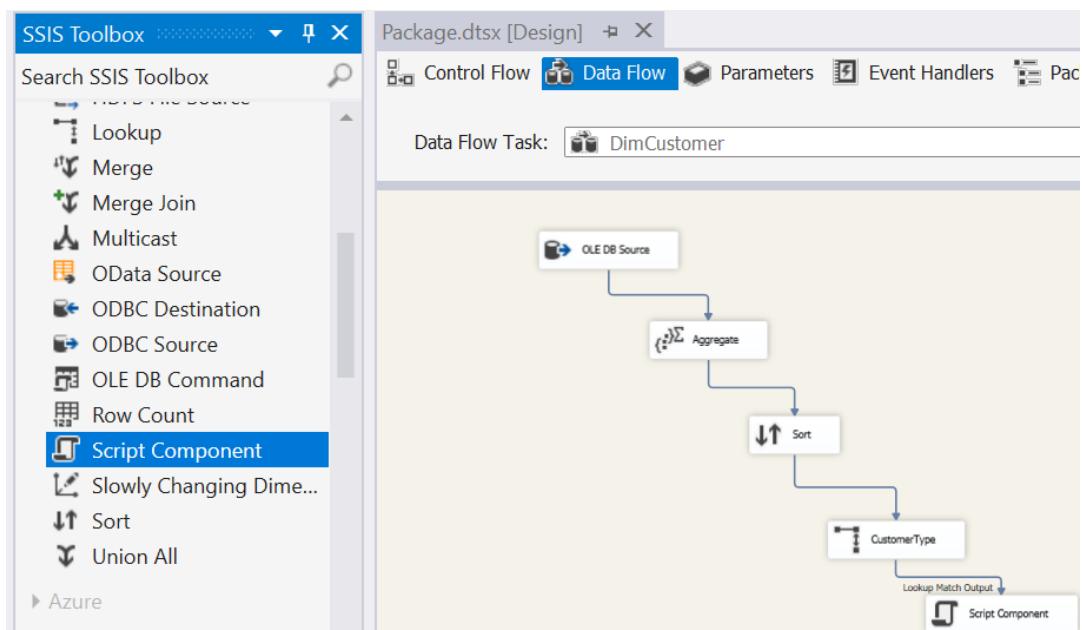


## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

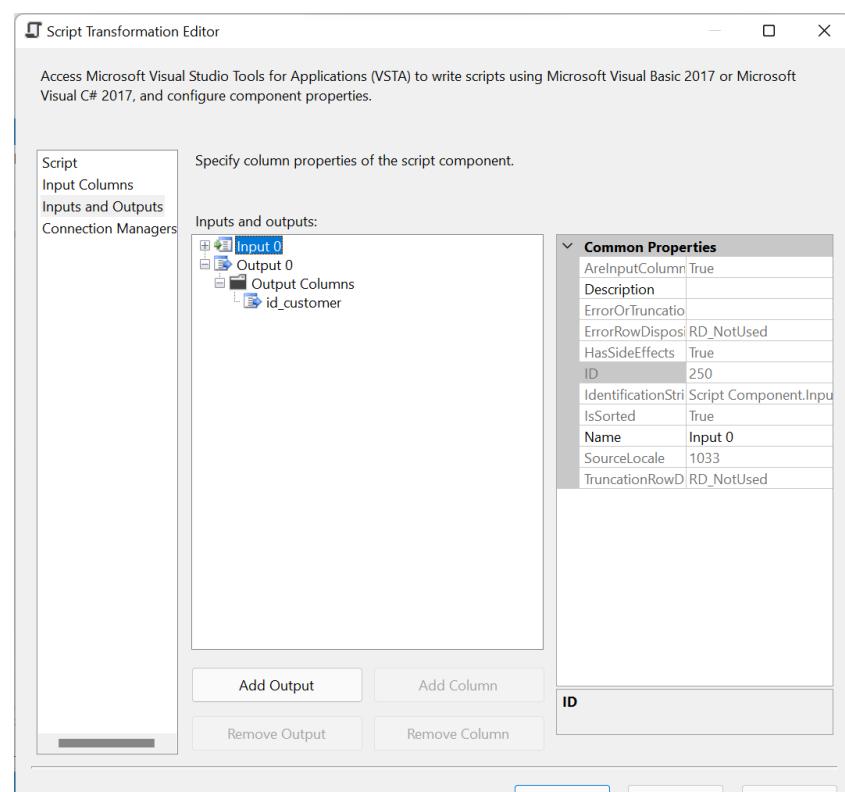
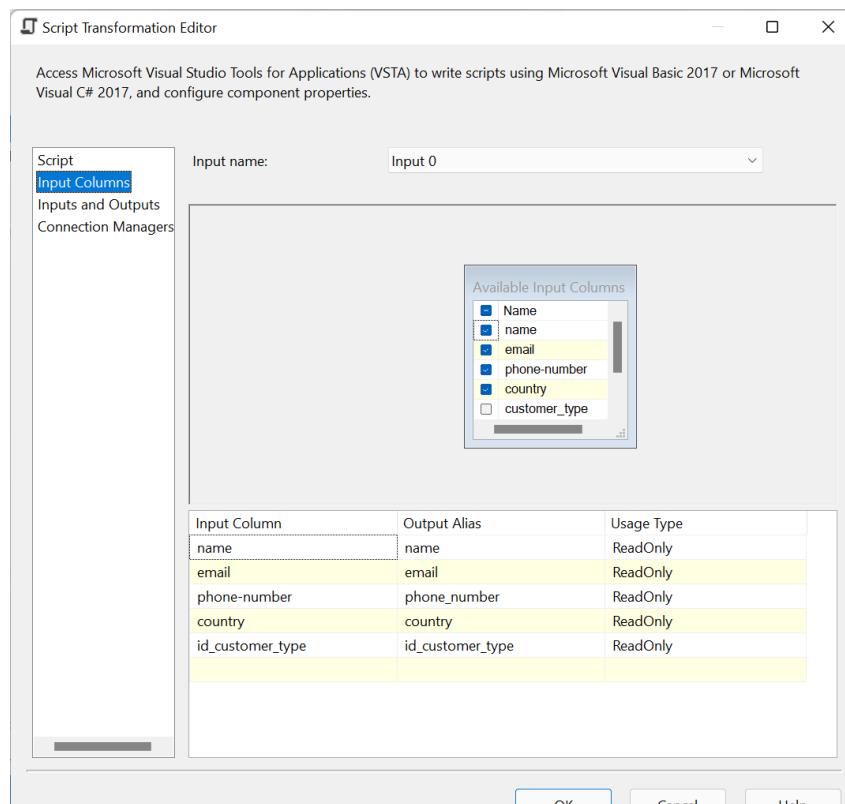


Hình 2. 120.1.,2.,3.,4.Sử dụng Lookup để tạp khóa ngoại đến bảng Customer

- Bước 8: Kéo thả công cụ Script Component để tạo khóa chính id\_customer.



## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



```

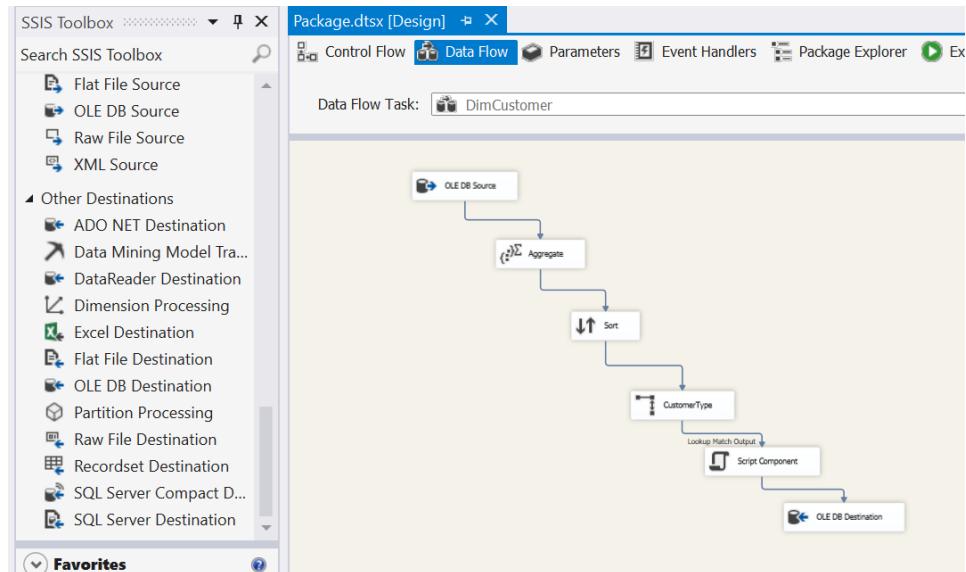
main.cs*  X
SC_35199b012973434bb17e529d2;  ScriptMain  Input0_Prc

18  /// of this class.
19  /// <summary>
20  /// [Microsoft.SqlServer.Dts.Pipeline.SSIScriptComponentEntryPointAttribute]
21  /// </summary>
22  public class ScriptMain : UserComponent
23  {
24      [Help: Using Integration Services variables and parameters]
25      [Help: Using Integration Services Connection Managers]
26      [Help: Firing Integration Services Events]
27      int count = 1;
28      /// <summary>
29      /// This method is called once, before rows begin to be processed in the data flow.
30      ///
31      /// You can remove this method if you don't need to do anything here.
32      /// </summary>
33      protected override void PreExecute()
34      {
35          base.PreExecute();
36          /*
37          * Add your code here
38          */
39      }
40
41      /// <summary>
42      /// This method is called after all the rows have passed through this component.
43      ///
44      /// You can delete this method if you don't need to do anything here.
45      /// </summary>
46      protected override void PostExecute()
47      {
48          base.PostExecute();
49          /*
50          * Add your code here
51          */
52      }
53
54      /// <summary>
55      /// This method is called once for every row that passes through the component from Input0.
56      ///
57      /// Example of reading a value from a column in the row:
58      /// string zipCode = Row.ZipCode
59      ///
60      /// Example of writing a value to a column in the row:
61      /// Row.ZipCode = zipCode
62      /// <summary>
63      /// <param name="Row">The row that is currently passing through the component</param>
64      /// </summary>
65      public override void Input0_ProcessInputRow(Input0Buffer Row)
66      {
67          Row.IdCustomer = count;
68          count++;
69      }
70  }

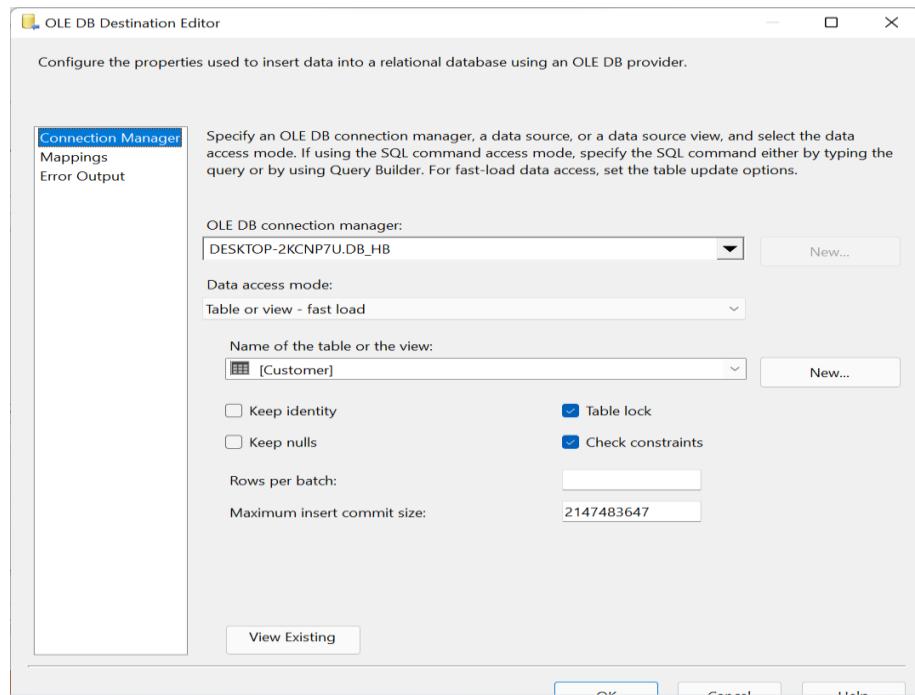
```

Hình 2. 121.1.,2.,3.,4. Sử dụng Script Component để tạo khóa chính id\_customer

- Bước 9: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là Customer.

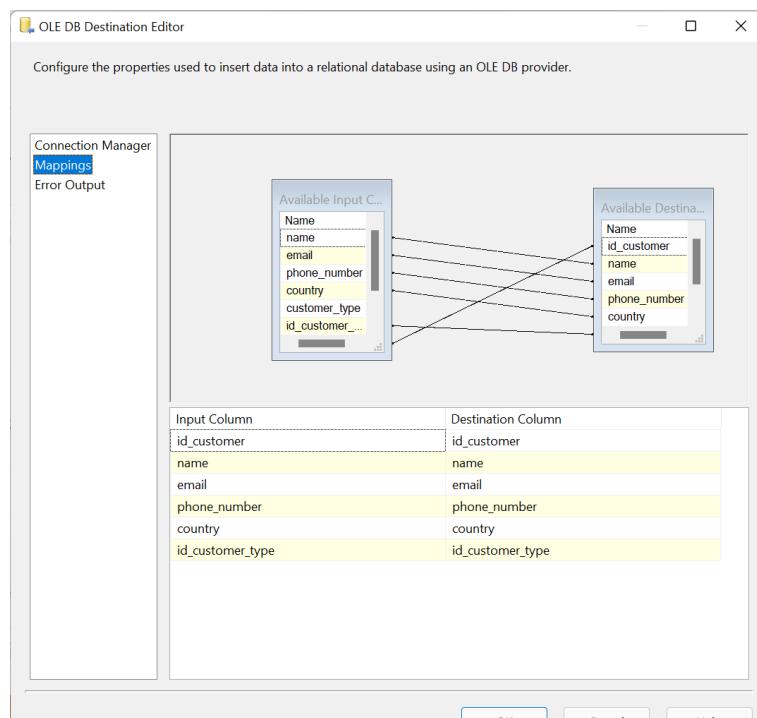


Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 2. 122.1,.2.Tạo bảng Customer

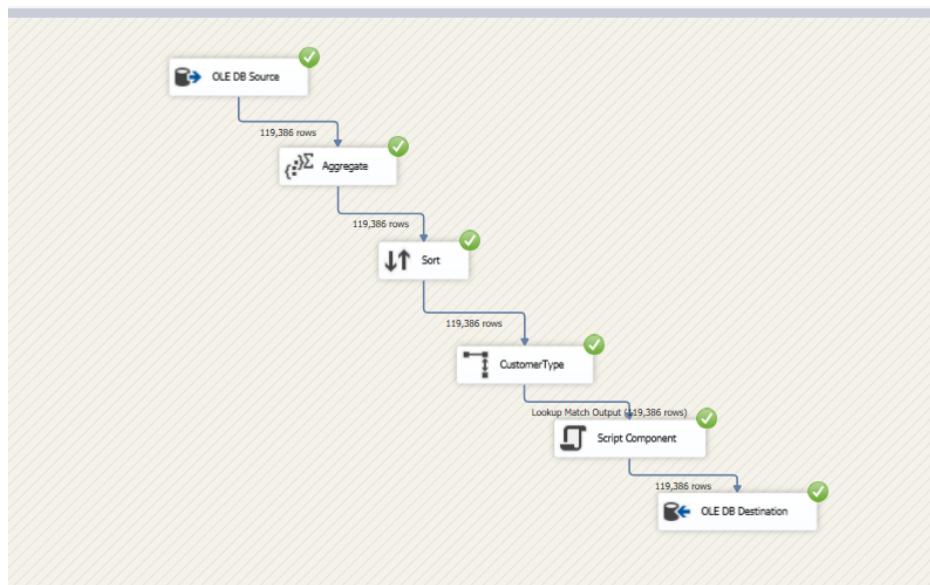
- Bước 10: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



Hình 2. 123. Quá trình Mappings dữ liệu bảng Customer

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 11: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



Hình 2. 124.Hoàn thành đổ dữ liệu vào Customer trong kho dữ liệu

- Bước 12: Kiểm tra bảng Customer trên SQL Server.

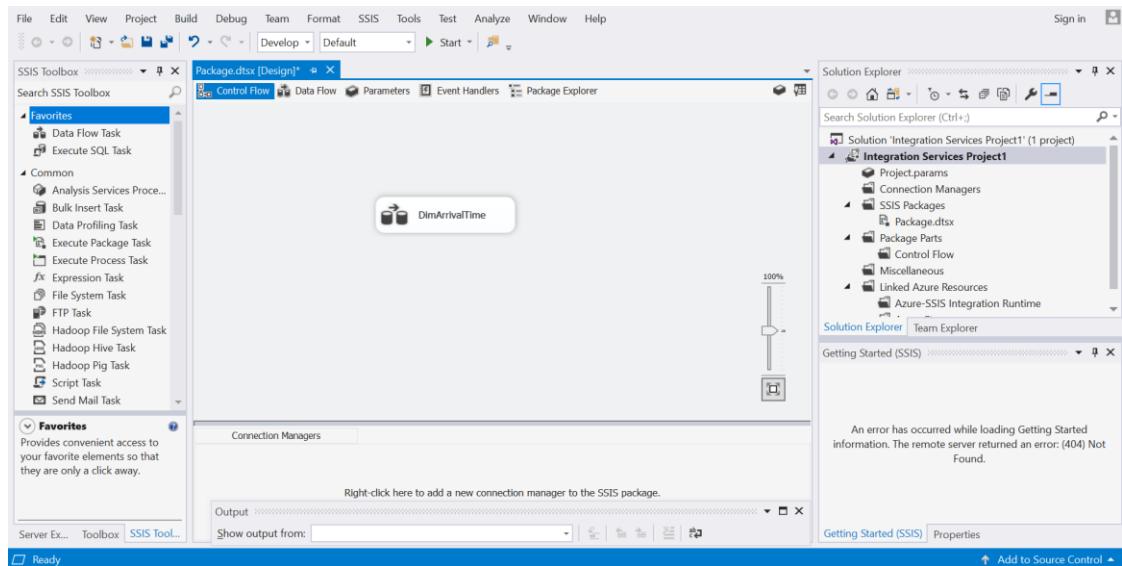
	id_customer	name	email	phone_number	country	id_customer_type
1	1	Aaron Acevedo	Aaron_A@outlook.com	200-008-5725	ITA	3
2	2	Aaron Acosta	Aaron_A@gmail.com	199-294-1491	PRT	3
3	3	Aaron Adams	ADams94@comcast.net	543-383-8949	CN	2
4	4	Aaron Adams	Aaron.Adams@zoho.com	340-616-3384	PRT	4
5	5	Aaron Adams	Aaron_A@verizon.com	110-703-6597	GBR	3
6	6	Aaron Adams	AaronAdams@zoho.com	574-840-4715	DEU	3
7	7	Aaron Aguilar	Aguilar_Aaron@verizon.com	844-200-6859	FRA	3
8	8	Aaron Alexander	Aaron_A26@zoho.com	802-241-1231	DEU	3
9	9	Aaron Allen	AAllen@mail.com	550-241-5926	GBR	3
10	10	Aaron Allen	Aaron_A6@outlook.com	793-389-6561	PRT	3
11	11	Aaron Allen MD	AaronMD@verizon.com	939-944-5115	PRT	4
12	12	Aaron Allison	Aaron_A@gmail.com	129-141-5099	ESP	1
13	13	Aaron Alvarez	Aaron_Alvarez@verizon.com	685-517-4925	LUX	3
14	14	Aaron Anderson	Aaron.Anderson@yandex.com	773-162-2625	ESP	3
15	15	Aaron Anderson	Aaron_Anderson@gmail.com	923-822-3426	ITA	4
16	16	Aaron Anderson	Aaron_Anderson96@yahoo.com	115-976-4411	GBR	4
17	17	Aaron Andrews	Andrews.Aaron33@aol.com	349-236-7542	CHN	3
18	18	Aaron Austin	AAustin@aol.com	348-629-6654	PRT	3
19	19	Aaron Avila	AaronAvila@yandex.com	417-646-9506	PRT	3
20	20	Aaron Bailey	Aaron.Bailey@aol.com	602-442-2678	PRT	3

Query executed successfully. DESKTOP-2KCNP7U (15.0 RTM) DESKTOP-2KCNP7U\phuong ... DB\_HB 00:00:01 119,386 rows

Hình 2. 125.Kiểm tra bảng Customer trên SQL Server

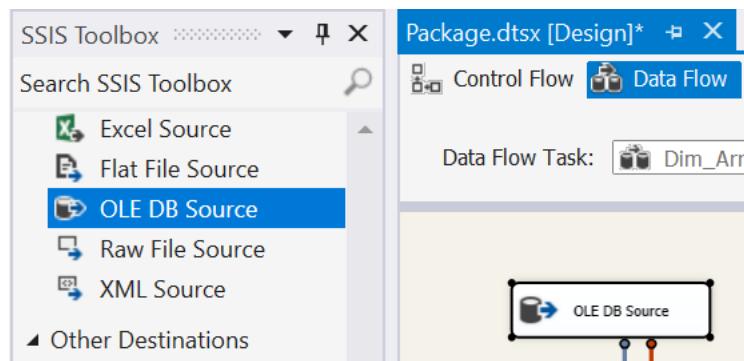
### 5.8. Bảng DimArrivalTime

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimArrivalTime.



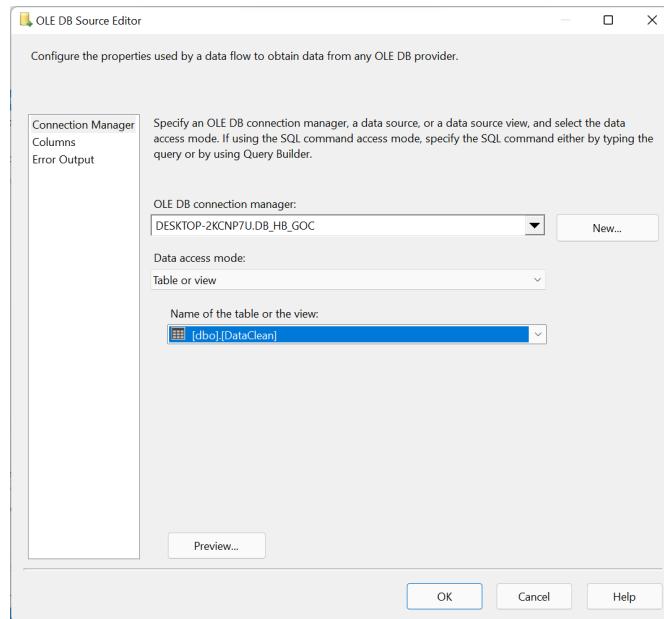
*Hình 2. 126.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimArrivalTime*

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



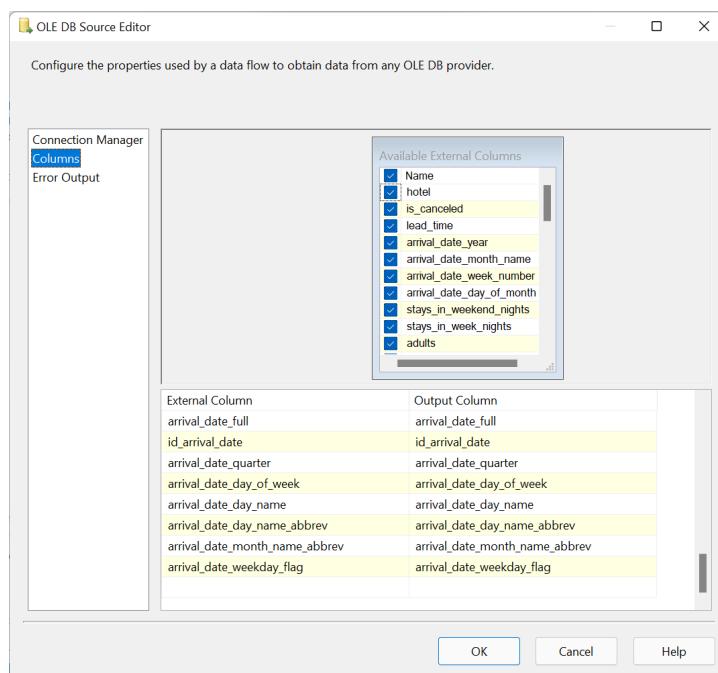
*Hình 2. 127.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



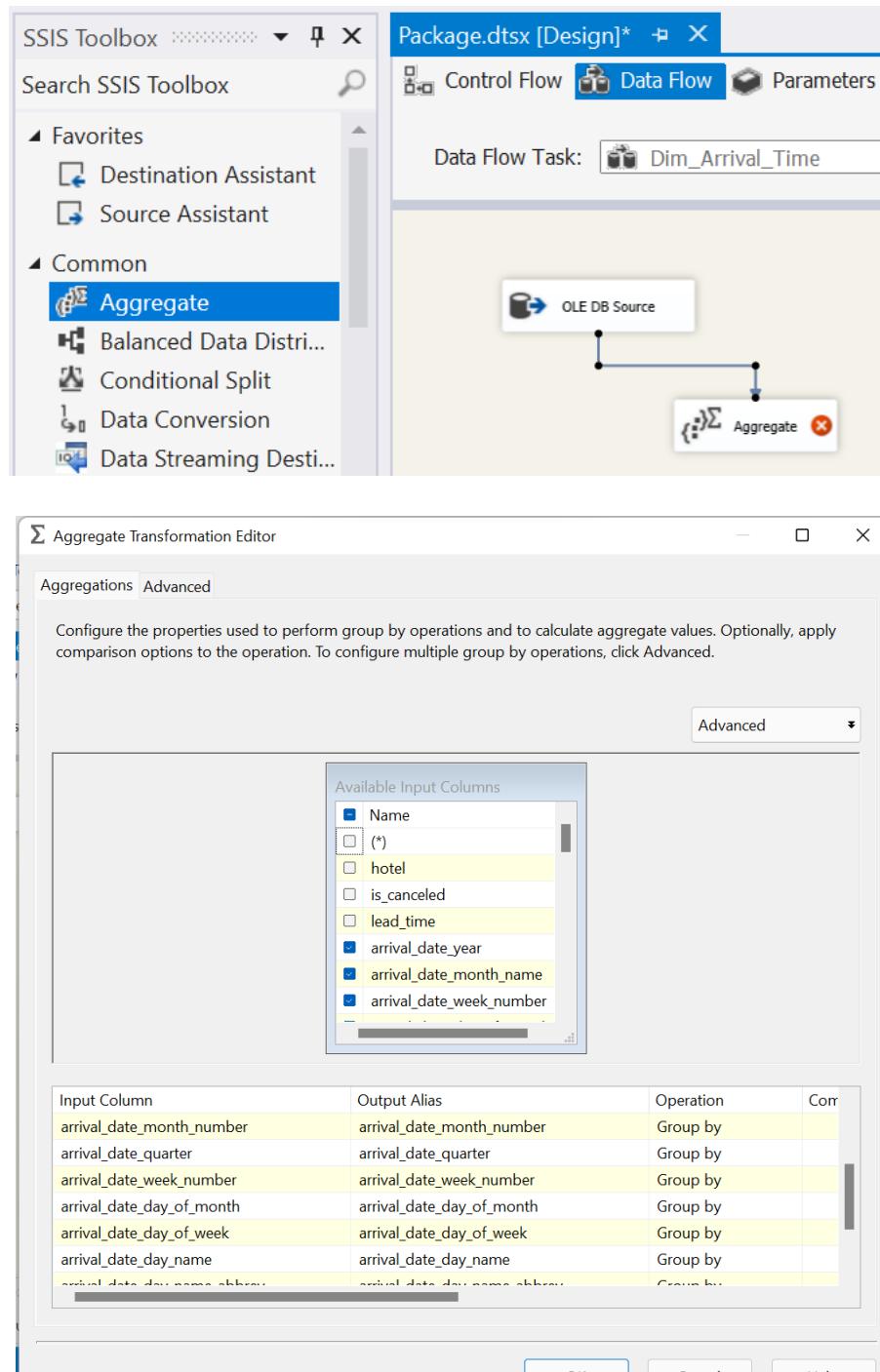
*Hình 2. 128.Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.



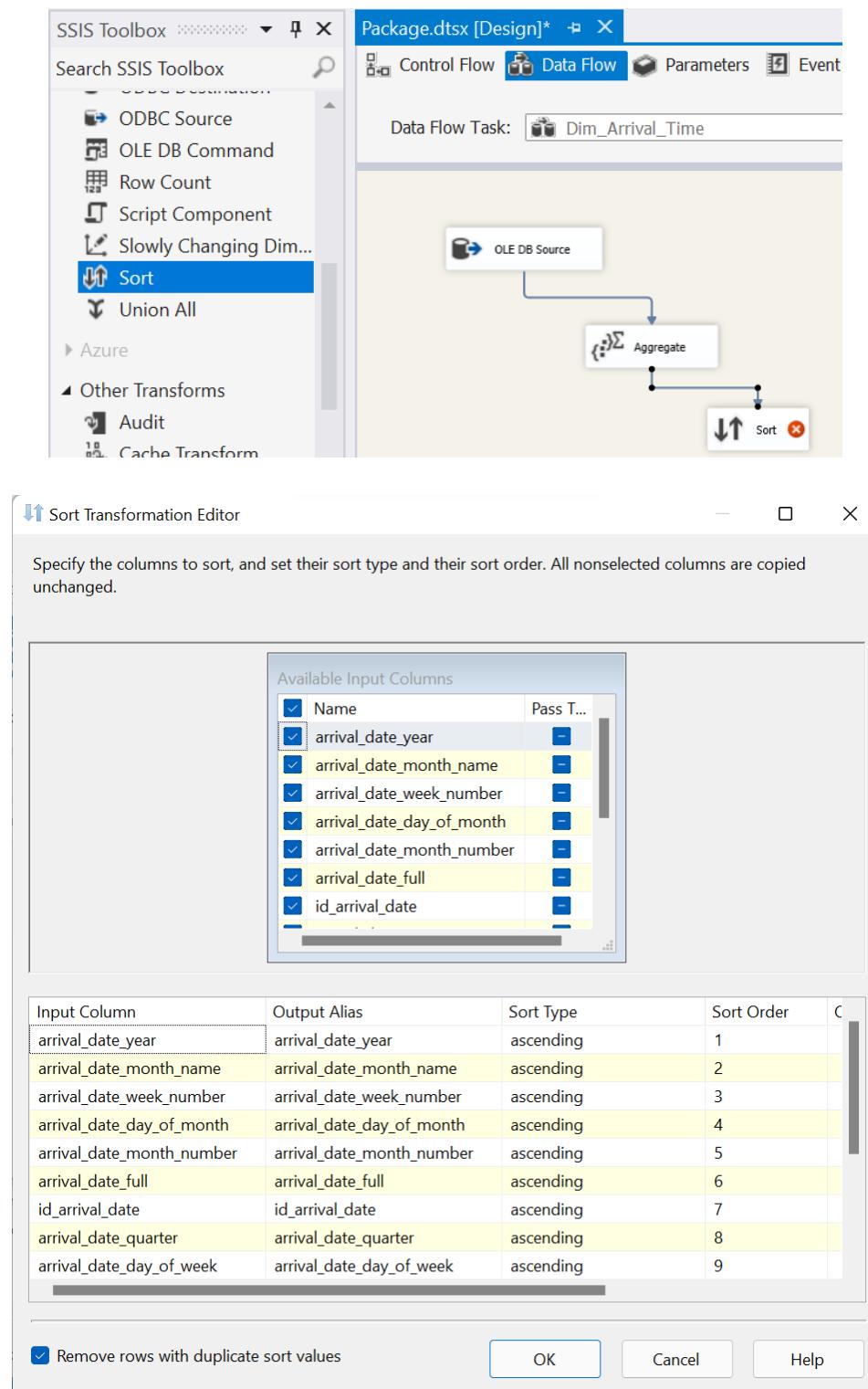
*Hình 2. 129.Chọn các columns cần sử dụng*

- Bước 5: Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau lên thuộc tính sử dụng là id\_arrival\_date, arrival\_date\_full, arrival\_date\_year, ...



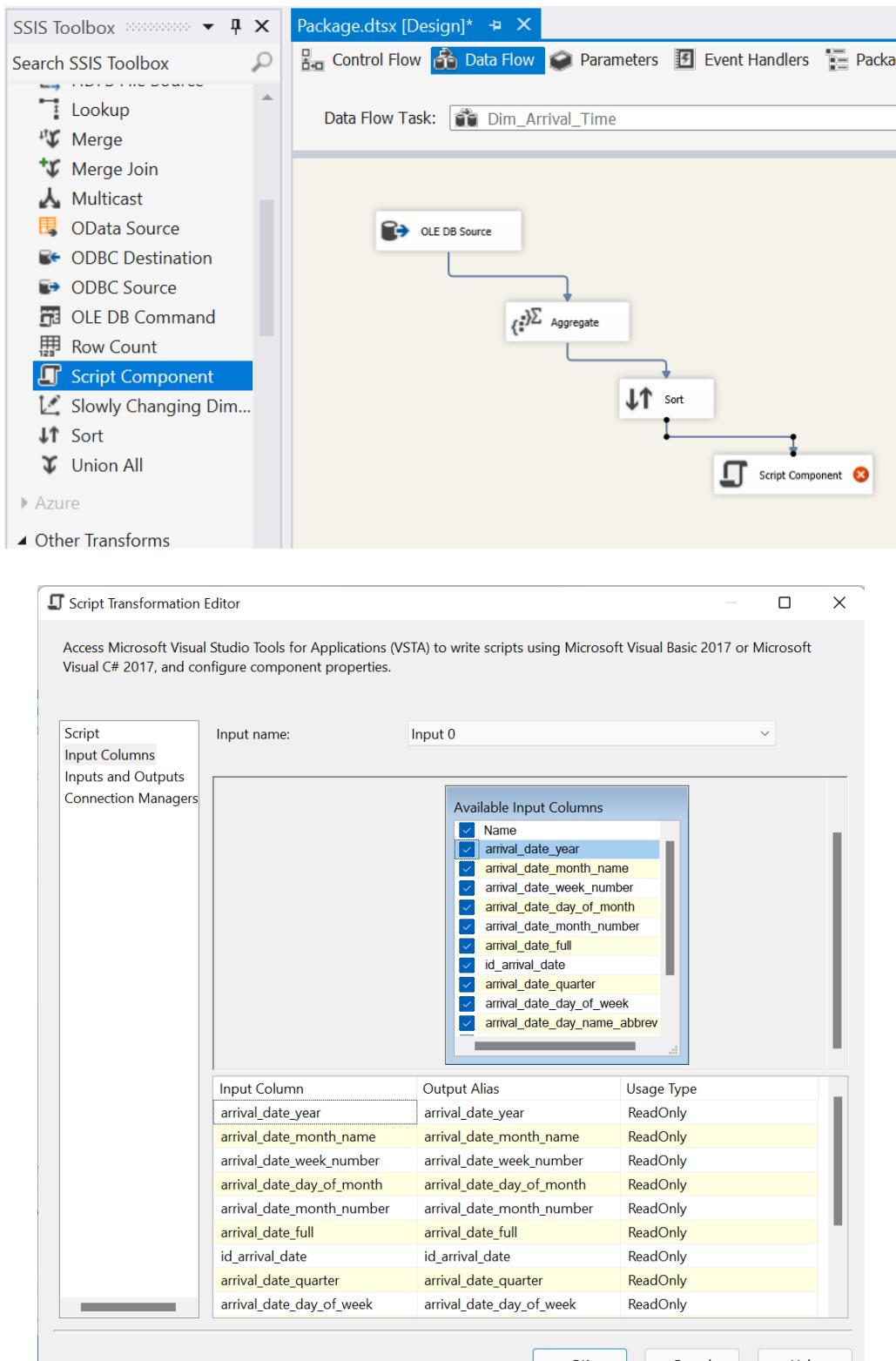
*Hình 2. 130.1..2.Dùng công cụ Aggregate dùng để thực hiện lọc các thuộc tính trùng nhau*

- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.



Hình 2. 131. Sắp xếp các giá trị theo chiều tăng dần

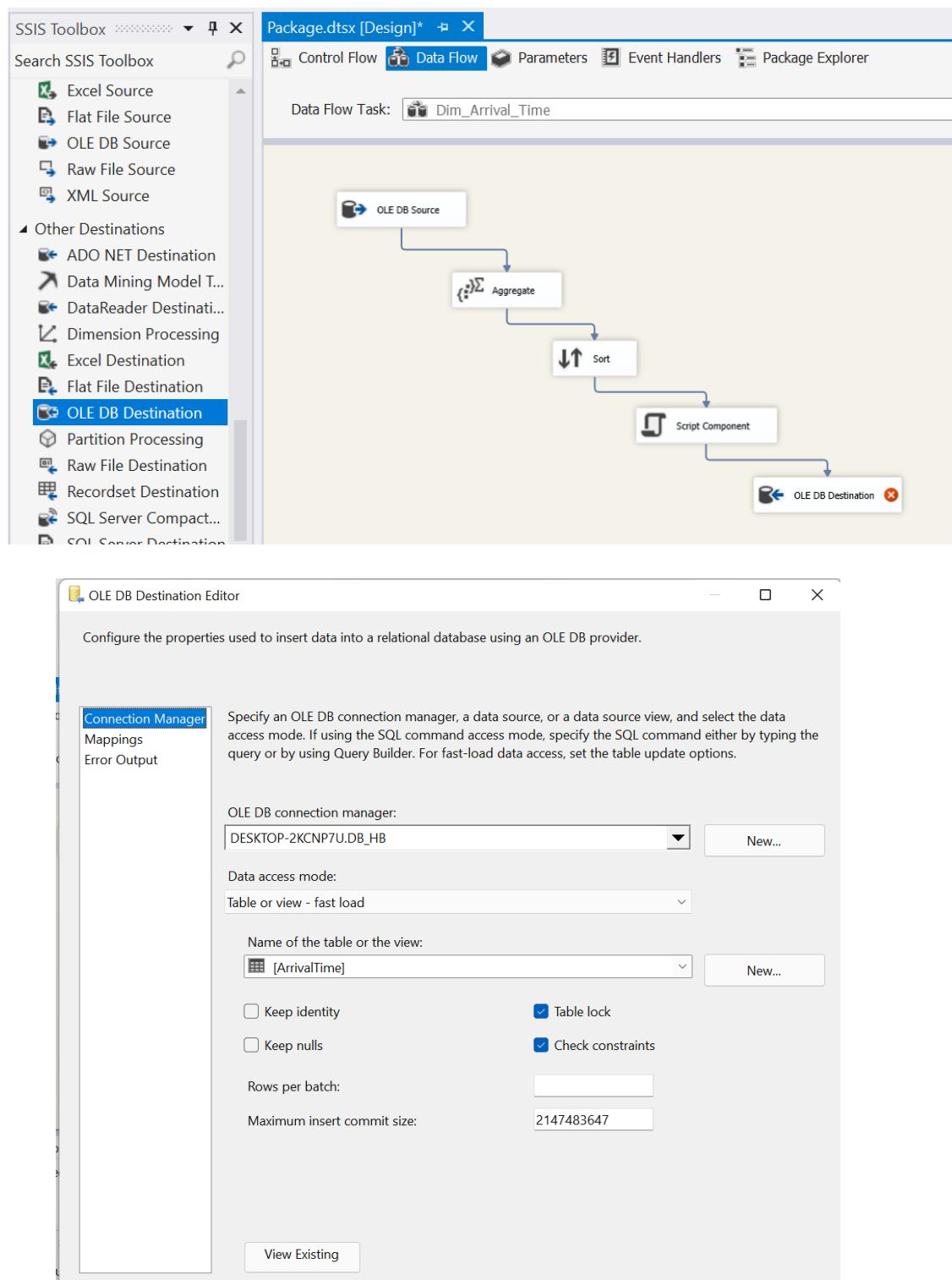
- Bước 7: Kéo thả công cụ Script Component.



*Hình 2. 132. Sử dụng Script Component để tạo khóa chính*

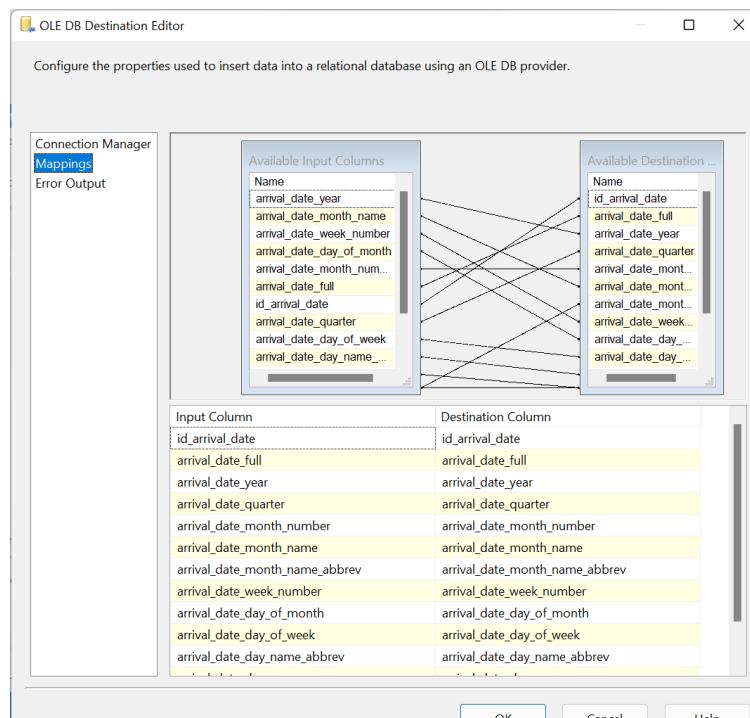
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 8: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là ArrivalTime.



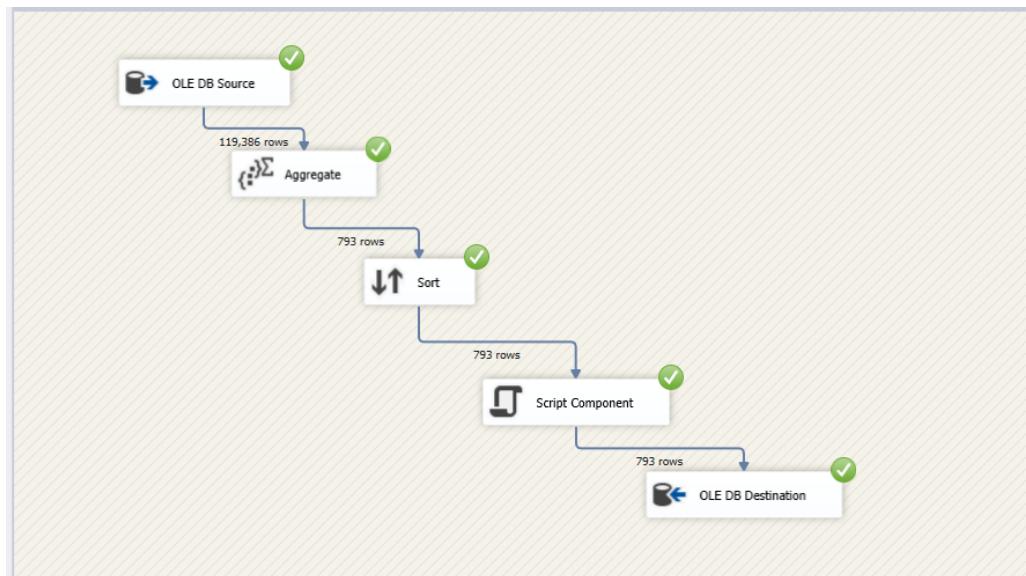
*Hình 2. 133. Tạo bảng ArrivalTime*

- Bước 9: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



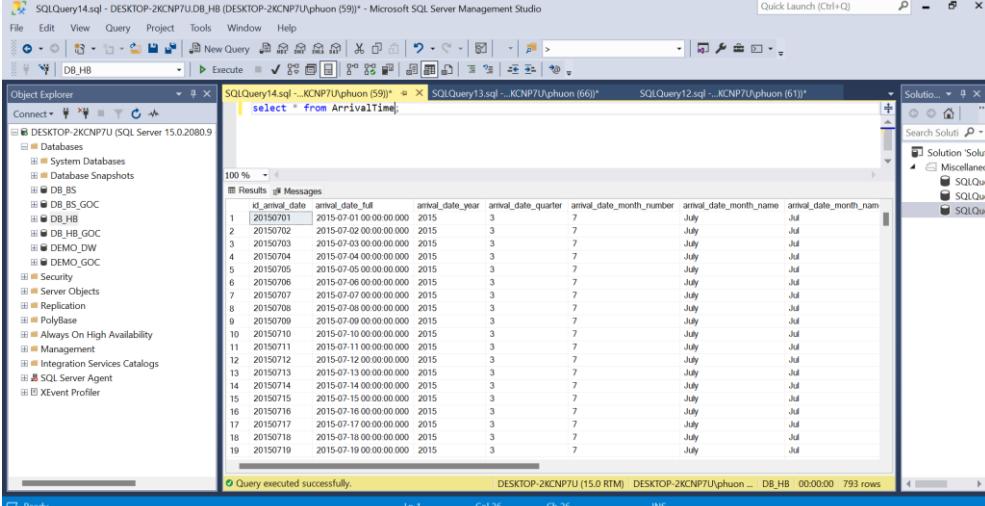
Hình 2. 134. Quá trình Mappings dữ liệu bảng ArrivalTime

- Bước 10: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



Hình 2. 135. Hoàn thành đổ dữ liệu vào ArrivalTime trong kho dữ liệu

- Bước 11: Kiểm tra bảng DistributionChannel trên SQL Server.



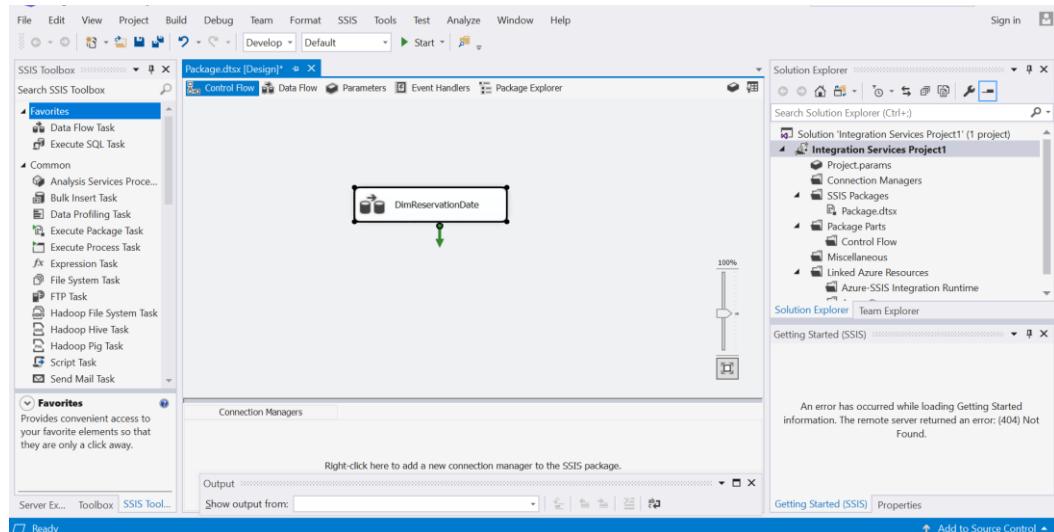
The screenshot shows the Microsoft SQL Server Management Studio interface. In the center, there is a results grid titled 'ArrivalTime' with 793 rows. The columns are: id\_arrival\_date, arrival\_date\_full, arrival\_date\_year, arrival\_date\_quarter, arrival\_date\_month\_number, arrival\_date\_month\_name, and arrival\_date\_month\_name. The data shows various dates from July 2015. At the bottom of the results grid, it says 'Query executed successfully.' To the right of the results grid, there is a Solution Explorer window showing a solution named 'Solution1' with several projects and files.

id_arrival_date	arrival_date_full	arrival_date_year	arrival_date_quarter	arrival_date_month_number	arrival_date_month_name	arrival_date_month_name
1	20150701	2015-07-01 00:00:00.000	2015	3	7	July
2	20150702	2015-07-02 00:00:00.000	2015	3	7	July
3	20150703	2015-07-03 00:00:00.000	2015	3	7	July
4	20150704	2015-07-04 00:00:00.000	2015	3	7	July
5	20150705	2015-07-05 00:00:00.000	2015	3	7	July
6	20150706	2015-07-06 00:00:00.000	2015	3	7	July
7	20150707	2015-07-07 00:00:00.000	2015	3	7	July
8	20150708	2015-07-08 00:00:00.000	2015	3	7	July
9	20150709	2015-07-09 00:00:00.000	2015	3	7	July
10	20150710	2015-07-10 00:00:00.000	2015	3	7	July
11	20150711	2015-07-11 00:00:00.000	2015	3	7	July
12	20150712	2015-07-12 00:00:00.000	2015	3	7	July
13	20150713	2015-07-13 00:00:00.000	2015	3	7	July
14	20150714	2015-07-14 00:00:00.000	2015	3	7	July
15	20150715	2015-07-15 00:00:00.000	2015	3	7	July
16	20150716	2015-07-16 00:00:00.000	2015	3	7	July
17	20150717	2015-07-17 00:00:00.000	2015	3	7	July
18	20150718	2015-07-18 00:00:00.000	2015	3	7	July
19	20150719	2015-07-19 00:00:00.000	2015	3	7	July

Hình 2. 136.Kiểm tra bảng ArrivalTime trên SQL Server

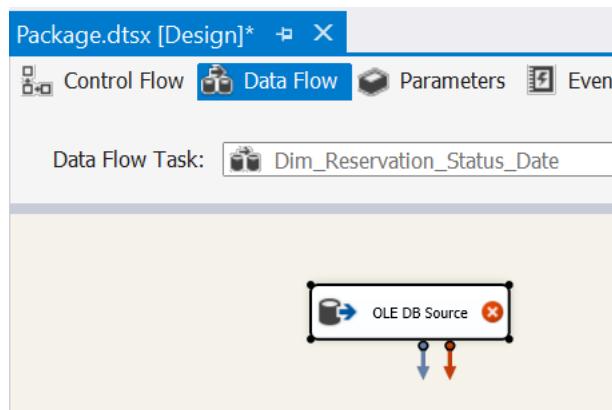
### 5.9. Bảng DimReservationDate

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành DimReservationDate.



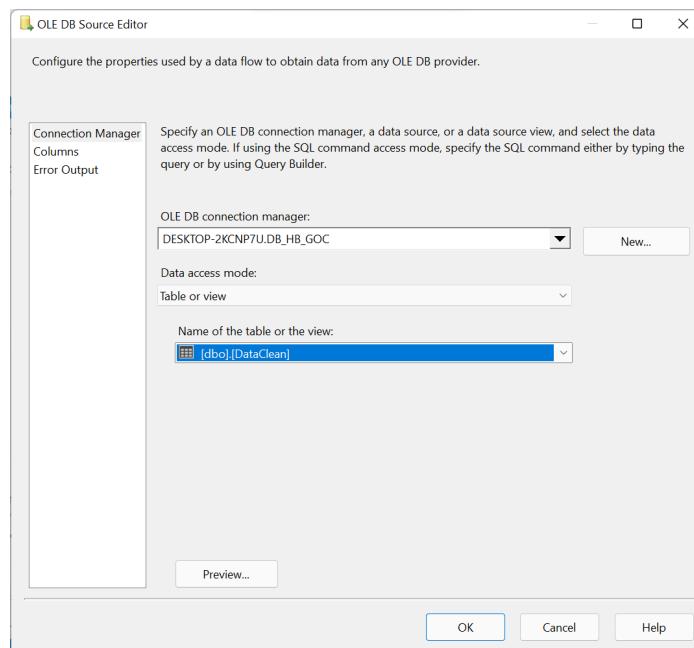
Hình 2. 137.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành DimReservationDate

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



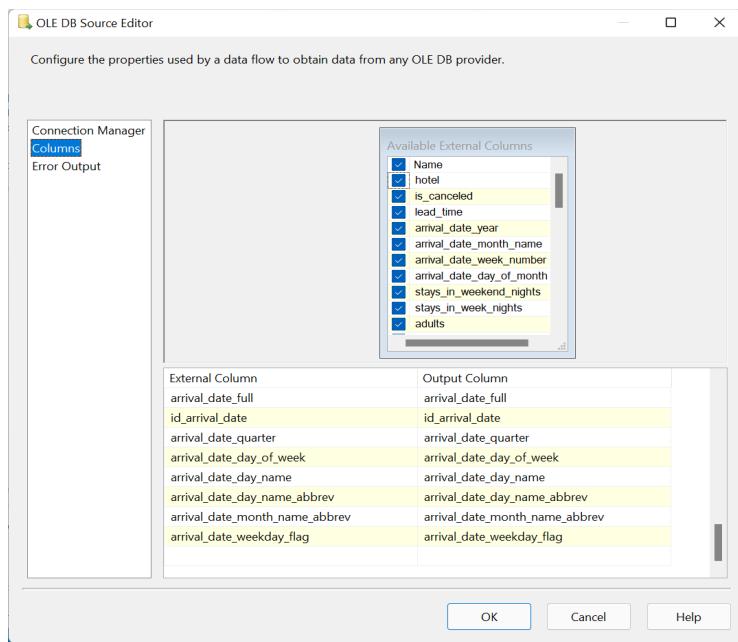
*Hình 2. 138.Kéo chức năng OLE DB Source vào màn hình làm việc*

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



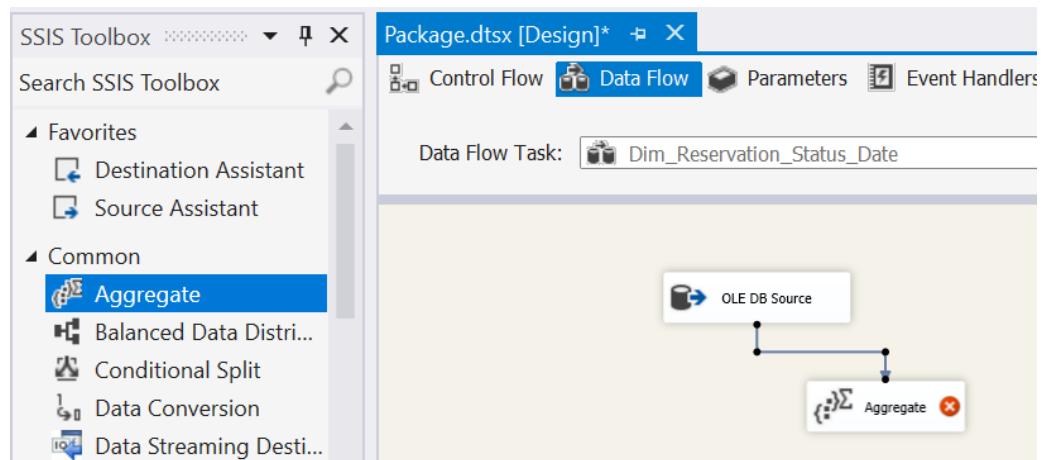
*Hình 2. 139.Chọn table sẽ lấy dữ liệu*

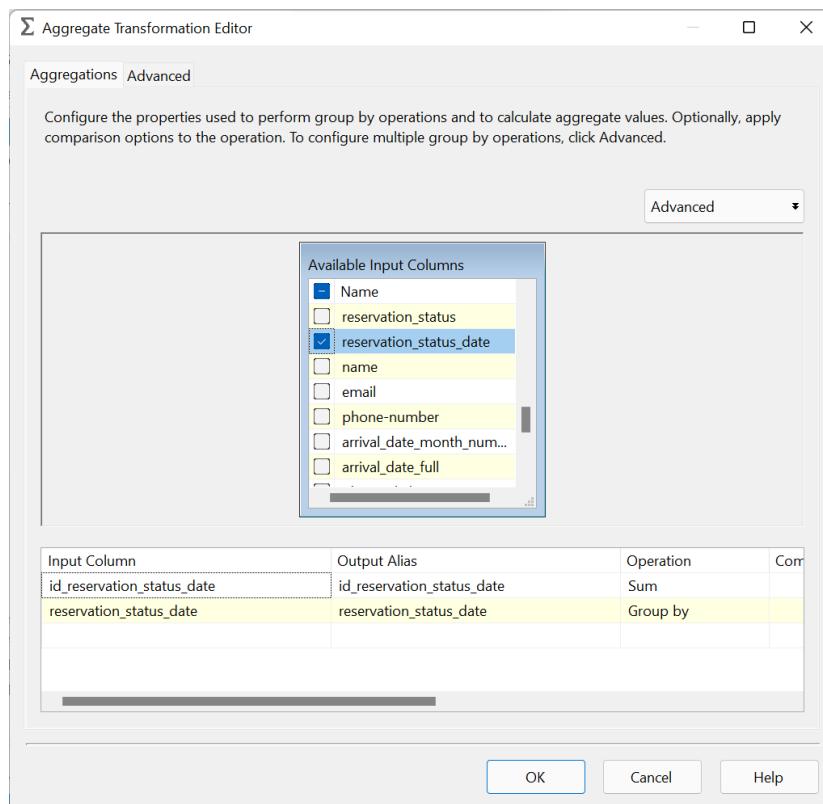
- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.



Hình 2. 140. Chọn các columns cần sử dụng

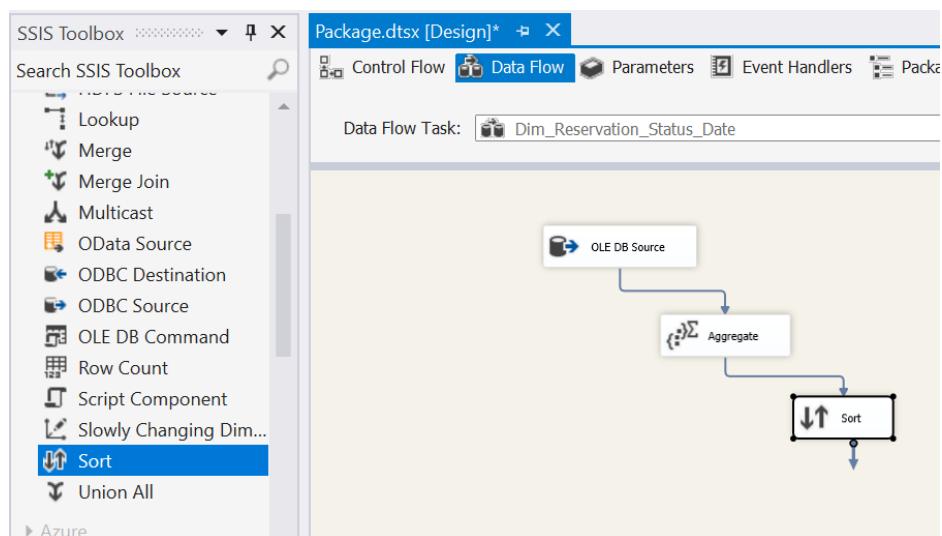
- Bước 5: Chọn các thuộc tính cần gom nhóm lên thuộc tính sử dụng là id\_reservation\_status\_date, reservation\_status\_date.

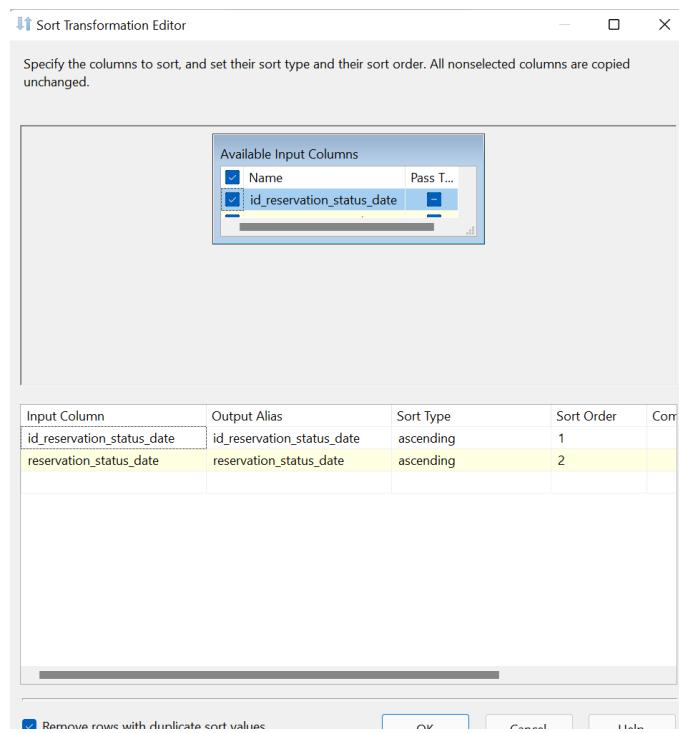




*Hình 2. 141.1.,2 Chọn các thuộc tính cần gom nhóm*

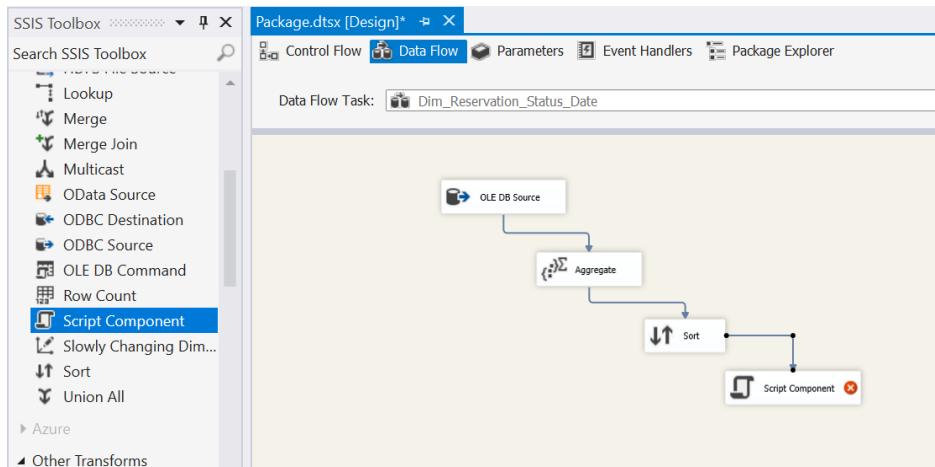
- Bước 6: Dùng công cụ Sort để sắp xếp các giá trị theo chiều tăng dần.

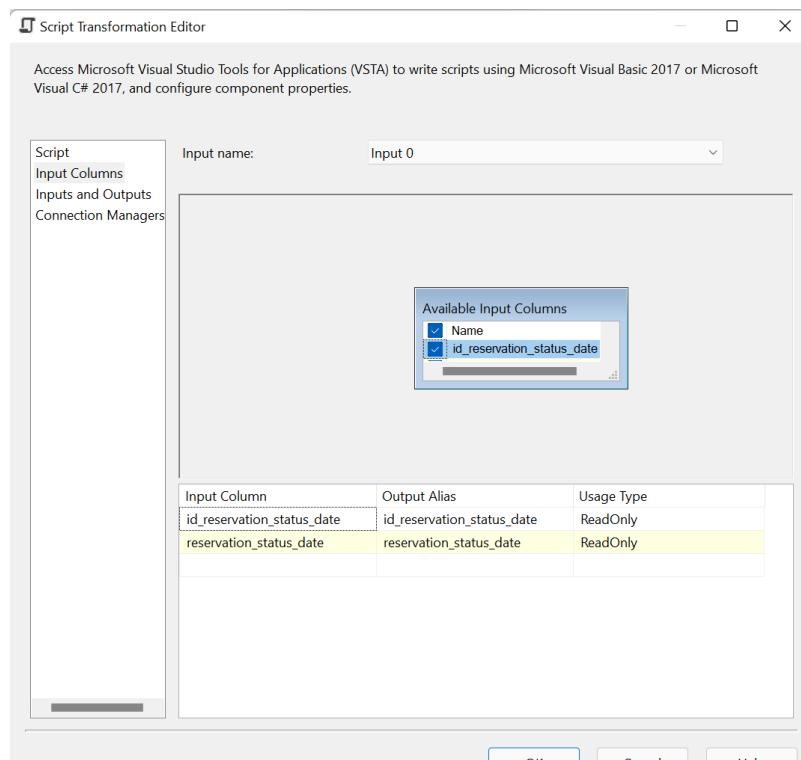




*Hình 2. 142.1,.2.Sắp xếp các giá trị theo chiều tăng dần*

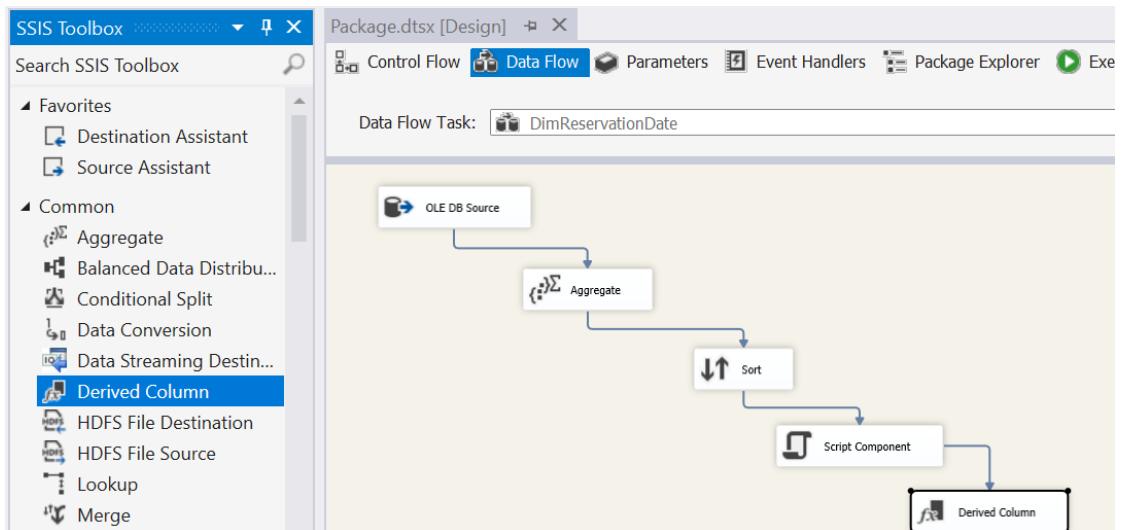
- **Bước 7: Kéo thả công cụ Script Component.**

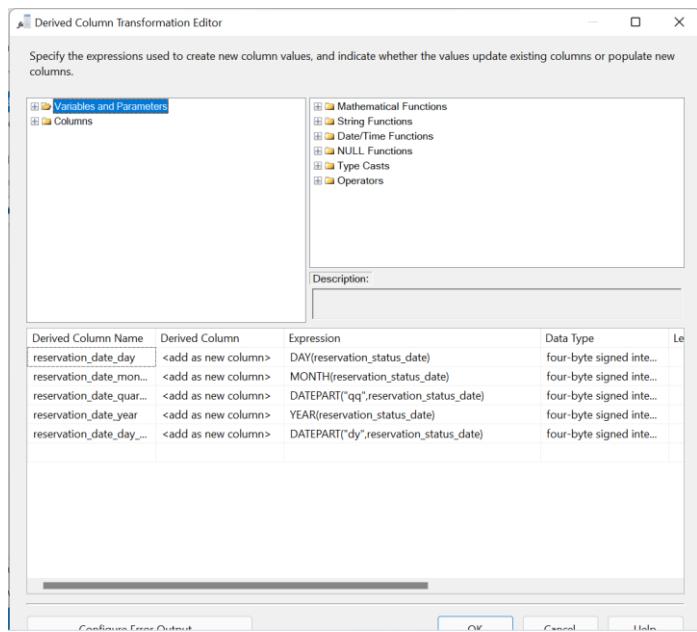




Hình 2. 143. Sử dụng Script Component để tạo khóa chính id\_reservation\_status\_date

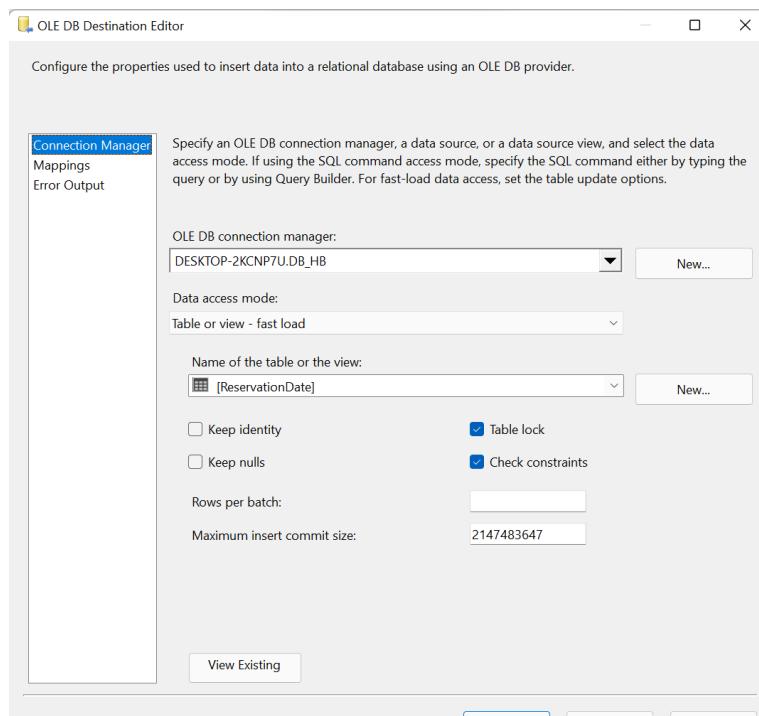
- Bước 8: Tiếp tục chọn công cụ Derived Column để tiến hành tách reservation\_status\_date thành reservation\_date\_day, reservation\_date\_month, reservation\_date\_quarter, reservation\_date\_year, reservation\_date\_day\_of\_year.





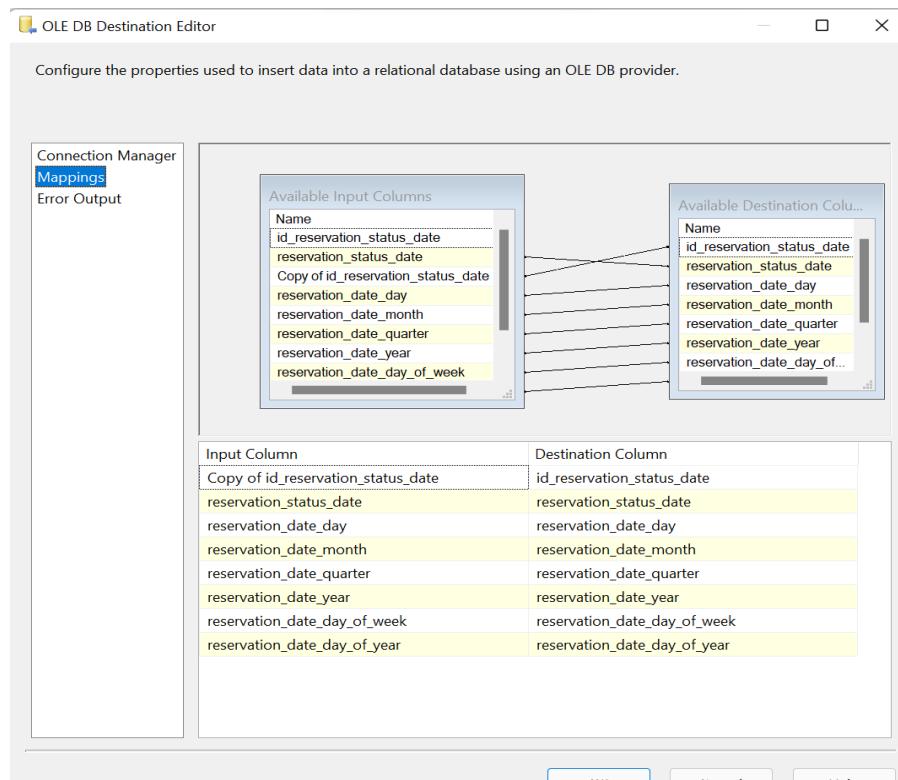
*Hình 2. 144.1,.2 Sử dụng Derives Column để tách cột resevation\_status\_date*

- Bước 9: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là ReservationDate.



*Hình 2. 145. Tạo bảng ReservationDate*

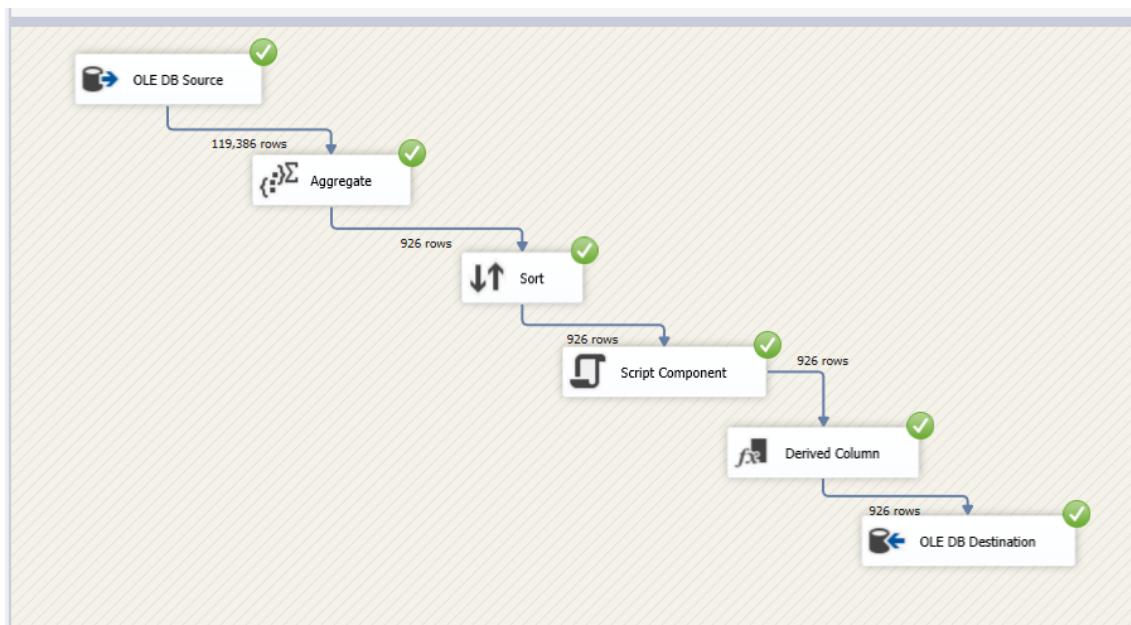
- Bước 10: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



*Hình 2. 146. Quá trình Mappings dữ liệu bảng ReservationDate*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 11: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 147.Hoàn thành đổ dữ liệu vào ReservationDate trong kho dữ liệu*

- Bước 12: Kiểm tra bảng ReservationDate trên SQL Server.

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the database 'DB\_HB' is selected. In the center pane, a query window displays the following SQL code and results:

```
select * from ReservationDate;
```

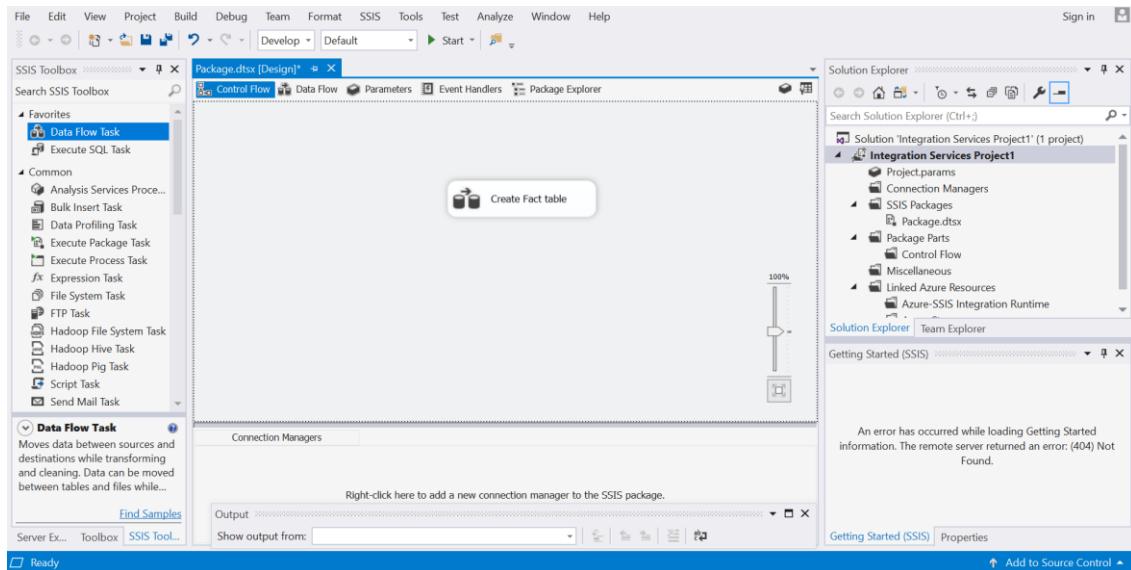
id_reservation_status_date	reservation_status_date	reservation_date_day	reservation_date_month	reservation_date_quarter	reservation_date_year	reservation	
1	20141017	2014-10-17 00:00:00.000	17	10	4	2014	290
2	20141118	2014-11-18 00:00:00.000	18	11	4	2014	322
3	20150101	2015-01-01 00:00:00.000	1	1	1	2015	1
4	20150102	2015-01-02 00:00:00.000	2	1	1	2015	2
5	20150118	2015-01-18 00:00:00.000	18	1	1	2015	18
6	20150120	2015-01-20 00:00:00.000	20	1	1	2015	20
7	20150121	2015-01-21 00:00:00.000	21	1	1	2015	21
8	20150122	2015-01-22 00:00:00.000	22	1	1	2015	22
9	20150128	2015-01-28 00:00:00.000	28	1	1	2015	28
10	20150129	2015-01-29 00:00:00.000	29	1	1	2015	29
11	20150130	2015-01-30 00:00:00.000	30	1	1	2015	30
12	20150202	2015-02-02 00:00:00.000	2	2	1	2015	33
13	20150205	2015-02-05 00:00:00.000	5	2	1	2015	36
14	20150206	2015-02-06 00:00:00.000	6	2	1	2015	37
15	20150209	2015-02-09 00:00:00.000	9	2	1	2015	40
16	20150210	2015-02-10 00:00:00.000	10	2	1	2015	41
17	20150211	2015-02-11 00:00:00.000	11	2	1	2015	42
18	20150212	2015-02-12 00:00:00.000	12	2	1	2015	43
19	20150217	2015-02-17 00:00:00.000	17	2	1	2015	48

Below the results, a message states: "Query executed successfully." The status bar at the bottom indicates: DESKTOP-2KCNP7U (15.0 RTM) DESKTOP-2KCNP7U\phuong ... DB\_HB 00:00:00 926 rows".

*Hình 2. 148.Kiểm tra bảng ReservationDate trên SQL Server*

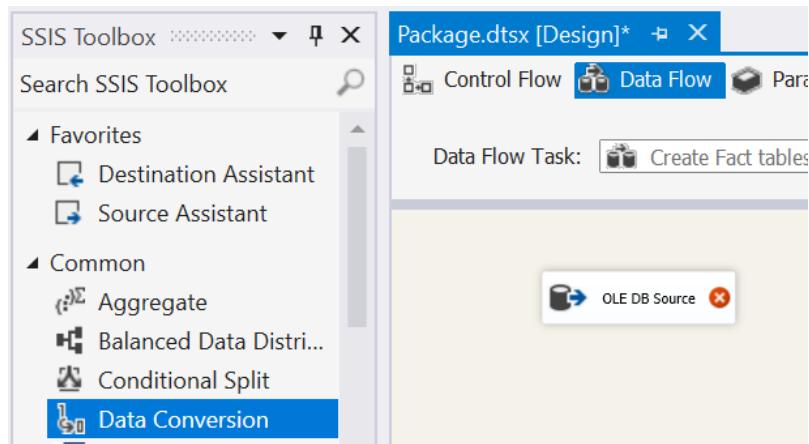
### 5.10. Bảng FactHotelBooking

- Bước 1: Kéo chức năng Data Flow Task từ cột trái sang màn hình làm việc và đổi tên thành Create Fact table.



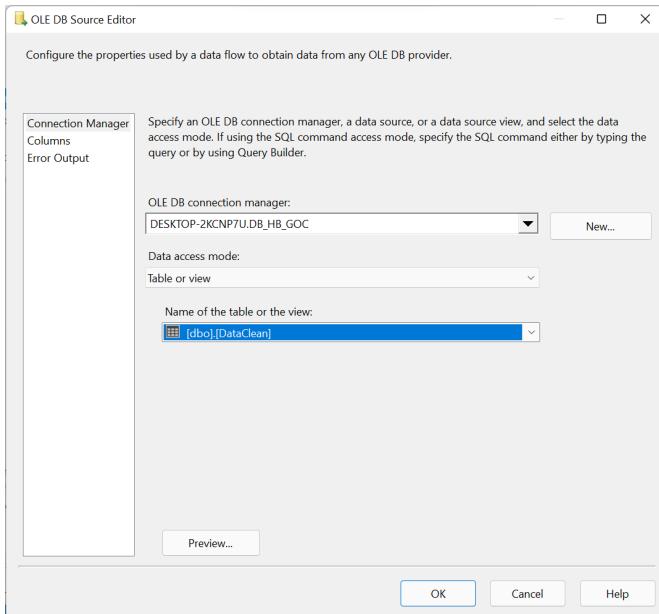
Hình 2. 149.Kéo Data Flow Task vào màn hình làm việc và đổi tên thành Create Fact table

- Bước 2: Nhấn double vào Data Flow Task vừa tạo và tìm kiếm chức năng OLE DB Source, kéo vào màn hình làm việc.



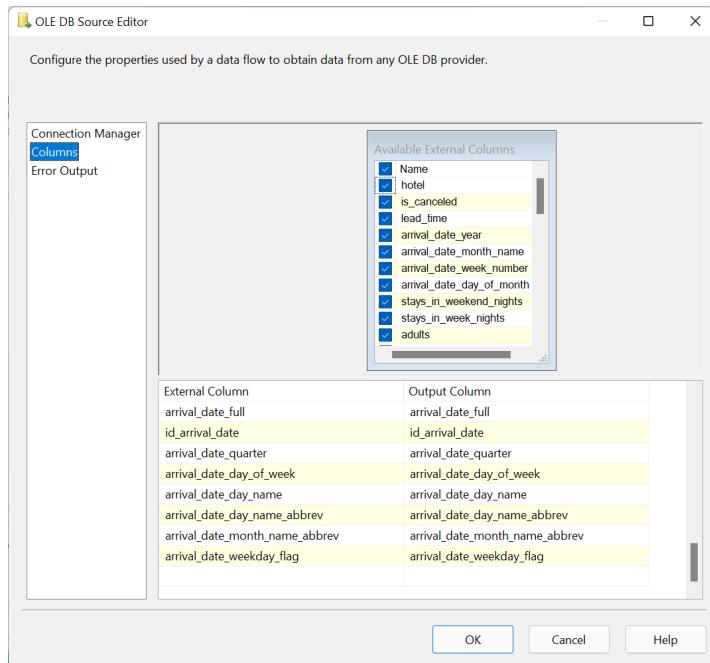
Hình 2. 150.Kéo chức năng OLE DB Source vào màn hình làm việc

- Bước 3: Click double vào OLE DB Source và dẫn đến DB\_HB\_GOC. Sau đó chọn Name of the table or the view là [dbo].DataClean (đã tạo ở chương 2 ý 3 bước 7).



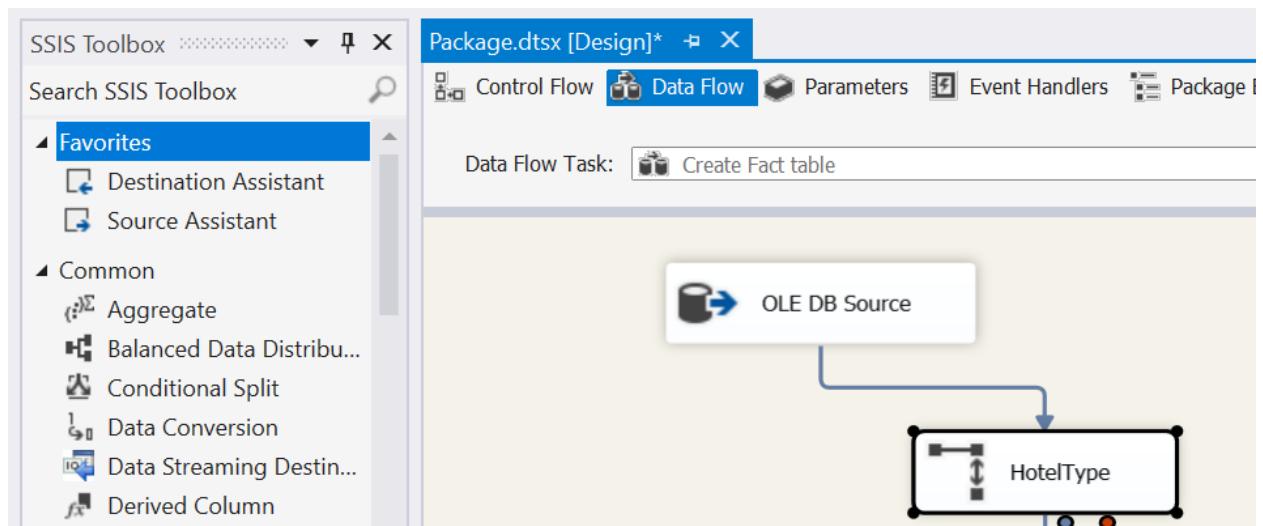
*Hình 2. 151.Chọn table sẽ lấy dữ liệu*

- Bước 4: Chọn Columns để lựa chọn các cột cần sử dụng và nhấn OK.

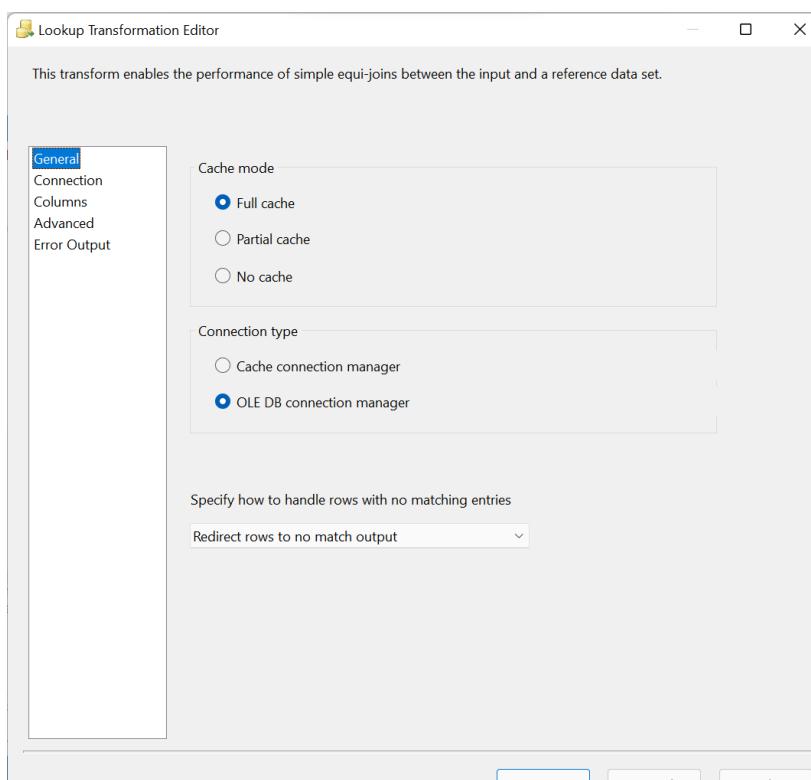


*Hình 2. 152.Chọn các columns cần sử dụng*

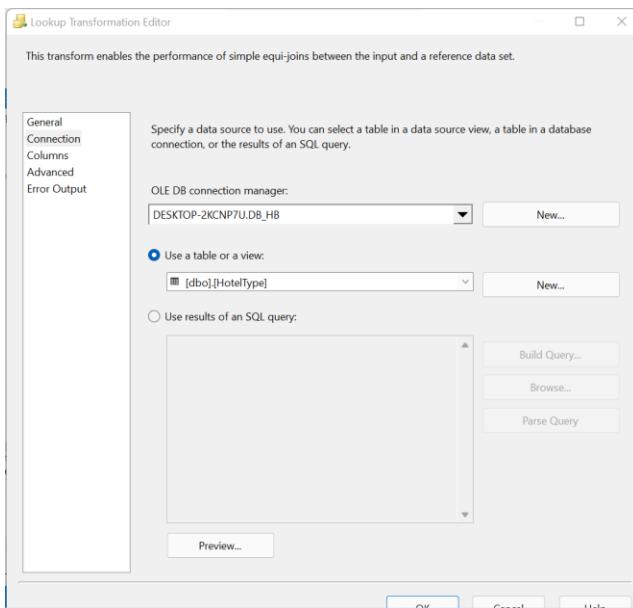
- Bước 5: Dùng công cụ Lookup và đổi tên thành CustomerType để tạo khóa ngoại id\_hotel\_type đến bảng Hotel.



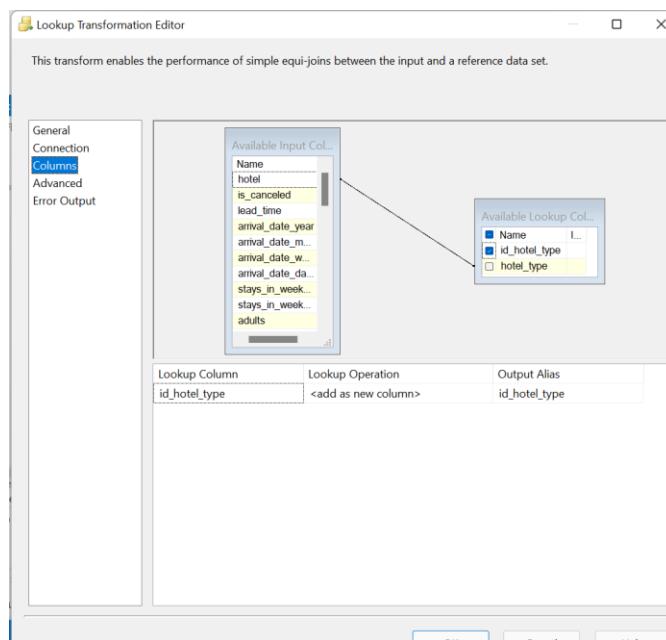
- + Đổi thành Redirects row to no match output



+ Tại Connection, chọn bảng HotelType đã tạo ở phần 4 chương 2

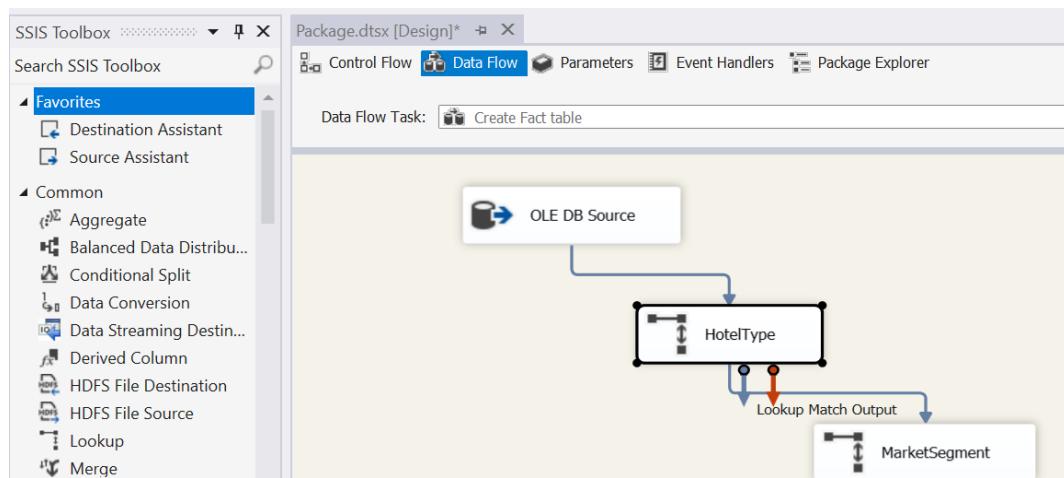


+ Trong khung Columns, nôi thuộc tính hotel với hotel\_type của bảng HotelType đã chọn ở bước trên và click vào id\_hotel\_type để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với Hotel Type.

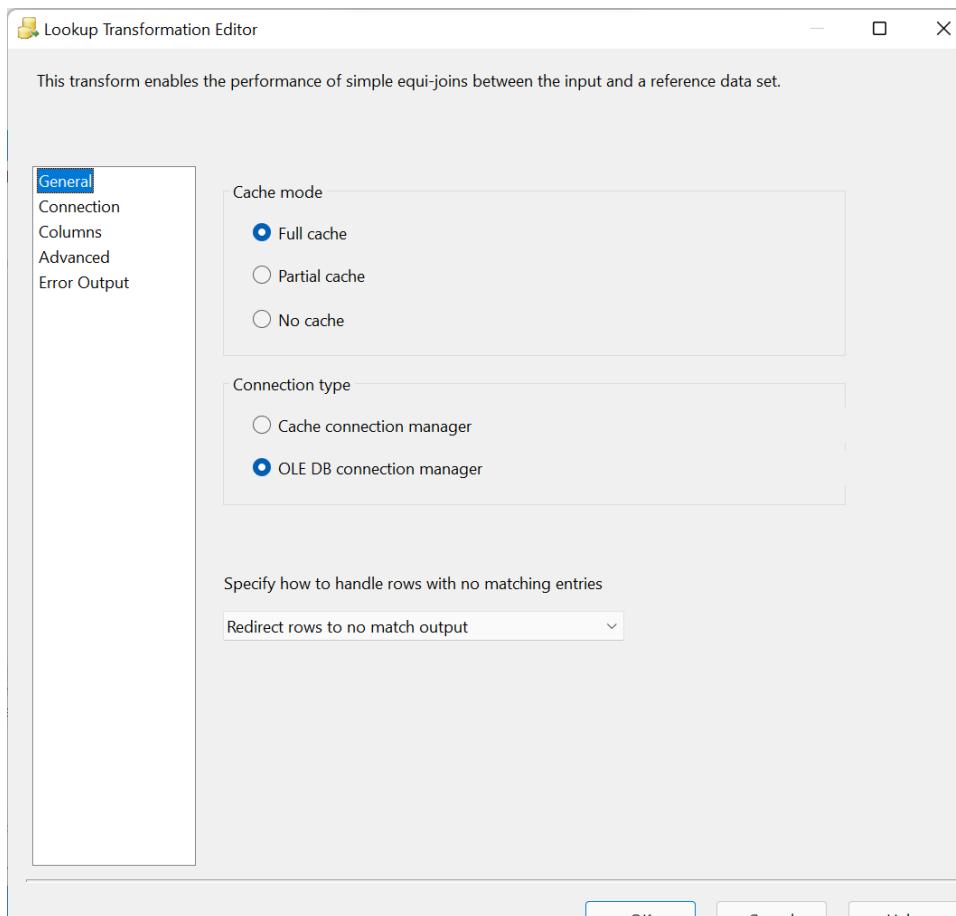


*Hình 2. 153.1.,2.,3.,4.Tạo khóa ngoại id\_hotel\_type đến bảng Hotel bằng công cụ Lookup*

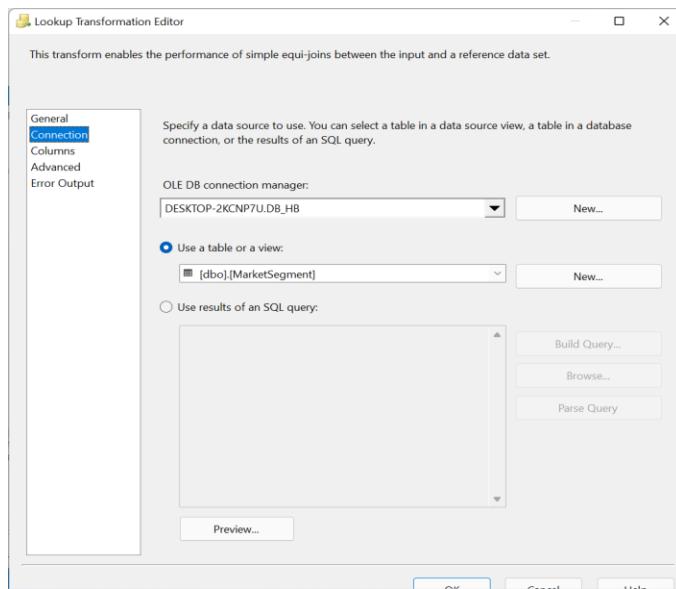
- Bước 6: Tìm kiếm công cụ Lookup, đổi tên thành MarketSegment để tạo khóa ngoại id\_market\_segment đến bảng MarketSegment và chọn đường kết nối giữ hai Lookup là Lookup Match Output.



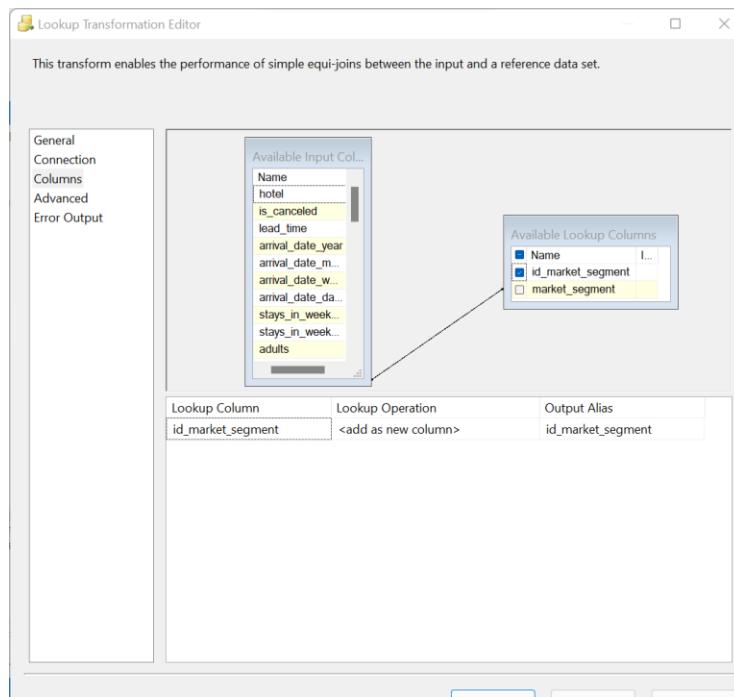
+ Đổi thành Redirects row to no match output



+ Tại Connection, chọn bảng MarketSegment đã tạo ở phần 4 chương 2



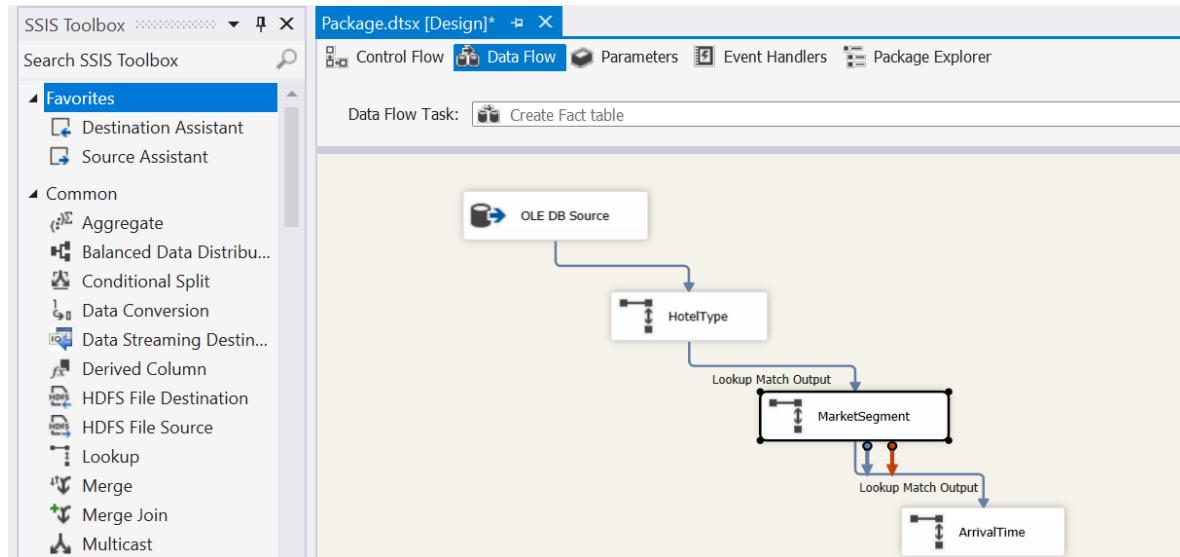
+ Trong khung Columns, nối thuộc tính market\_segment với market\_segment của bảng MarketSegment đã chọn ở bước trên và click vào id\_market\_segment để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với MarketSegment.



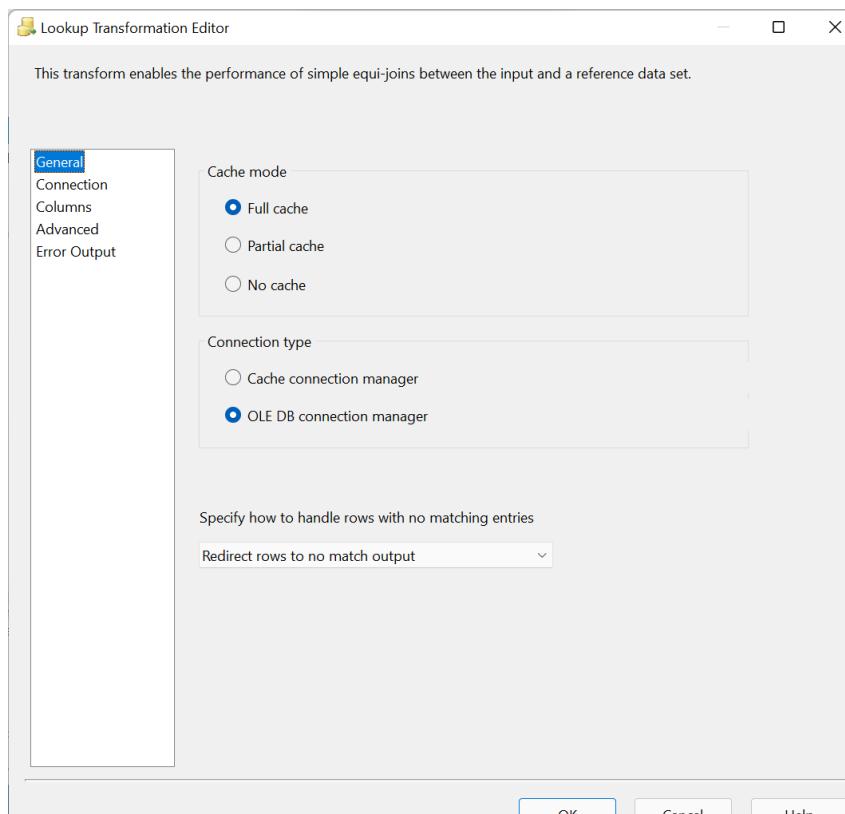
*Hình 2. 154.1.,2.,3.,4. Tạo khóa ngoại với bảng MarketSegment bằng công cụ Lookup*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

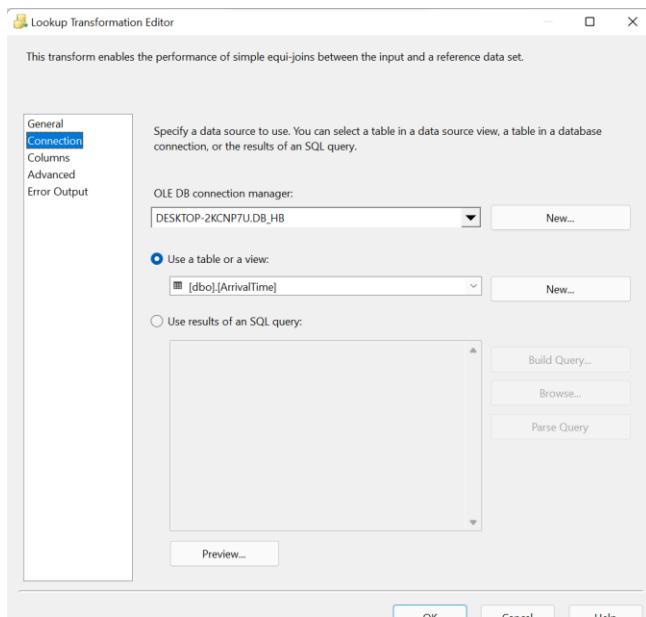
- Bước 7: Tìm kiếm công cụ Lookup, đổi tên thành ArrivalTime để tạo khóa ngoại id\_arrival\_date đến bảng ArrivalTime và chọn đường kết nối hai Lookup là Lookup Match Output.



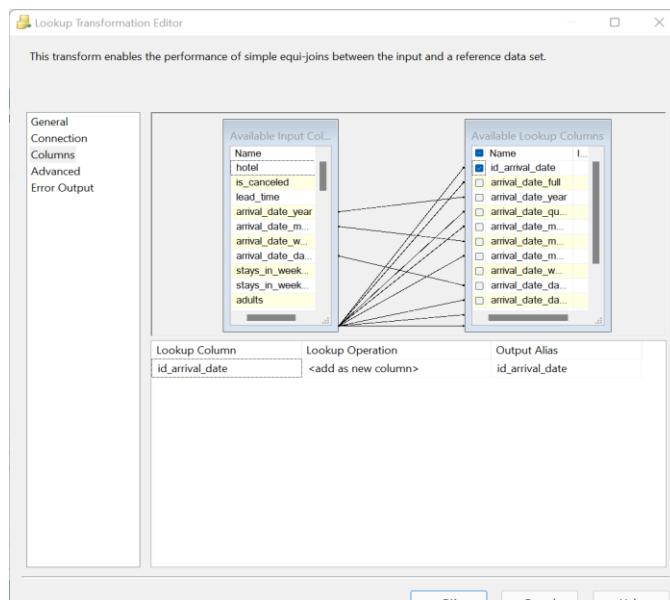
+ Đổi thành Redirects row to no match output



+ Tại Connection, chọn bảng ArrivalTime đã tạo ở phần 4 chương 2

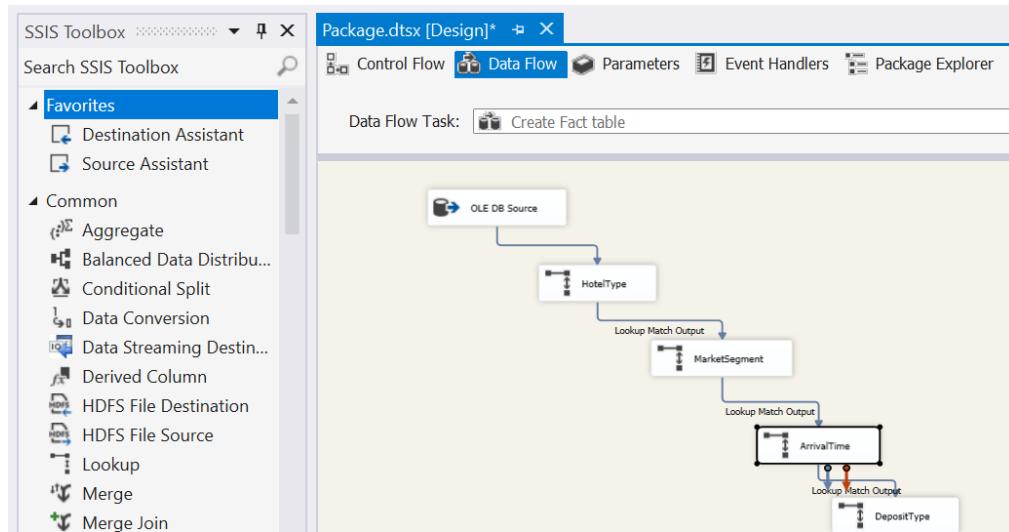


+ Trong khung Columns, nội thuộc tính id\_arrival\_date – id arrival\_date, arrival\_date\_full – arrival\_date\_full, arrival\_date\_year – arrival\_date\_year, .... của bảng ArrivalTime đã chọn ở bước trên và click vào id\_ arrival\_date để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với ArrivalTime.

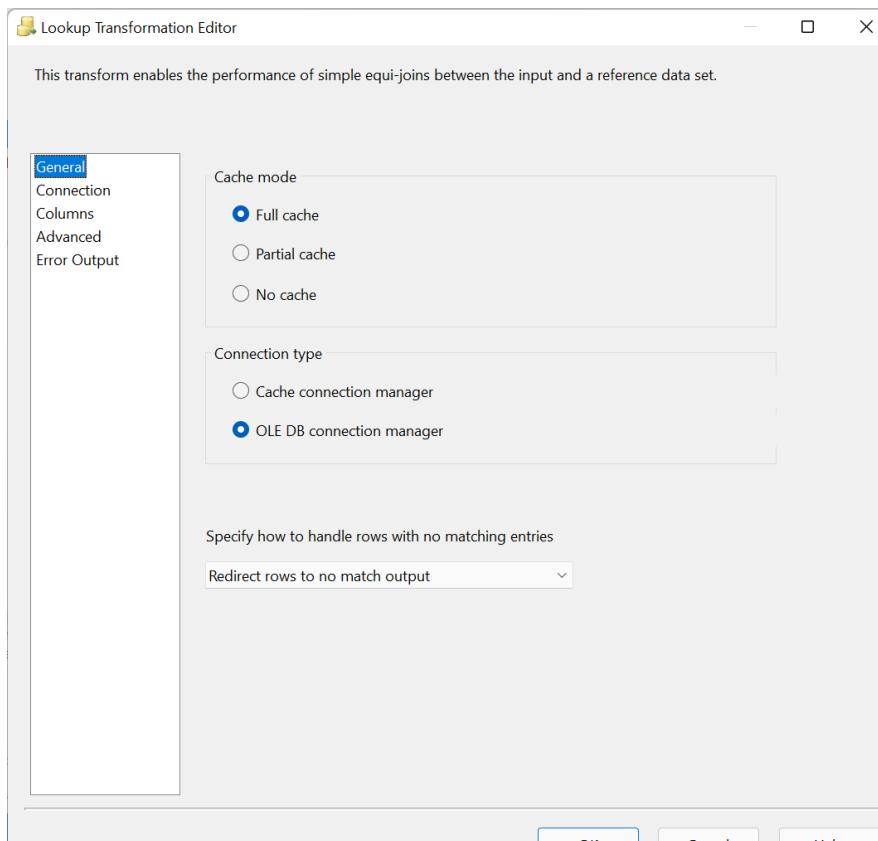


*Hình 2. 155.1.,2.,3.,4. Tạo khóa ngoại với bảng ArrivalTime bằng công cụ Lookup*

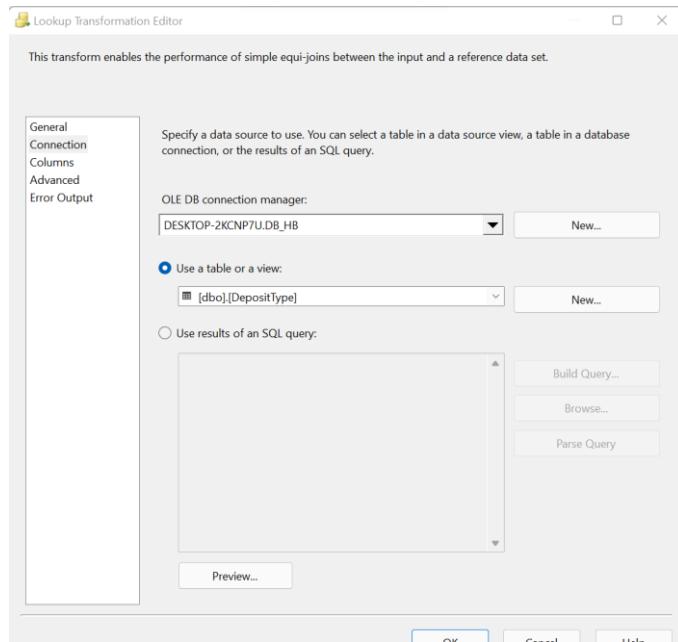
- Bước 8: Chọn công cụ Lookup, đổi tên thành DepositType để tạo khóa ngoại id\_deposit\_type đến bảng DepositType và chọn đường kết nối hai Lookup là Lookup Match Output.



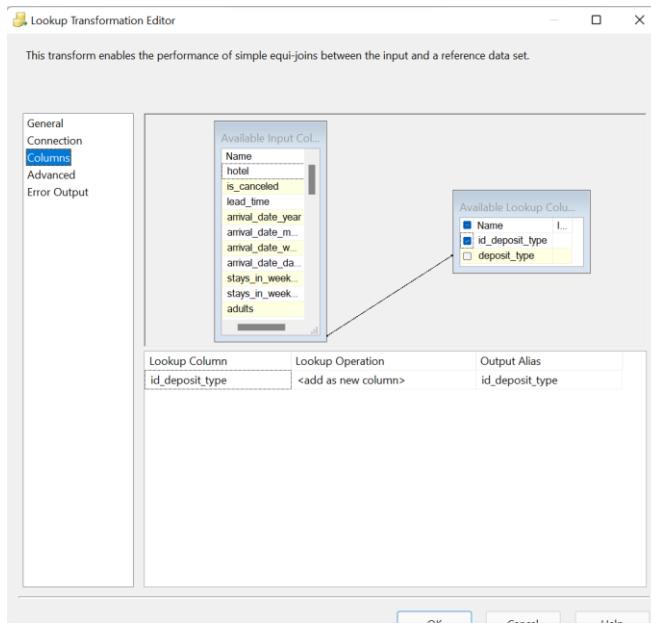
+ Đổi thành Redirects row to no match output



- + Tại Connection, chọn bảng DepositType đã tạo ở phần 4 chương 2

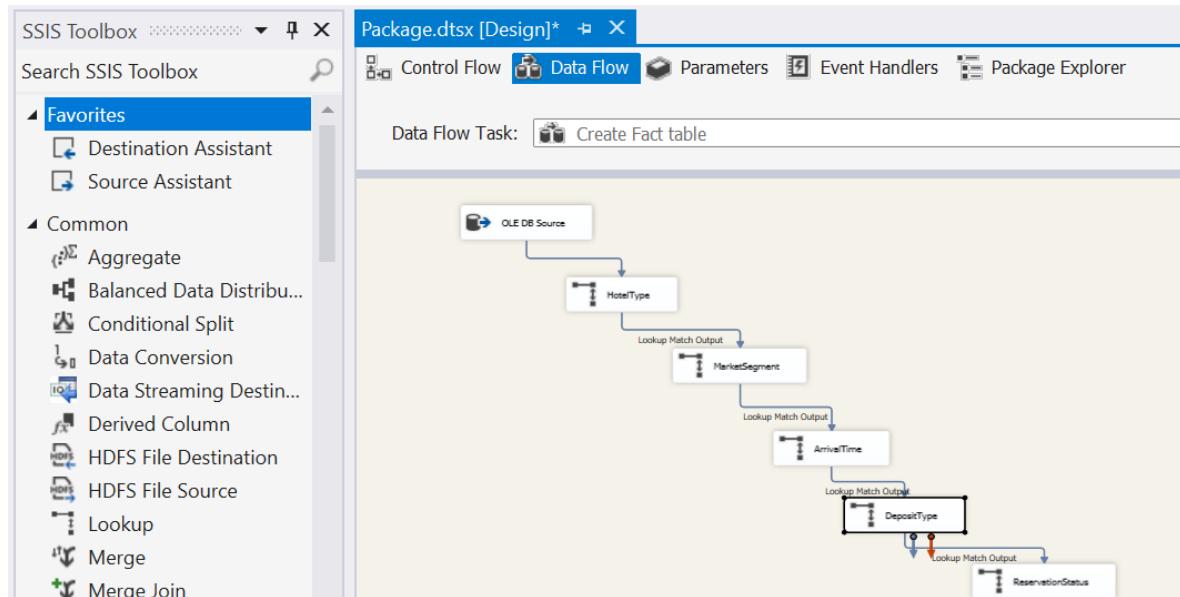


- + Trong khung Columns, nối thuộc tính deposit\_type với thuộc tính deposit\_type của bảng DepositType đã chọn ở bước trên và click vào id\_deposit\_type để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với DepositType.

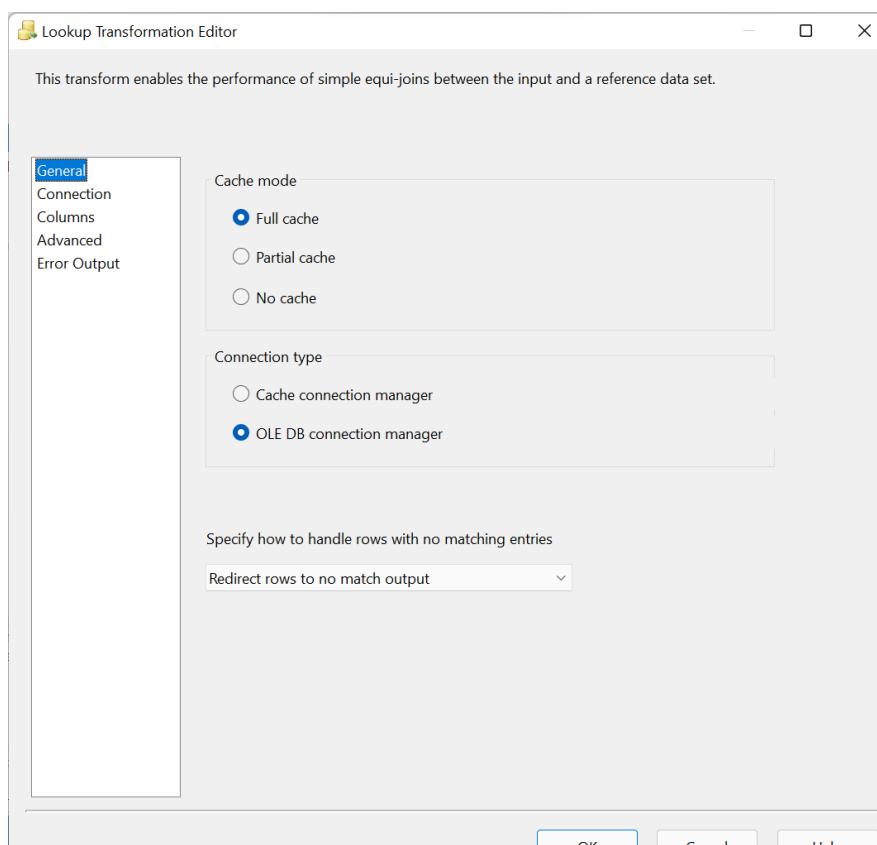


Hình 2. 156.1.,2.,3.,4. Tạo khóa ngoại với bảng Deposit Type bằng công cụ Lookup

- Bước 9: Chọn công cụ Lookup, đổi tên thành ReservationStatus để tạo khóa ngoại id\_reservation\_status đến bảng ReservationStatus và chọn đường kết nối hai Lookup là Lookup Match Output.

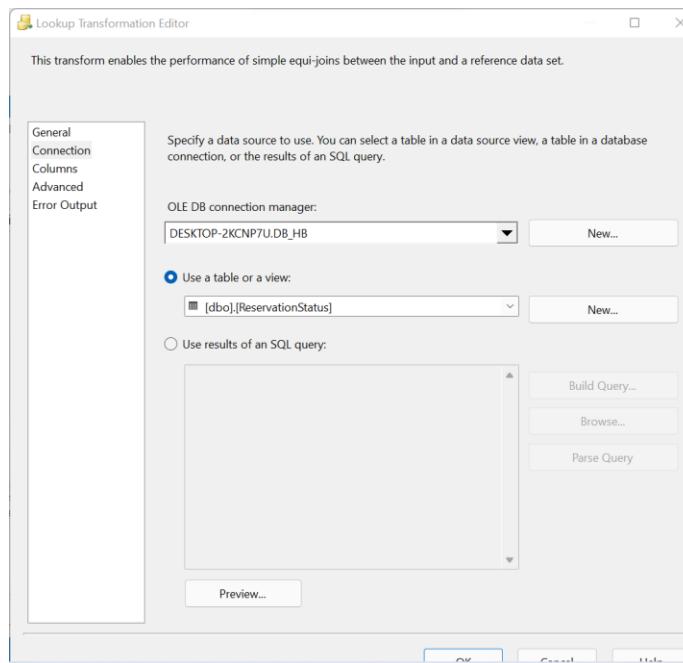


+ Đổi thành Redirects row to no match output

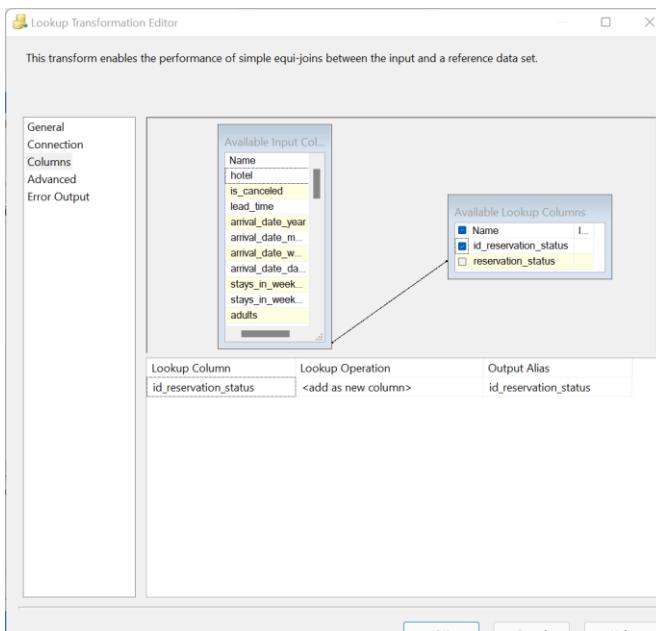


## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

+ Tại Connection, chọn bảng ReservationStatus đã tạo ở phần 4 chương 2



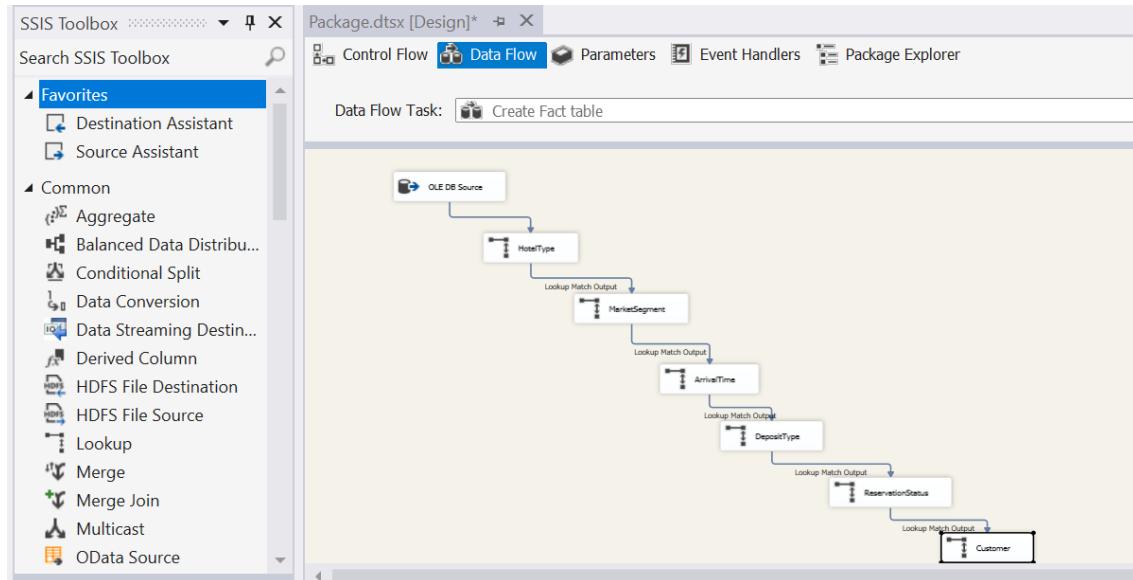
+ Trong khung Columns, nối thuộc tính reservation\_status với thuộc tính reservation\_status của bảng ReservationStatus đã chọn ở bước trên và click vào id\_reservation\_status để chọn làm thuộc tính khóa ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với ReservationStatus.



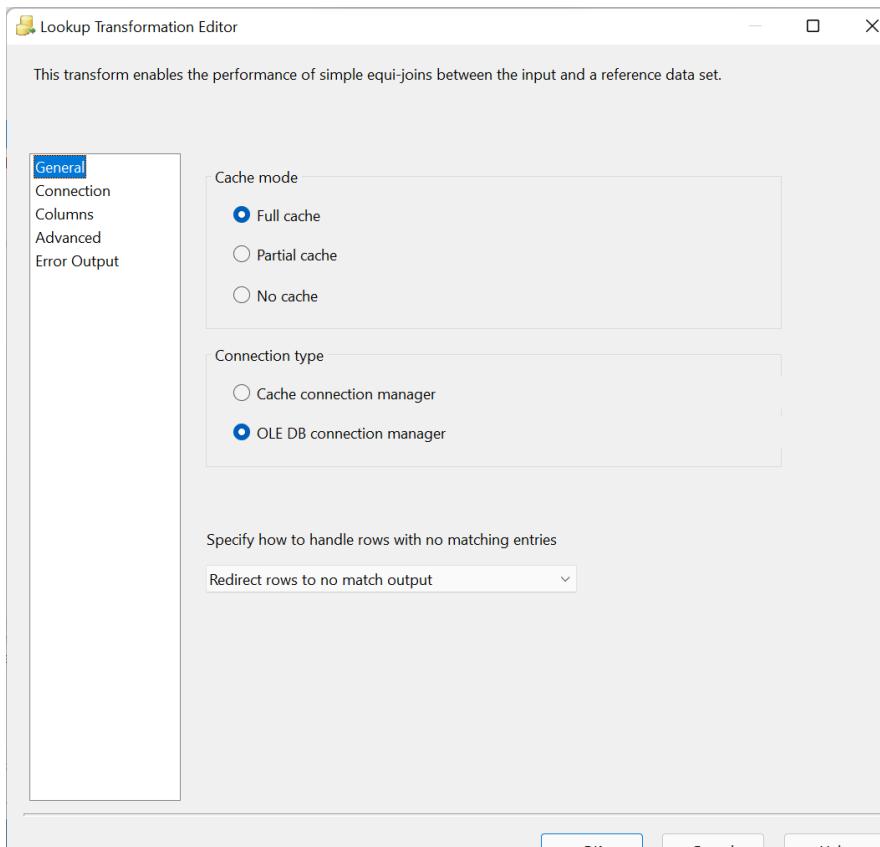
Hình 2. 157.1.,2.,3.,4. Tạo khóa ngoại với bảng ReservationStatus bằng công cụ Lookup

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: Chọn công cụ Lookup, đổi tên thành Customer để tạo khóa ngoại id\_customer đến bảng Customer và chọn đường kết nối hai Lookup là Lookup Match Output.

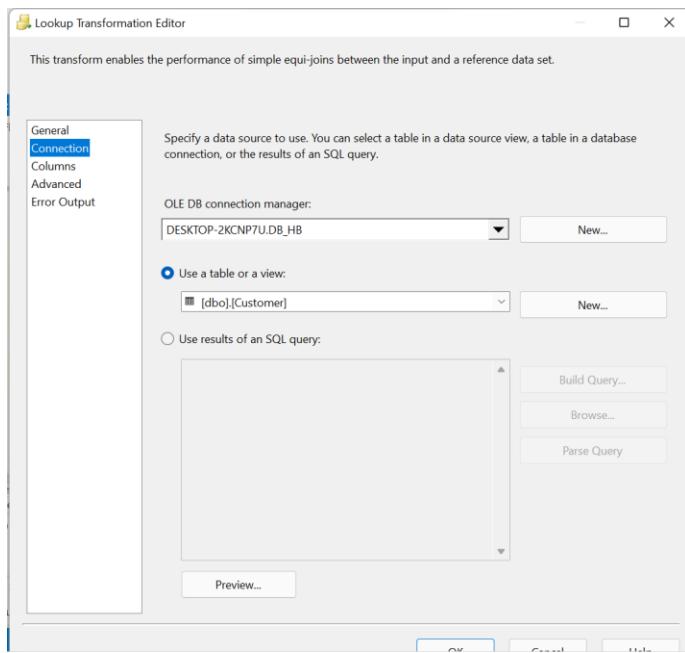


+ Đổi thành Redirects row to no match output

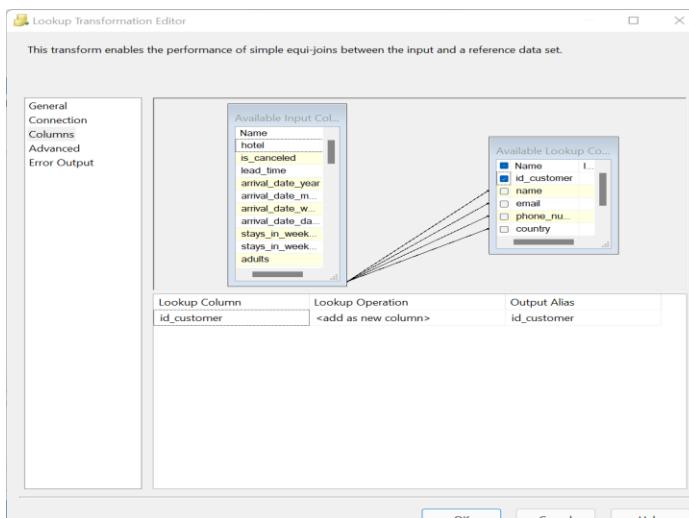


## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- + Tại Connection, chọn bảng Customer đã tạo ở phần 4 chương 2

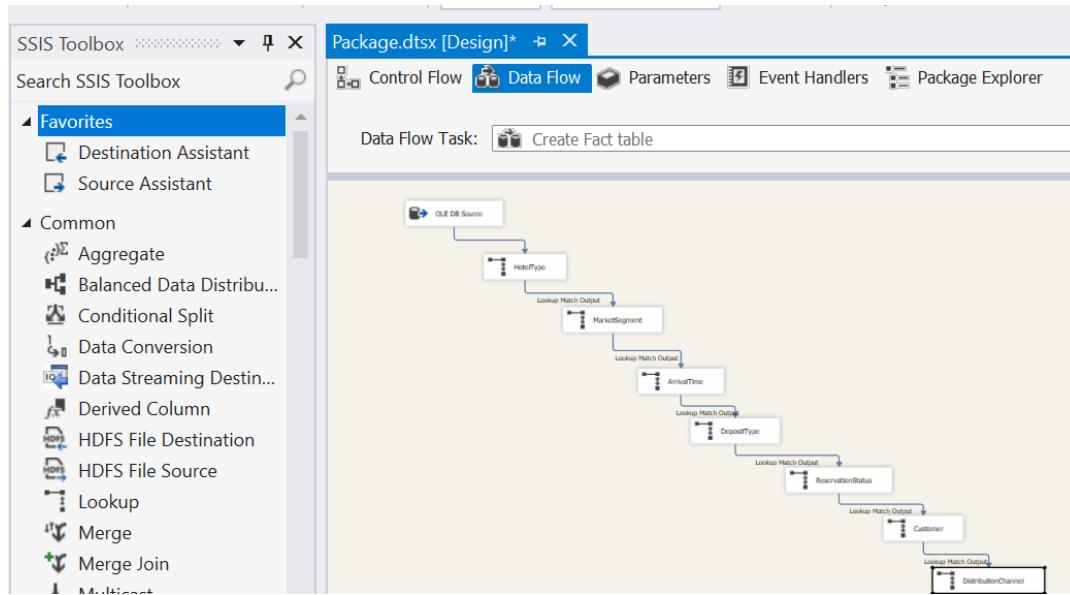


- + Trong khung Columns, nối thuộc tính name với thuộc tính name, email với email, phone-number với phone\_number và country với country của bảng Customer đã chọn ở bước trên và click vào id\_customer để chọn làm thuộc tính khóa ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với Customer.

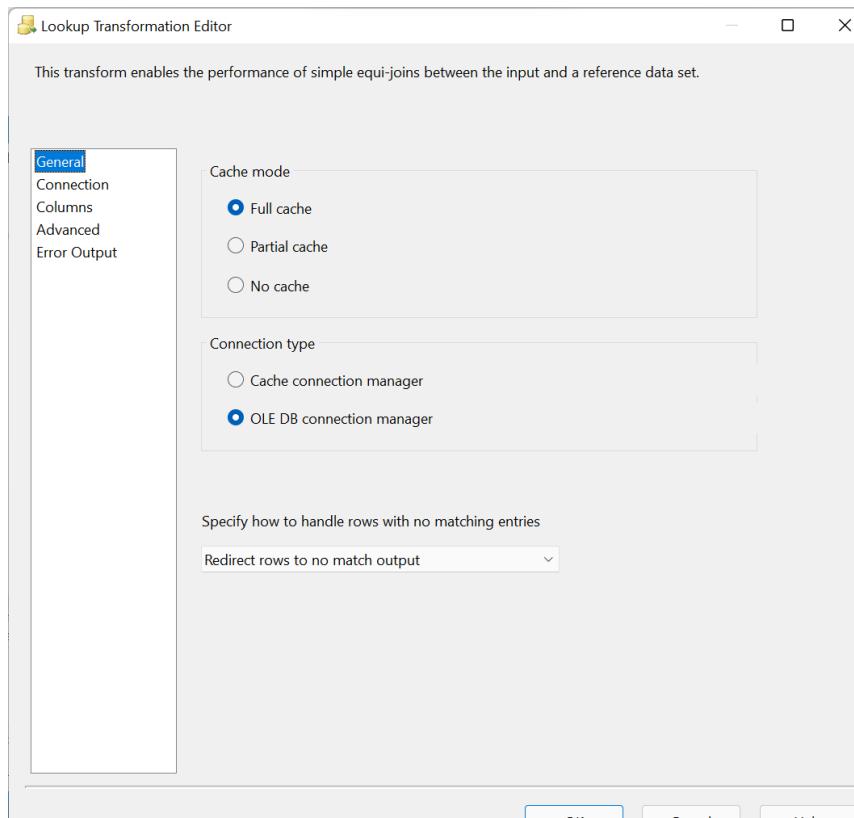


*Hình 2. 158.1.,2.,3.,4. Tạo khóa ngoại với bảng Customer bằng công cụ Lookup*

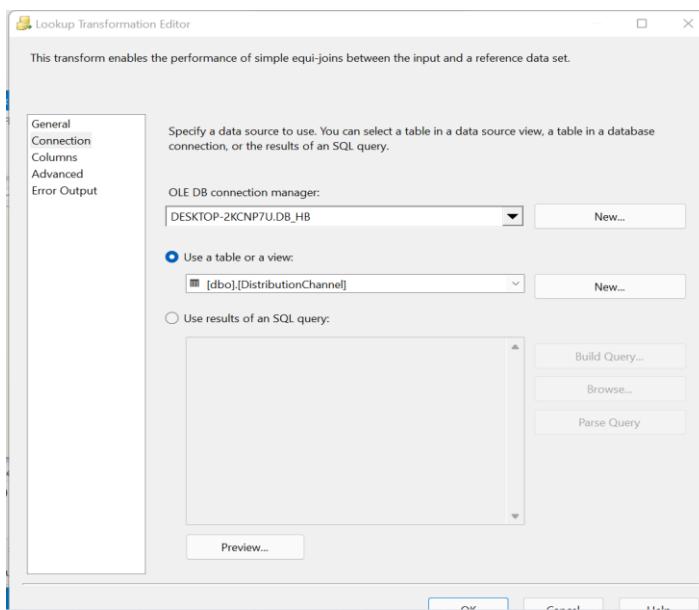
- Bước 11: Chọn công cụ Lookup, đổi tên thành DistributionChannel để tạo khóa ngoại id\_distribution\_channel đến bảng DistributionChannel và chọn đường kết nối giữ hai Lookup là Lookup Match Output.



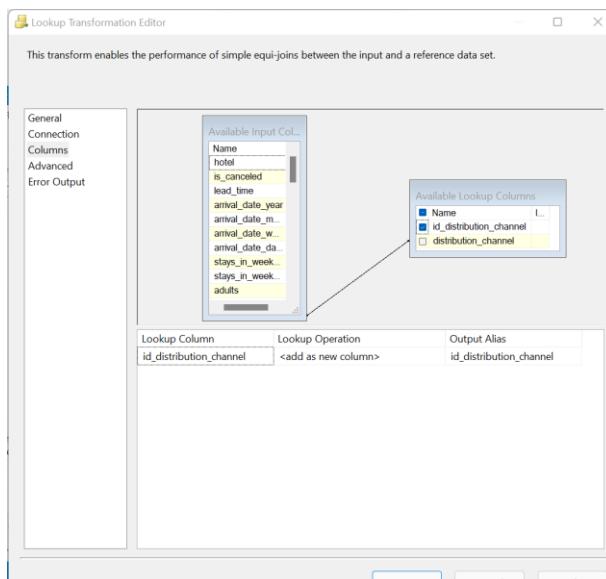
+ Đổi thành Redirects row to no match output



+ Tại Connection, chọn bảng DistributionChannel đã tạo ở phần 4 chương 2

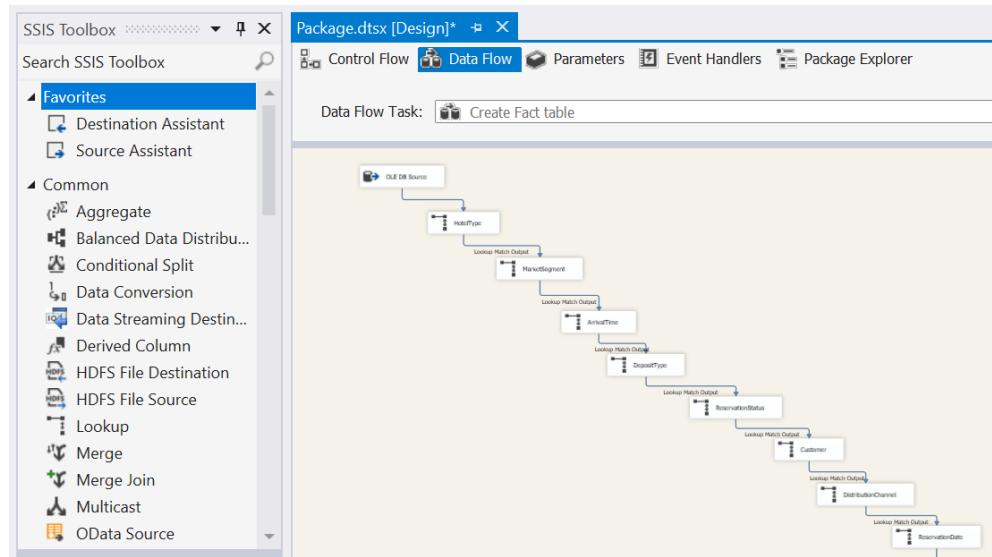


+ Trong khung Columns, nối thuộc tính distribution\_channel với thuộc tính distribution\_channel của bảng DistributionChannel đã chọn ở bước trên và click vào id\_distribution\_channel để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với DistributionChannel.

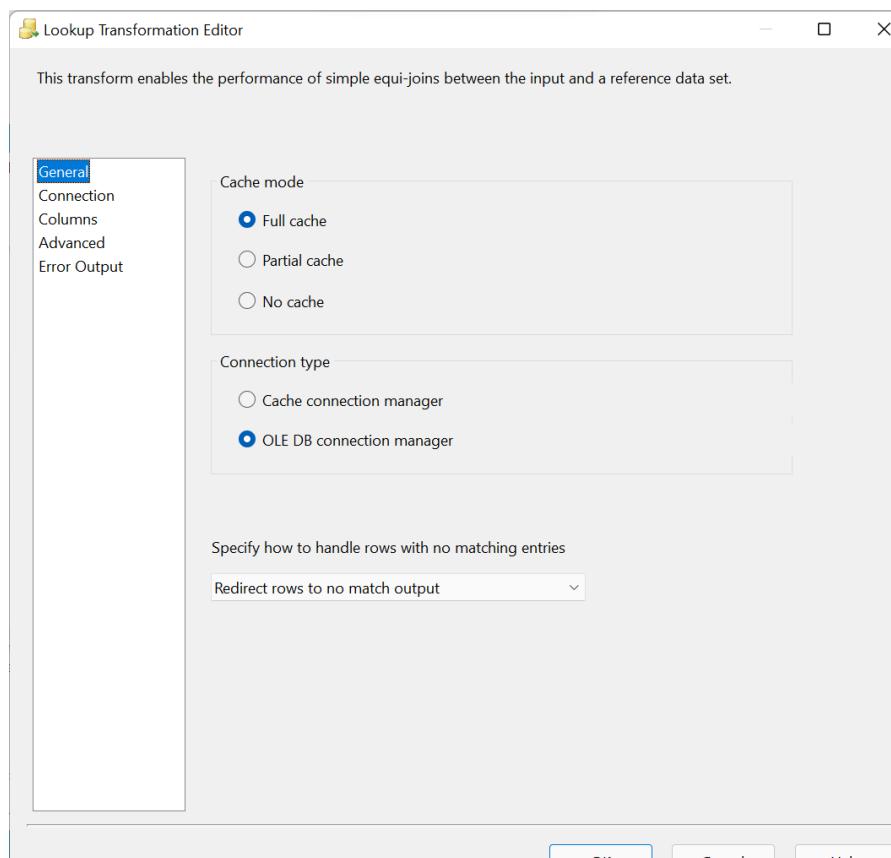


*Hình 2. 159.1.,2.,3.,4. Tạo khóa ngoại với bảng DistributionChannel bằng công cụ Lookup*

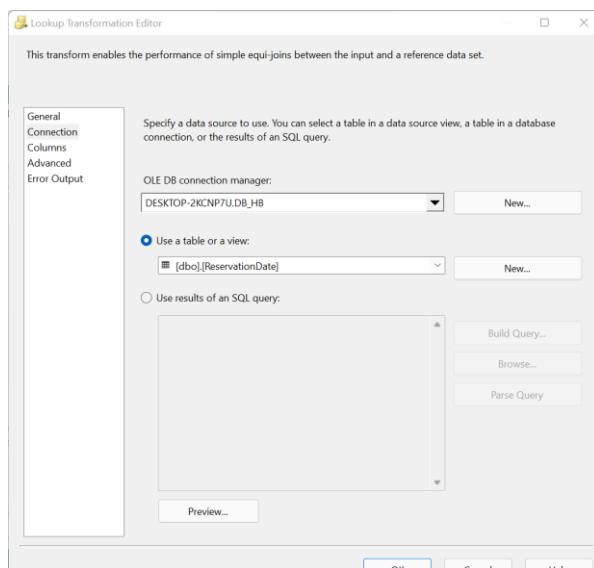
- Bước 12: Chọn công cụ Lookup, đổi tên thành ReservationDate để tạo khóa ngoại id\_reservation\_status\_date đến bảng ReservationDate và chọn đường kết nối giữ hai Lookup là Lookup Match Output.



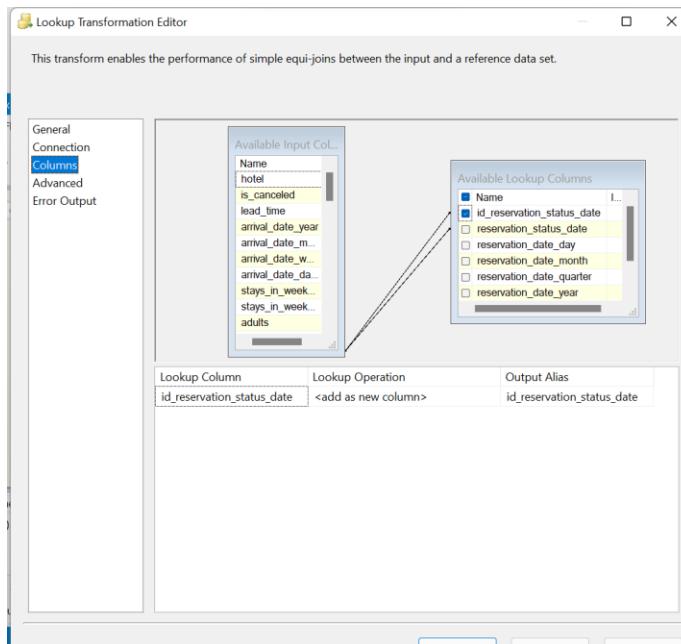
+ Đổi thành Redirects row to no match output



+ Tại Connection, chọn bảng ReservationDate đã tạo ở phần 4 chương 2

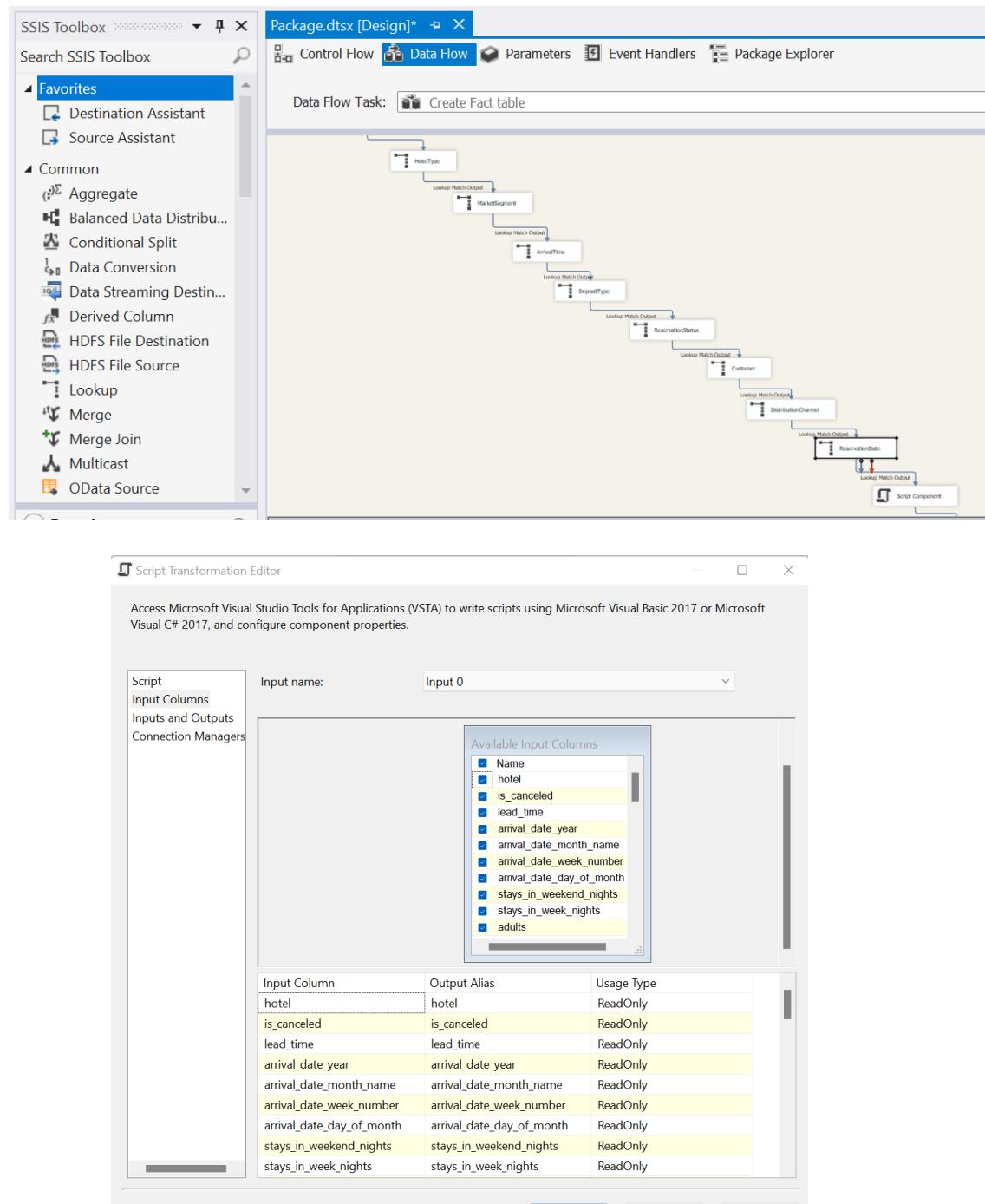


+ Trong khung Columns, nối thuộc tính reservation\_status\_date với thuộc tính reservation\_status\_date,id\_reservation\_status\_date với thuộc tính id\_reservation\_status\_date của bảng ReservationDate đã chọn ở bước trên và click vào id\_reservation\_status\_date để chọn làm thuộc tính khoá ngoại đối với bảng OLE DB Source này. Tiếp tục nhấn OK để hoàn thành kết nối giữa OLE DB Source với ReservationDate.

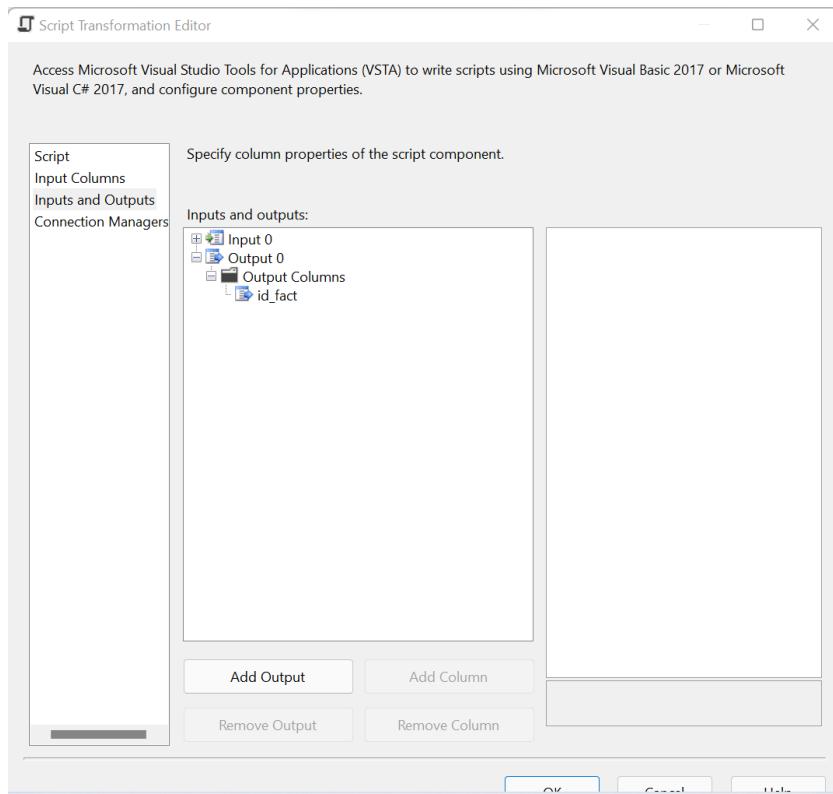


*Hình 2. 160.1.,2.,3.,4. Tạo khóa ngoại với bảng ReservationDate bằng công cụ Lookup*

- Bước 13: Kéo thả công cụ Script Component để tạo khóa chính id\_fact.



## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



main.cs

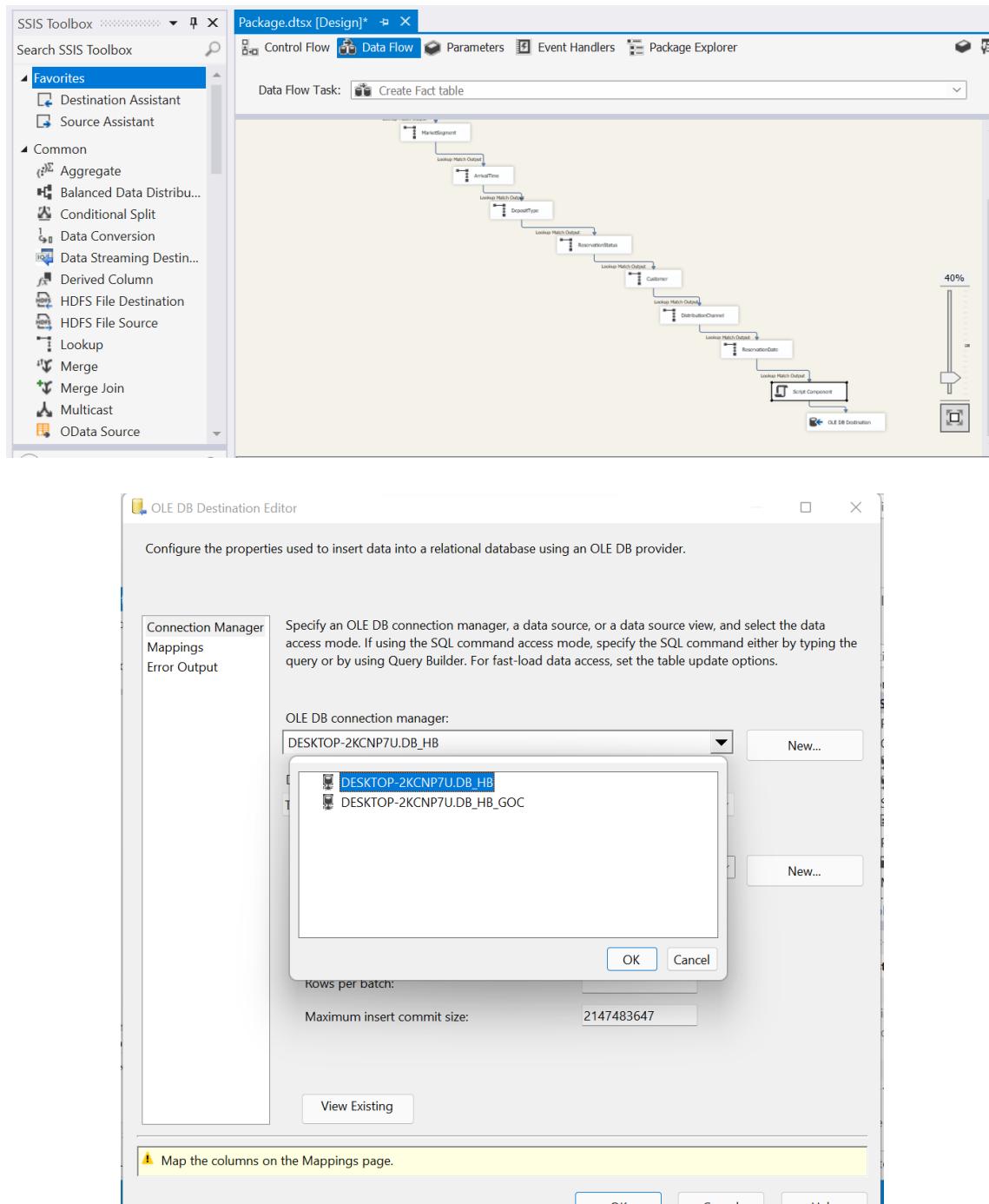
```

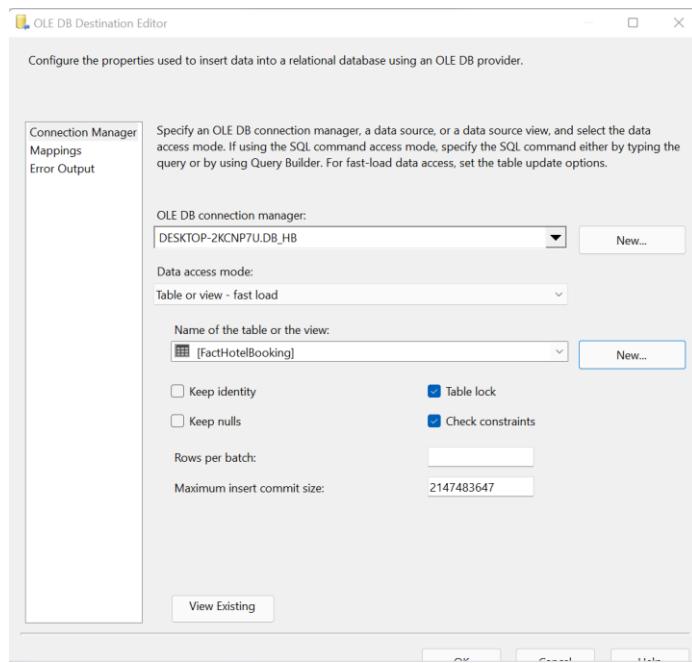
1 // Help: Introduction to the Script Component
2
3 //Namespace
4
5 //<summary>
6 // This is the class to which to add your code. Do not change the name, attributes, or parent
7 // of this class.
8 //</summary>
9 [Microsoft.SqlServer.Dts.Pipeline.SSIScriptComponentEntryPointAttribute]
10
11 public class ScriptMain : UserComponent
12 {
13     // Help: Using Integration Services variables and parameters
14     // Help: Using Integration Services Connection Managers
15     // Help: Firing Integration Services Events
16     //<summary>
17     // This method is called once, before rows begin to be processed in the data flow.
18     // You can remove this method if you don't need to do anything here.
19     //</summary>
20
21     public override void PreExecute()
22     {
23         base.PreExecute();
24         /*
25          * Add your code here
26          */
27     }
28
29     //<summary>
30     // This method is called after all the rows have passed through this component.
31     // You can delete this method if you don't need to do anything here.
32     //</summary>
33
34     public override void PostExecute()
35     {
36         base.PostExecute();
37         /*
38          * Add your code here
39          */
40     }
41
42     //<summary>
43     // This method is called once for every row that passes through the component from Input0.
44     // Example of reading a value from a column in the row:
45     // string zipCode = Row.ZipCode
46     // Example of writing a value to a column in the row:
47     // Row.ZipCode = zipCode
48     //</summary>
49     // param name="Row">The row that is currently passing through the component./param>
50
51     public override void Input0_ProcessingRow(InputBuffer Row)
52     {
53         Row.idfact = count;
54         count++;
55     }
56 }
57
58 
```

Hình 2. 161.1,.2,3,4.Sử dụng Script Component để tạo khóa chính id\_fact

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

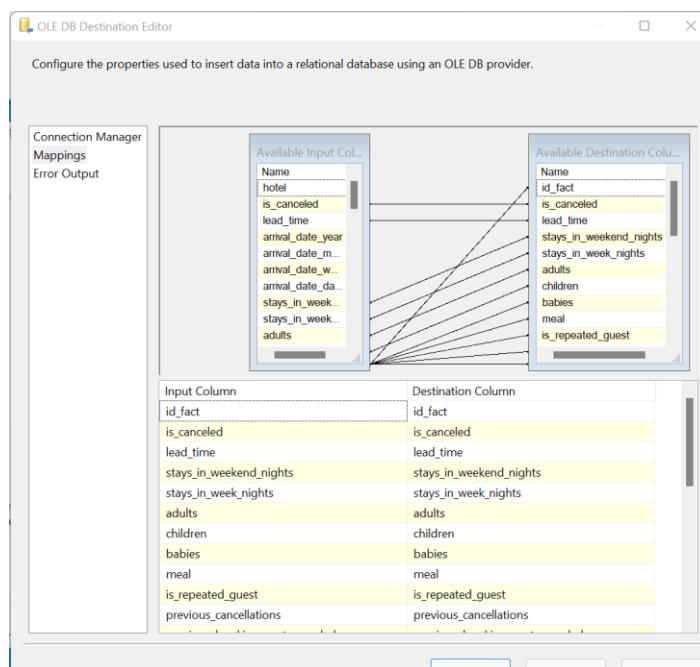
- Bước 14: Chọn công cụ OLE DB Destination để tiến hành tạo bảng trong cơ sở dữ liệu DB\_HB với tên gọi là FactHotelBooking.





*Hình 2. 162.1,2,3.Tạo bảng FactHotelBooking*

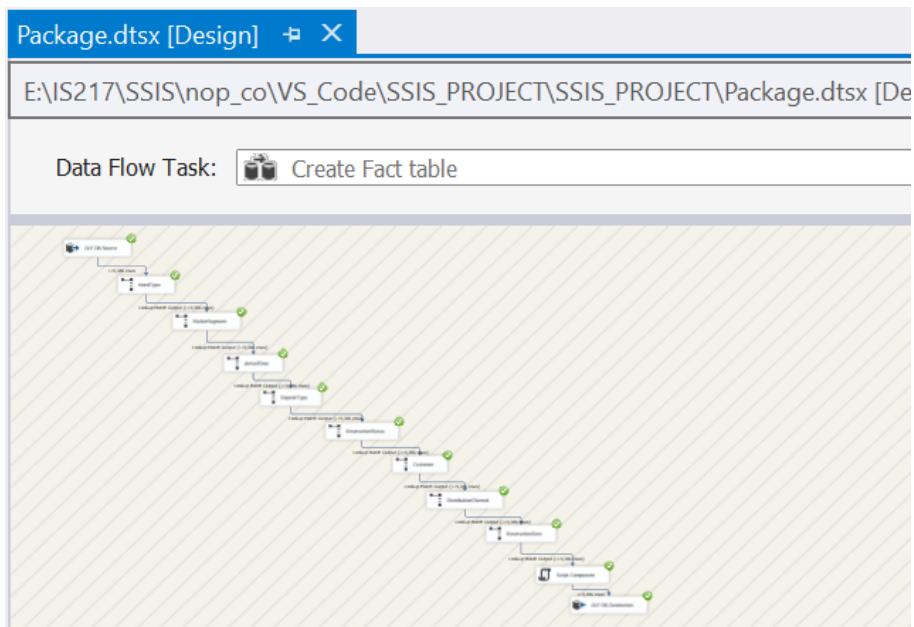
- Bước 15: Nhấn Mappings để kiểm tra kết nối và nhấn Ok để hoàn tất quá trình tạo bảng.



*Hình 2. 163..Quá trình Mappings dữ liệu bảng FactHotelBooking*

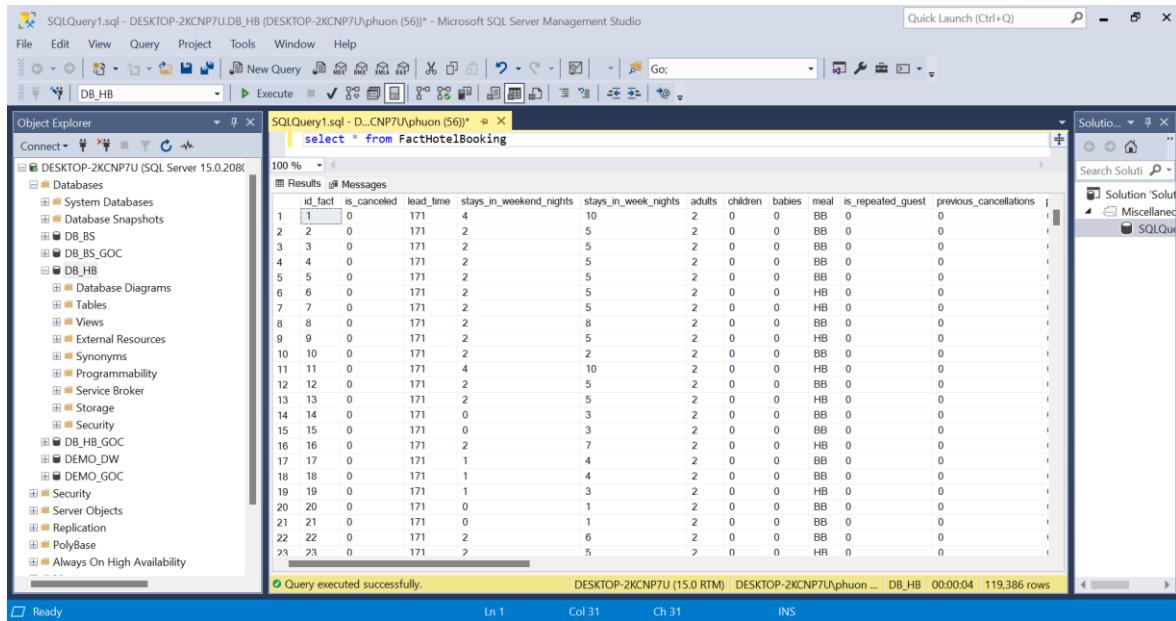
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 16: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:



*Hình 2. 164..Hoàn thành đổ dữ liệu vào FactHotelBooking trong kho dữ liệu*

- Bước 17: Kiểm tra bảng FactHotelBooking trên SQL Server.



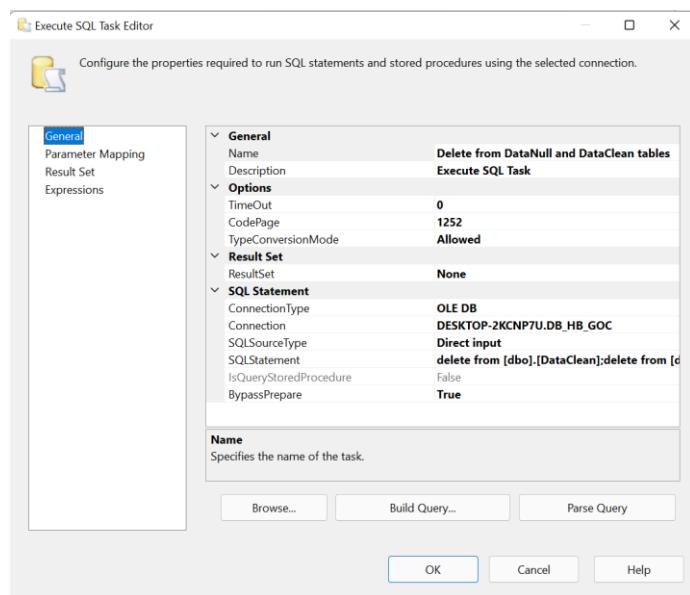
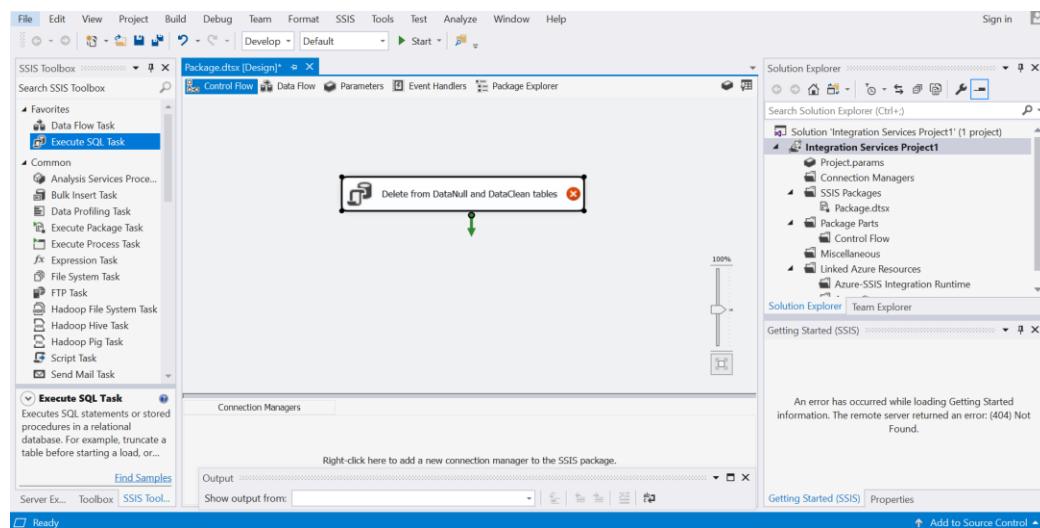
1	1	0	171	4	10	2	0	0	BB	0	0
2	2	0	171	2	5	2	0	0	BB	0	0
3	3	0	171	2	5	2	0	0	BB	0	0
4	4	0	171	2	5	2	0	0	BB	0	0
5	5	0	171	2	5	2	0	0	BB	0	0
6	6	0	171	2	5	2	0	0	HB	0	0
7	7	0	171	2	5	2	0	0	HB	0	0
8	8	0	171	2	8	2	0	0	BB	0	0
9	9	0	171	2	5	2	0	0	HB	0	0
10	10	0	171	2	2	2	0	0	BB	0	0
11	11	0	171	4	10	2	0	0	HB	0	0
12	12	0	171	2	5	2	0	0	BB	0	0
13	13	0	171	2	5	2	0	0	HB	0	0
14	14	0	171	0	3	2	0	0	BB	0	0
15	15	0	171	0	3	2	0	0	BB	0	0
16	16	0	171	2	7	2	0	0	HB	0	0
17	17	0	171	1	4	2	0	0	BB	0	0
18	18	0	171	1	4	2	0	0	BB	0	0
19	19	0	171	1	3	2	0	0	HB	0	0
20	20	0	171	0	1	2	0	0	BB	0	0
21	21	0	171	0	1	2	0	0	BB	0	0
22	22	0	171	2	6	2	0	0	BB	0	0
23	23	0	171	2	5	2	0	0	HR	0	0

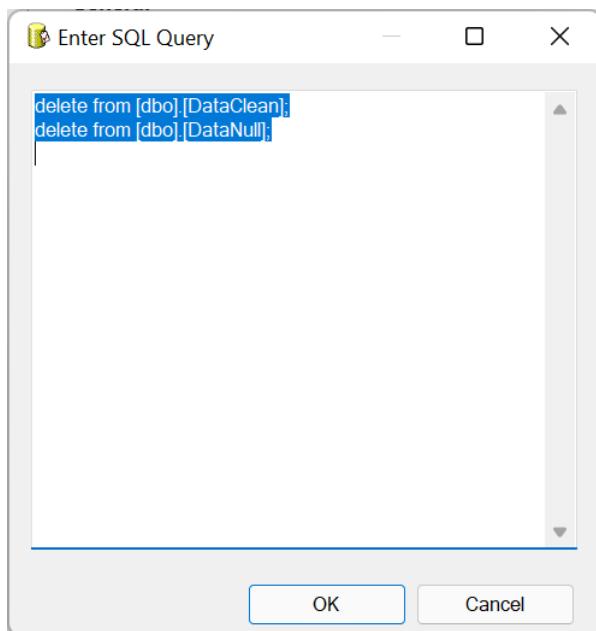
*Hình 2. 165.Kiểm tra bảng FactHotelBooking trên SQL Server*

## 6. Tạo Execute SQL Task

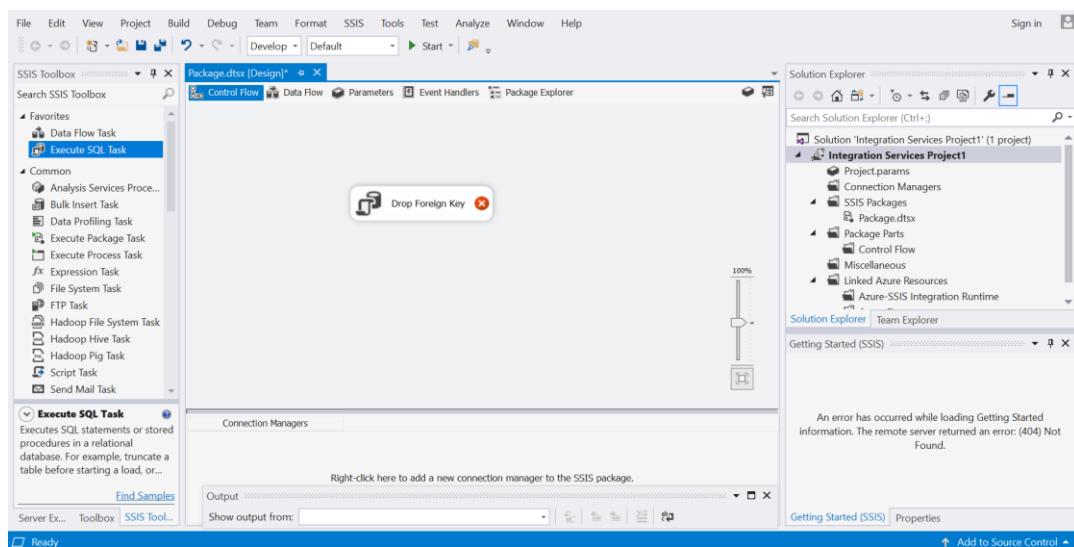
❖ **Execute SQL Task** : Thực thi chạy các câu lệnh SQL hoặc các procedure được lưu trữ từ package. Tác vụ có thể chứa lệnh SQL đơn hoặc nhiều câu lệnh SQL chạy tuân tự.

- Bước 1: Kéo thả các Execute SQL Task và lần lượt đặt tên là Delete from DataNull and DataClean tables, Drop Foreign key, Delete from Dimension and Fact tables, Create Foreign key. Bổ sung các câu lệnh SQL để ngăn ngừa việc trùng khi đổ dữ liệu vào những lần sau và tạo kết nối giữa các bảng Dimension và bảng Fact.

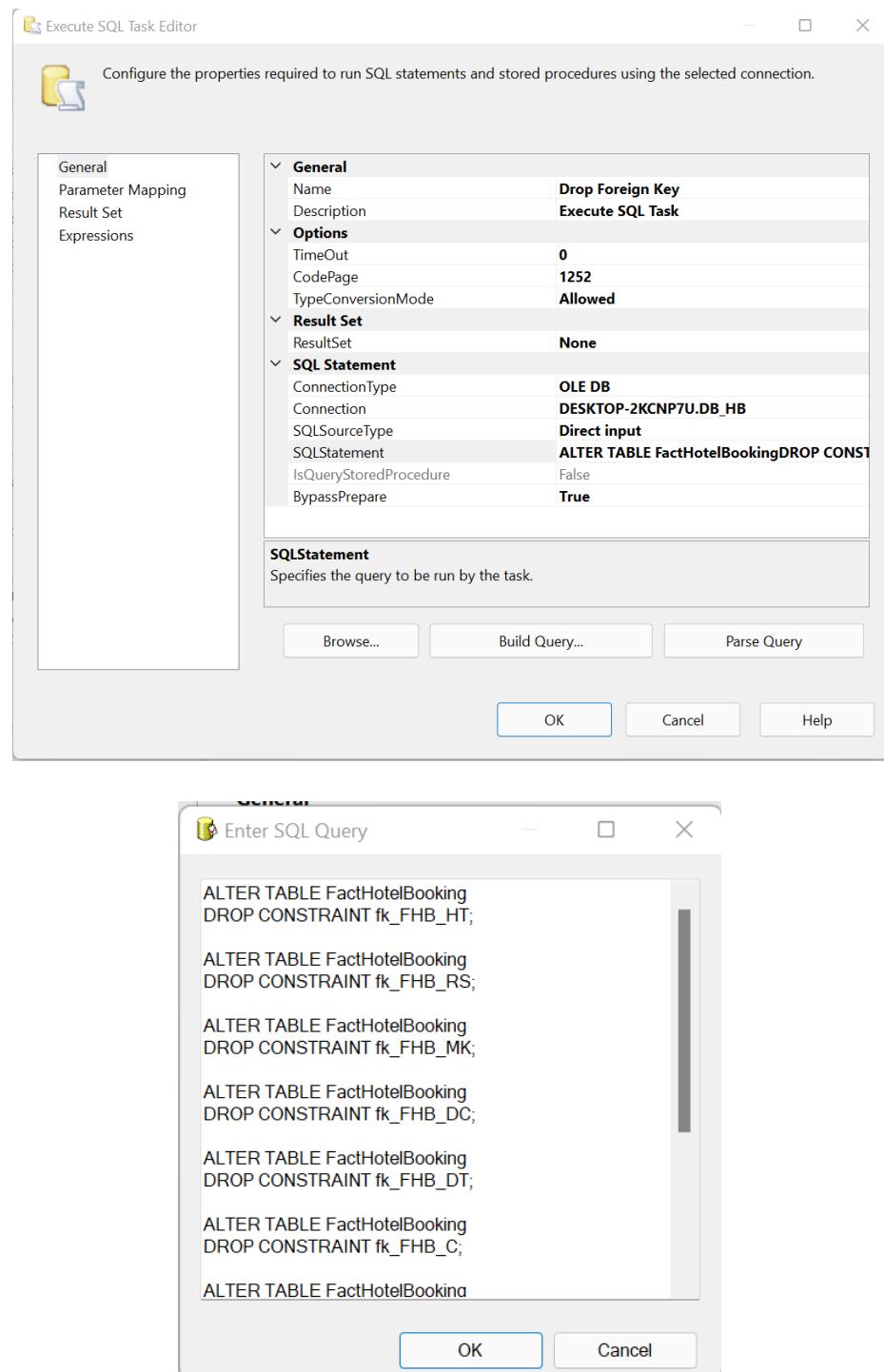




Hình 2. 166.1,2,3. Viết câu lệnh loại bỏ trùng sau mỗi lần dò dữ liệu của bảng DataNull và Data Clean

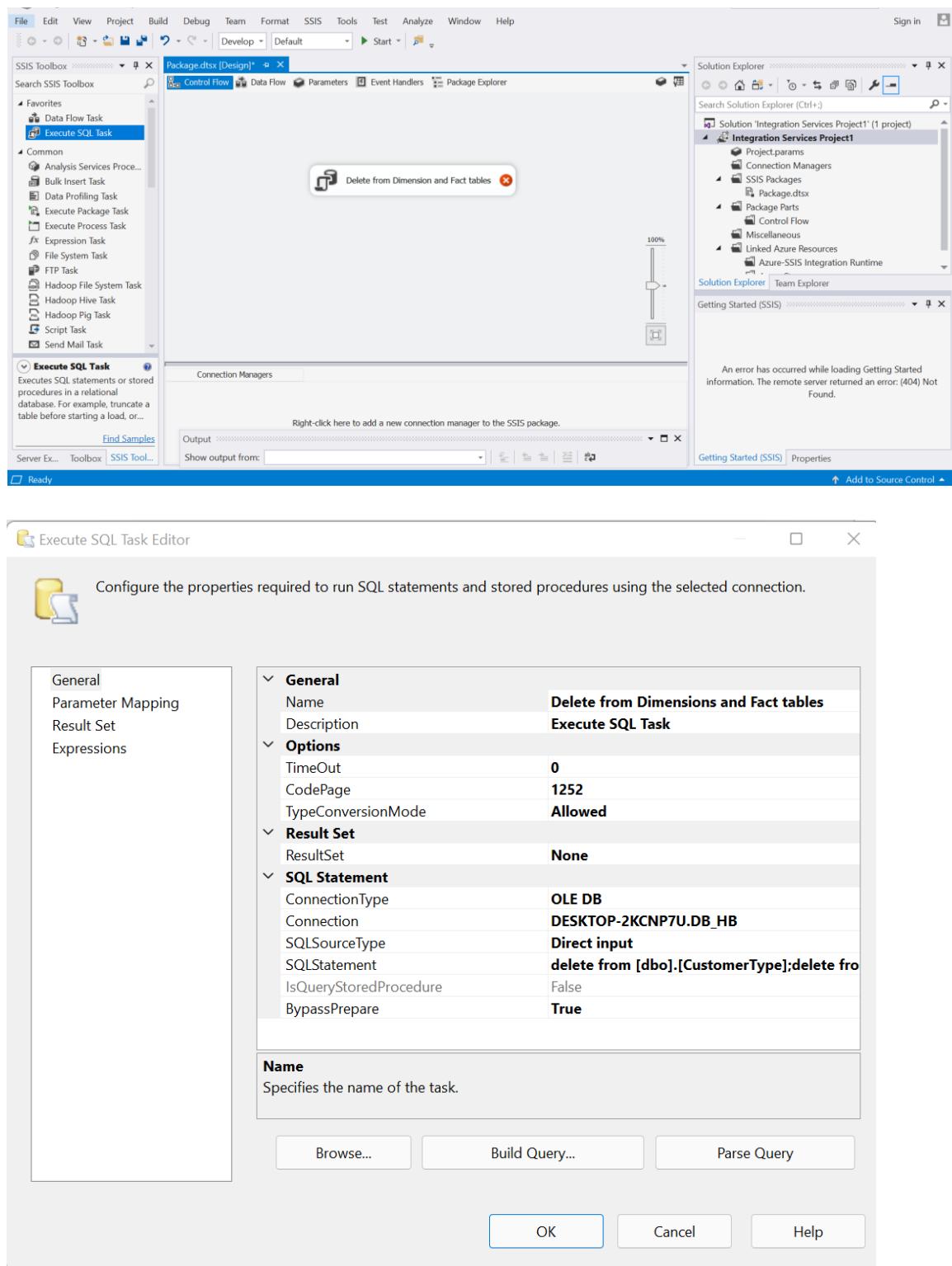


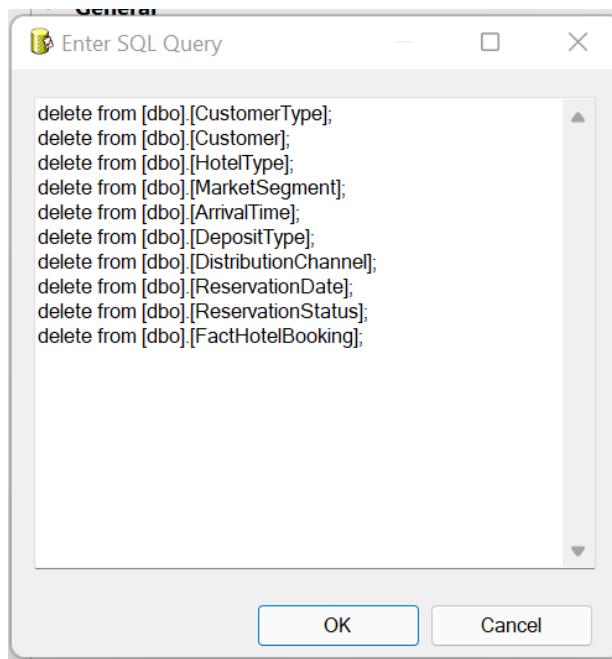
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



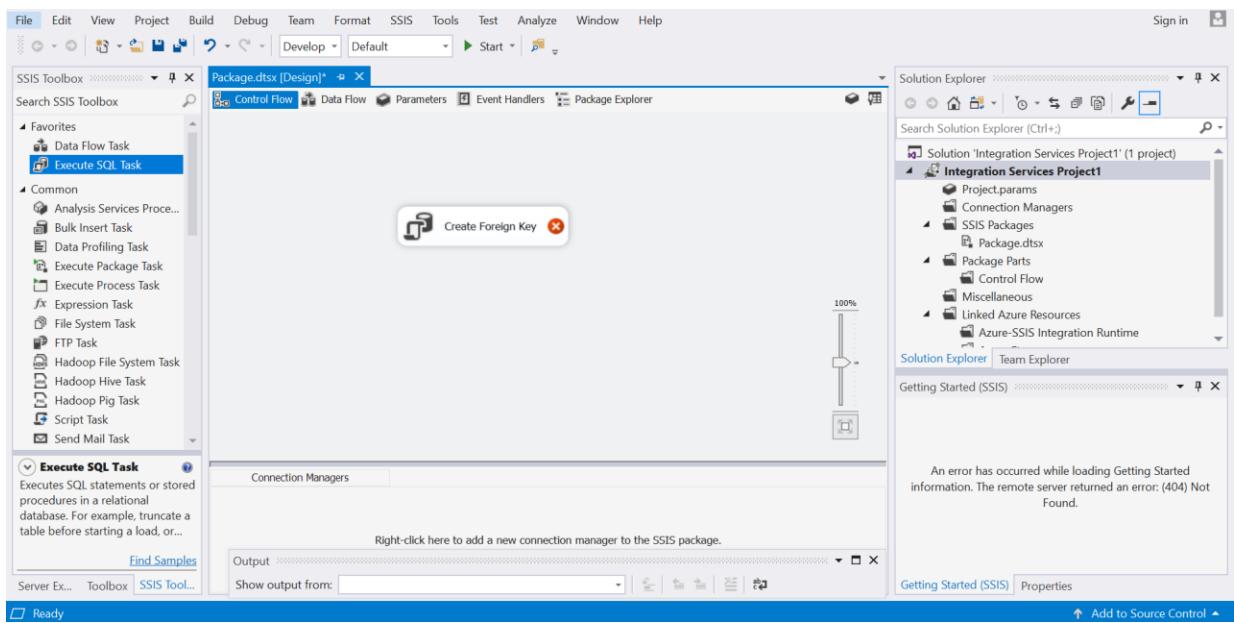
*Hình 2. 167.1,2,3. Viết câu lệnh loại bỏ trùng sau mỗi lần đổ dữ liệu của bảng FactHotelBooking*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

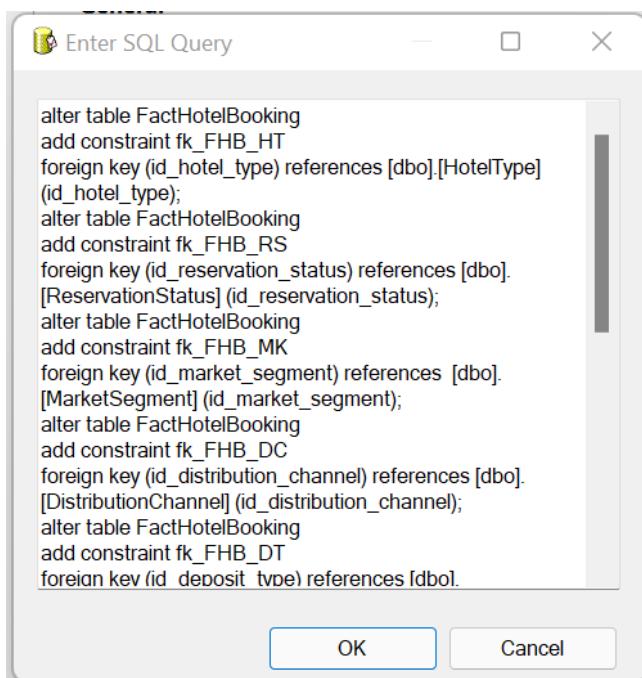
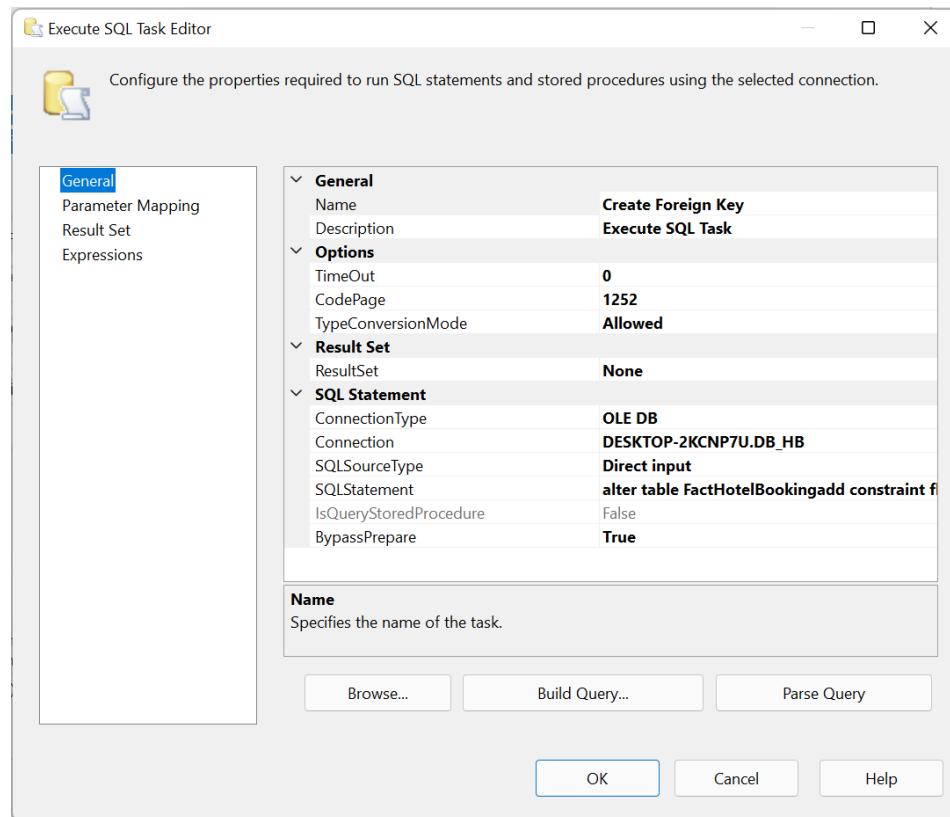




*Hình 2. 168.1,2,3. Viết câu lệnh loại bỏ trùng sau mỗi lần đổ dữ liệu của bảng Delete from Dimension and Fact*



## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



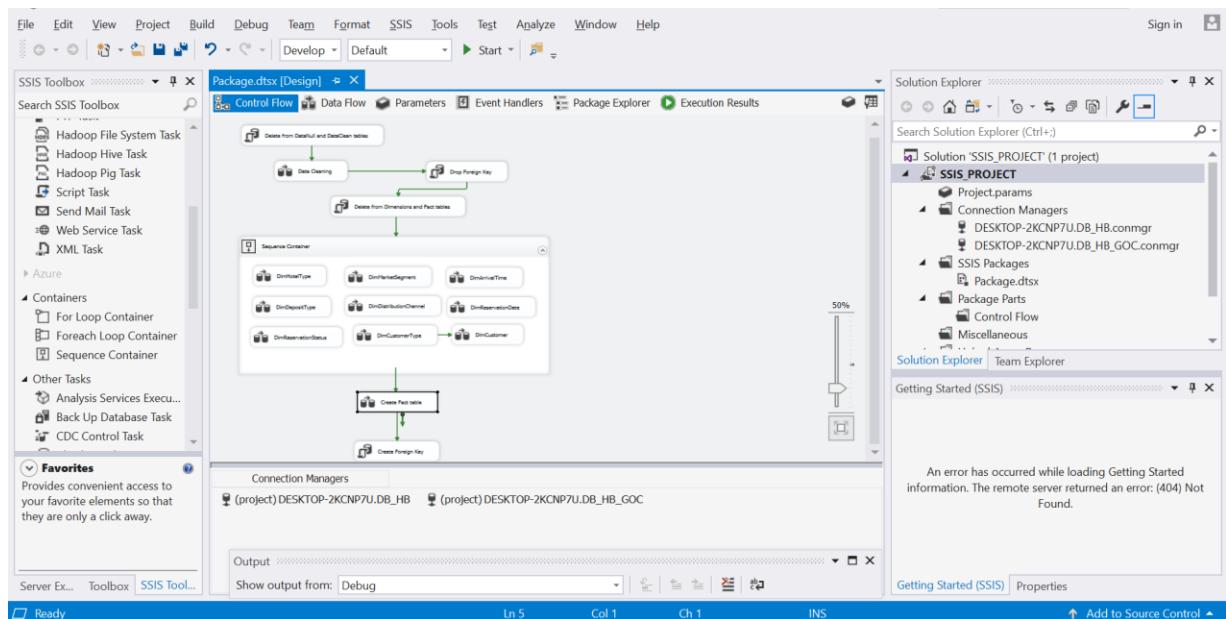
Hình 2. 169.1,2,3. Viết câu lệnh loại bỏ trùng sau mỗi lần đổ dữ liệu của bảng Create Foreign Key

- Bước 2: Chọn Sequence Container và chuyển các Dimension tables vào trong đó.



*Hình 2. 170. Các bảng Dimensions trong Sequence Container*

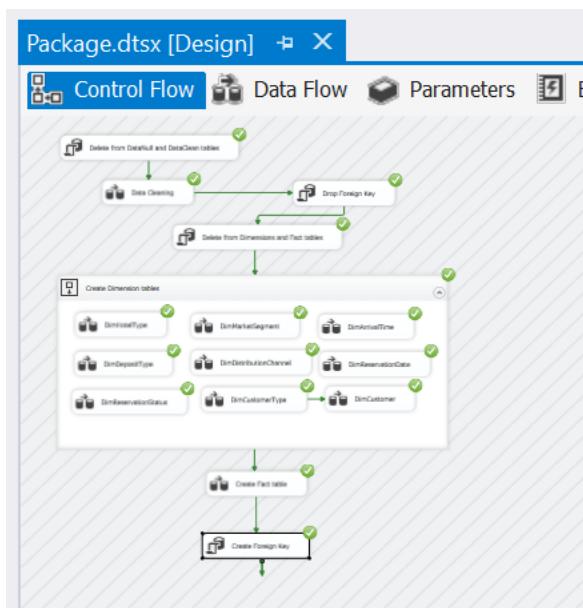
- Bước 3: Tạo đường kết nối giữa các thành phần với nhau theo thứ tự



*Hình 2. 171. Các thành phần được kết nối với nhau theo thứ tự*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 4: Chọn Start để bắt đầu quá trình đổ dữ liệu. Khi đổ thành công, ta có kết quả như sau:

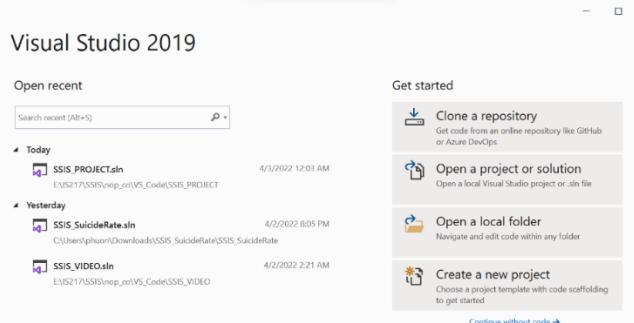


*Hình 2. 172. Hoàn thành đổ dữ liệu vào kho dữ liệu*

### CHƯƠNG III : PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS)

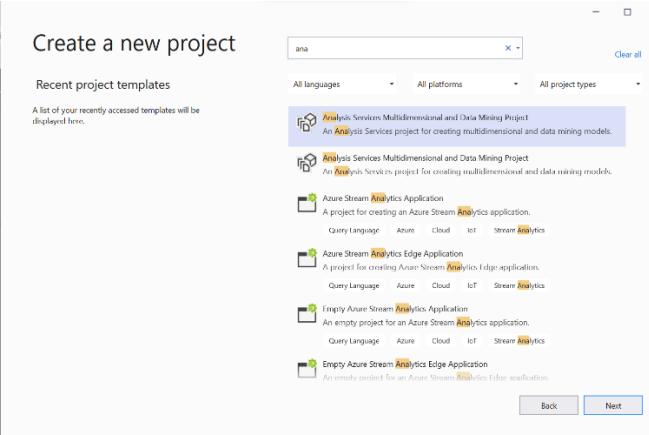
#### 1. Tạo project SSAS trong Visual Studio 2019

- Bước 1: Chọn phần mềm Visual Studio 2019 → Create a new project như hình:



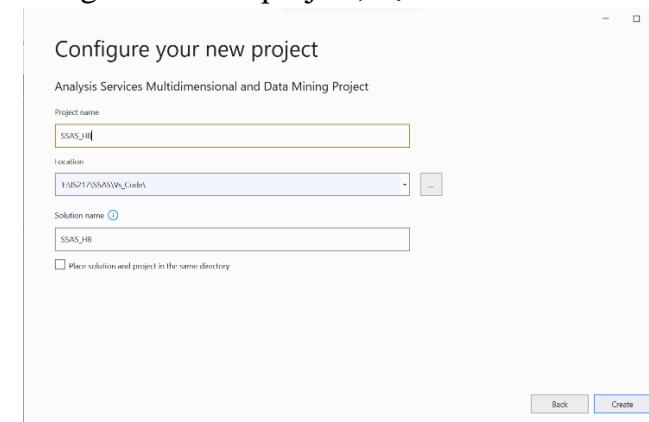
Hình 3.1: Giao diện khi sử dụng Visual Studio 2019

- Bước 2: Tìm project có công cụ Analysis Services Multidimensional and Data Mining project → Next như hình:



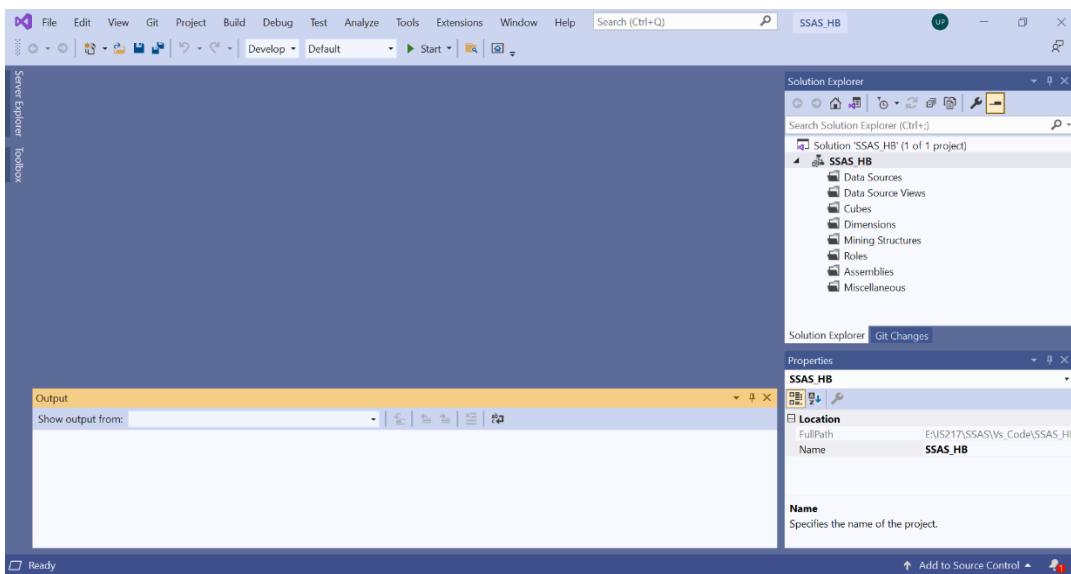
Hình 3.2 Tìm kiếm một Project SSAS của Visual Studio

- Bước 3: Chọn đường dẫn để lưu project, đặt tên và nhấn OK.



Hình 3.3 Giao diện cấu hình tên file của SSAS

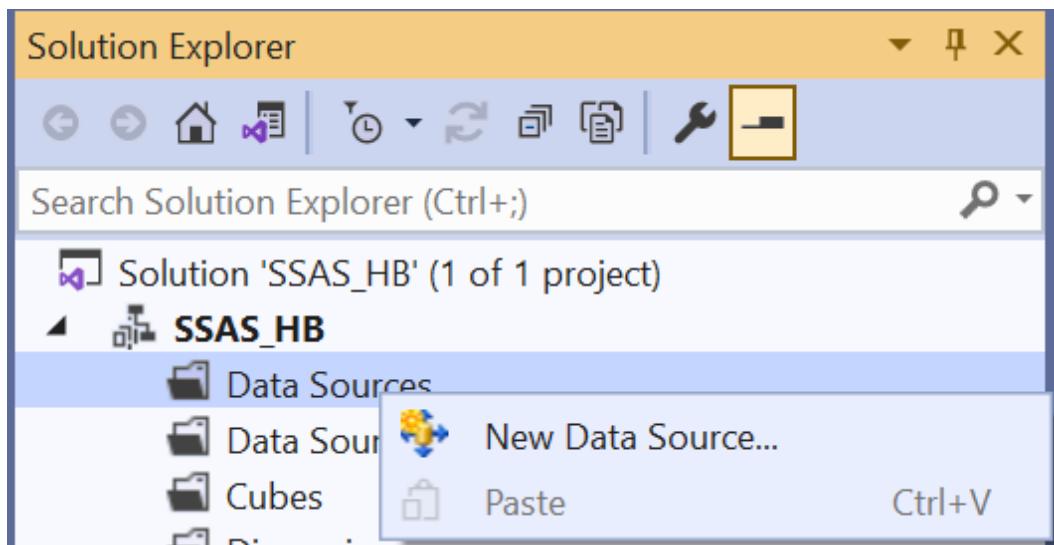
- Bước 4: Giao diện quá trình SSAS của công cụ BI hiện ra.



*Hình 3.4. Giao diện của công cụ BI quá trình SSAS*

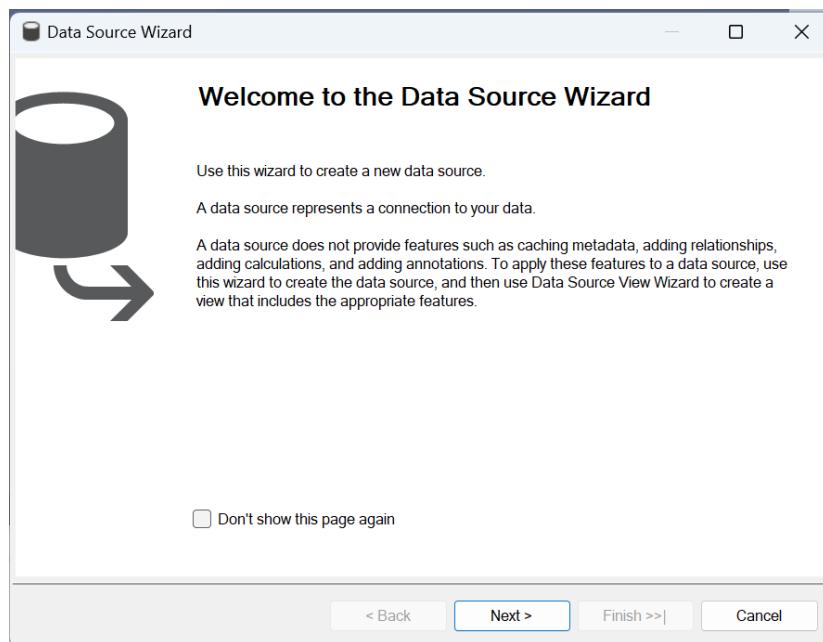
## **2. Xác định dữ liệu nguồn (DEFINE DATA SOURCE)**

- Bước 1: Bên gốc phải phần Solution Explorer nhấn chuột phải vào Data Source và chọn New Data Source.



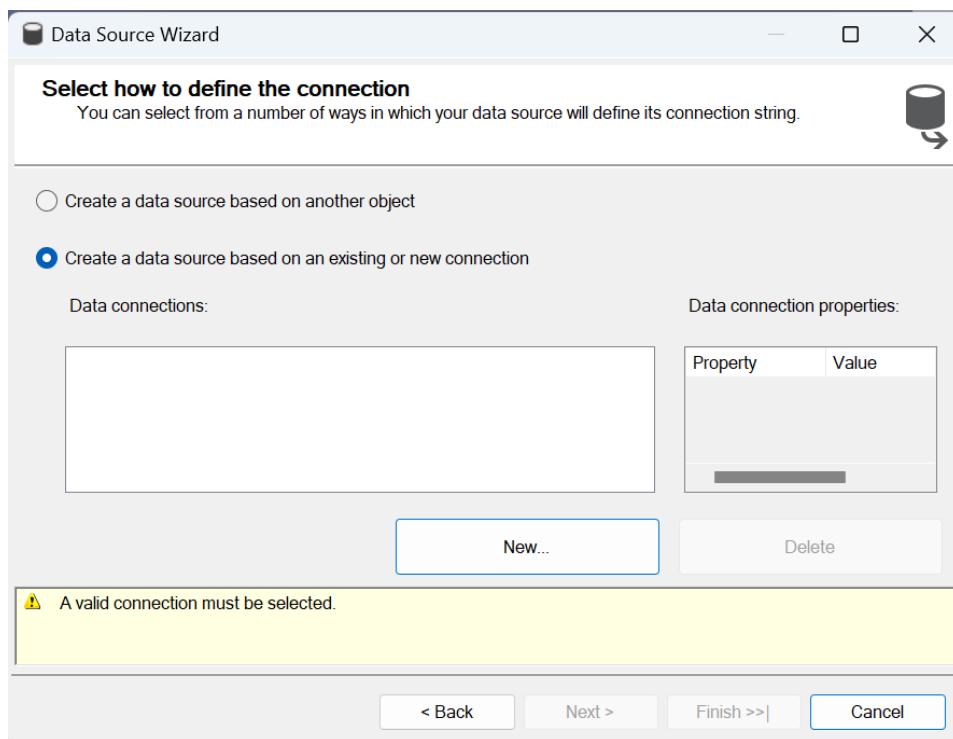
*Hình 3.5. Tạo một Data Source cho quá trình SSAS*

- Bước 2: Một cửa sổ hiện ra tiếp tục nhấn Next để tiếp tục.



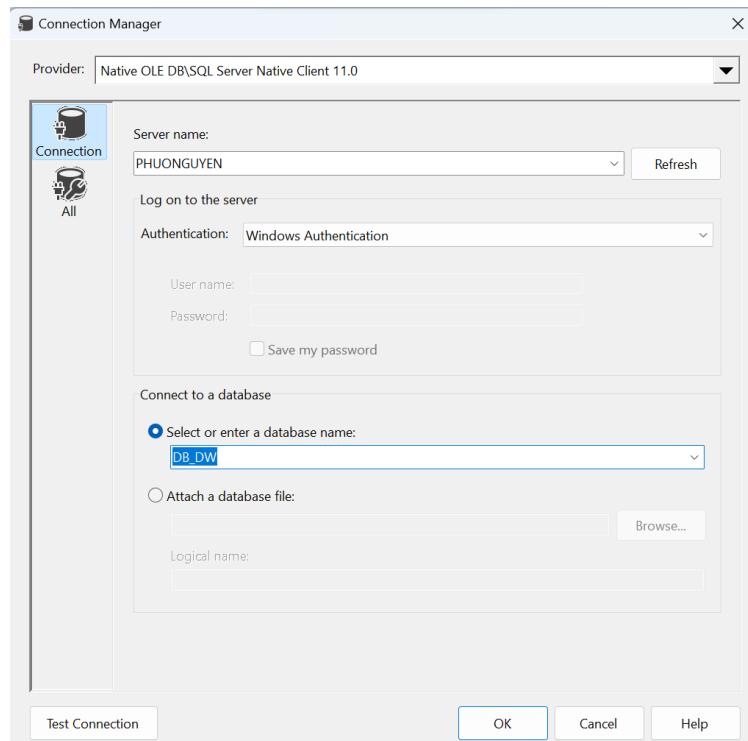
*Hình 3.6. Phần giới thiệu Data Source Wizard*

- Bước 3: Chọn kết nối đã tạo từ quá trình SSIS rồi tiếp tục nhấn Next.

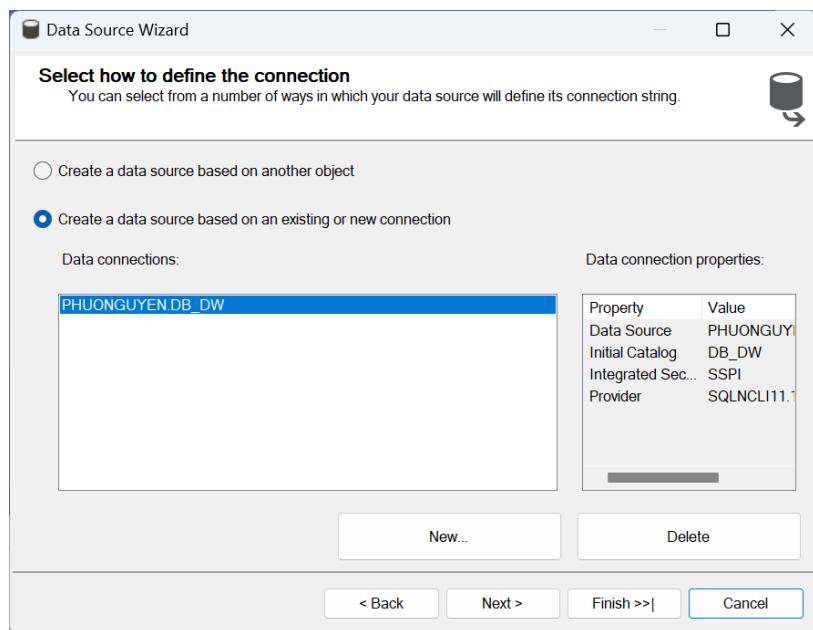


*Hình 3.7. Chọn kết nối với Data Source đã có sẵn hoặc tạo kết nối mới*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

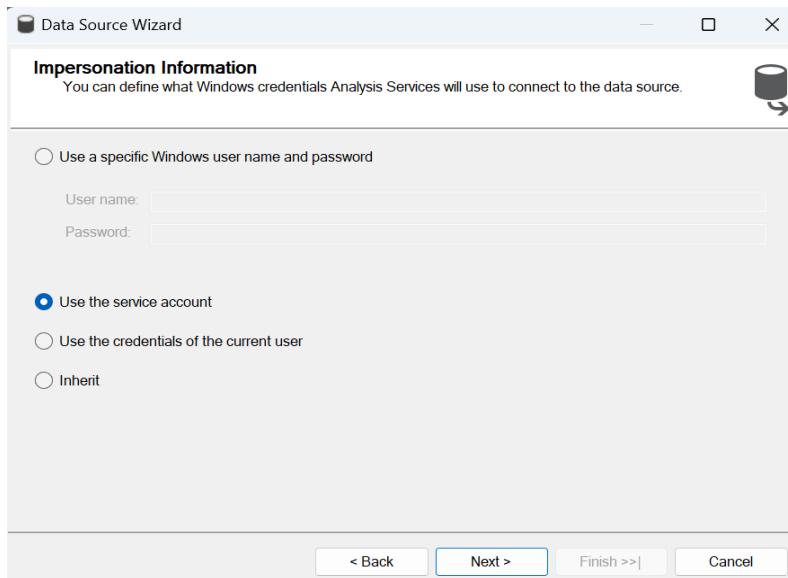


*Hình 3.8. Tao kết nối mới khi project chạy lần đầu*



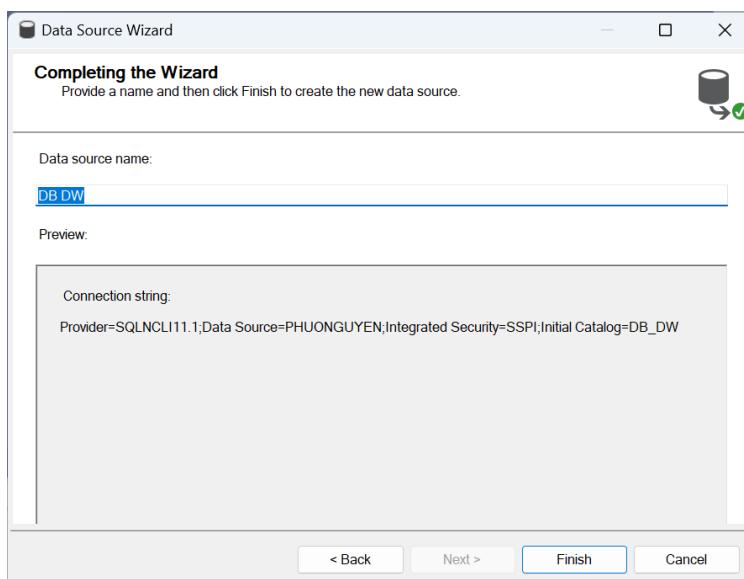
*Hình 3.9. Chọn kết nối với Data Source*

- Bước 4: Tại trang Impersonation Information, định nghĩa security credentials cho Analysis Services để kết nối tới data source. Impersonation ảnh hưởng đến tài khoản Windows dùng để kết nối đến data source khi Windows Authentication được chọn. Analysis Services thì không hỗ trợ impersonation để xử lý OLAP objects. Chọn Use the service account nhấn Next



*Hình 3.10. Lựa chọn tài khoản phù hợp*

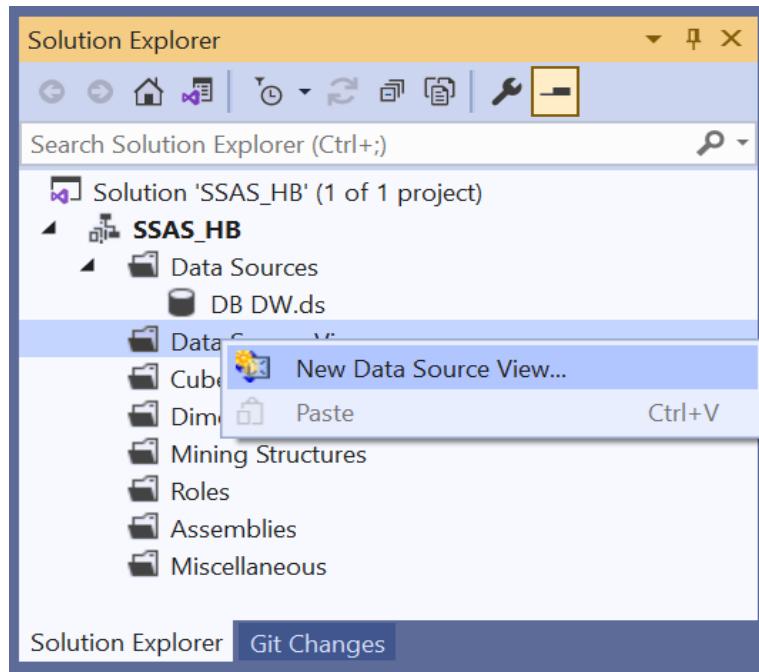
- Bước 5: Nhấn Finish để kết thúc quá trình xác định dữ liệu nguồn



*Hình 3.11. Kết thúc quá trình Define Data Source*

### **3. Xác định khung nhìn dữ liệu nguồn ( DEFINE DATA SOURCE VIEW)**

- Bước 1: Bên gốc phải phần Solution Explorer nhấn chuột phải vào Data Source View và chọn New Data Source View.



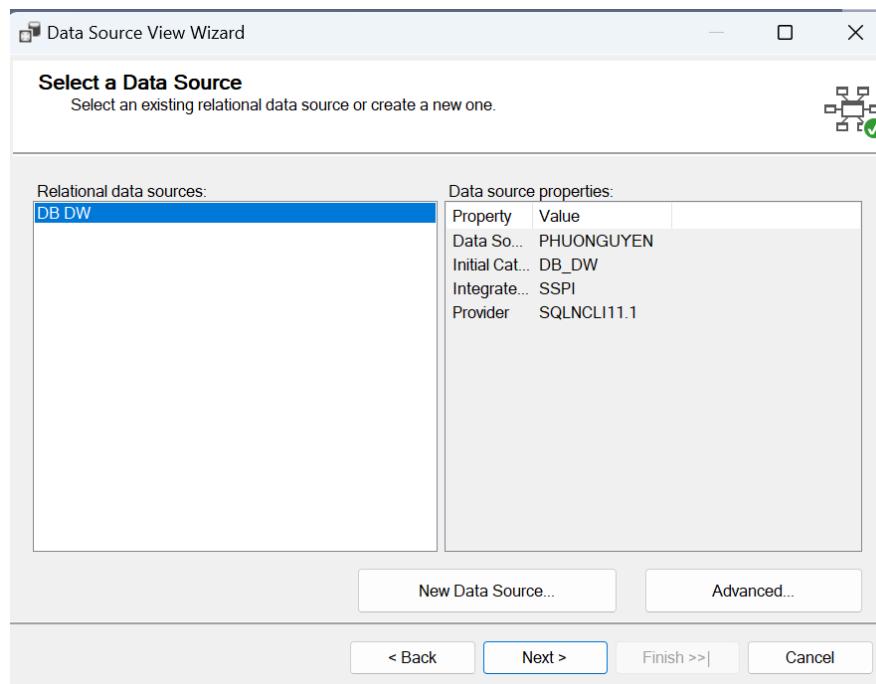
*Hình 3.12. Tạo Data Source View cho quá trình SSAS*

- Bước 2: Một cửa sổ hiện ra tiếp tục nhấn Next để tiếp tục.



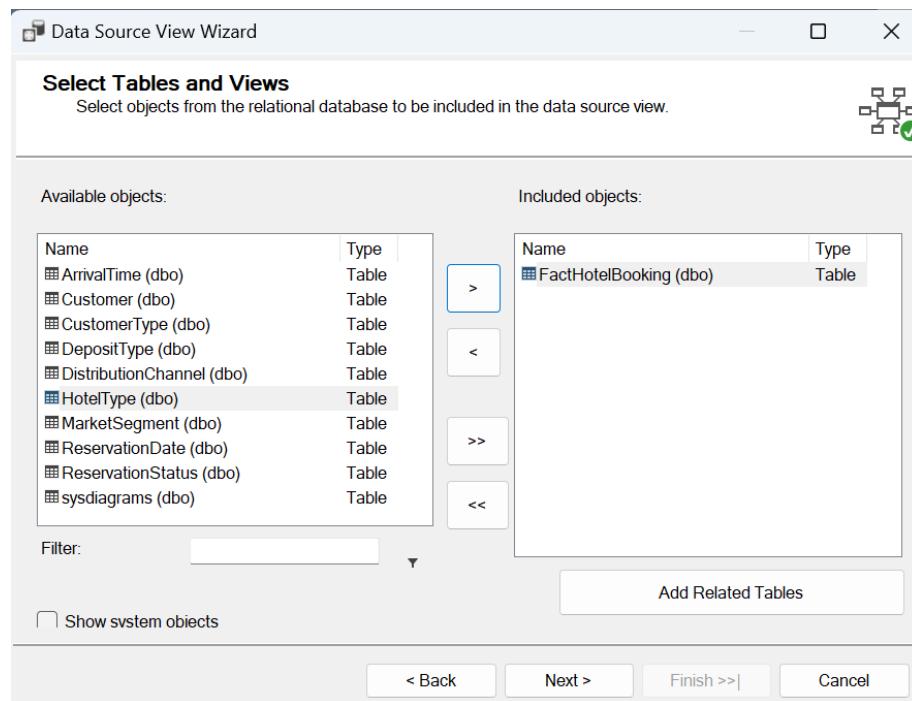
*Hình 3.13. Phần giới thiệu Data Source View Wizard*

- Bước 3: Cho kho dữ liệu DB DW rồi nhấn Next để tiếp tục

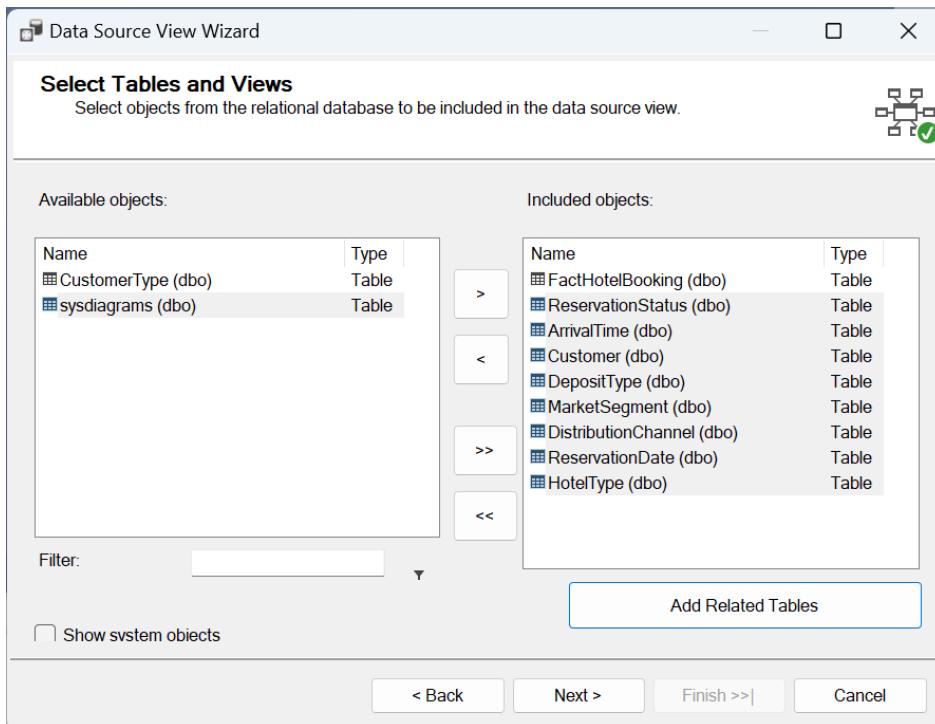


*Hình 3.14. Chọn kho dữ liệu cho quá trình SSAS*

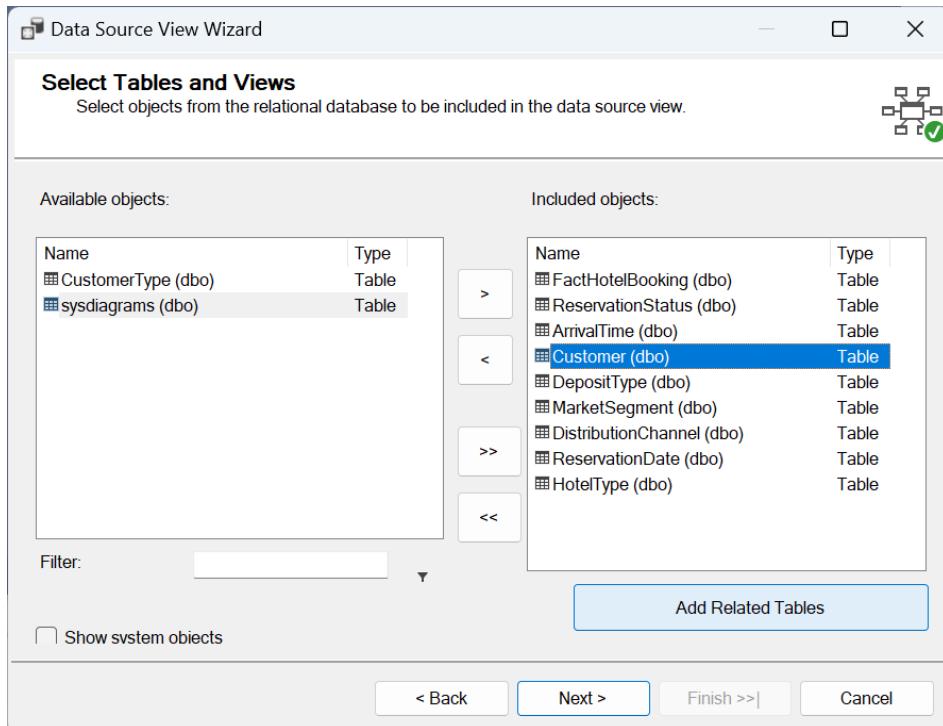
- Bước 4: Chọn bảng FACT và DIM cho quá trình phân tích sau đó nhấn Next.



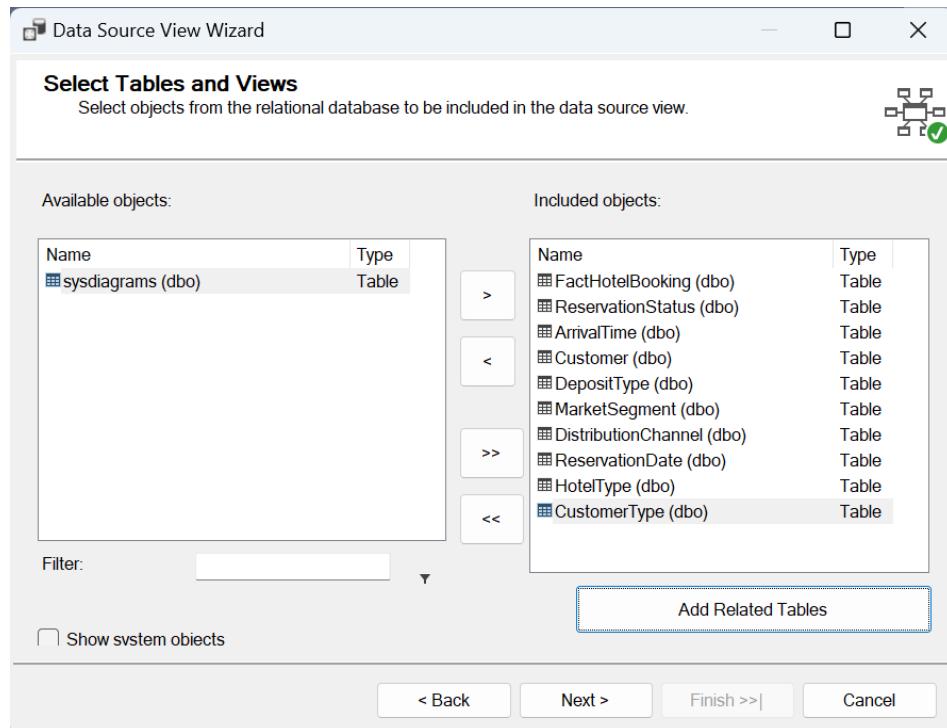
*Hình 3.15. Chọn bảng dữ liệu Fact cho quá trình phân tích*



*Hình 3.16. Chọn bảng Dimension có liên quan đến Fact*

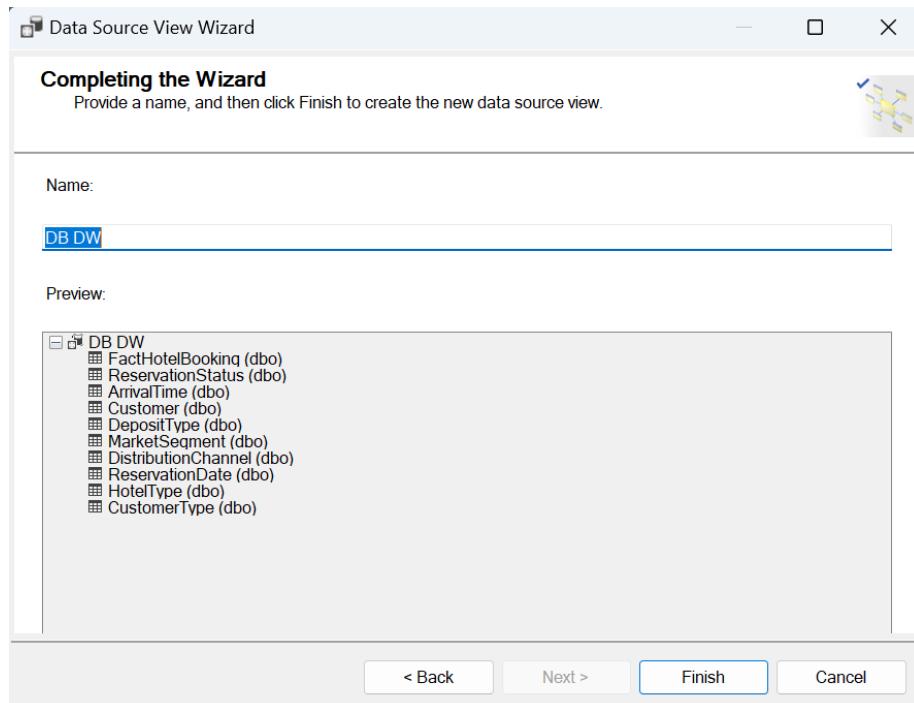


*Hình 3.17. Chọn bảng Customer và tiếp tục chọn Add related Table để lấy các bảng Dimension có liên quan với Customer*



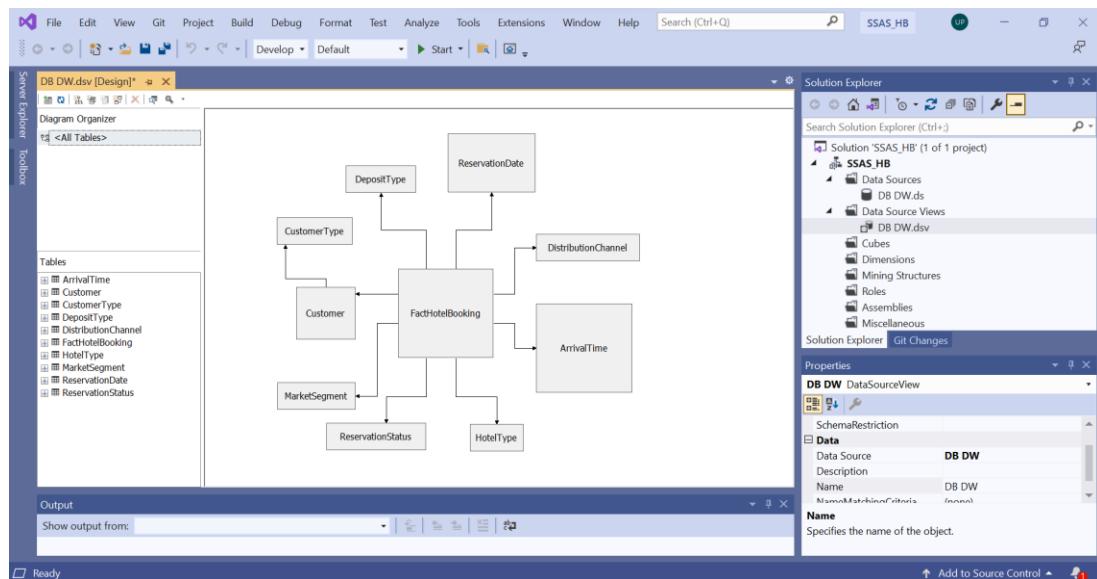
*Hình 3.18. Các bảng dữ liệu cho quá trình phân tích*

- Bước 6: Nhấn Next:



*Hình 3.19. Các bảng dữ liệu được tạo*

- Bước 7: Nhấn Finish để kết thúc quá trình tạo Data Source View và kết sẽ được bảng như hình

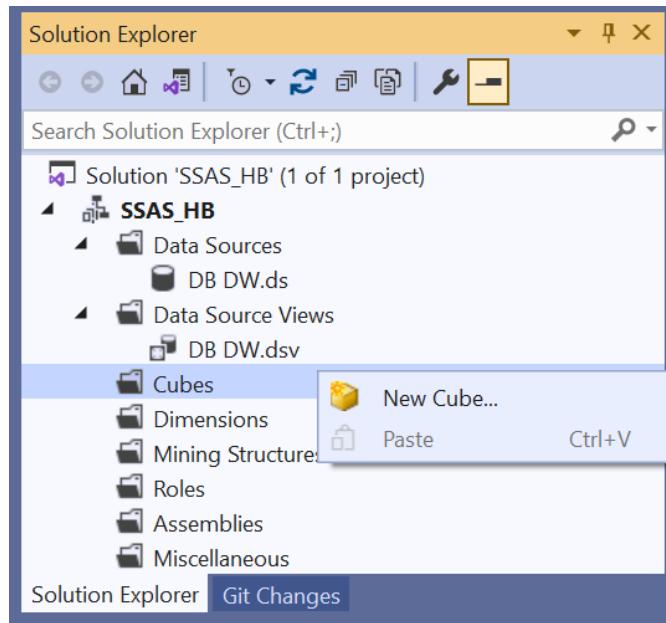


Hình 3.20. Hoàn thành quá trình tạo Date Source View

#### 4. Xác định Cube và tạo mới các Measures

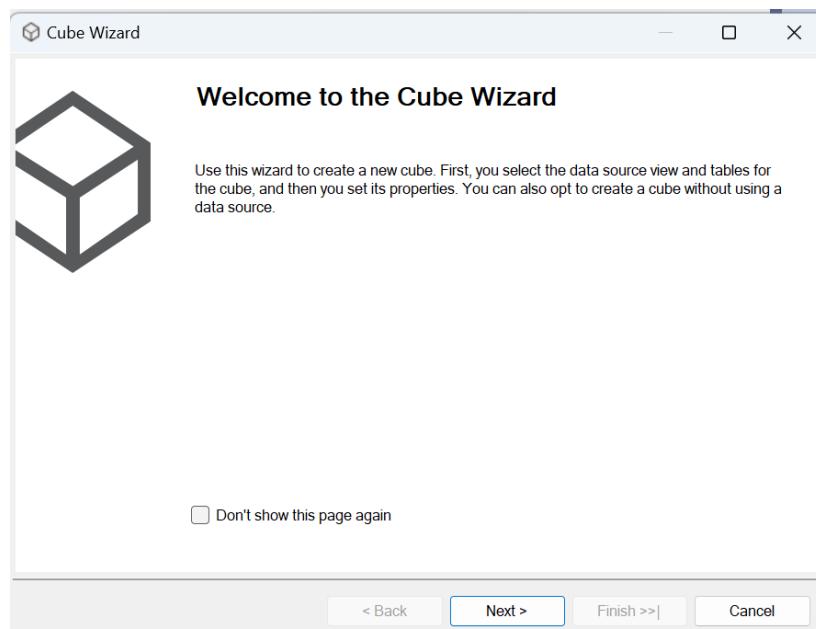
##### 4.1. Xác định Cube và các bảng chiều

- Bước 1: Bên gốc phải phần Solution Explorer nhấp chuột phải vào Cube nhấp tiếp New Cube



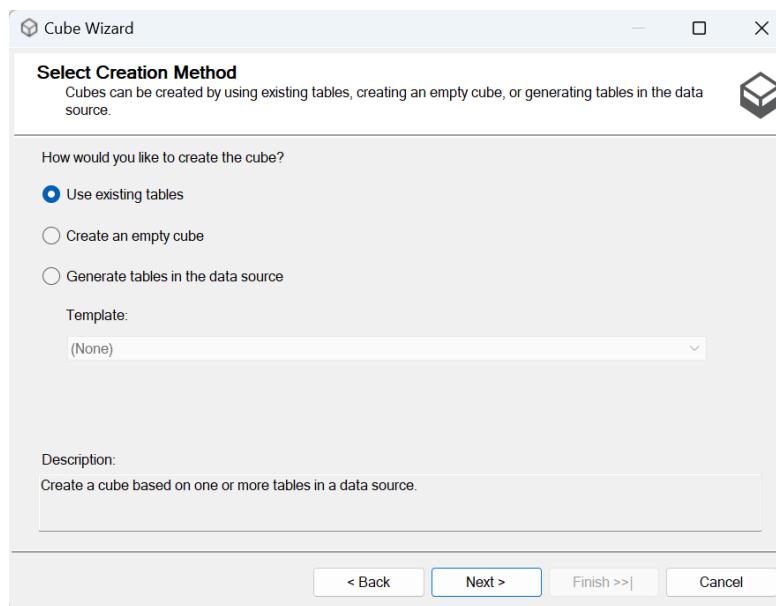
Hình 3.21. Tạo khối cube cho quá trình SSAS

- Bước 2: Cửa sổ Cube Wizard hiện ra tiếp tục nhấn Next để tiếp tục.



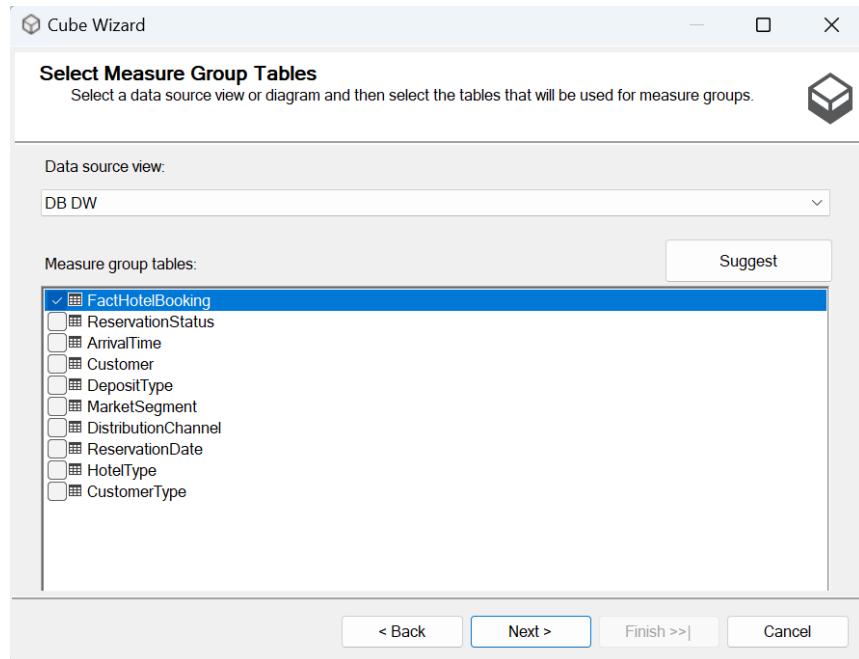
*Hình 3.22 Phản giới thiệu Cube Wizard*

- Bước 3: Trong cửa sổ Select Createion Method chọn use existing tables sau đó nhấn Next.



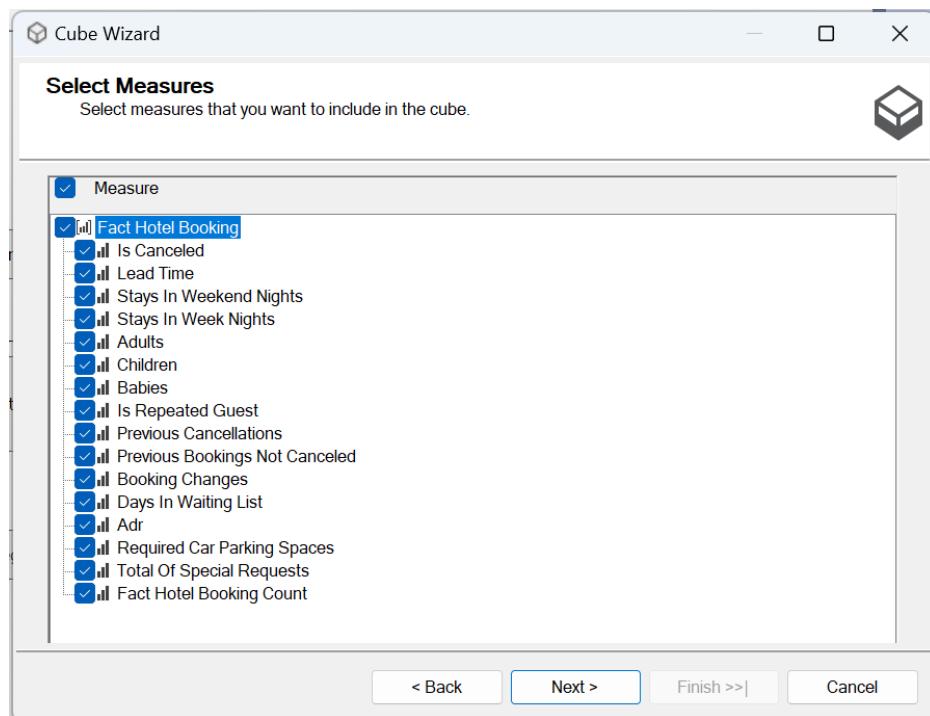
*Hình 3.23. Tùy chọn tạo Cube*

- Bước 4: Chọn FactHotelBooking chứa các thuộc tính độ đo để phân chia các measure group



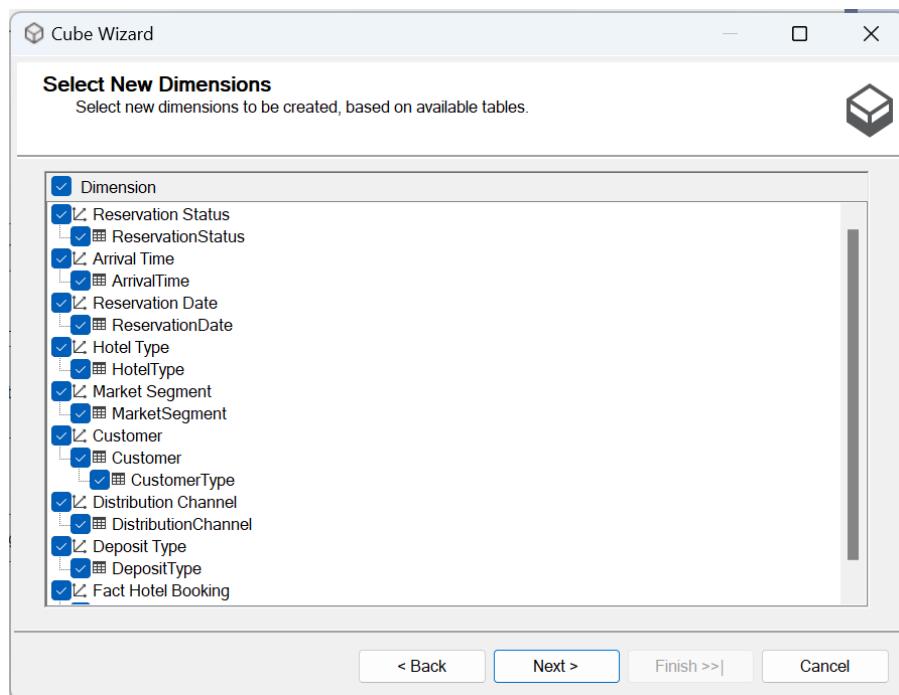
*Hình 3.24. Tùy chọn Measure Group cho cube*

- Bước 5: Chọn những thuộc tính định lượng để xuất. Nhấn Next để tiếp tục.



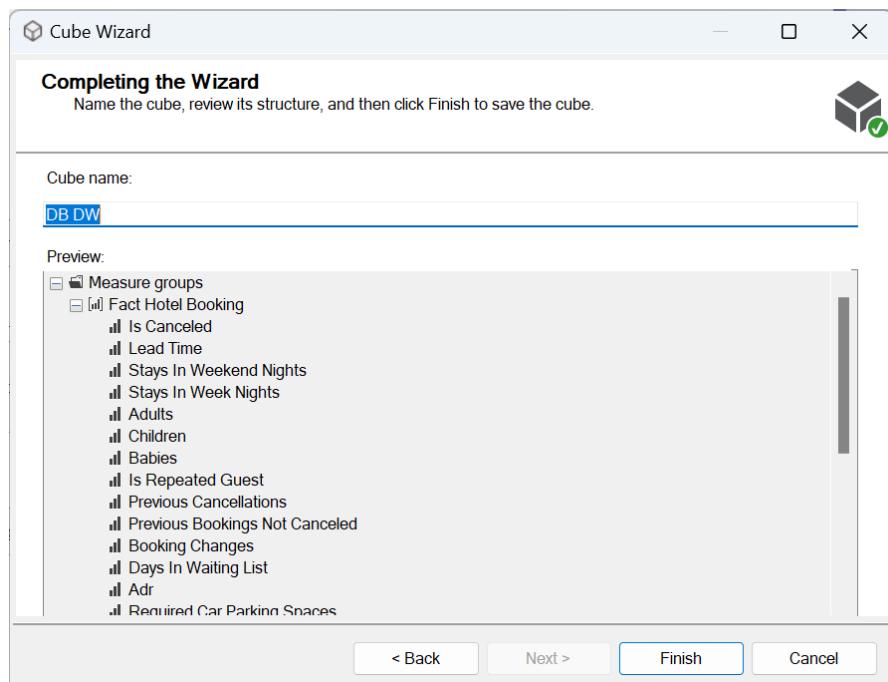
*Hình 3.25 Tùy chọn những thuộc tính định lượng*

- Bước 6: Chọn danh sách Dimension nhấn Next để tiếp tục.



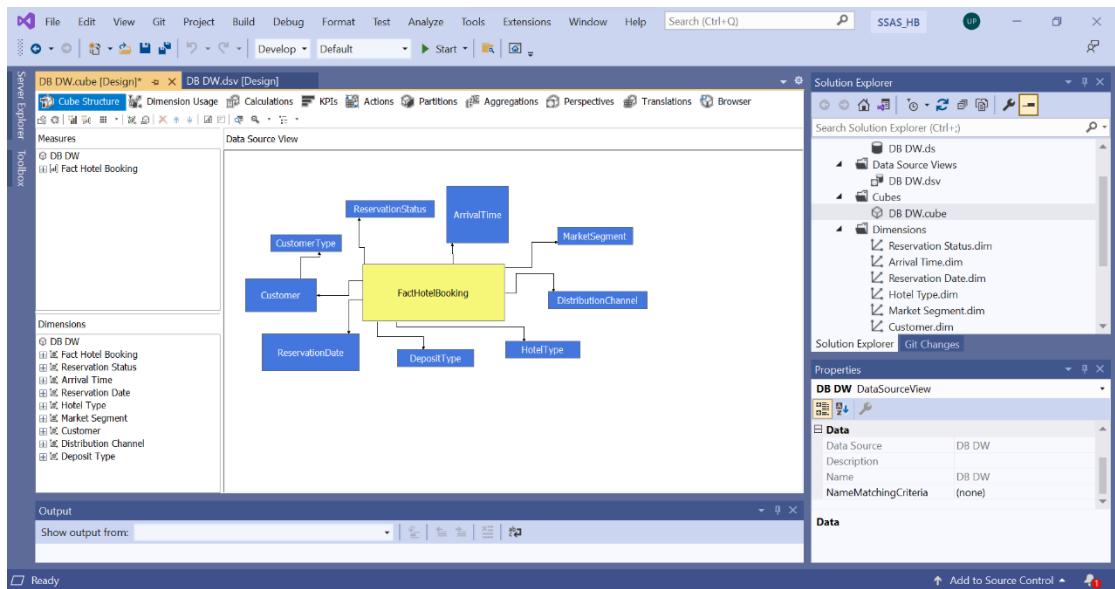
*Hình 3.26 Chọn các bảng Dimension cho các cube*

- Bước 7: Nhấn Finish để hoàn thành quá trình tạo cube



*Hình 3.27 Xem lại các thuộc tính độ đo và các bảng chiều được tạo*

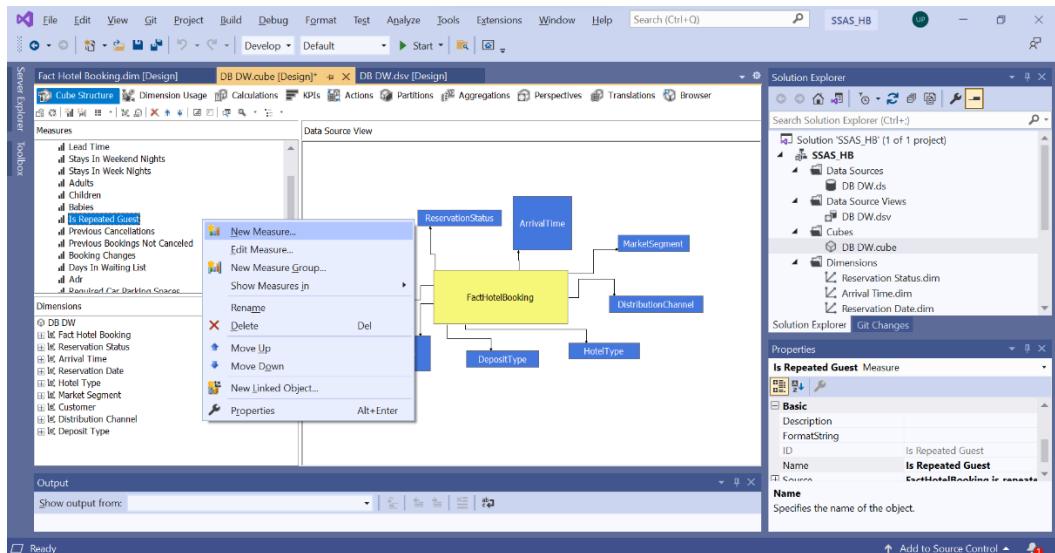
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



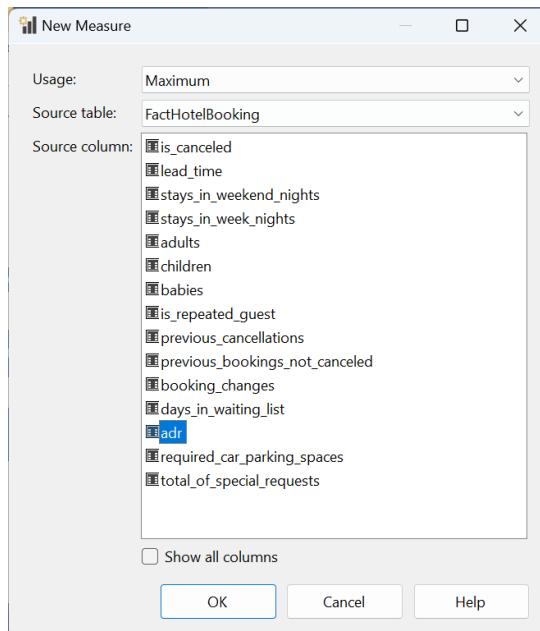
Hình 3.28. Kết thúc quá trình tạo Cube

### 4.2. Xác định các Measures

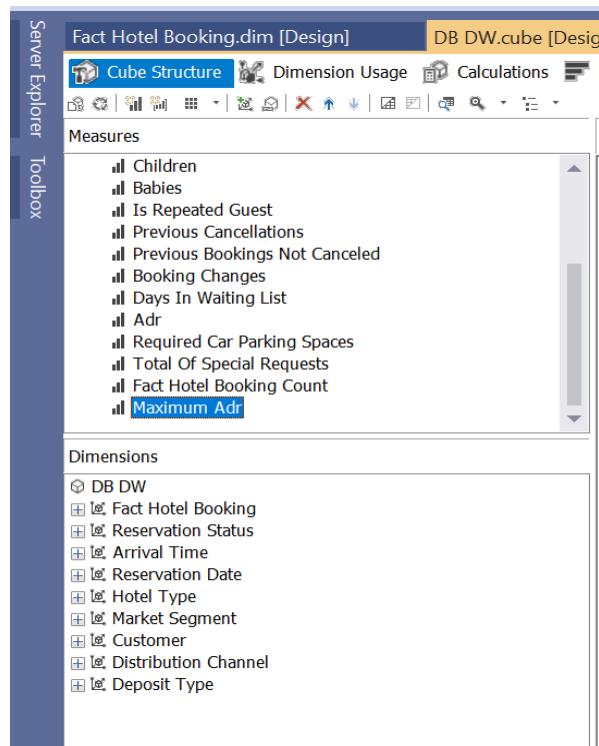
- Bước 1: Để đổi tên và thêm thuộc tính các độ đo trong phần Measure ta chọn Show MeasuresTree
- Bước 2: Thêm mới các độ đo
  - + Trong Measure, click chuột phải chọn New Measure:



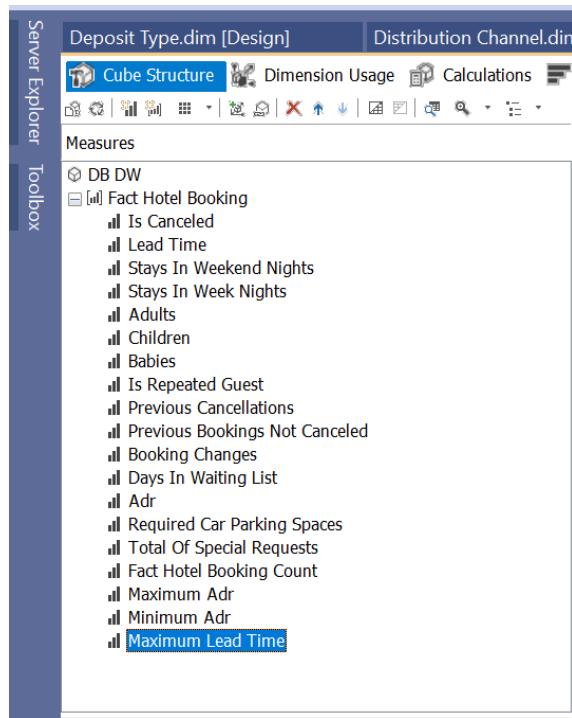
Hình 3.29. Minh họa bước tạo New Measure



*Hình 3.30 Chọn các độ đo cần tạo mới và cách thống kê*



*Hình 3.31 Độ đo Maximum Adr được tạo*



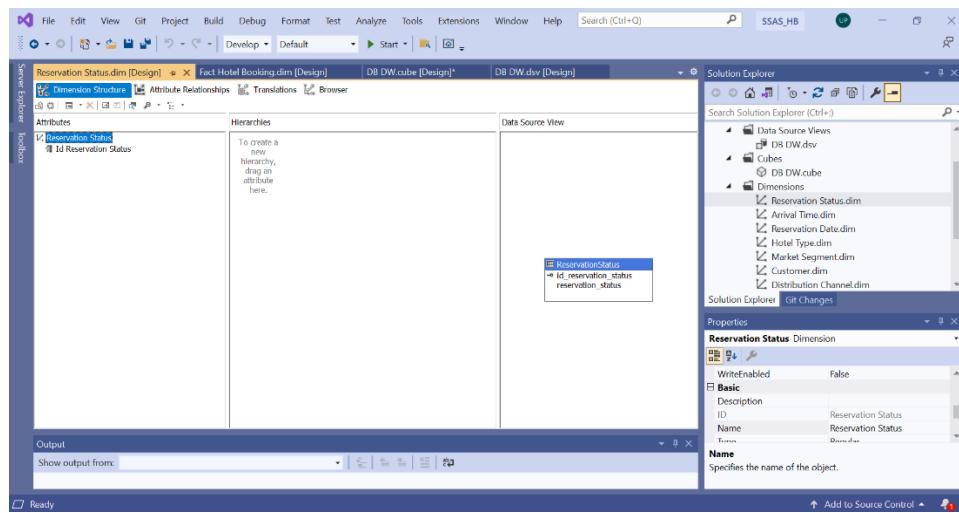
Hình 3.32 Các độ đo khác được tạo

## 5. Tính chất của các bảng Dimensions và Attribution nâng cao

### 5.1. Định nghĩa các tính chất của bảng Dimension:

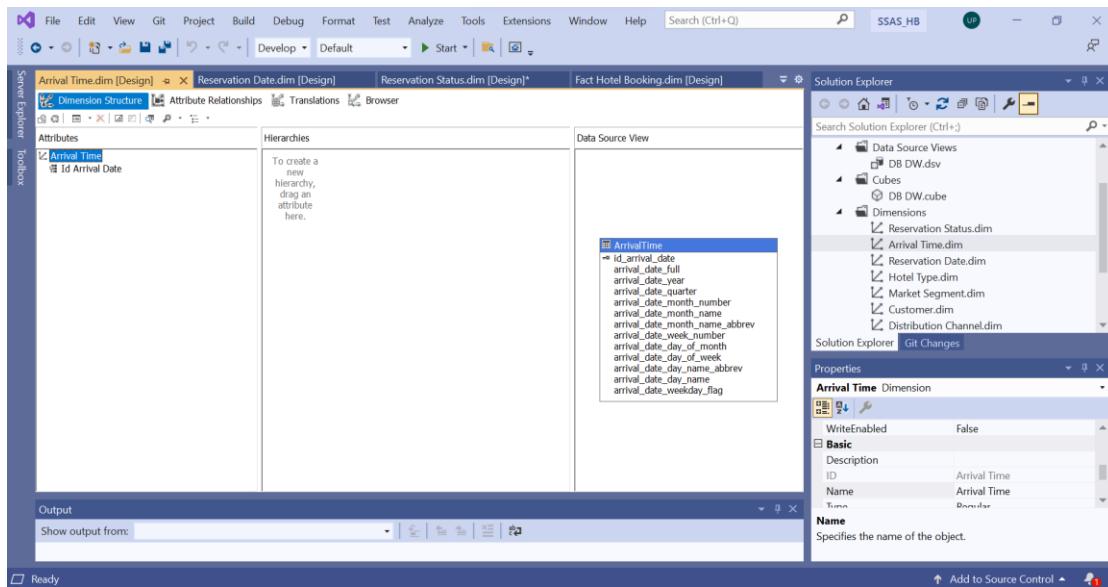
#### 5.1.1. Reservation Status

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Reservation Status để tiến hành thêm các thuộc tính:



Hình 3.33. Các bảng Dimension trong Solution Explorer

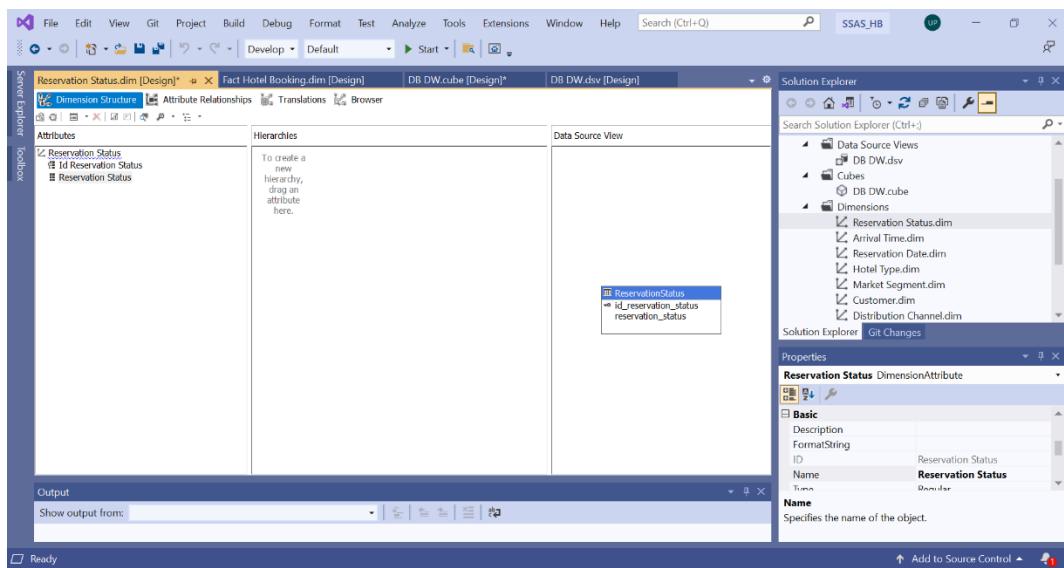
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.34. Kết quả sau khi kéo thả các thuộc tính

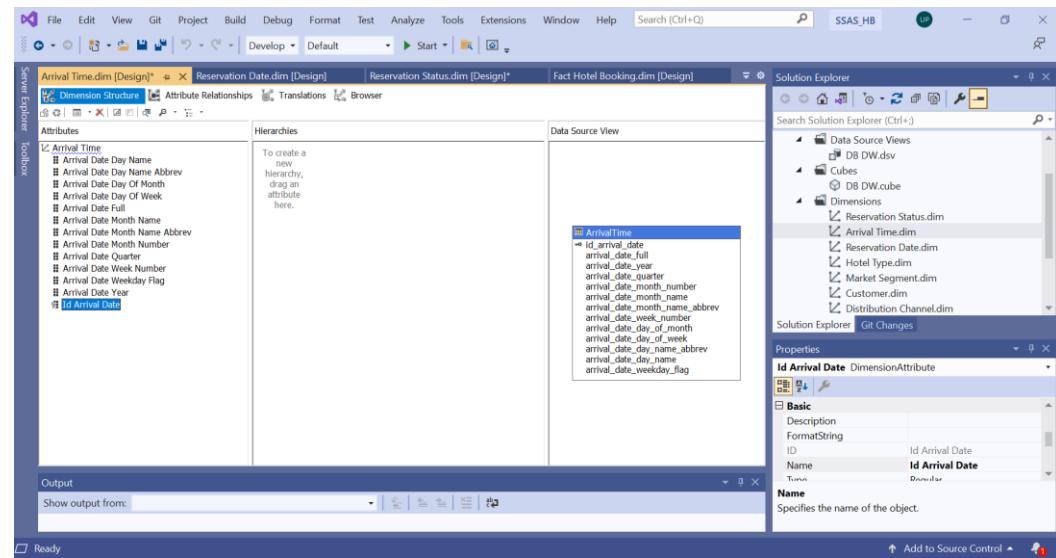
### 5.1.2. Arrival Time

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Arrival Time để tiến hành thêm các thuộc tính:



Hình 3.35 Các bảng Dimension trong Solution Explorer

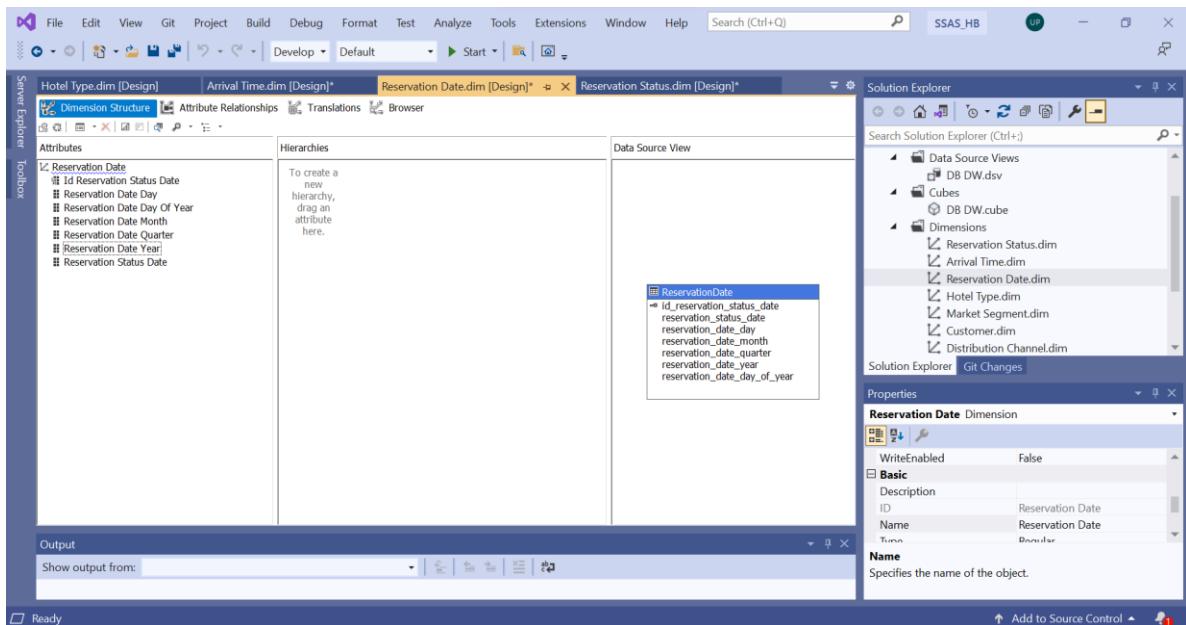
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.36. Kết quả sau khi kéo thả các thuộc tính

### 5.1.3. Reservation Date

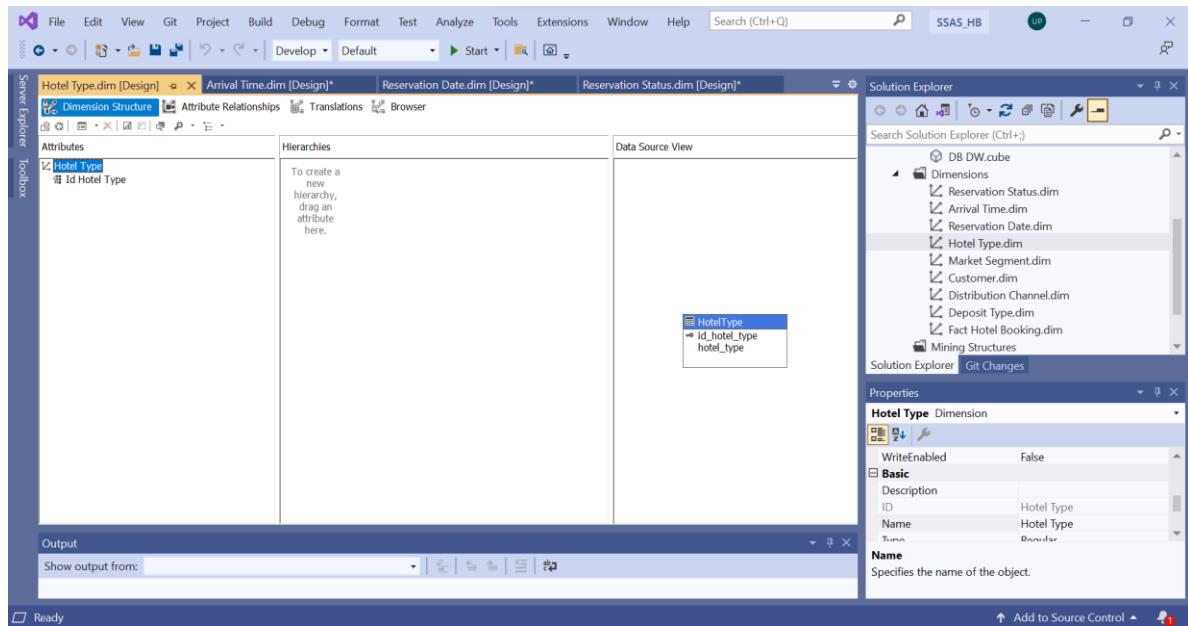
- Bước 1: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.37. Kết quả sau khi kéo thả các thuộc tính

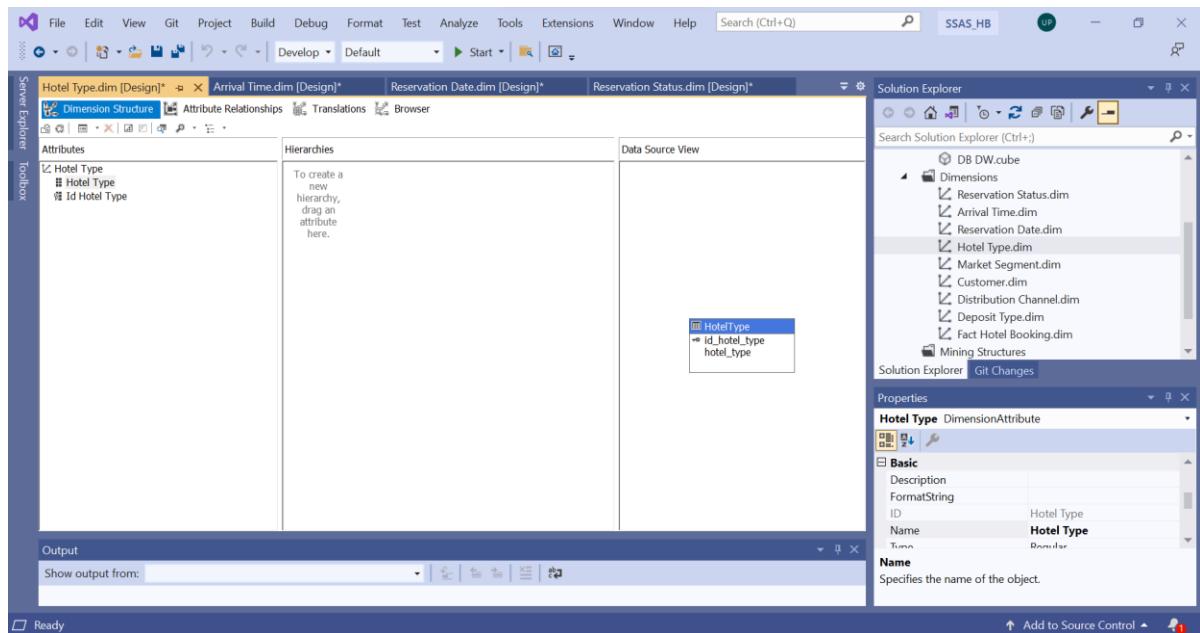
#### 5.1.4. Hotel Type

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Hotel Type để tiến hành thêm các thuộc tính:



Hình 3.38. Các bảng Dimension trong Solution Explorer

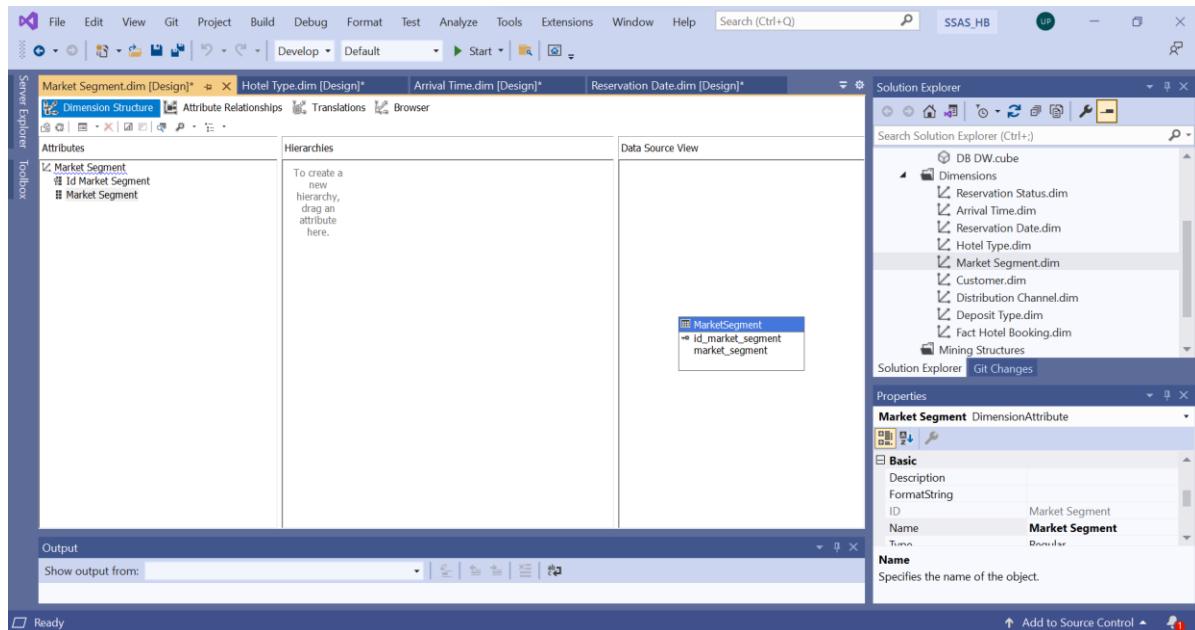
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.39. Kết quả sau khi kéo thả các thuộc tính

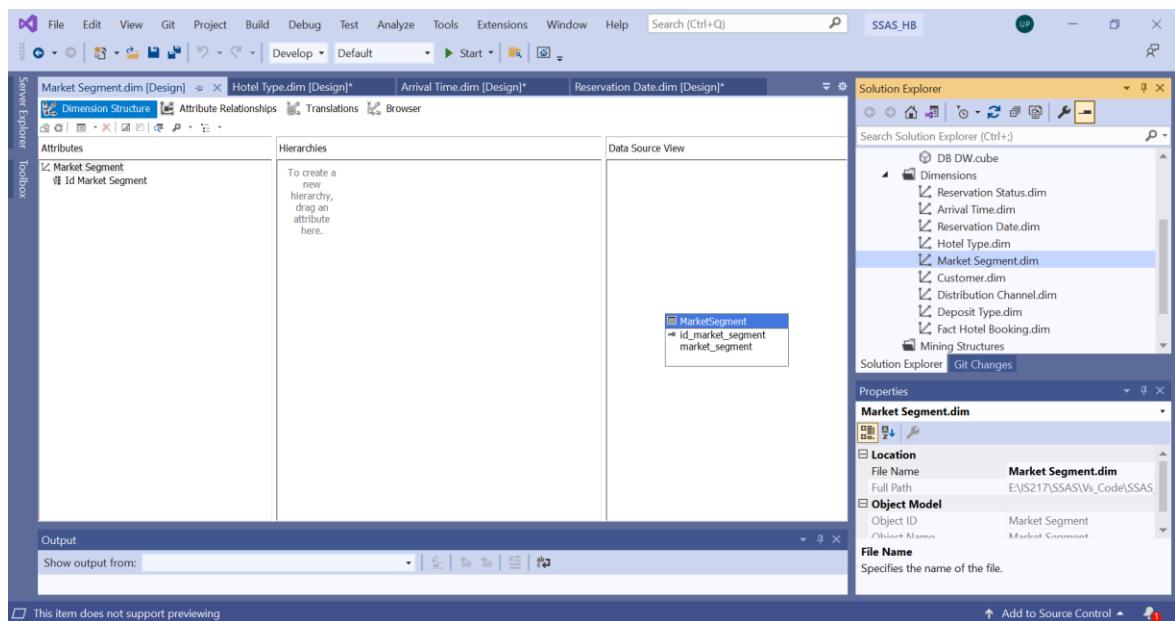
### 5.1.5. Market Segment

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Market Segment để tiến hành thêm các thuộc tính:



Hình 3.40. Các bảng Dimension trong Solution Explorer

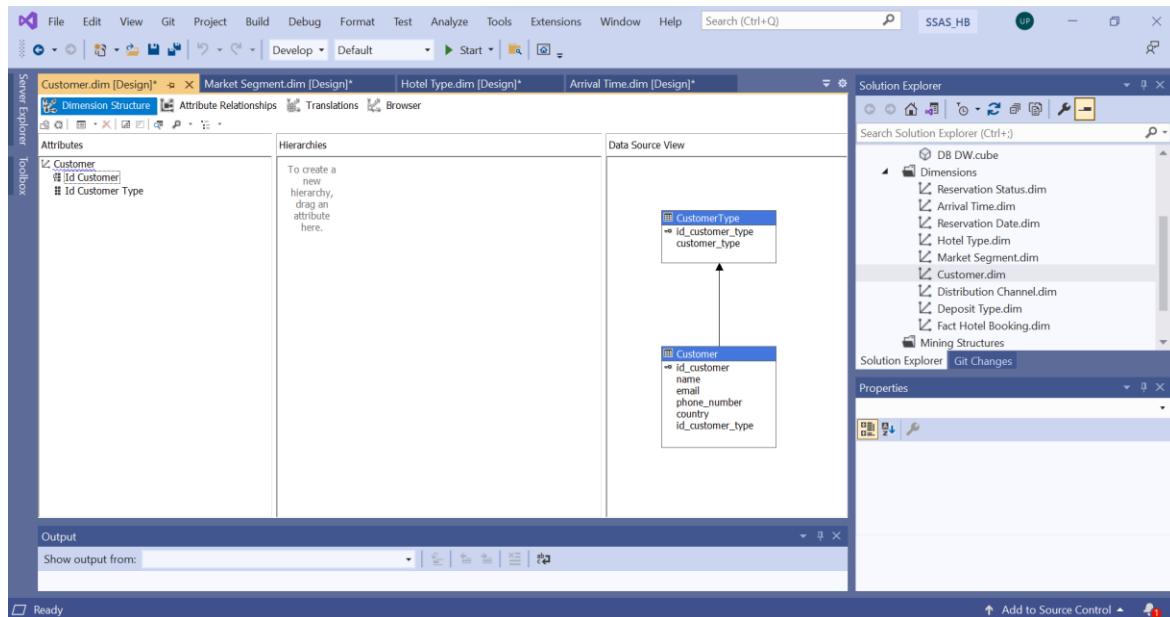
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.41. Kết quả sau khi kéo thả các thuộc tính

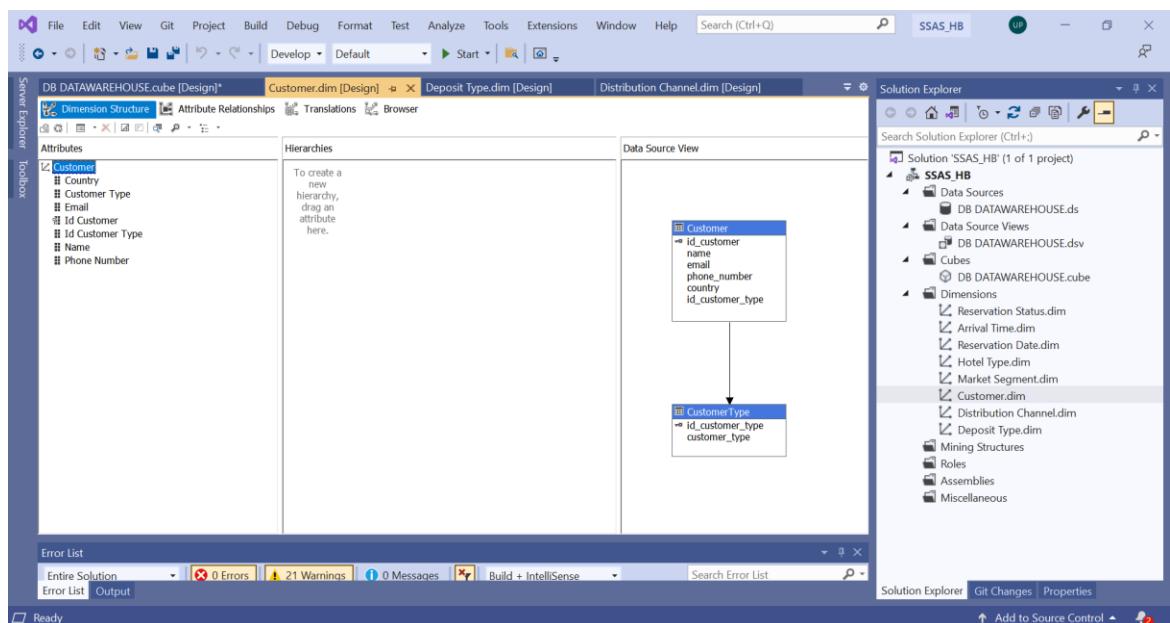
### 5.1.6. Customer và Customer Type

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Customer và Customer Type để tiến hành thêm các thuộc tính:



*Hình 3.42. Các bảng Dimension trong Solution Explorer*

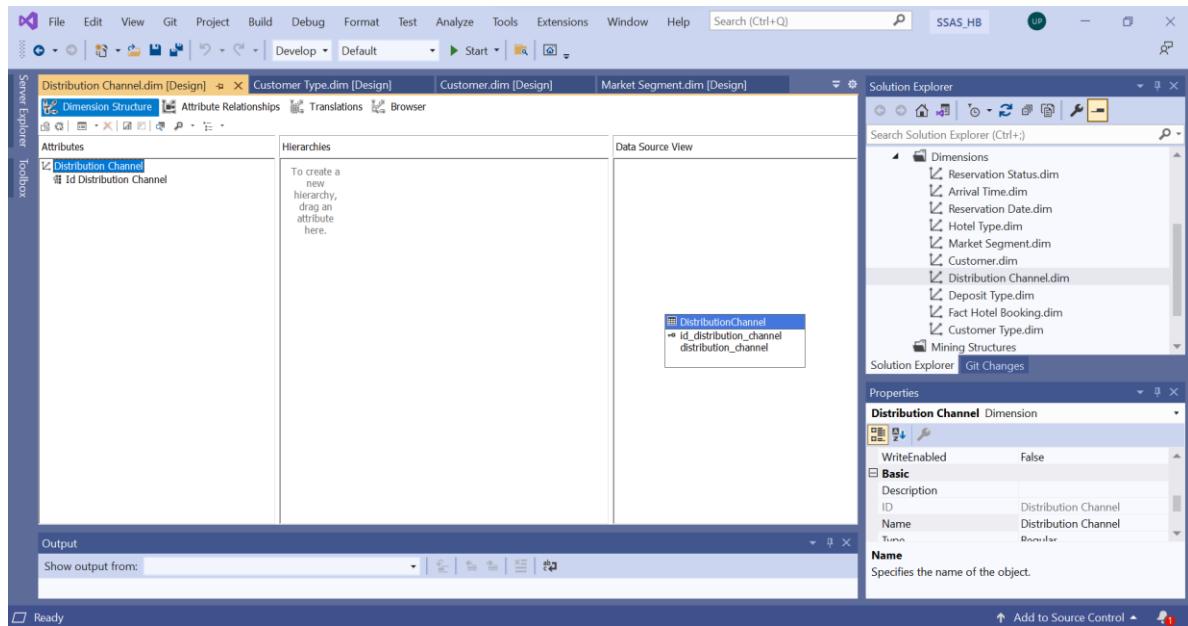
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



*Hình 3.43. Kết quả sau khi kéo thả các thuộc tính*

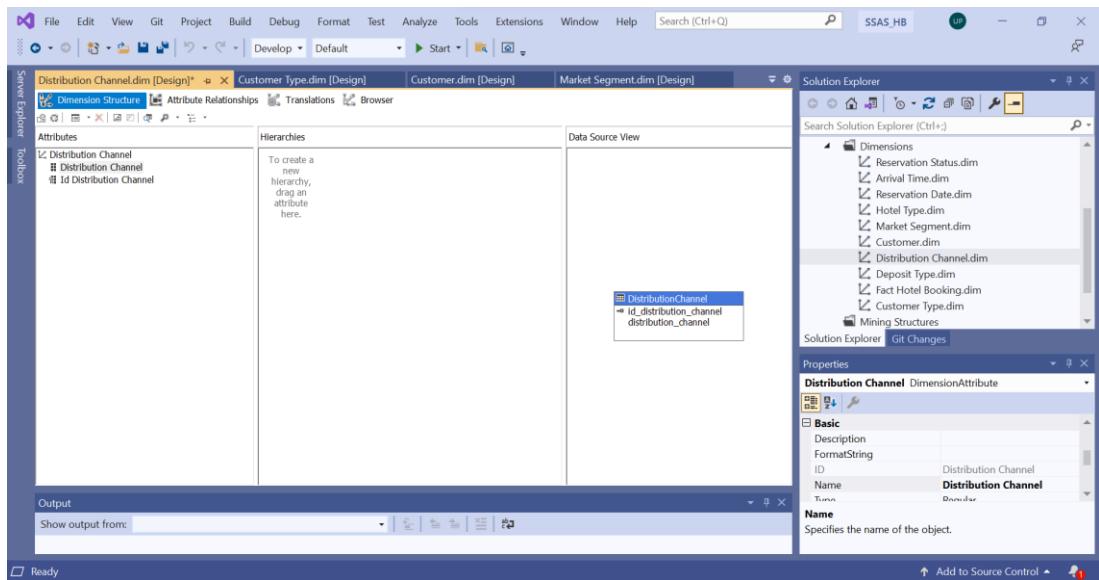
### 5.1.7. Distribution Channel

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Distribution Channel để tiến hành thêm các thuộc tính:



*Hình 3.44. Các bảng Dimension trong Solution Explorer*

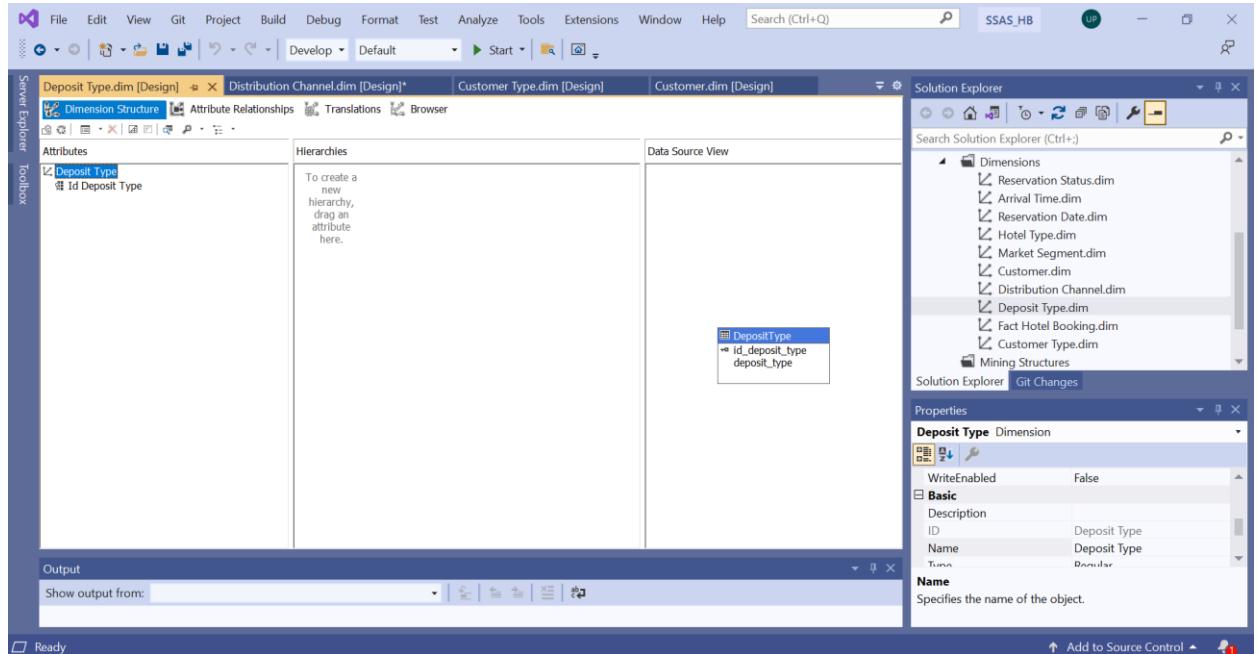
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



*Hình 3.45. Kết quả sau khi kéo thả các thuộc tính*

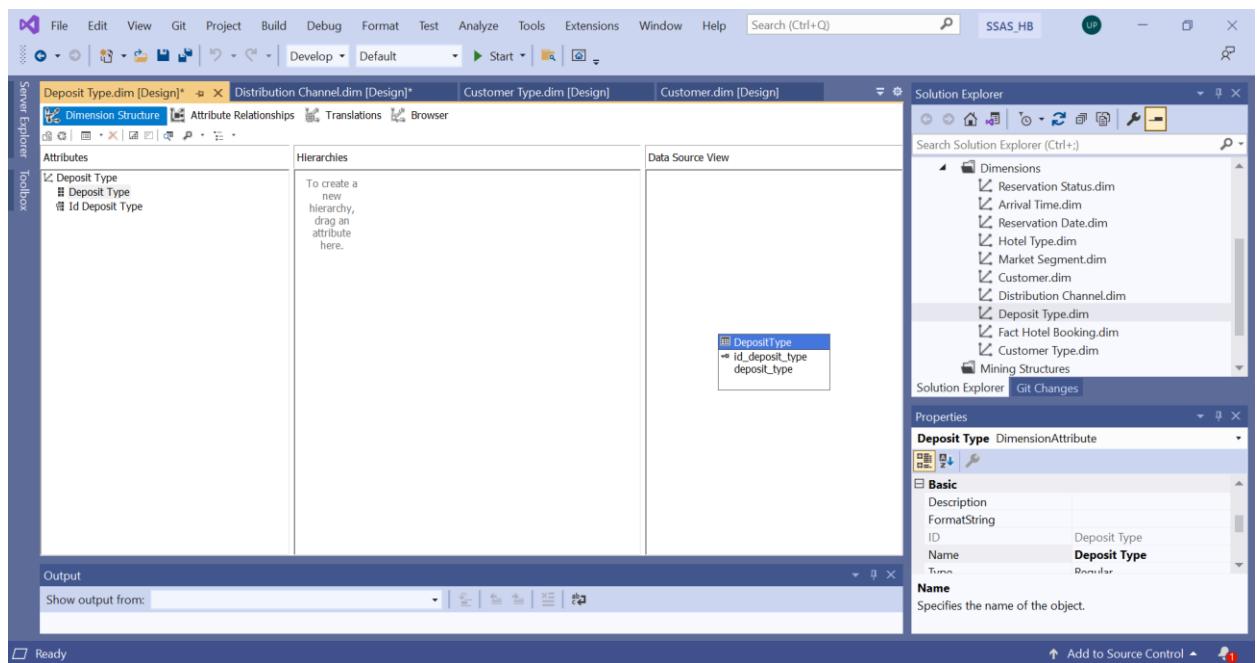
### 5.1.8. Deposite Type

- Bước 1: Ở Solution Explorer, lần lượt double-click vào từng dimension Deposit Type để tiến hành thêm các thuộc tính:



Hình 3.46. Các bảng Dimension trong Solution Explorer

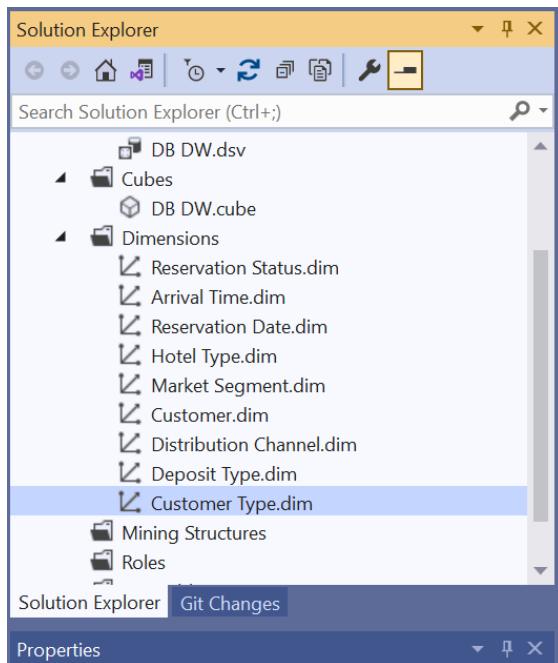
- Bước 2: Kéo các thuộc tính sử dụng từ Data Source View sang Attributes



Hình 3.47. Kết quả sau khi kéo thả các thuộc tính

### 5.1.9. Xóa bảng Fact Hotel Booking trong Dimension

- Bước 1: Click chuột phải vào Fact Hotel Booking chọn Delete:

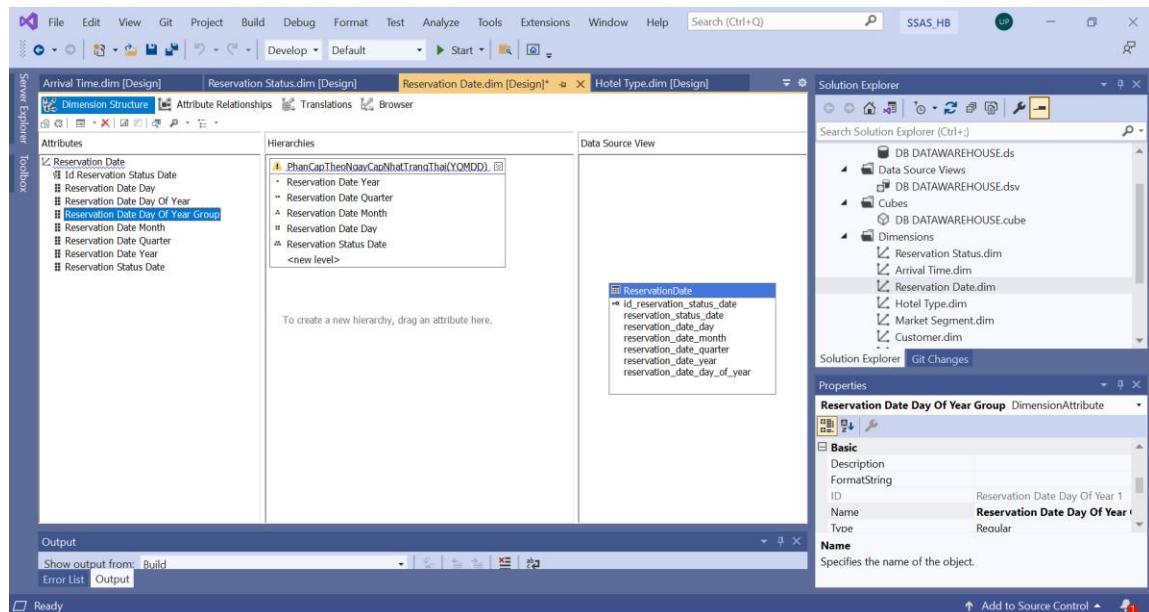


Hình 3.45. Kết quả của mục Dimensions sau khi xóa bảng Fact Hotel Booking

### 5.2. Gom nhóm các thuộc tính

#### 5.2.1. Reservation Date

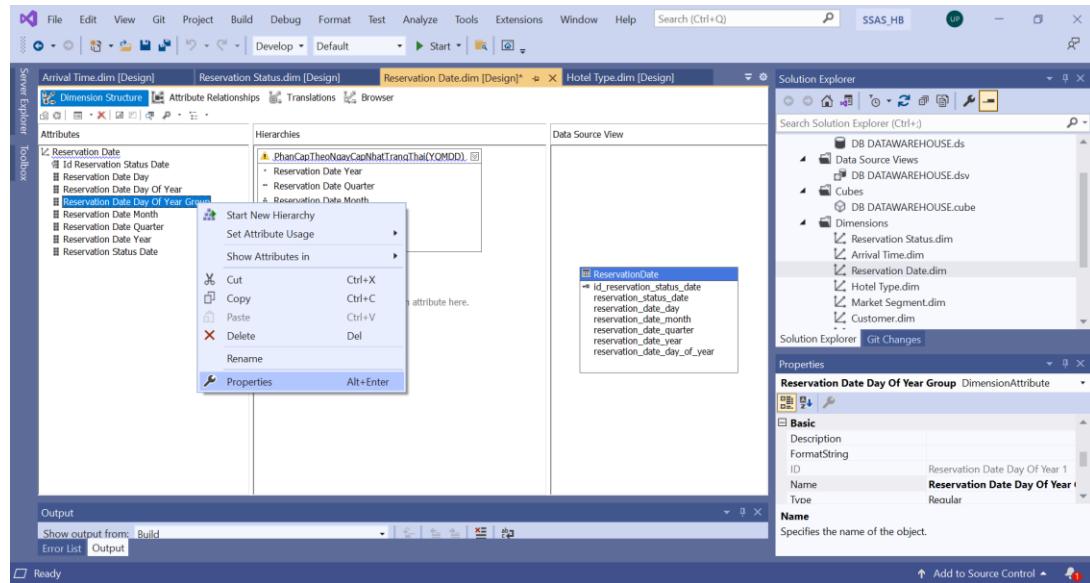
- Bước 1: Kéo thả thuộc tính Reservation\_Date\_Day\_Of\_Year từ Data Source view vào Attributes và đặt tên là Reservation Date Day Of Year Group



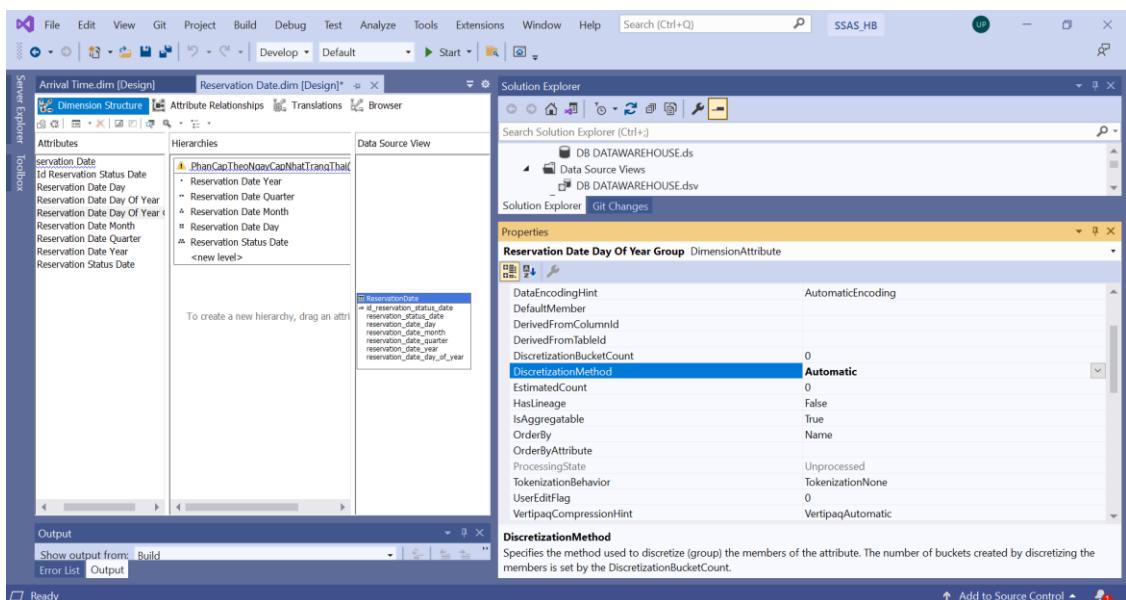
Hình 3.46..Kết quả sau khi kéo thả Reservation\_Date\_Day\_Of\_Year

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 2: Chọn Reservation Date Day Of Year Group → Properties → DiscretizationMethod → Automatic

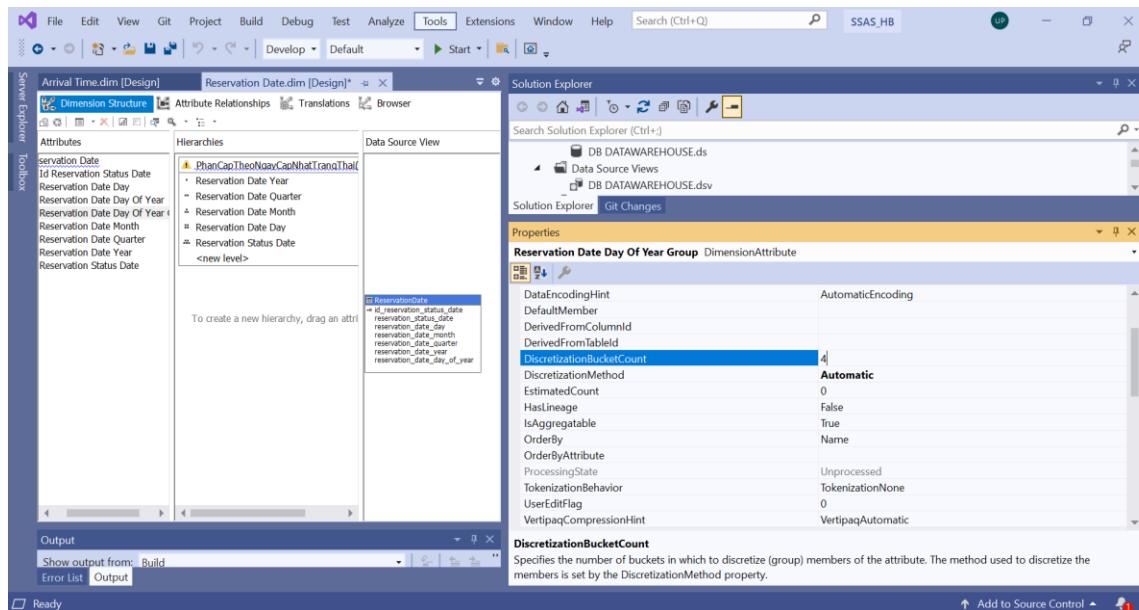


Hình 3.47. Chọn Properties



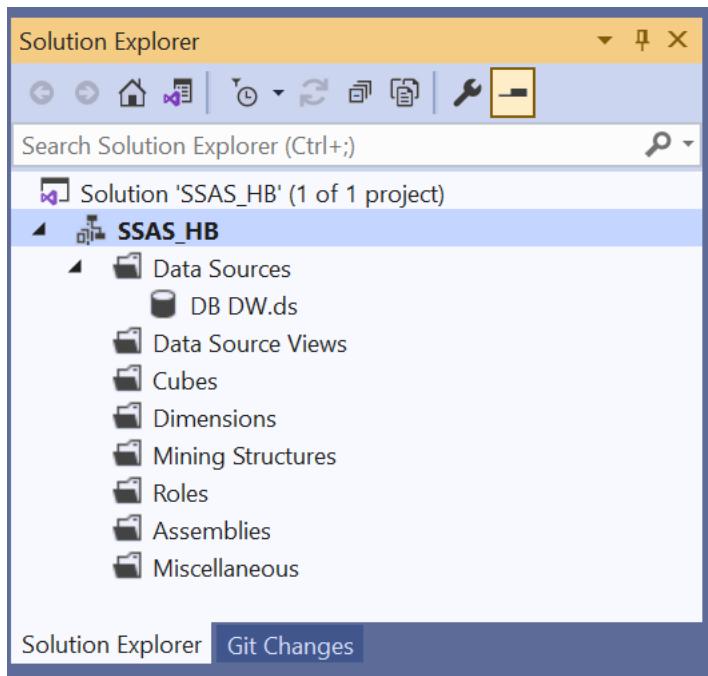
Hình 3.48. Cấu hình lại DiscretizationMethod

- Bước 3: Tại DiscretizationBucketCount, điền số lượng nhóm cần gom:



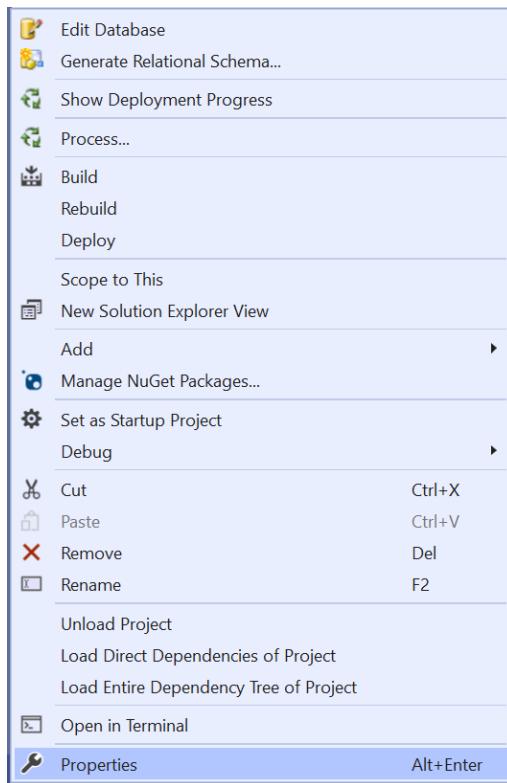
Hình 3.49. Cấu hình lại DiscretizationBucketCount.

- Bước 4: Chạy Process toàn bộ project

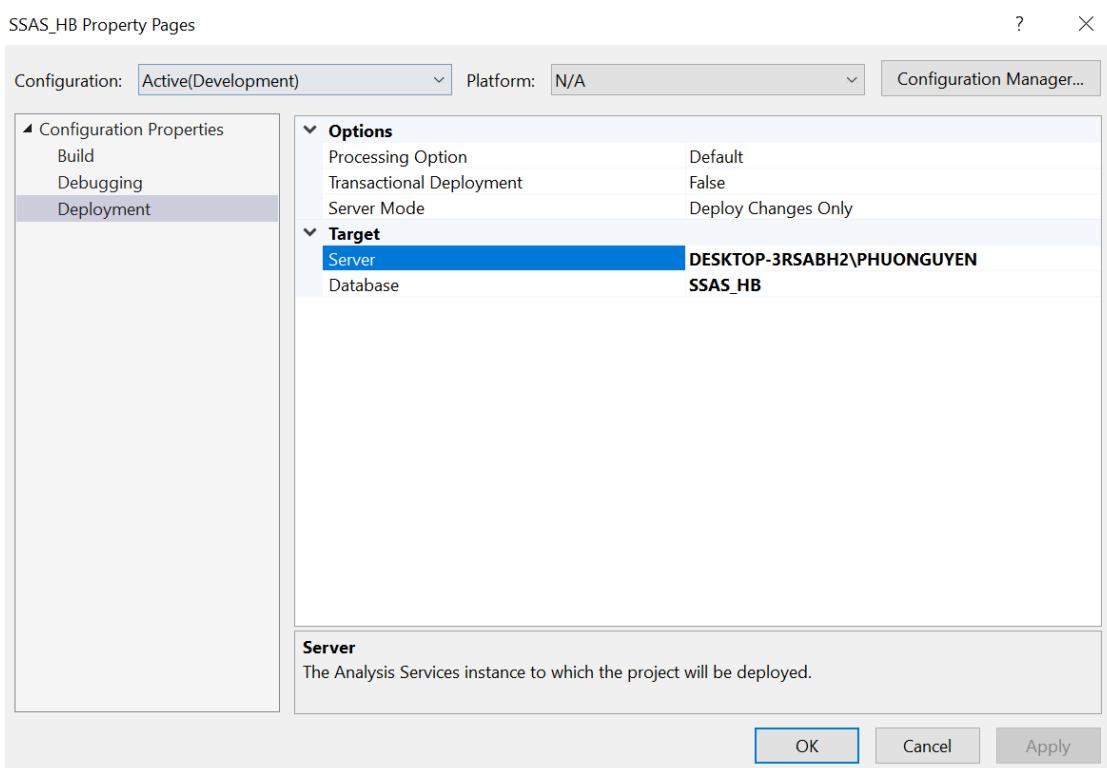


Hình 3.50. Tại khung sửa só Solution Explorer, tìm kiếm tên đồ án

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

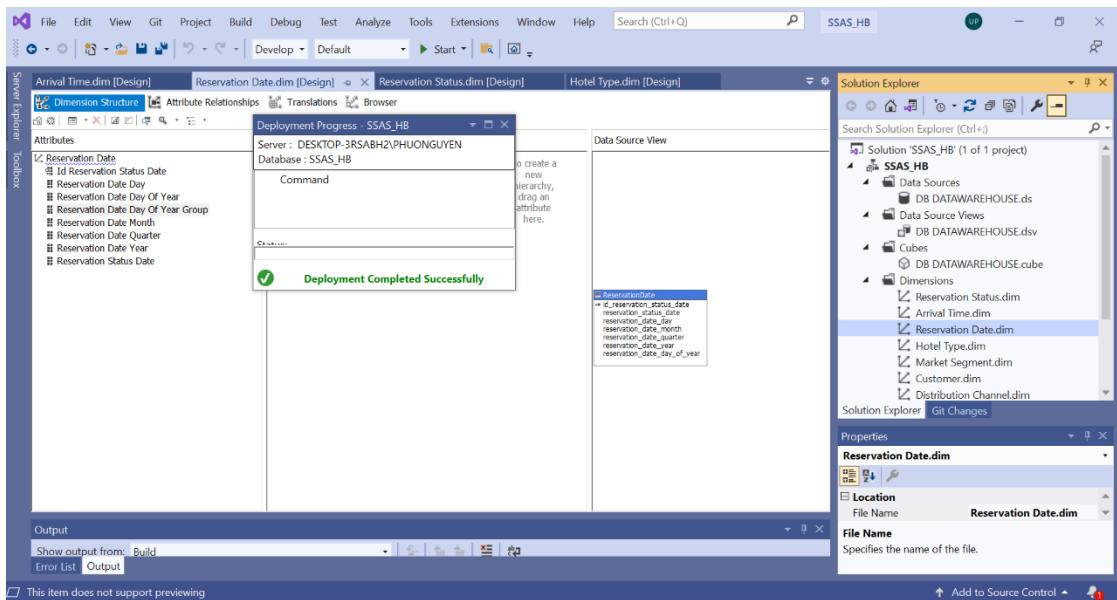


*Hình 3.51. Click chuột phải vào tên project → Properties*



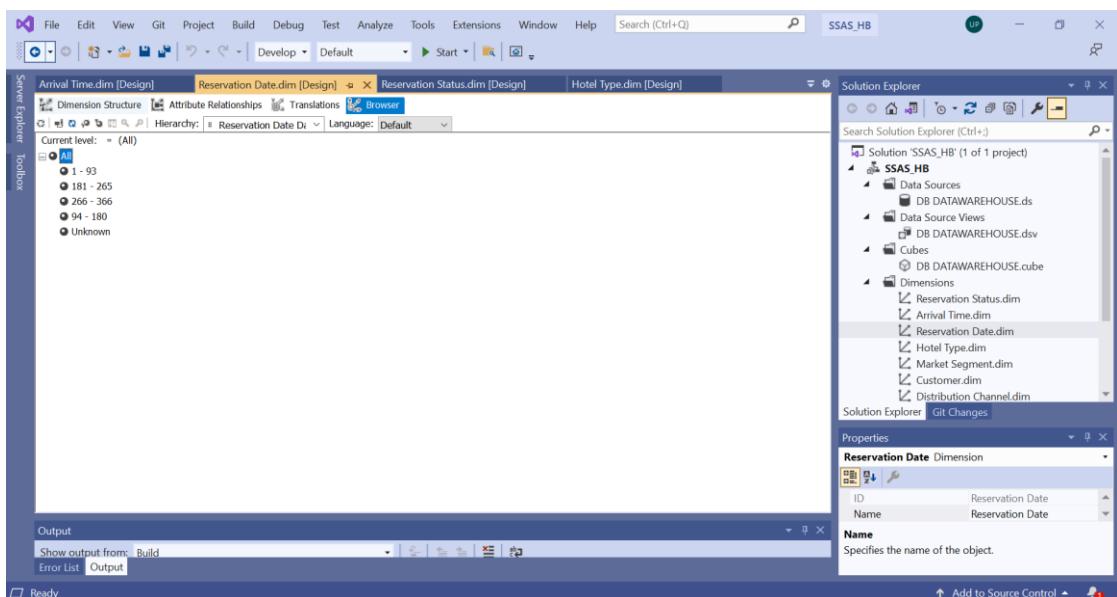
*Hình 3.51. Tại Deployment, đổi tên server*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.52. Chạy Process

### - Bước 5: Kiểm tra tại Browser

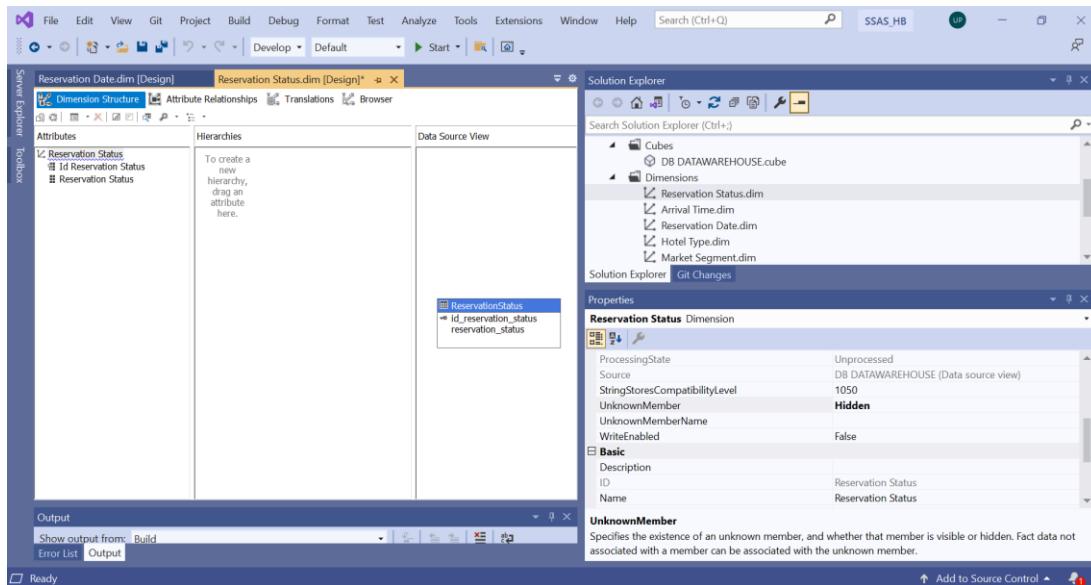


Hình 3.53. Kết quả sau khi gom nhóm dữ liệu

### 5.3. Xóa dòng dữ liệu Unknown

#### 5.3.1. Reservation Status

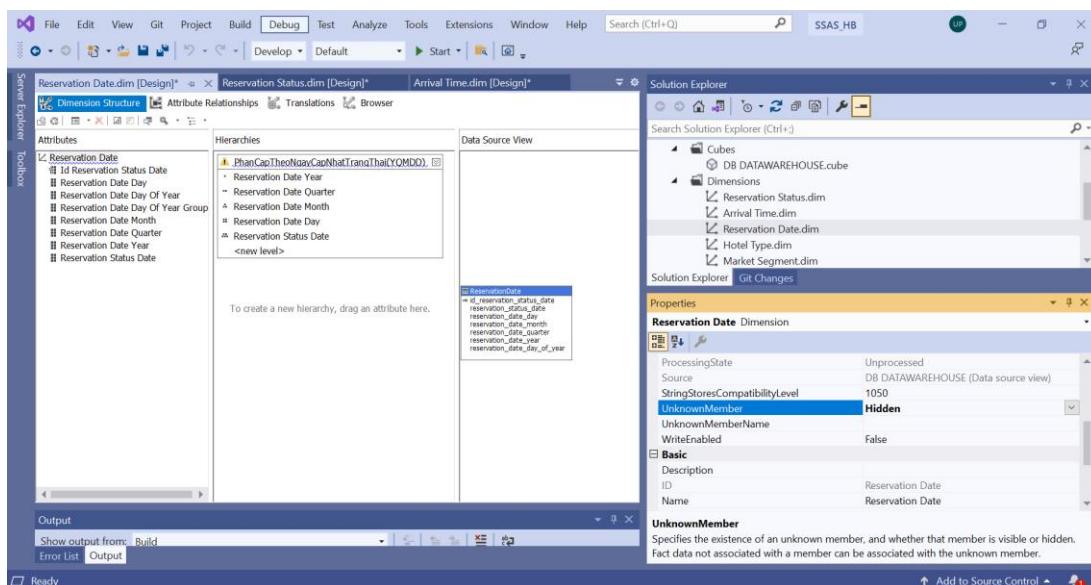
- Bước 1: Double click bảng Reservation Status, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.54. Kết quả sau khi chỉnh sửa UnknownMember

#### 5.3.2. Arrival time

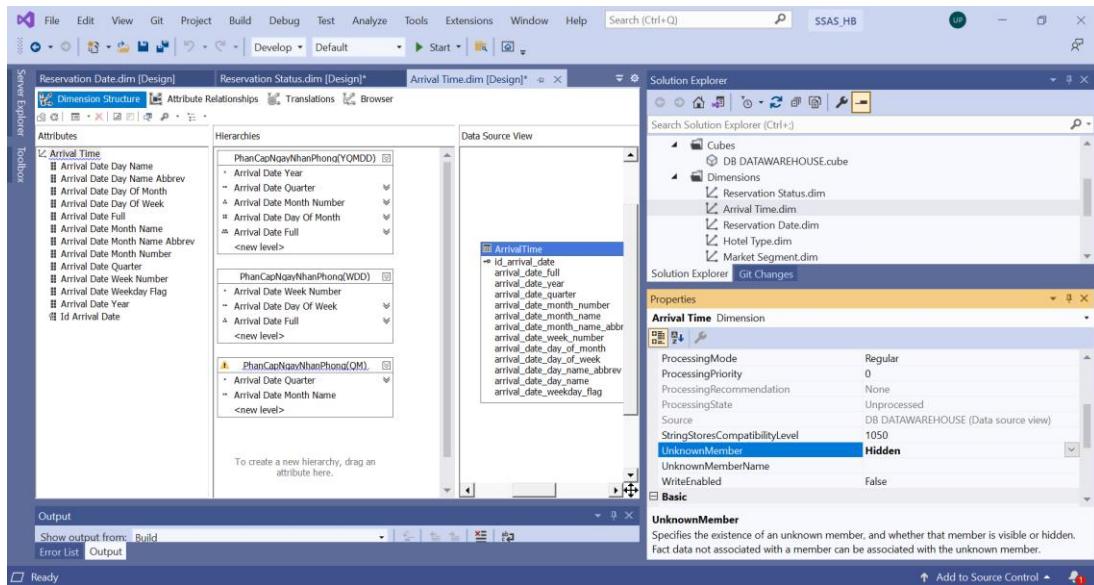
- Bước 1: Double click bảng Arrival time, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.55. Kết quả sau khi chỉnh sửa UnknownMember

### 5.3.3. Reservation Date

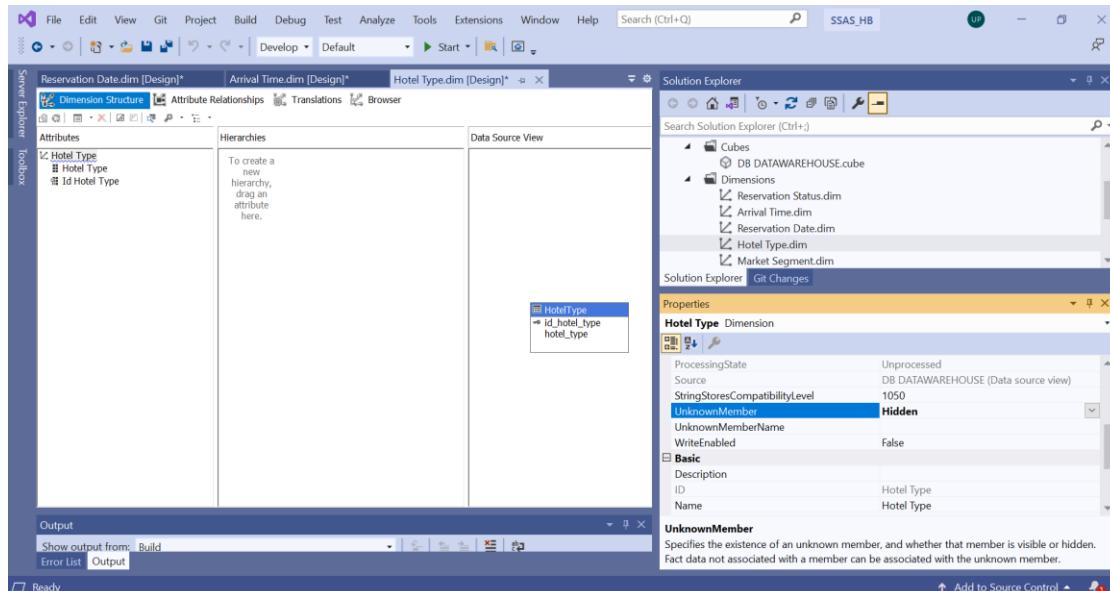
- Bước 1: Double click bảng Reservation Date, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.56. Kết quả sau khi chỉnh sửa UnknownMember

### 5.3.4. Hotel Type

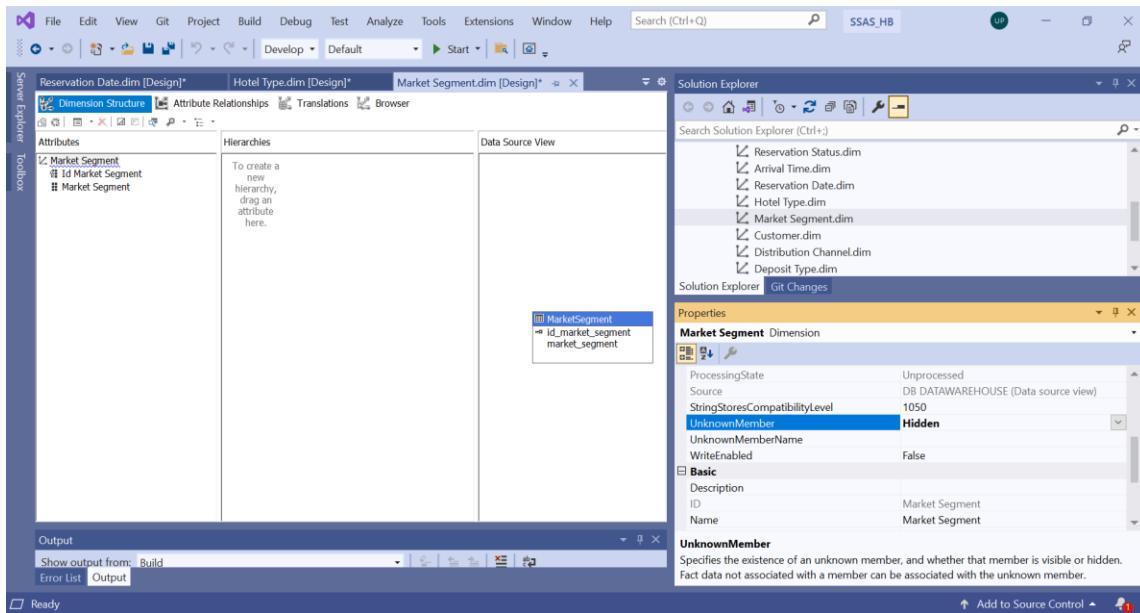
- Bước 1: Double click bảng Hotel Type, tại cửa sổ Solution Explore, tab Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.56. Kết quả sau khi chỉnh sửa UnknownMember

### 5.3.5. Market Segment

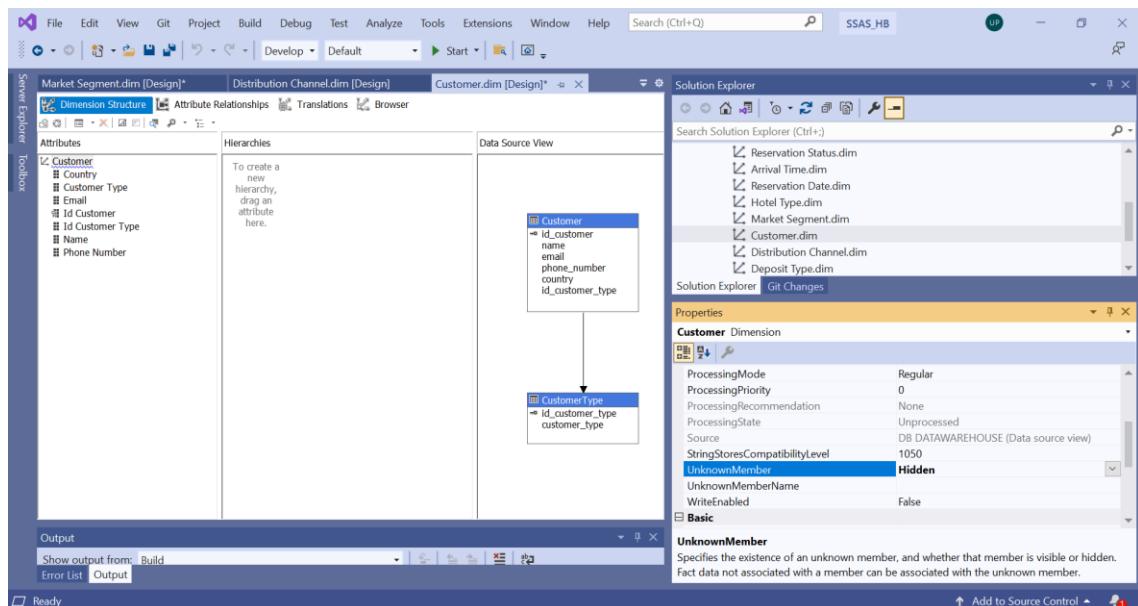
- Bước 1: Double click bảng Market Segment, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.57 .Kết quả sau khi chỉnh sửa UnknownMember.

### 5.3.6. Customer

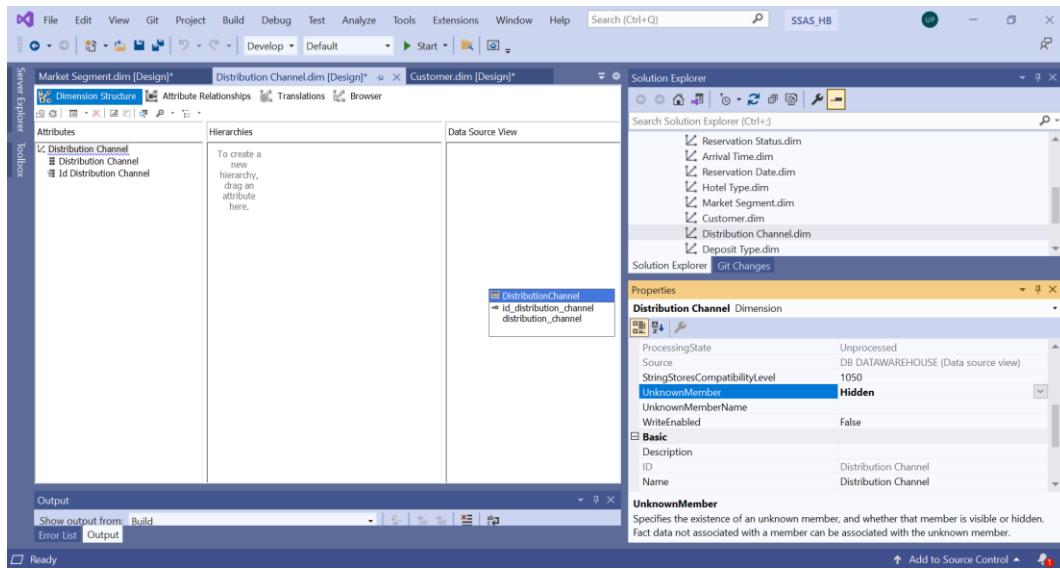
- Bước 1: Double click bảng Customer, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.58 .Kết quả sau khi chỉnh sửa UnknownMember.

### 5.3.7. Distribution Channel

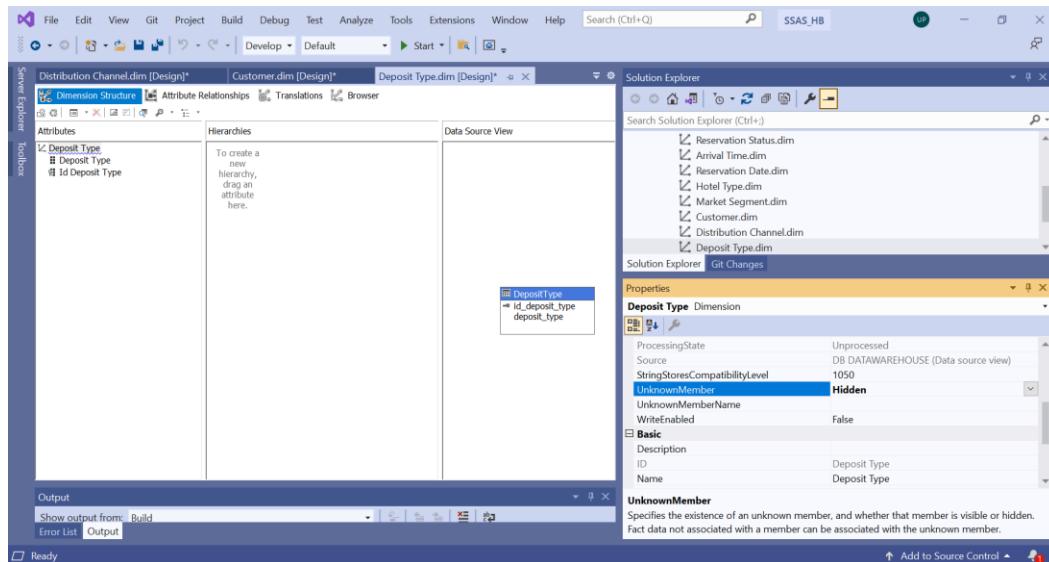
- Bước 1: Double click bảng Distribution Channel, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



Hình 3.59 .Kết quả sau khi chỉnh sửa UnknownMember.

### 5.3.8. Deposit Type

- Bước 1: Double click bảng Deposit Type, tại cửa sổ Properties, thay đổi thông tin UnknownMember từ Visible thành Hidden



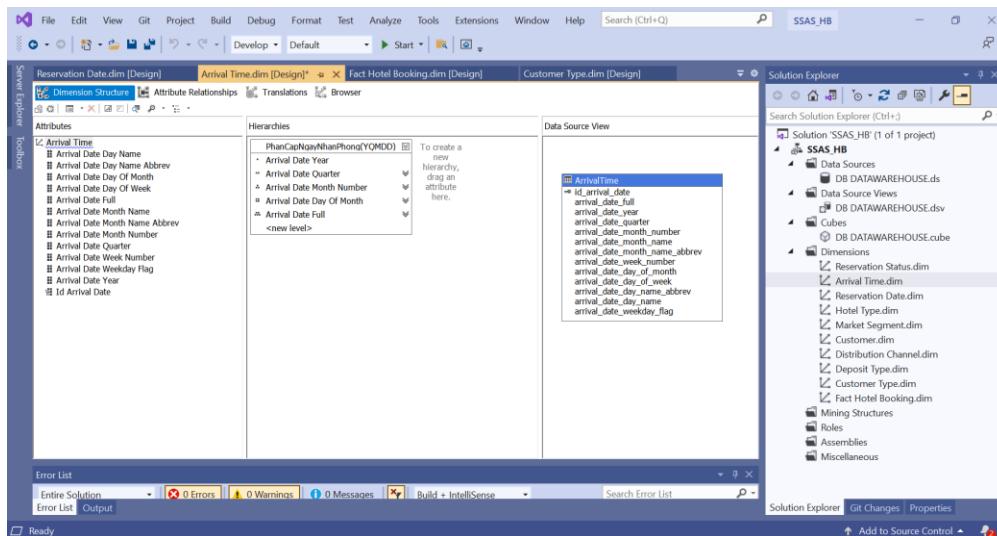
Hình 3.60..Kết quả sau khi chỉnh sửa UnknownMember.

## 6. Phân cấp trong các bảng chiều

### 6.1. Arrival Time

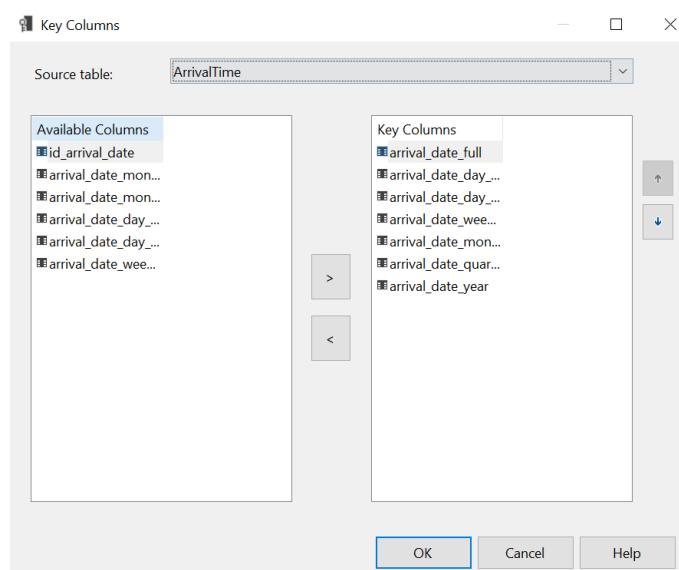
#### 6.1.1. PhanCapNgayNhanPhong(YQMDD)

- Bước 1: Kéo thả các thuộc tính phân cấp từ Attributes sang Hierarchies và đặt tên cho Hierarchies đó:



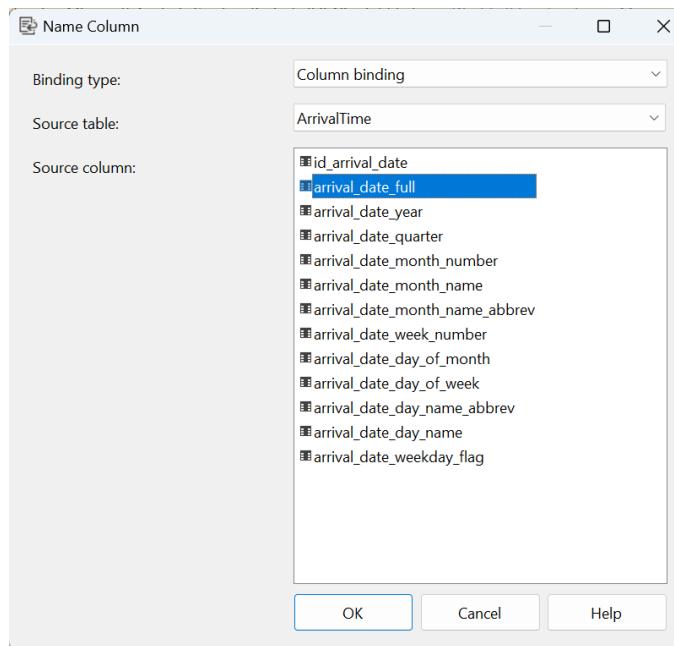
Hình 3.61. Phân cấp ngày nhận phòng (YQMDD)

- Bước 2: Chính khóa dòng và tên dòng của thuộc tính Arrival Date Full
  - + Vì thuộc tính Arrival Date Full là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



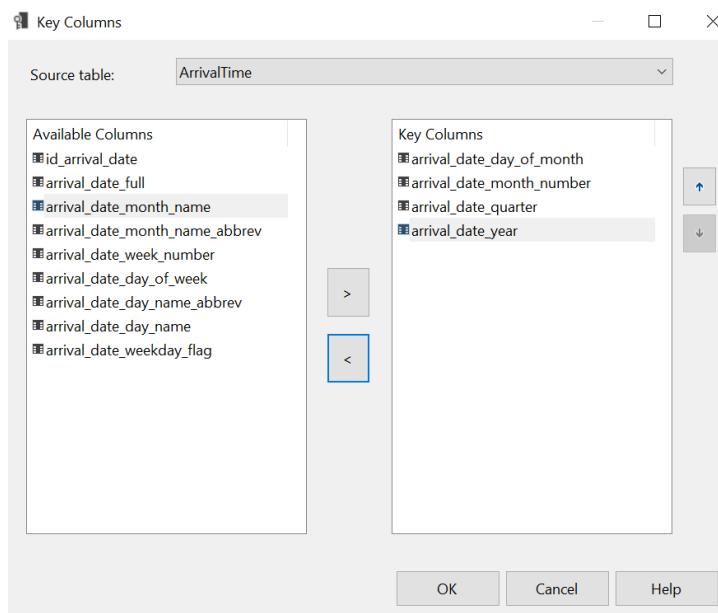
Hình 3.62. Khóa dòng cho thuộc tính Arrival Date Full

- + Tạo tên dòng của thuộc tính Arrival Date Full



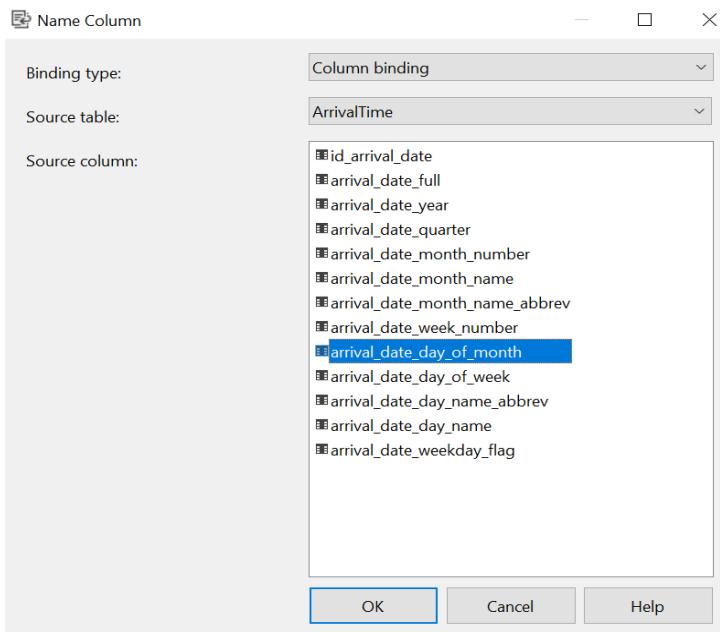
*Hình 3.63. Tạo tên dòng cho thuộc tính Arrival Date Full*

- Bước 3: Chính khóa dòng và tên dòng của thuộc tính Arrival Date Day Of Month
  - + Vì thuộc tính Arrival Date Day Of Month là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



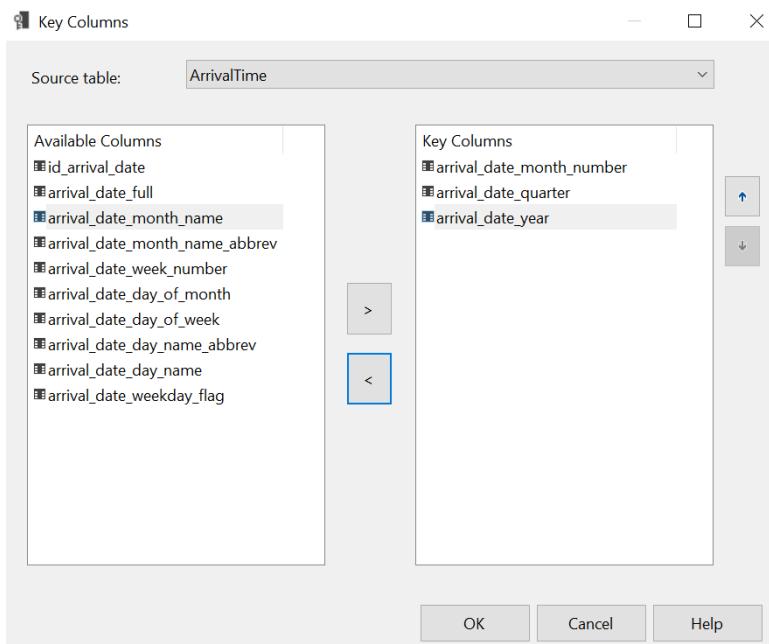
*Hình 3.64. Khóa dòng cho thuộc tính Arrival Date Day Of Month*

- + Tạo tên dòng của thuộc tính Arrival Date Day Of Month



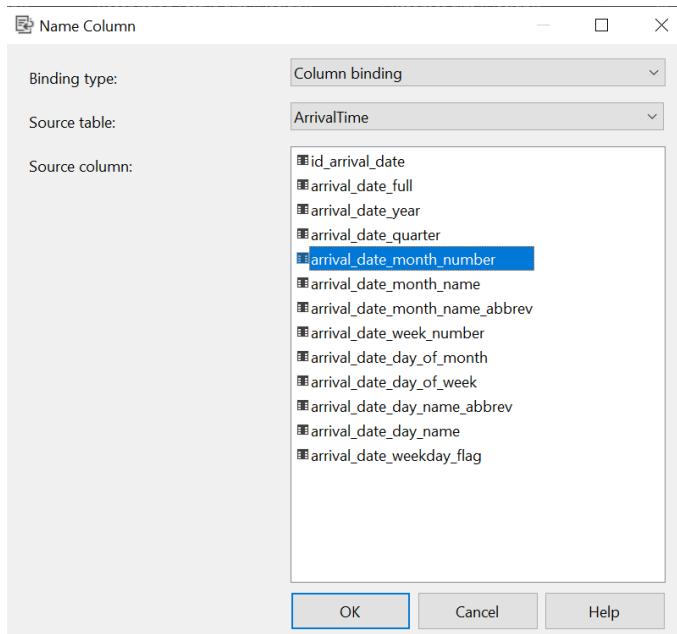
*Hình 3.65. Tạo tên dòng cho thuộc tính Arrival Date Day Of Month*

- Bước 4: Chính khóa dòng và tên dòng của thuộc tính Arrival Date Month Number
  - + Vì thuộc tính Arrival Date Month Number là thuộc tính cấp nhõ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



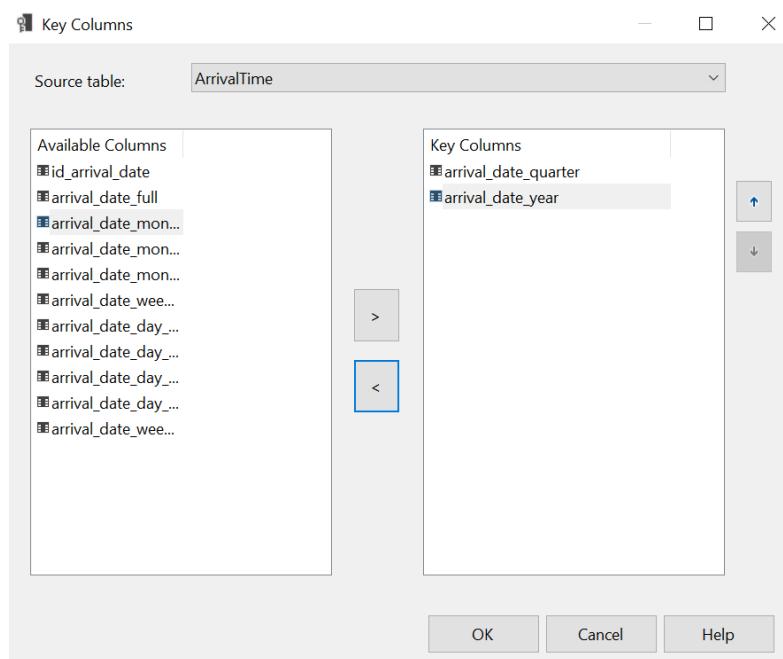
*Hình 3.66. Khóa dòng cho thuộc tính Arrival Date Month Number*

+ Tạo tên dòng của thuộc tính Arrival Date Month Number



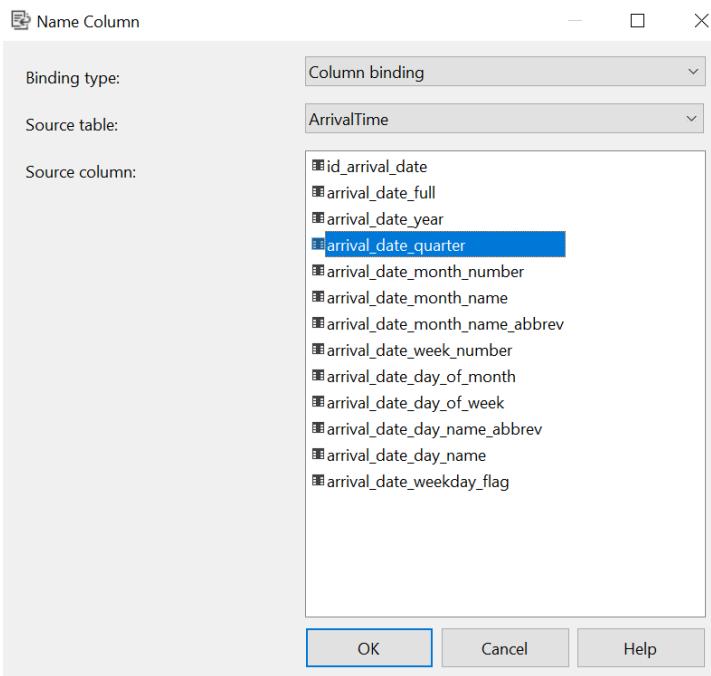
*Hình 3.67. Tạo tên dòng cho thuộc tính Arrival Date Month Number*

- Bước 5: Chỉnh khóa dòng và tên dòng của thuộc tính Arrival Date Quarter
  - + Vì thuộc tính Arrival Date Quarter là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



*Hình 3.68. Khóa dòng cho thuộc tính Arrival Date Quarter*

+ Tạo tên dòng của thuộc tính Arrival Date Quarter



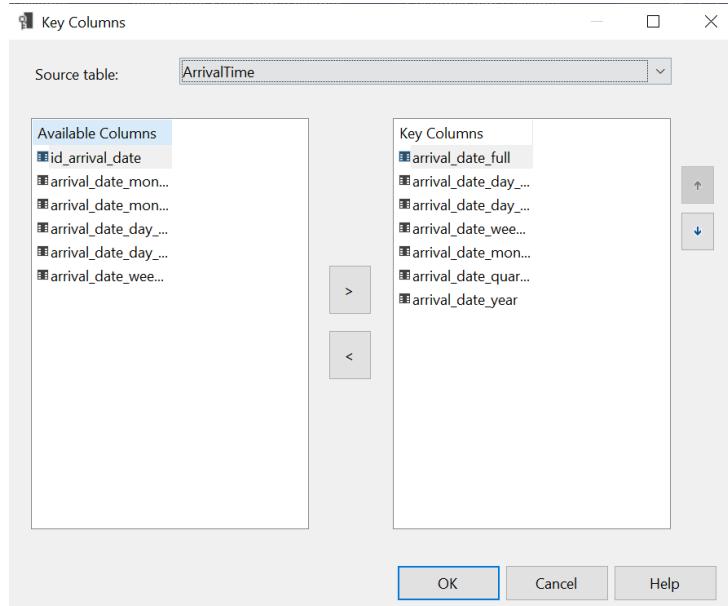
*Hình 3.69. Tạo tên dòng cho thuộc tính Arrival Date Quarter*

#### 6.1.2. Phân Cấp Ngày Nhận Phòng (WDD)

- Bước 1: Kéo thả các thuộc tính phân cấp từ Attributes sang Hierarchies và đặt tên cho Hierarchies đó:

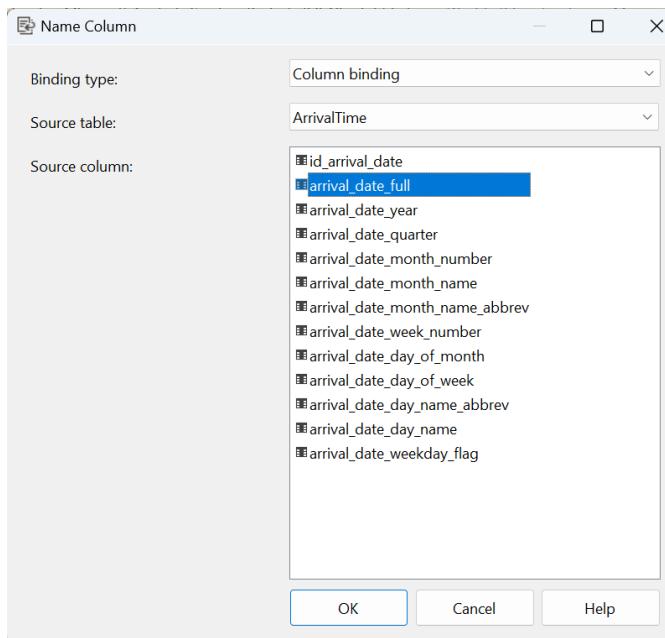
*Hình 3.70. Phân cấp ngày nhận phòng (WDD)*

- Bước 2: Chỉnh khóa dòng và tên dòng của thuộc tính Arrival Date Day Of Week
  - + Vì thuộc tính Arrival Date Day Of Week là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



*Hình 3.71. Khóa dòng cho thuộc tính Arrival Date Day Of Week*

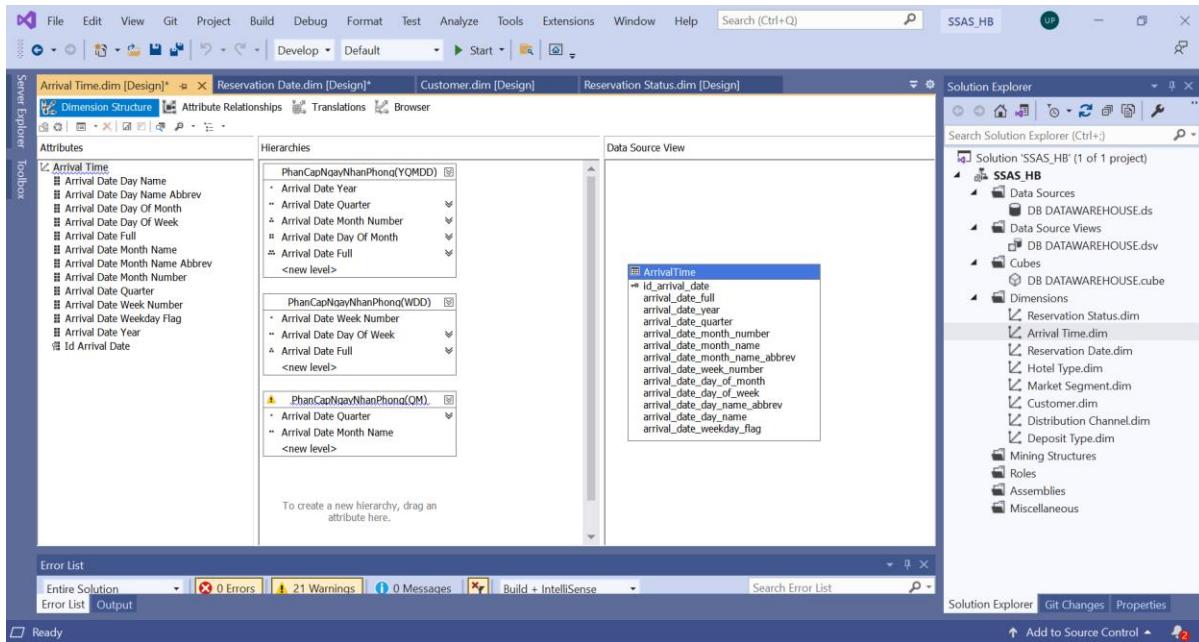
- + Tạo tên dòng của thuộc tính Arrival Date Day Of Week



*Hình 3.72. Tạo tên dòng cho thuộc tính Arrival Date Day Of Week*

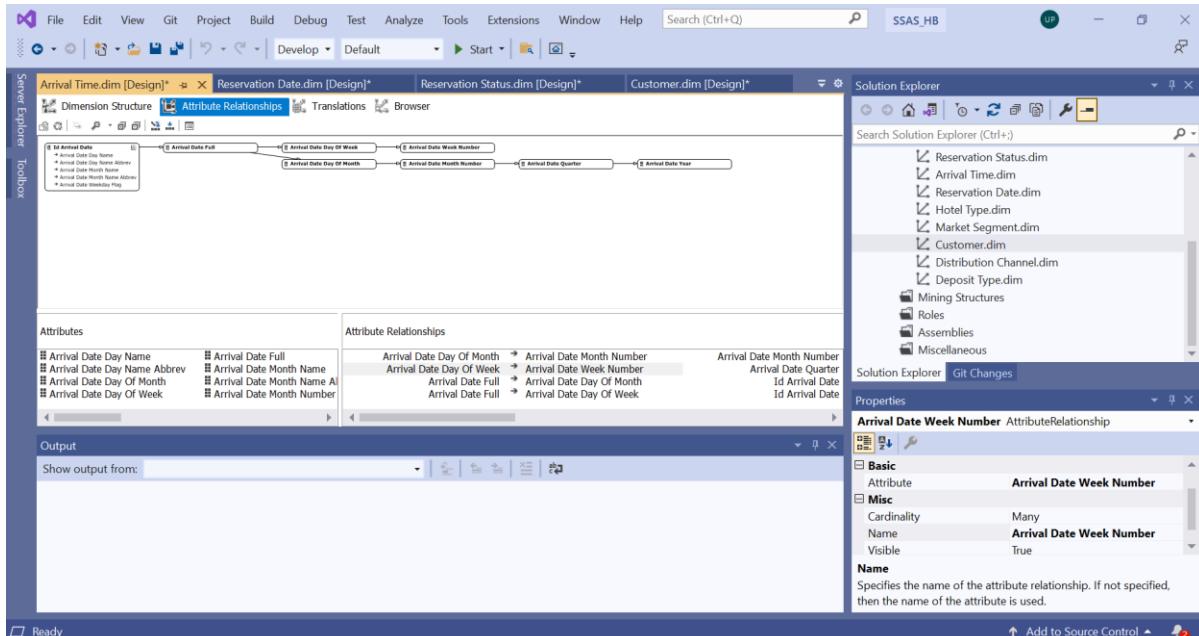
### 6.1.3. PhanCapNgayNhanPhong(QM)

- + Bước 1: Kéo thả các thuộc tính phân cấp từ Attributes sang Hierarchies và đặt tên cho Hierarchies đó:



Hình 3.73. Phân cấp ngày nhận phòng (QM)

### 6.1.4. Sang tab Attribute Relationships tạo liên kết các quan hệ phân cấp và tiến hành phân cấp

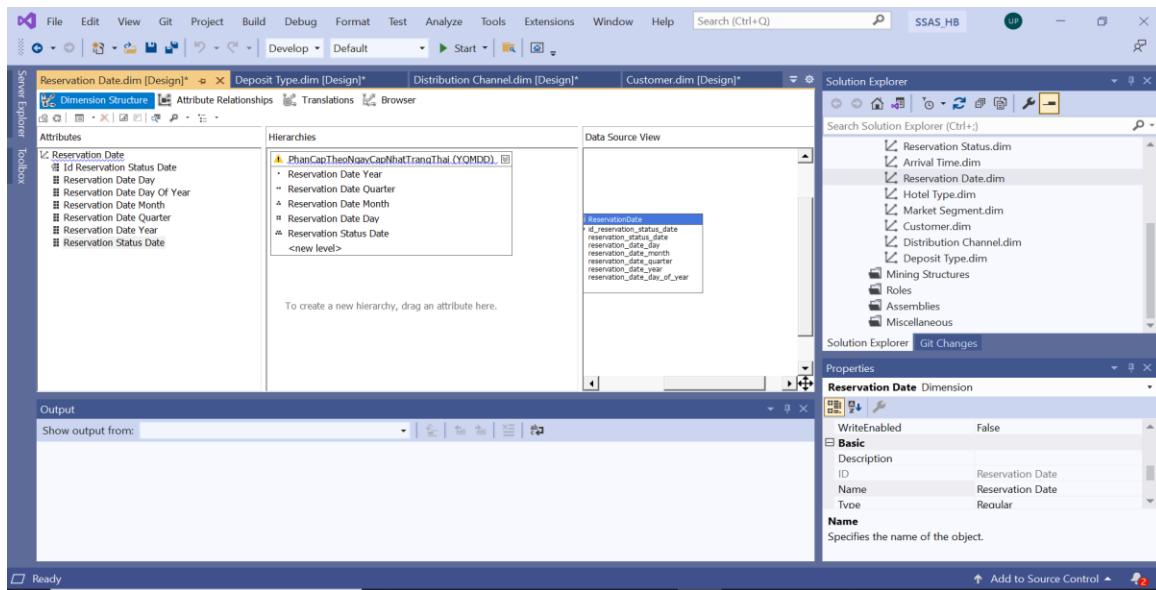


Hình 3.74. Phân cấp thuộc tính theo ngày nhận phòng

## 6.2. Reservation Date

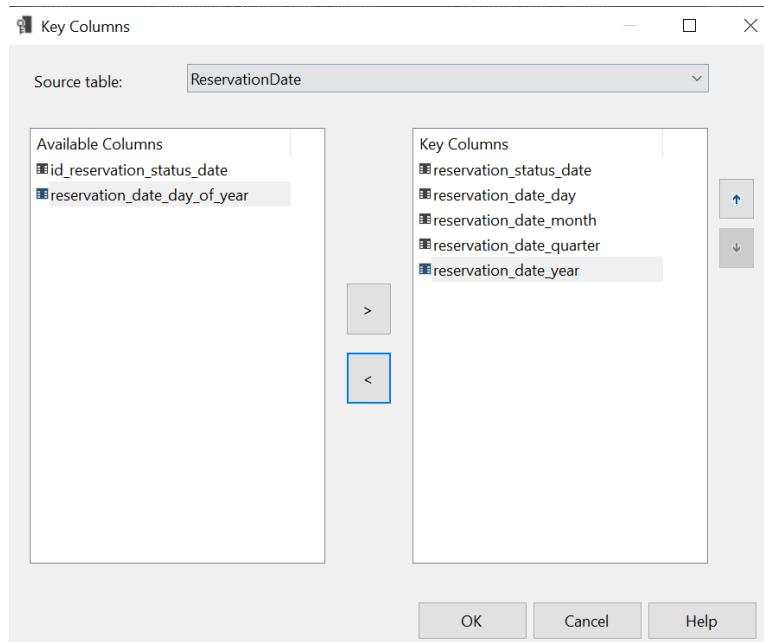
### 6.2.1. Phân Cấp Theo Ngày Cập Nhật Trạng Thái (YQMDD)

- + Bước 1: Kéo thả các thuộc tính phân cấp từ Attributes sang Hierarchies và đặt tên cho Hierarchies đó:



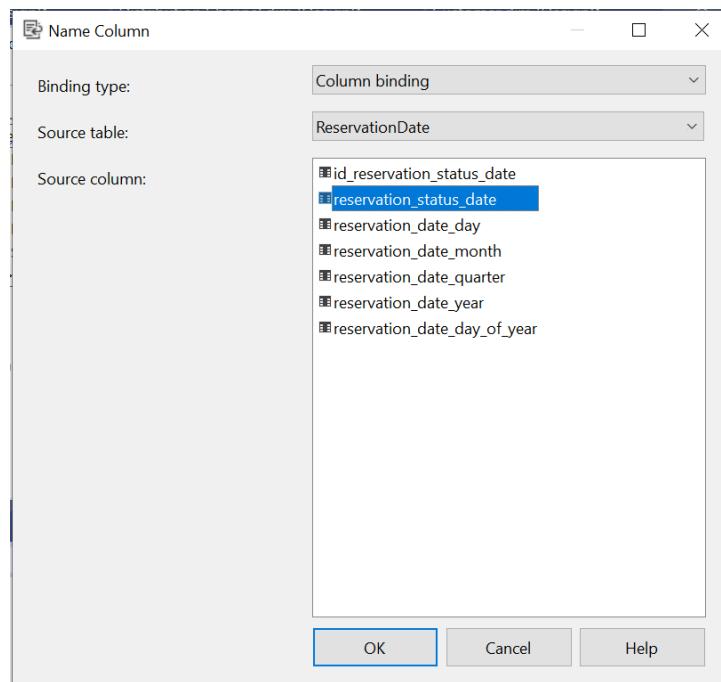
*Hình 3.74. Phân cấp ngày cập nhật trạng thái (YQMDD)*

- + Bước 2: Chỉnh khóa dòng và tên dòng của thuộc tính Reservation Status Date
- + Vì thuộc tính Reservation Status Date là thuộc tính cấp thấp nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



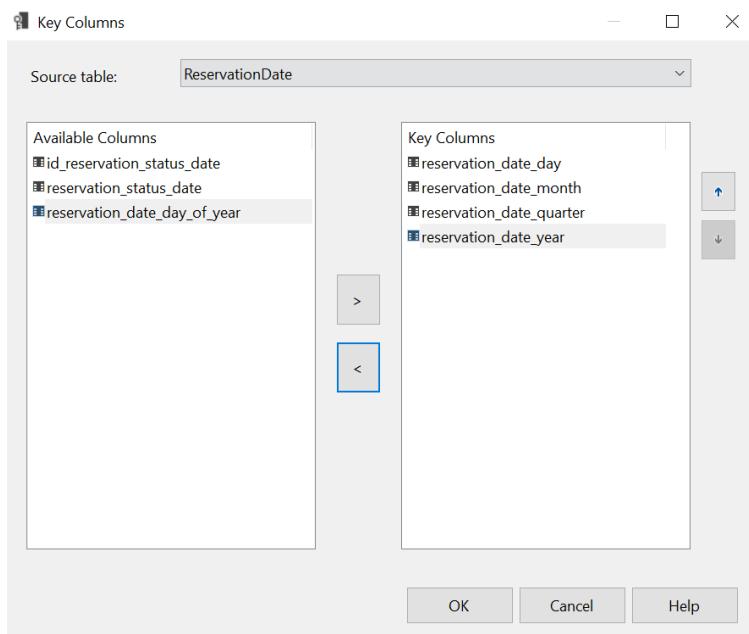
*Hình 3.75. Khóa dòng cho thuộc tính Reservation Status Date*

+ Tạo tên dòng của thuộc tính Reservation Status Date



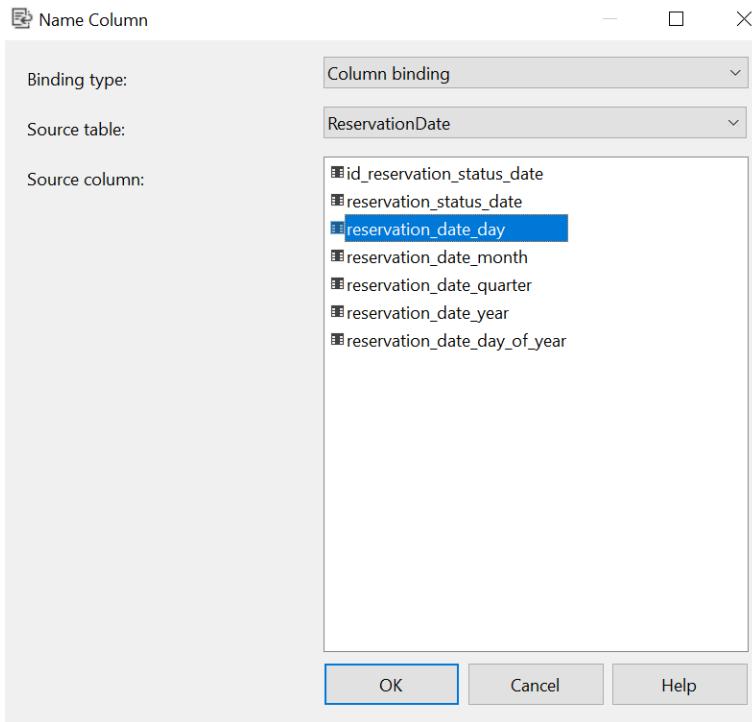
*Hình 3.76. Tạo tên dòng cho thuộc tính Reservation Status Date*

- Bước 3: Chính khóa dòng và tên dòng của thuộc tính Reservation Date Day
  - + Vì thuộc tính Reservation Date Day là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



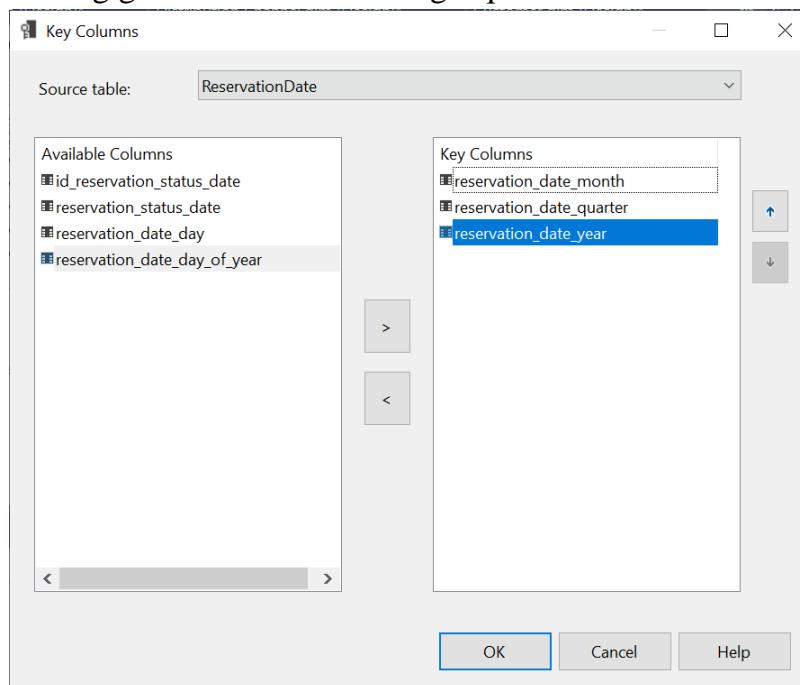
*Hình 3.77. Khóa dòng cho thuộc tính Reservation Date Day*

+ Tạo tên dòng của thuộc tính Reservation Date Day



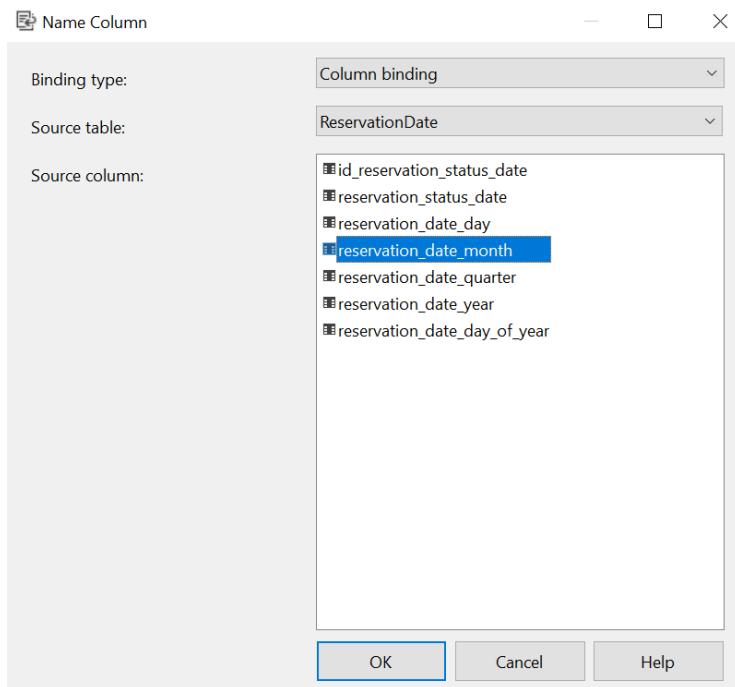
*Hình 3.78. Tạo tên dòng cho thuộc tính Reservation Date Day*

- Bước 4: Chính khóa dòng và tên dòng của thuộc tính Reservation Date Month
  - + Vì thuộc tính Reservation Date Month là thuộc tính cấp nhõ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



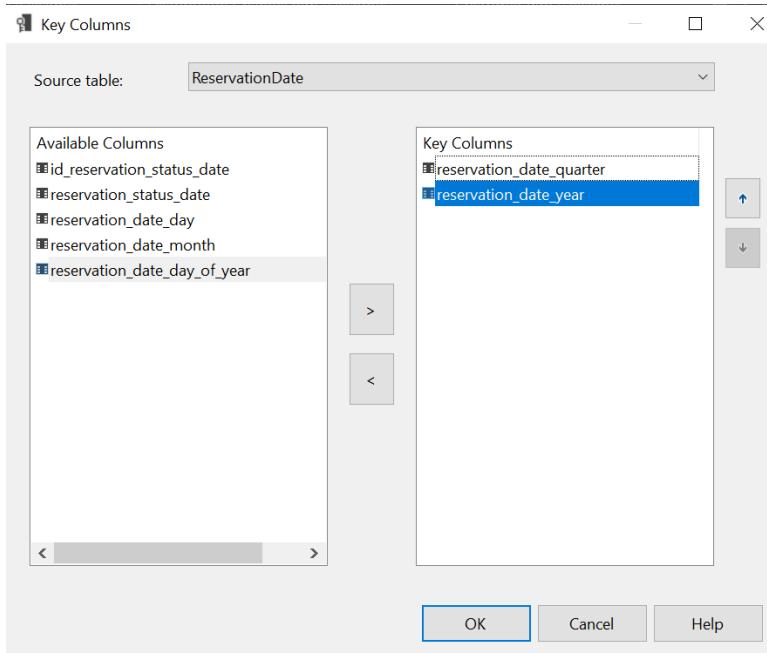
*Hình 3.79. Khóa dòng cho thuộc tính Reservation Date Month*

- + Tạo tên dòng của thuộc tính Reservation Date Month



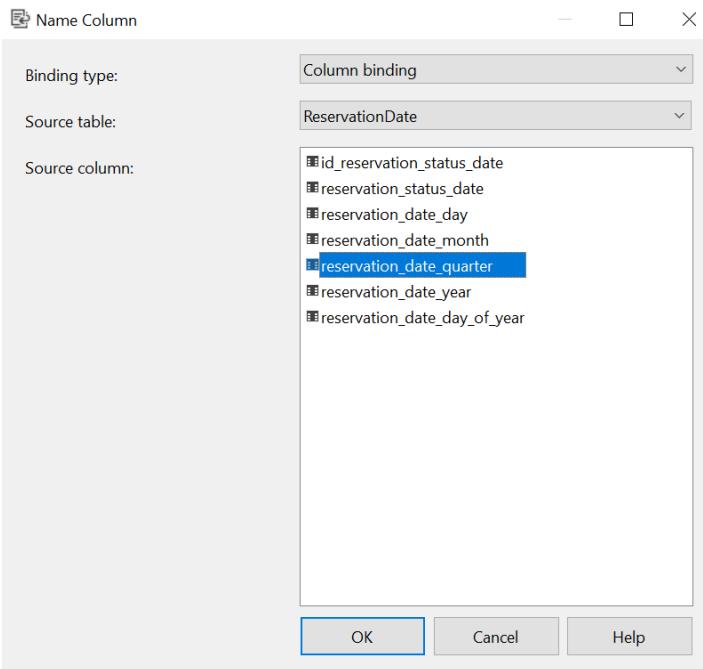
*Hình 3.80. Tạo tên dòng cho thuộc tính Reservation Date Month*

- Bước 5: Chỉnh khóa dòng và tên dòng của thuộc tính Reservation Date Quarter
  - + Vì thuộc tính Reservation Date Quarter là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những cấp cao hơn.



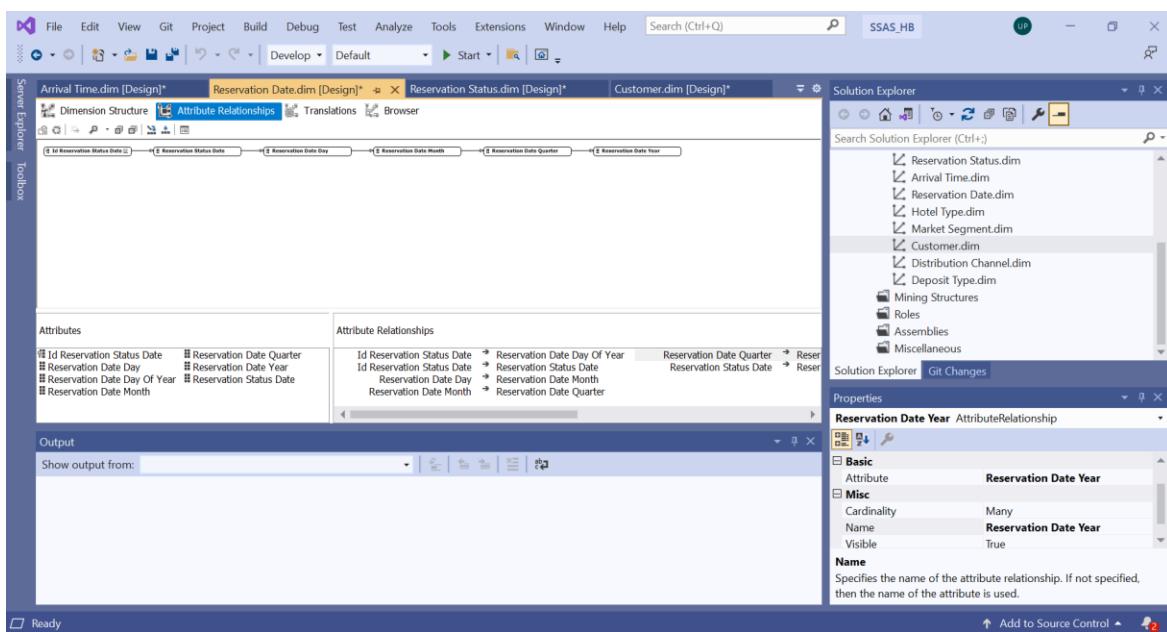
*Hình 3.81. Khóa dòng cho thuộc tính Reservation Date Quarter*

+ Tạo tên dòng của thuộc tính Reservation Date Quarter



Hình 3.82. Tạo tên dòng cho thuộc tính Reservation Date Quarter

6.2.2. Sang tab Attribute Relationships tạo liên kết các quan hệ phân cấp và tiến hành phân cấp

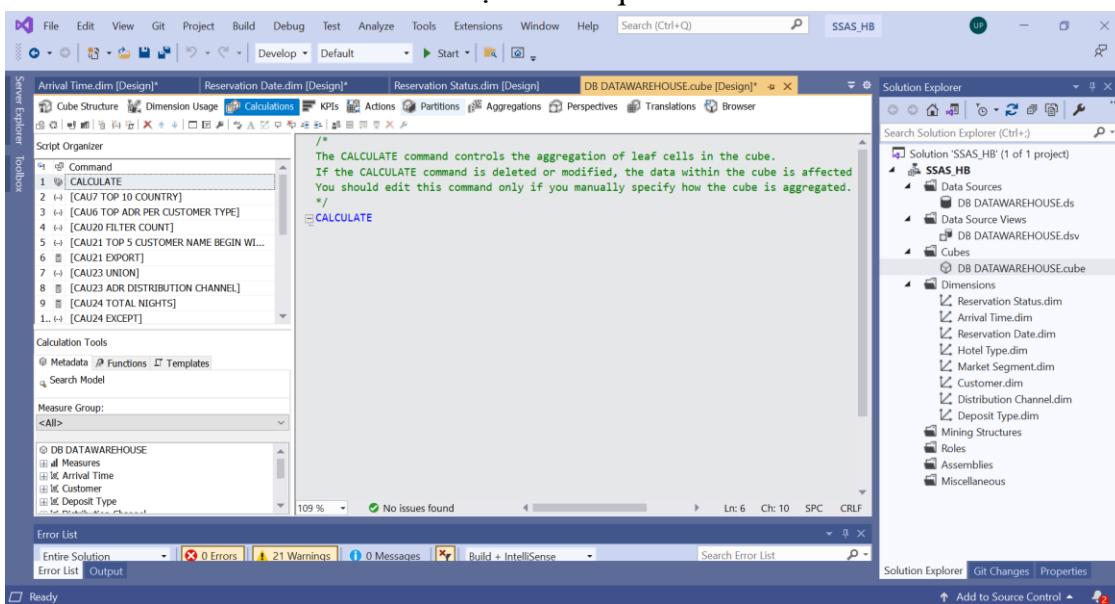


Hình 3.83. Phân cấp thuộc tính theo ngày cập nhật trạng thái (YQMDD)

## 7. Định nghĩa Calculate, Named Set và Key Performance Indicators (KPIs)

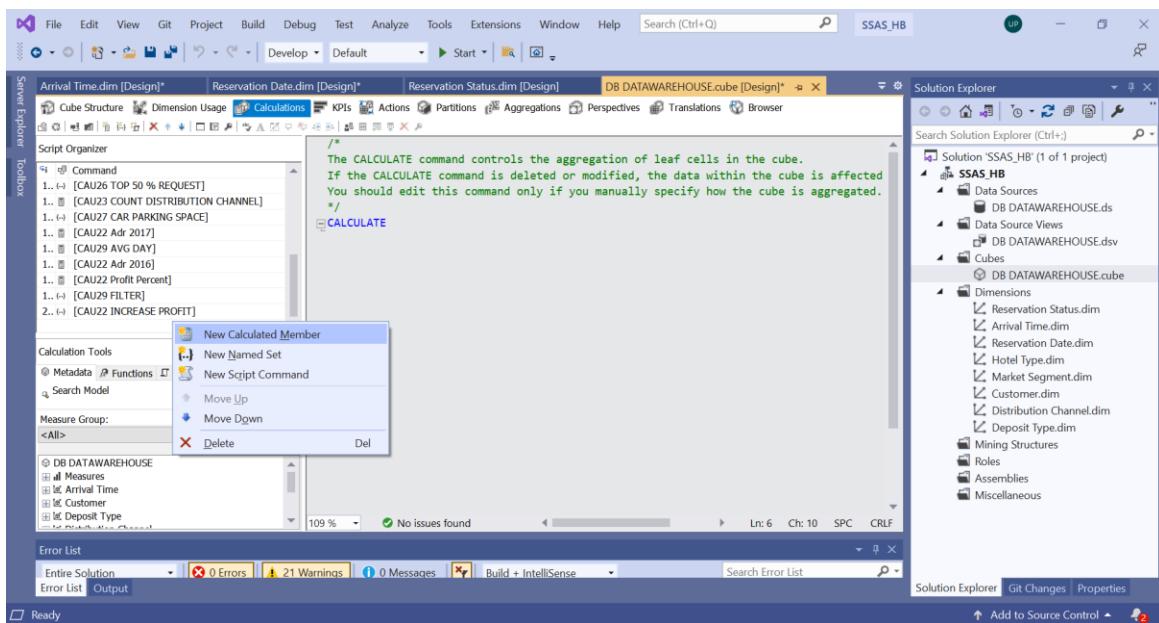
### 7.1. Calculate

- Khái niệm: Calculate sẽ tạo ra các thuộc tính độ đo thông qua việc tính toán từ các thuộc tính Measures khác trong bảng Fact.
- Bước 1: Mở khôi cube đã tạo lên → qua tab calculations



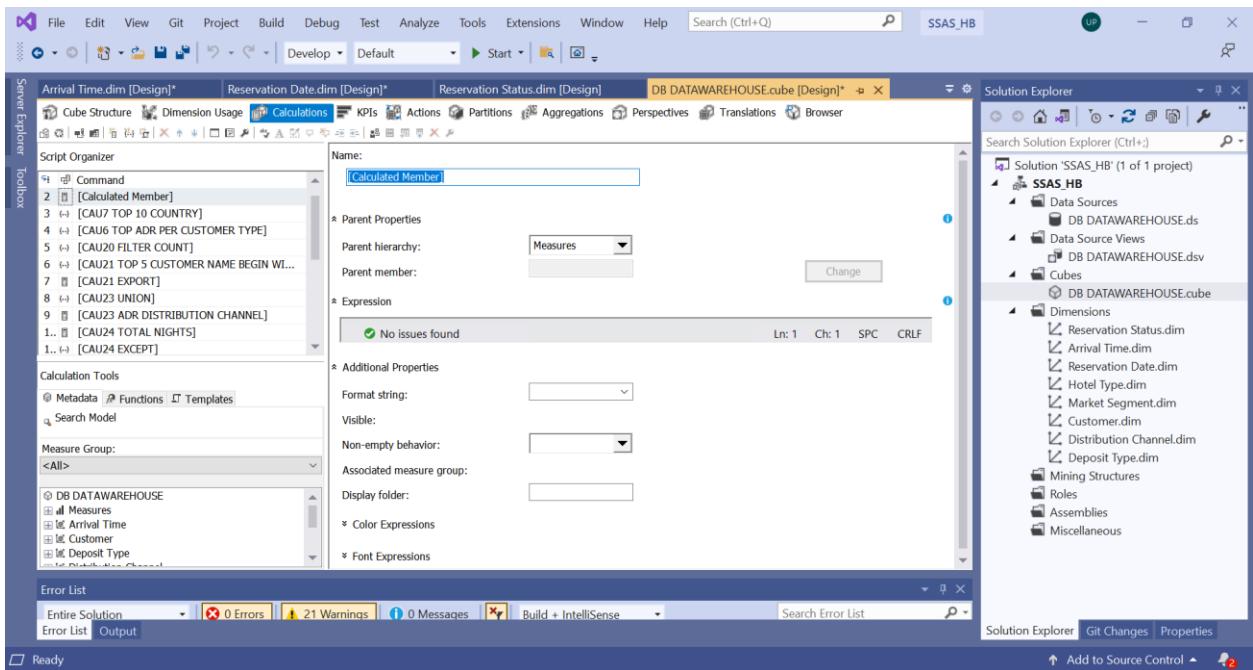
Hình 3.84. Giao diện Calculations

- Bước 2: Để tạo một Calculate mới trong bên trái phần Script Organizer ta chọn CALCULATE → click chuột phải vào New Calculated Members



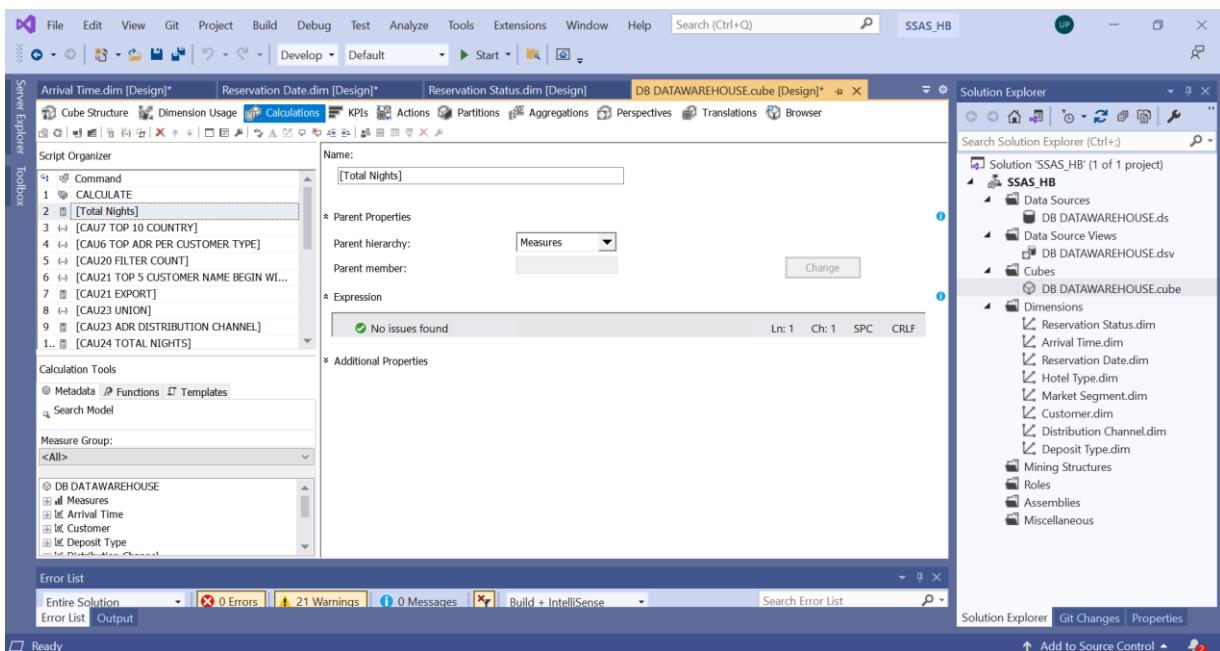
Hình 3.85. Giao diện tạo New Calculated Member

- Bước 3: Thành phần trong Calculated Member



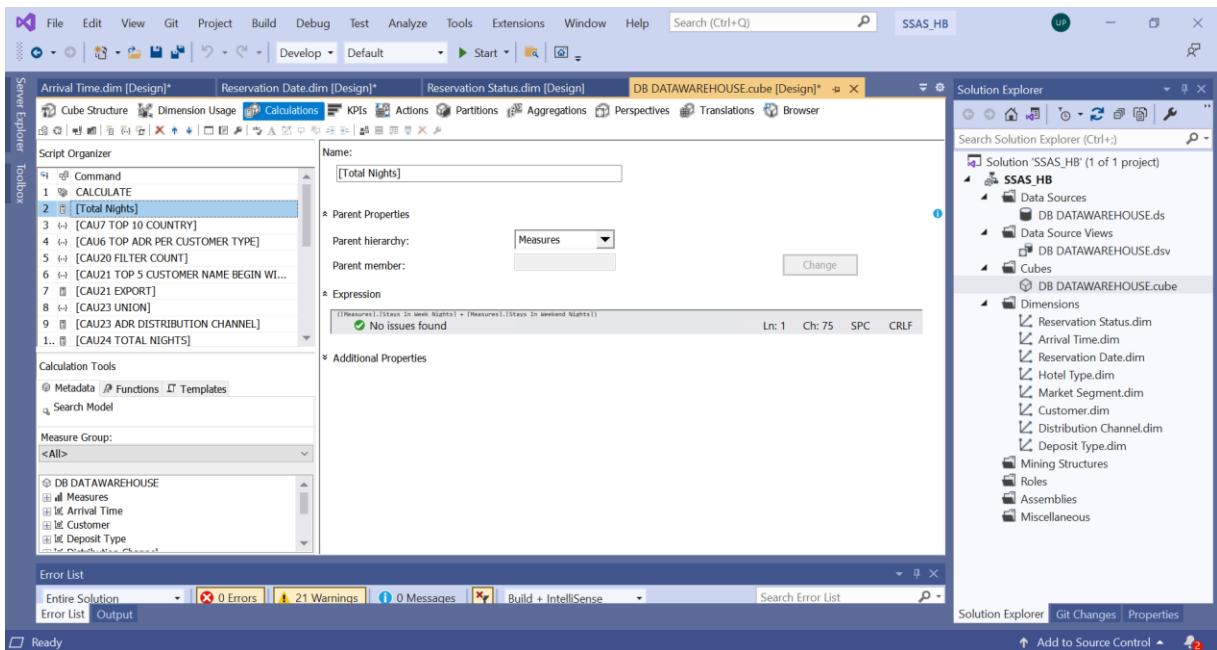
Hình 3.86. Giao diện một calculated Member

- + Bước 3.1: Name: Chứa tên của Calculate
- + Bước 3.2: Expression: Chứa các công thức tính toán
- + Bước 3.3: Format String: Chứa các định dạng của độ đo.
- + Bước 3.4: Ví dụ: Tính tổng số đêm mà khách hàng đặt



Hình 3.87. Đổi tên thành Total Nights

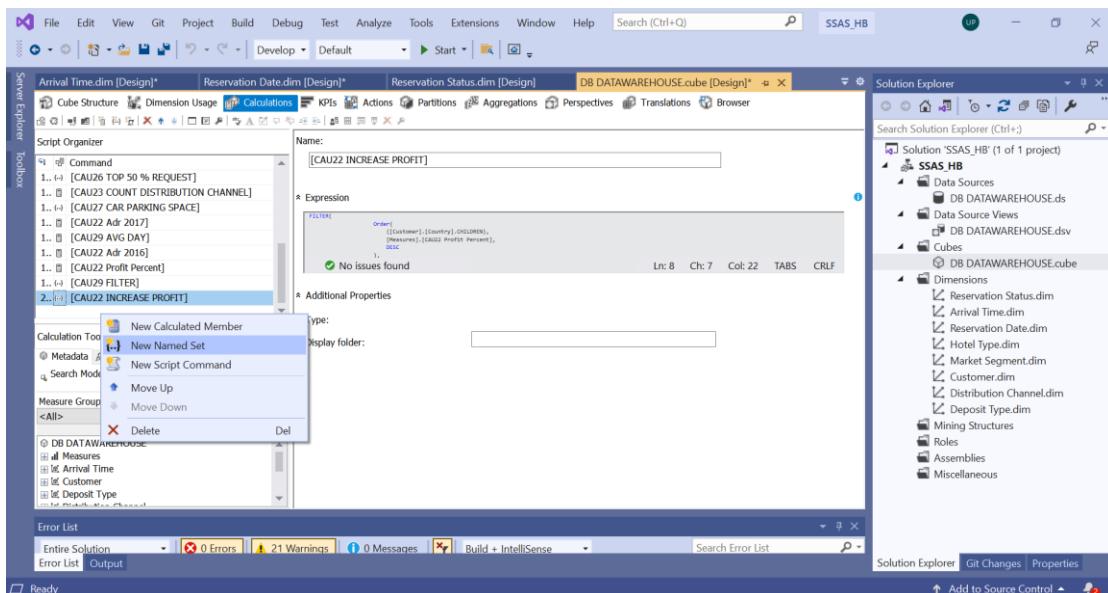
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.88. Total Nights = Stays In Weekend Nights + Stays In Week Nights

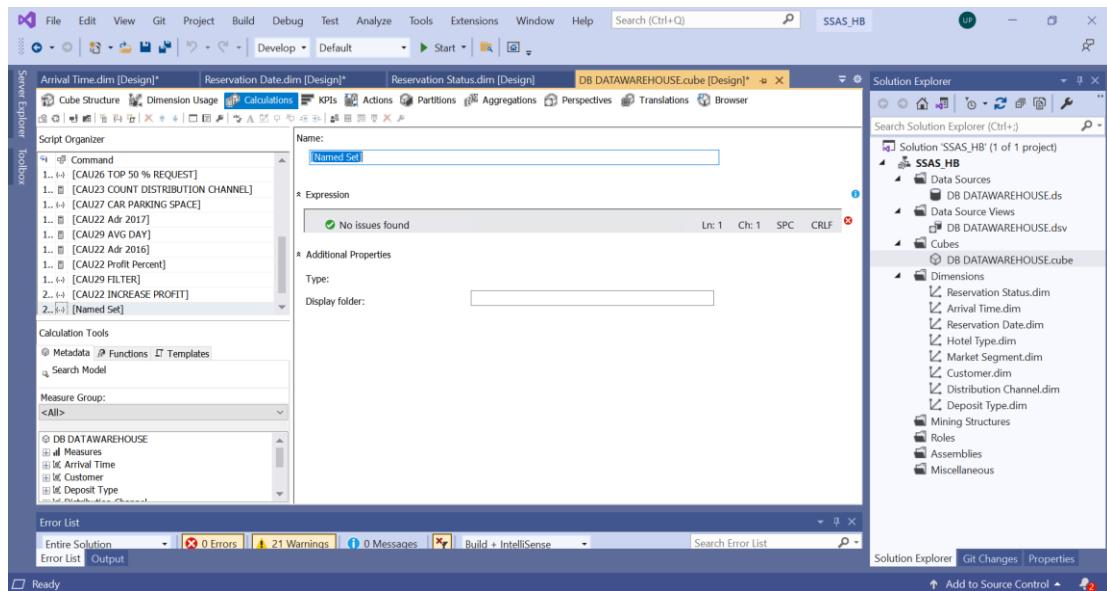
### 7.2. Named Set

- Khái niệm:
  - + Named Set là một thể hiện của ngôn ngữ MDX cái mà trả về một cái tập các thành viên trong một chiều
  - + Named Set giống như những operator tự lọc trước khi truy vấn MDX.
- Bước 1: Tạo mới Named Set bằng cách click chuột phải chọn New Named Set



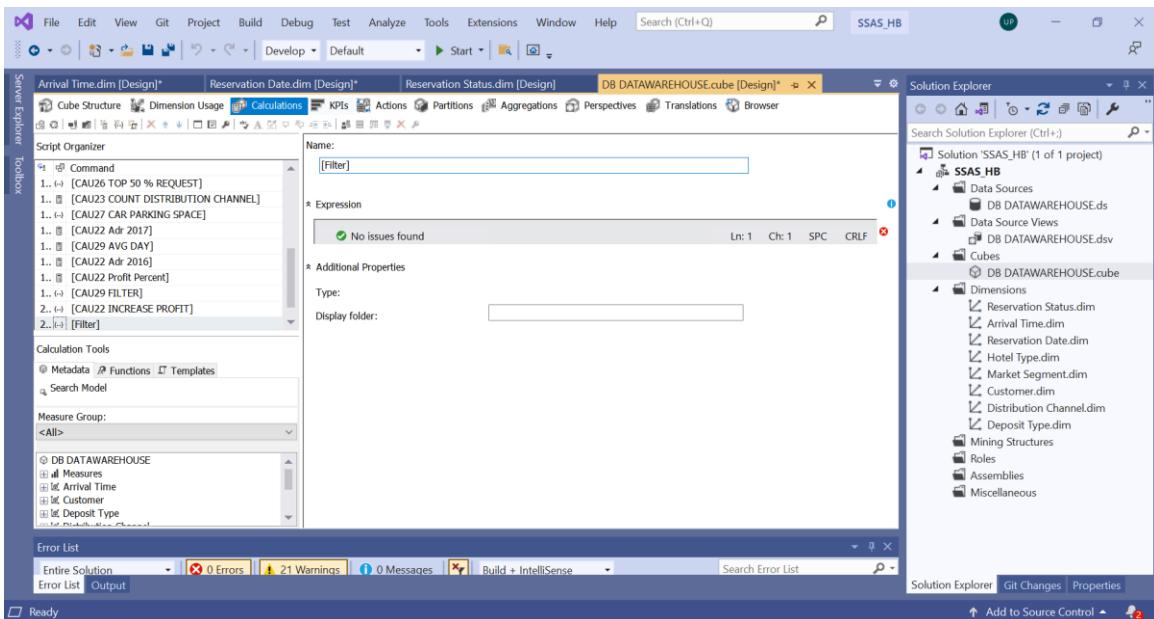
Hình 3.89. Tạo mới Named Set

- **Bước 2: Thành phần trong Named Set**



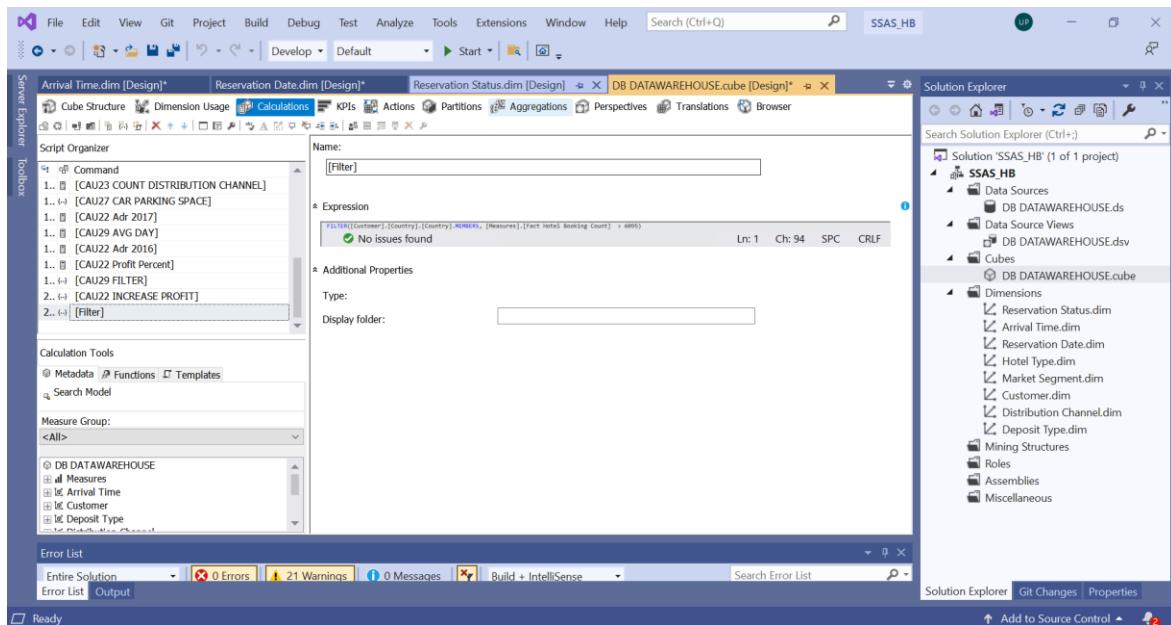
*Hình 3.90. Giao diện một Named Set*

- + **Bước 3.1:** Name: Chứa tên của Named Set
- + **Bước 3.2:** Expression: Chứa các hàm trong ngôn ngữ MDX
- + **Bước 3.3:** Ví dụ: Lọc ra các nước có số lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn



*Hình 3.91. Đổi tên thành Filter*

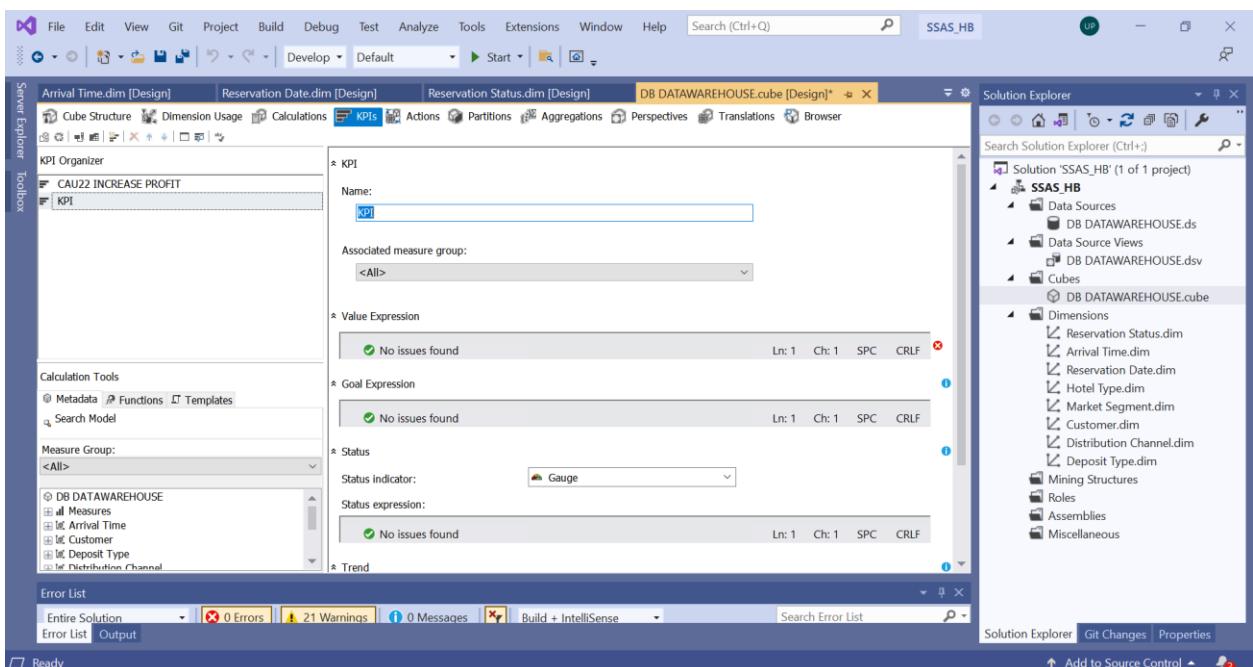
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.92. Dùng hàm Filter để lọc dữ liệu

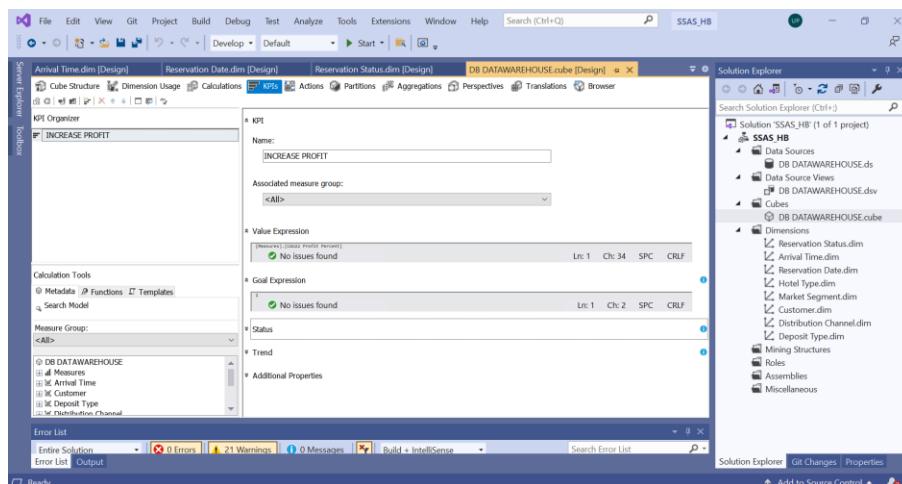
### 7.3. Key Performance Indicators (KPIs)

- Bước 1: Mở khôi cube đã tạo lên → qua tab KPIs
- Bước 2: Để tạo một KPIs mới trong bên trái phần Script Organizer → click chuột phải vào New KPI
- Bước 3: Thành phần trong KPI



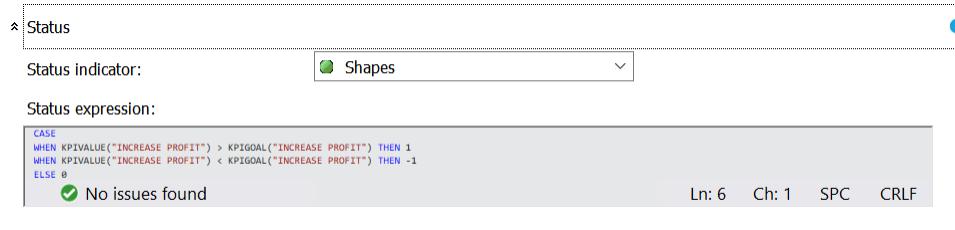
Hình 3.93. Giao diện một KPI

- + Bước 3.1: Name: Chứa tên của KPI
- + Bước 3.2: Associated measure group: Nơi chứa các thuộc tính đo lường
- + Bước 3.3: Value Expression: Giá trị hiện tại
- + Bước 3.4: Goal Expression: Mục tiêu, dung để so sánh với giá trị đã xác định trước đó (Goal Expression)
- + Bước 3.5: Status indicator: Icon để hiển thị tương ứng với giá trị trả về
- + Bước 3.6: Status Expression: Nơi chứa các đoạn mã để thể hiện status indicator
- + Bước 3.7: Trend indicator: Icon thể hiện tương ứng với giá trị trả về trong trend expression
- + Bước 3.8: Nơi chứa đoạn mã để mô tả cho Trend indicator
- + Bước 3.9: Ví dụ: Tạo một KPI thể hiện tỷ lệ lợi nhuận của từng quốc gia trong năm 2017 và 2016. So sánh tỷ lệ đó với 1 để xem doanh thu của các quốc gia tăng, giảm hay ổn định. Thể hiện xu hướng tỷ lệ doanh thu là tăng, giảm hay ổn định.

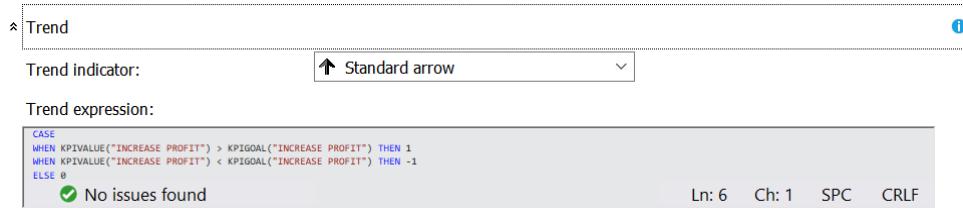


*Hình 3.93. Tạo một KPI tên INCREASE PROFIT, sử dụng các thuộc tính lấy từ bảng Fact Hotel Booking với giá trị hiện tại là CAU22 Profit Percent và giá trị mục tiêu là 1.*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.94. Xét điều kiện. Nếu trả về 1 thì icon màu xanh, trả về -1 thì icon màu đỏ và trả về 0 thì icon màu vàng.



Hình 3.95. Xét điều kiện. nếu trả về 1 thì mũi tên đi lên, nếu trả về -1 thì mũi tên đi xuống và trả về 0 thì mũi tên nằm ngang

Country	CAU22 Apr 2016	CAU22 Apr 2017	INCREASE PROFIT Status	INCREASE PROFIT Trend
ABW	158	99	-1	-1
AGO	21758	15295	-1	-1
AIR	0	0	-1	-1
ALB	486	353	1	1
AND	491	928	1	1
ARE	4506	2409	-1	-1
ARG	12587	11312	-1	-1
ARM	499	252	-1	-1
ATG	0	72	-1	1
ATA	176	93	-1	-1
ATF	0	126	-1	-1
AUS	19149	28137	1	1
AUT	68535	57671	-1	-1
AZE	762	958	1	1
BGR	46	20	1	1
BEL	121077	124859	1	1
DEU	476	746	1	1

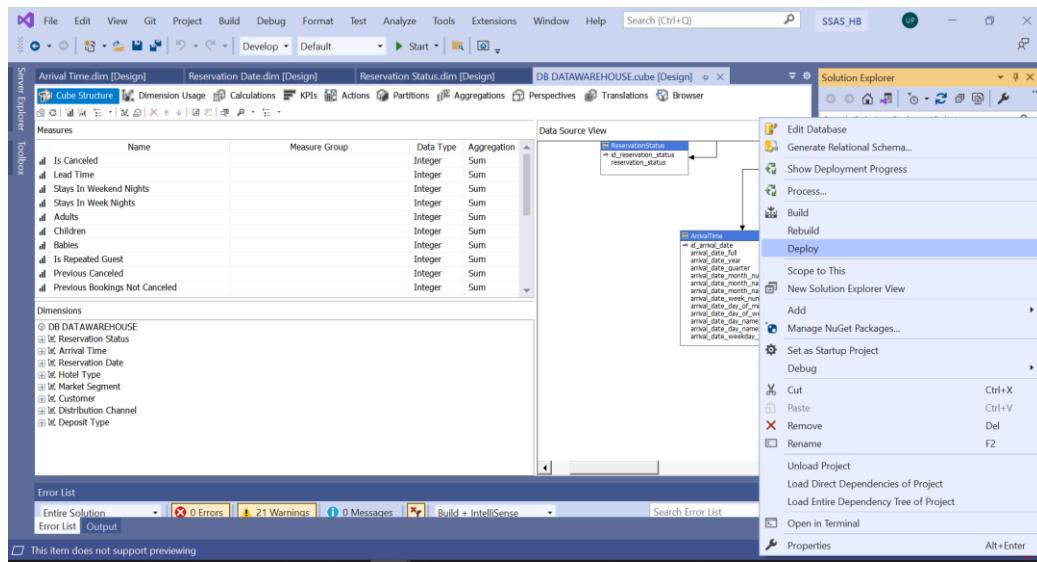
The cube has been reprocessed on the server. To prevent possible browsing errors, click Reconnect. To hide this message, Click here.

Error List: 0 Errors, 21 Warnings, 0 Messages

Hình 3.96. Kết quả khi chạy KPIs

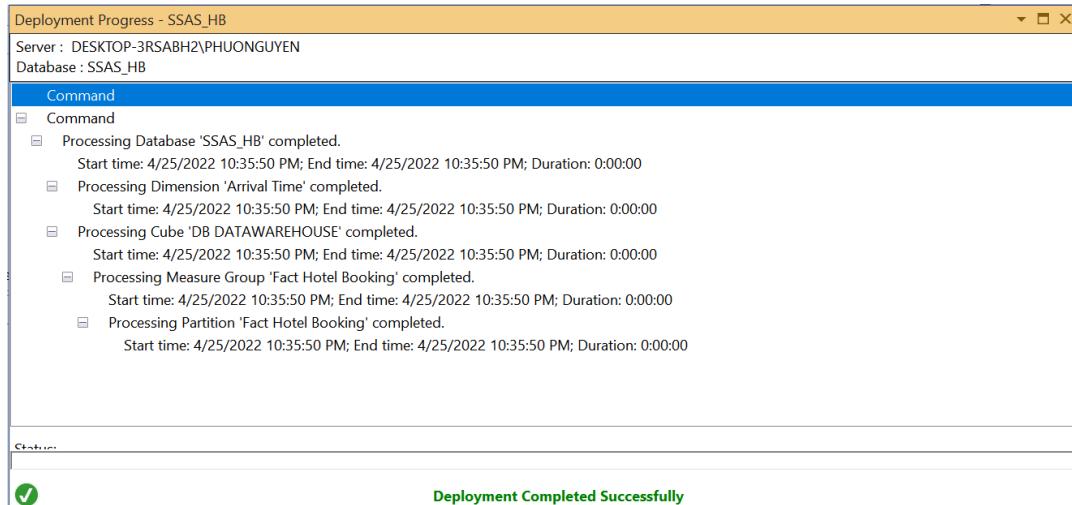
## 8. Chạy Project SSAS

- Bước 1: Nhấn chuột phải vào Project SSAS\_HB nhấp Deploy như hình minh họa:



Hình 3.98. Deploy Project SSAS

- Bước 2: Kết quả của quá trình Deploy thành công như hình minh họa:



Hình 3.99. Kết quả deploy thành công

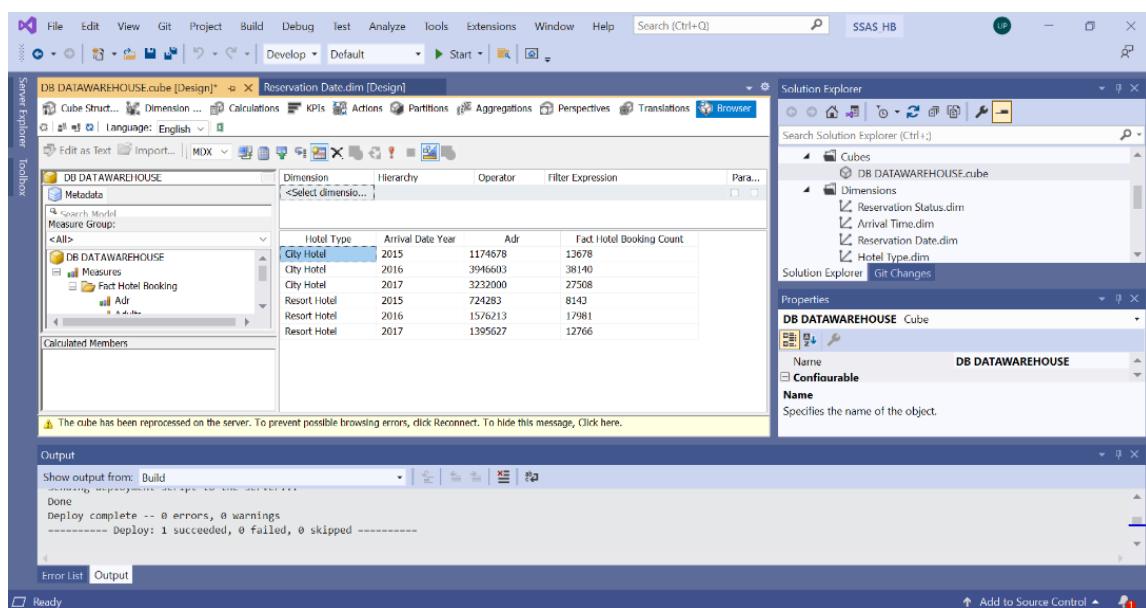
## 9. Quá trình phân tích dữ liệu bảng công cụ SSAS trên các khối Cube, công cụ Pivot Excel và ngôn ngữ MDX

9.1. Thống kê tổng số lượng khách hàng đặt phòng, tổng doanh thu của từng loại khách sạn theo từng năm.

*Ý nghĩa câu truy vấn:* Cho phép người quản lý có cái nhìn tổng quan tình hình kinh doanh của khách sạn từ đó đưa ra các chiến lược kinh doanh phù hợp hơn.

### 9.1.1. Công cụ SSAS trên các khối Cube

Kéo thuộc tính Hotel Type trong bảng Hotel type, thuộc tính Arrival Date year trong bảng Arrival Time, Adr và Fact Hotel Booking Count trong bảng Measures từ Measure Group sang cửa sổ lọc dữ liệu và cửa sổ thực thi. Ta thu được kết quả câu truy vấn



Hình 3.100. Kết quả câu truy vấn trong Visual Studio 2019

### 9.1.2. Công cụ Pivot Excel

Kéo thả Adr và Fact Hotel Booking từ bảng Measures vào bảng Values, thuộc tính Arrival Date Year thuộc bảng Arrival Time và thuộc tính Hotel Type của bảng Hotel Type vào bảng Rows ta được kết quả:

	Adr	Fact Hotel Booking Count
1 Row Labels		
2 City Hotel	2015	1174678
	2016	3946603
	2017	3232000
6 Resort Hotel	2015	724283
	2016	1576213
	2017	1395627
Grand Total	12049404	118216

Hình 3.101. Kết quả khi dùng công cụ Pivot Excel

### 9.1.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

```

WITH MEMBER [Measures].[Turnover]
as [Measures].[Adr],FORMAT_STRING = 'Currency'

Select {[Measures].[Turnover], [Measures].[Fact Hotel Booking Count]} on Rows ,
NON EMPTY{ [Hotel Type].[Hotel Type].Children * [Arrival Time].[Arrival Date Year].Children} on Columns
From [DB DATAWAREHOUSE];
  
```

	City Hotel	2015	2016	2017	Resort Hotel	2015	2016	2017
Turnover	\$1,174,678.00	\$3,946,603.00	\$3,232,000.00	\$724,283.00	\$1,576,213.00	\$1,395,627.00		
Fact Hotel Booking Count	13678	38140	27508	8143	17981	12766		

Hình 3.102. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

### ❖ Giải thích câu truy vấn MDX:

- Độ đo Turnover được dung để định dạng lại kiểu dữ liệu của độ đo Adr từ Standard(dạng chuẩn) sang Currency(đơn vị tiền tệ).

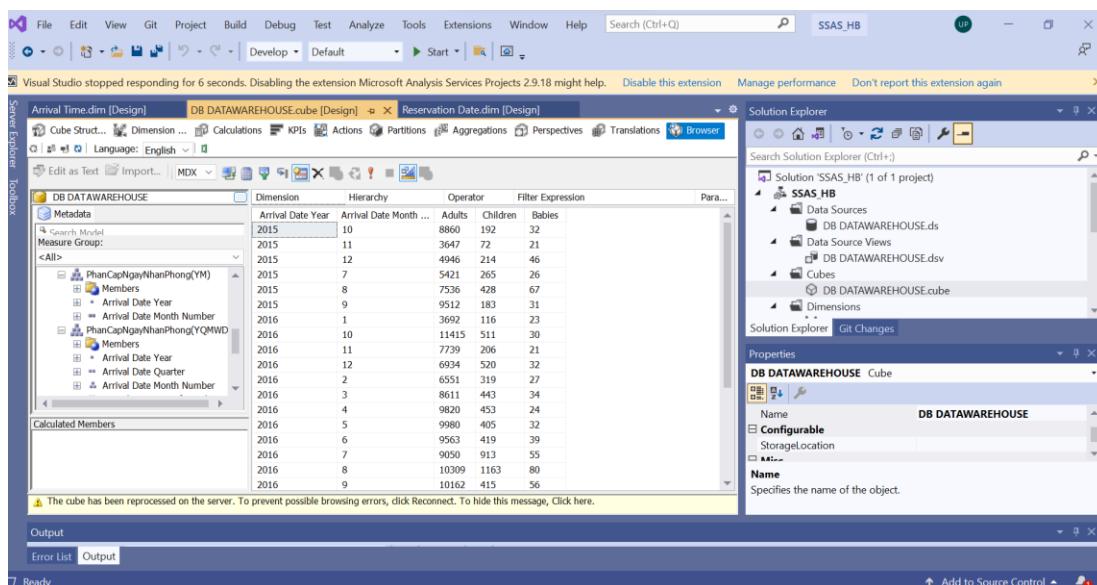
- Hàm Non Empty được sử dụng để loại bỏ các dữ liệu có giá trị null.

**9.2. Thống kê có bao nhiêu khách hàng là người lớn, bao nhiêu khách hàng là trẻ em và bao nhiêu khách hàng là em bé trong từng tháng của từng năm.**

*Ý nghĩa câu truy vấn:* Cho biết số lượng của từng nhóm khách hàng vào các thời điểm trong năm từ đó giúp khách sạn đưa ra các kế hoạch kinh doanh phù hợp đối với từng nhóm khách hàng.

#### 9.2.1. Công cụ SSAS trên các khối Cube

Kéo các thuộc tính Arrival Date Year, Arrival Date Month trong bảng Arrival Time, độ đo Adults, Children, Babies trong bảng Measures từ Measure Group sang cửa sổ lọc dữ liệu và cửa sổ thực thi. Ta thu được kết quả câu truy vấn



Hình 3.102. Kết quả câu truy vấn trong Visual Studio 2019

Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

### 9.2.2. Công cụ Pivot Excel

Kéo thả các độ đo Adults, Babies, Children từ bảng Measures vào bảng Values, Hierarchies PhanCapNgayNhanPhong(YM) thuộc bảng Arrival Time vào bảng Rows ta được kết quả:

Hình 3.103. Kết quả khi dung công cụ Pivot Excel

### 9.2.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio

18 ta được:

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "MDXQuery3.mdx - PHUONGUYEN\PHUONGUYEN.SSAS\_HB (PHUONGUYEN\phuong)" - Microsoft SQL Server Management Studio. The left pane is the Object Explorer, showing a connection to "SSAS\_HB" containing databases like "DB DATAREPOSITORY" and "DB DATAREPOSITORY". The right pane displays the results of an MDX query:

```
Select {[Measures].[Adults], [Measures].[Babies], [Measures].[Children]} on Columns,
NON EMPTY {[Arrival Time].[Arrival Date Year].Children
            ,[Arrival Time].[Arrival Date Month Number].Children} on Rows
From [DB DATAREPOSITORY];
```

The results grid has columns "Adults", "Babies", and "Children". The data is as follows:

	Adults	Babies	Children
2015	8860	32	192
2015	3647	21	72
2015	4946	46	214
2015	5421	26	265
2015	7536	67	428
2015	6512	31	183
2016	3692	23	116
2016	11415	30	511
2016	7739	21	206
2016	6934	32	520
2016	6551	27	319
2016	9611	34	443
2016	9820	24	453
2016	9980	32	405
2016	9563	39	419
2016	9050	55	913
2016	10309	80	1163
2016	10162	56	415
2017	6197	25	336
2017	7439	32	464

At the bottom, a status bar shows "Query executed successfully." and the session details "PHUONGUYEN\PHUONGUYEN | PHUONGUYEN\phuong | SSAS\_HB | 00:00:01".

Hình 3.104. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

**❖ Giải thích câu truy vấn MDX:**

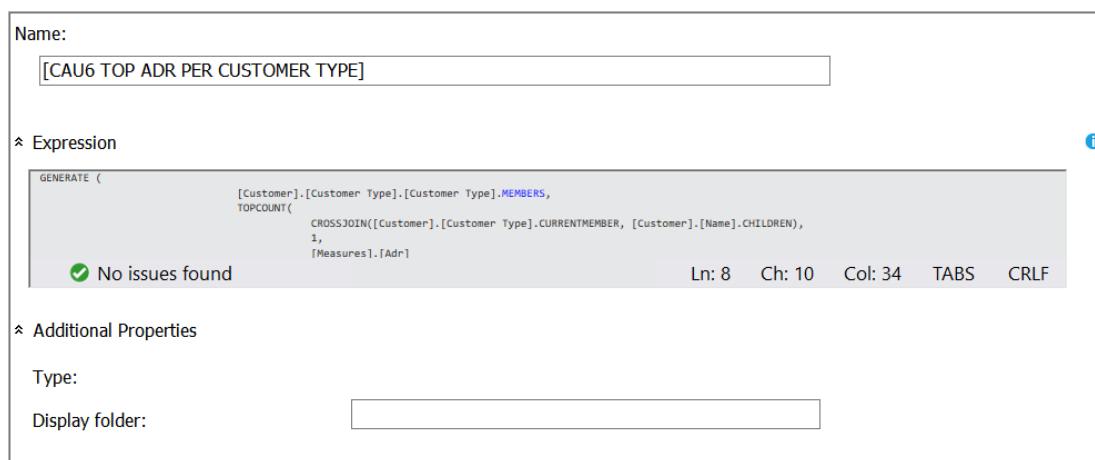
- Hàm Non Empty được sử dụng để loại bỏ các dữ liệu null.
- Operator “\*” có chức năng như hàm CrossJoin, được dùng để kết các thuộc tính của các bảng chiều.

**9.3. Liệt kê khách hàng của từng loại khách hàng có doanh số cao nhất trong năm 2017.**

*Ý nghĩa câu truy vấn:* Cho biết các khách hàng tiềm năng của khách sạn từ đó đưa ra các chương trình khuyến mãi đặc biệt nhằm tri ân khách hàng.

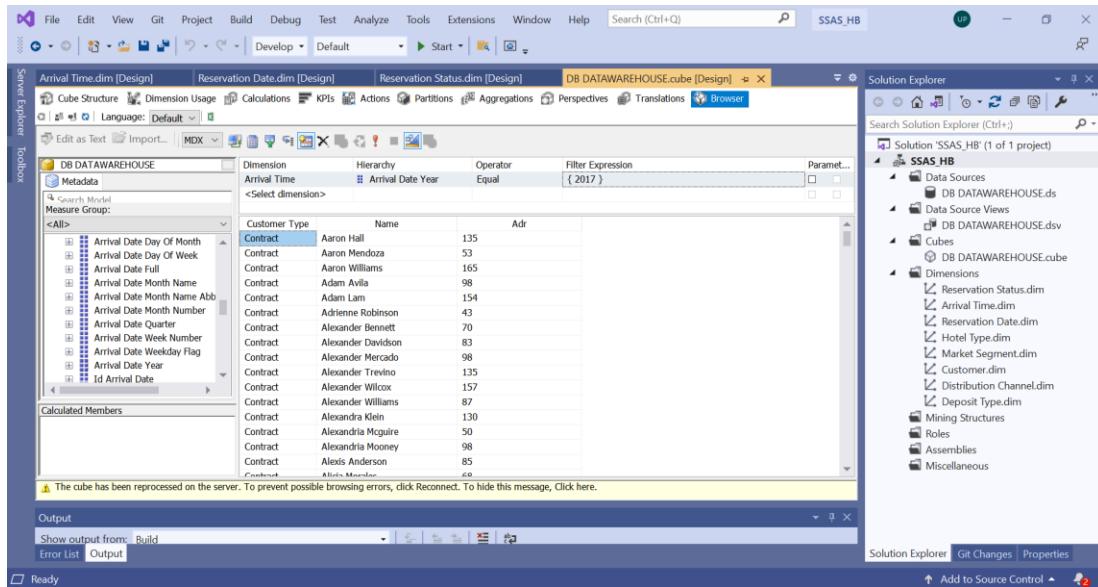
*9.3.1. Công cụ SSAS trên các khối Cube*

Tạo Named Set có tên là [CAU6 TOP ADR PER CUSTOMER TYPE] dùng để lấy thông tin khách hàng trong từng loại khách hàng đem đến doanh thu cao nhất cho khách sạn



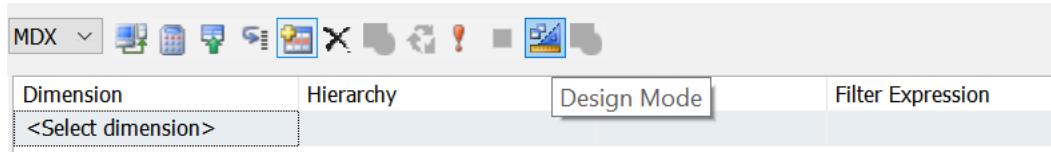
*Hình 3.105. Tạo Named Set*

Kéo thuộc tính Customer Type và Name trong bảng Customer và độ đo Adr trong bảng Measures từ Measure Group sang cửa sổ lọc dữ liệu và cửa sổ thực thi. Ta thu được kết quả câu truy vấn



*Hình 3.106. Kết quả câu truy vấn trong Visual Studio 2019*

Chuyển từ Design Mode sang Script Mode bằng các nhấp chuột vào biểu tượng Design Mode như hình minh họa:



*Hình 3.107. Design Mode trên thanh công cụ*

```

SELECT NON EMPTY {
    [Measures].[Adr]
} ON COLUMNS,
NON EMPTY {
    (
        [Customer].[Customer Type].[Customer Type].ALLMEMBERS
        [Customer].[Name].[Name].ALLMEMBERS
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE]
WHERE (
    [Arrival Time].[Arrival Date Year].&[2017]
)
    CELL_PROPERTIES VALUE,
    BACK_COLOR,
    FORE_COLOR,
    FORMATTED_VALUE,
    FORMAT_STRING,
    FONT_NAME,
    FONT_SIZE,
    FONT_FLAGS

```

*Hình 3.108. Câu lệnh truy vấn ban đầu*

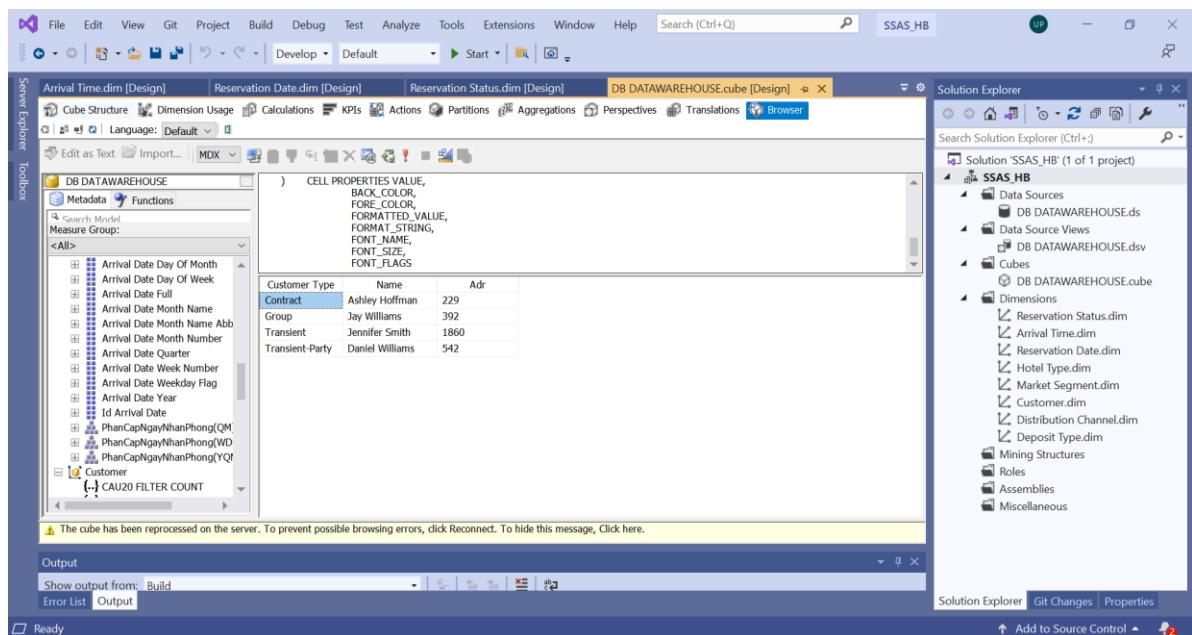
```

SELECT NON EMPTY {
    [Measures].[Adr]
} ON COLUMNS,
NON EMPTY {
    GENERATE (
        [Customer].[Customer Type].[Customer Type].MEMBERS,
        TOPCOUNT(
            CROSSJOIN(
                [Customer].[Customer Type].CURRENTMEMBER,
                [Customer].[Name].CHILDREN
            ),
            1,
            [Measures].[Adr]
        )
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE]
WHERE (
    [Arrival Time].[Arrival Date Year].&[2017]
)
    CELL_PROPERTIES VALUE,
    BACK_COLOR,
    FORE_COLOR,
    FORMATTED_VALUE,
    FORMAT_STRING,
    FONT_NAME,
    FONT_SIZE,
    FONT_FLAGS

```

*Hình 3.109. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.110. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.3.2. Công cụ Pivot Excel

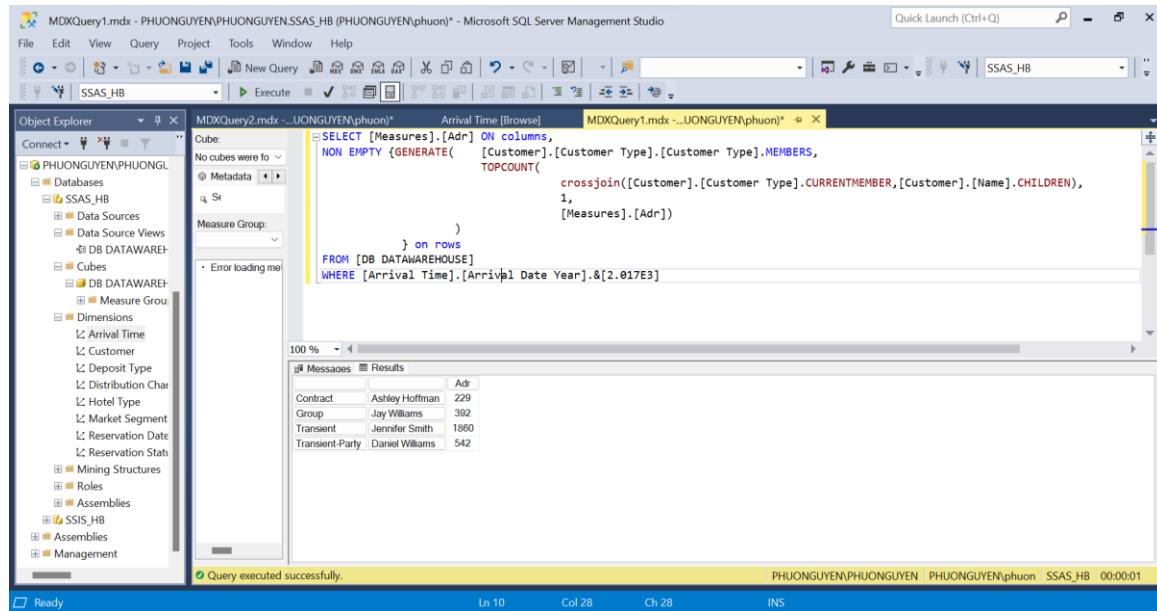
Kéo thả Adr từ bảng Measures vào bảng Values, Named Set CAU6 TOP ADR PER CUSTOMER TYPE của bảng Measures vào bảng Rows và sử dụng thuộc tính Arrival Date Year để lọc ra dữ liệu của năm 2017 ta được kết quả:

Customer Type	Name	Adr
Contract	Ashley Hoffman	229
Group	Jay Williams	392
Transient	Jennifer Smith	1860
Transient-Party	Daniel Williams	542

Hình 3.111. Kết quả khi dùng công cụ Pivot Excel

### 9.3.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:



The screenshot shows the Microsoft SQL Server Management Studio interface. In the center, there is a query results window titled "MDXQuery1.mdx - PHUONGUYEN\PHUONGUYEN.SSAS\_HB (PHUONGUYEN\phuong)". The query results are displayed in a table with columns "Contract", "Group", "Transient", and "Transient-Party". The data is as follows:

	Contract	Group	Transient	Transient-Party
1	Ashley Hoffman	Jay Williams	Jennifer Smith	Daniel Williams
2	229	392	1860	

Below the table, a message says "Query executed successfully." At the bottom of the screen, the status bar shows "PHUONGUYEN\PHUONGUYEN PHUONGUYEN\phuong SSAS\_HB 00:00:01".

Hình 3.112. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích câu truy vấn MDX:

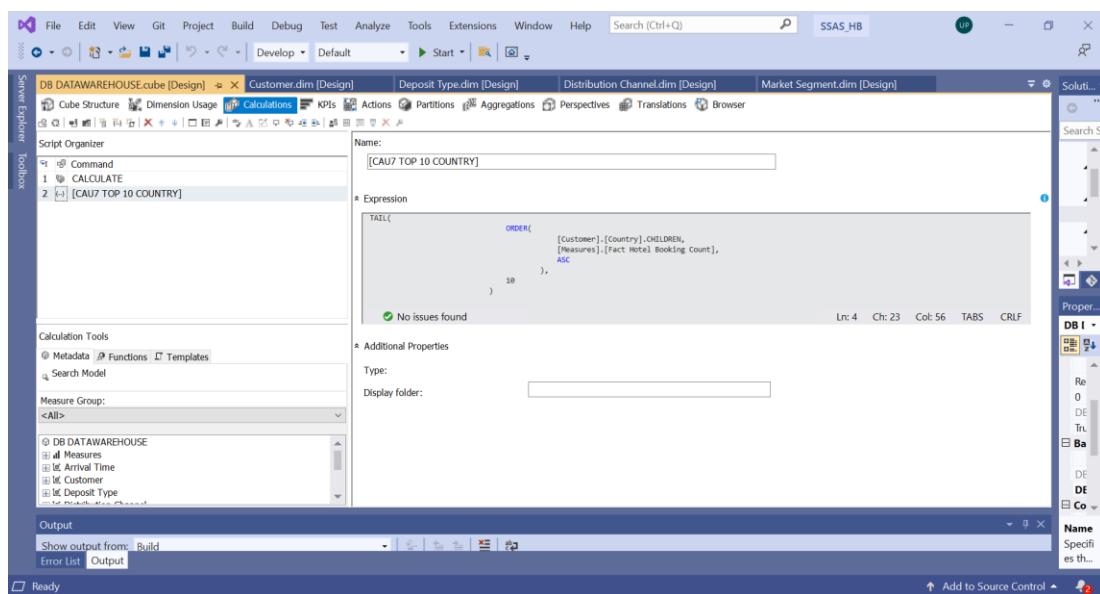
- Hàm CrossJoin dùng để tích chéo hai cột thuộc tính Customer Type (thể loại khách hàng) và Name (tên khách hàng).
- Hàm TopCount dùng để sắp xếp các dữ liệu của kết quả sau khi tích chéo hai thuộc tính bằng hàm CrossJoin theo thứ tự sắp xếp độ đo Adr (doanh thu) giảm dần và trả về dòng đầu tiên có giá trị cao nhất.
- Hàm Generate dùng để duyệt từng giá trị trong Customer Type ở đối số thứ nhất. Tại đối số thứ hai, hàm Generate lấy ra từng loại khách hàng hiện tại thông qua CurrentMember và dùng hàm TopCount để lấy thông tin khách hàng chi tiêu cao nhất.
- Where dùng để lọc ra toàn bộ dữ liệu chính trong năm 2017.

#### 9.4. Cho biết top 10 quốc gia có lượng khách du lịch cao nhất trong năm 2016, sắp xếp các giá trị theo thứ tự tăng dần.

Ý nghĩa câu truy vấn: Thể hiện các quốc gia thu hút khách du lịch nhất để có thể có những quyết định kinh doanh hoặc đầu tư phù hợp.

##### 9.4.1. Công cụ SSAS trên các khối Cube

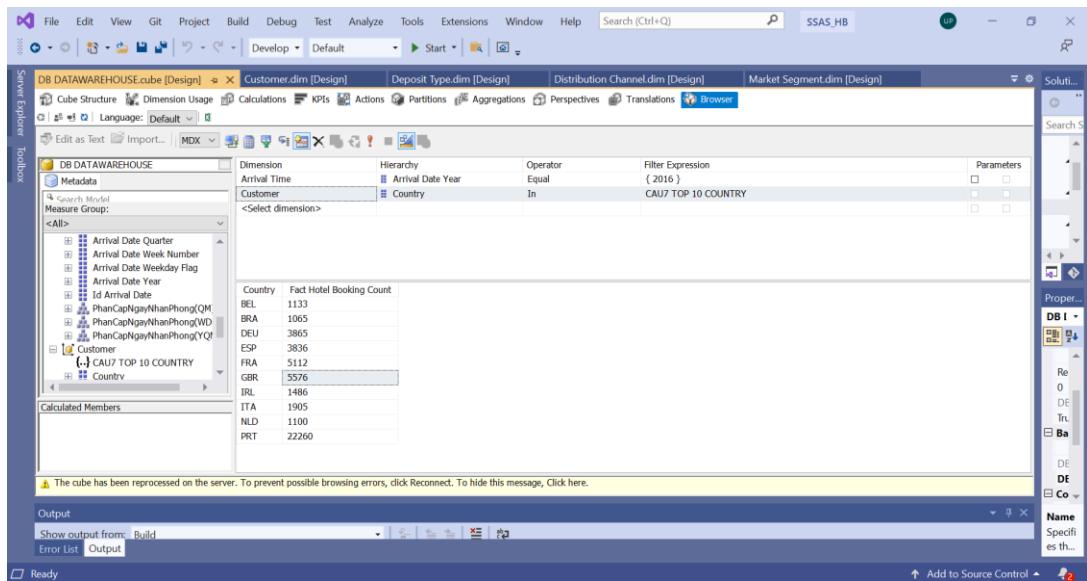
Tạo Named Set có tên là [CAU7 TOP 10 COUNTRY] dùng để lấy thông 10 quốc gia đầu tiên có lượng khách du lịch cao nhất trong năm 2016 và sắp xếp các giá trị theo thứ tự tăng dần



Hình 3.113.Tạo Named Set

Kéo thuộc tính Country trong bảng Customer và độ đo Fact Hotel Booking Count trong bảng Measures từ Measure Group sang cửa sổ thực thi và Named Set [CAU7 TOP 10 COUNTRY], thuộc tính Arrival Date Year với giá trị lọc là 2016 từ bảng chi tiêu Arrival Time sang cửa sổ lọc dữ liệu. Ta được kết quả câu truy vấn

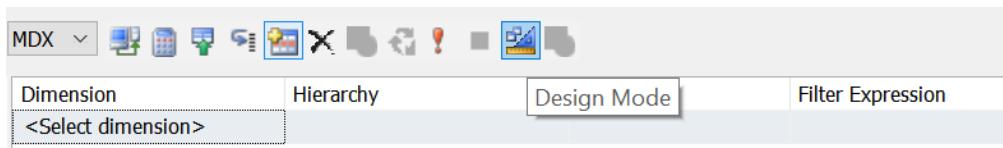
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.114. Kết quả câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng các nhấp chuột vào biểu tượng

Design Mode như hình minh họa:



Hình 3.115. Design Mode trên thanh công cụ

```

SELECT NON EMPTY {
    [Measures].[Fact Hotel Booking Count]
} ON COLUMNS,
NON EMPTY {
    {[Customer].[Country].[Country].ALLMEMBERS }
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM (
    SELECT (
        [CAU7 TOP 10 COUNTRY]
    ) ON COLUMNS
    FROM (
        SELECT (
            { [Arrival Time].[Arrival Date Year].&[2016] }
        ) ON COLUMNS
        FROM [DB DATAWAREHOUSE]
    )
    WHERE (
        [Arrival Time].[Arrival Date Year].&[2016]
    )
    CELL PROPERTIES VALUE,
    BACK_COLOR,
    FORE_COLOR,
    FORMATTED_VALUE,
    FORMAT_STRING,
    FONT_NAME,
    FONT_SIZE,
    FONT_FLAGS
)

```

Hình 3.116. Câu lệnh truy vấn ban đầu

```

SELECT  {
            [Measures].[Fact Hotel Booking Count]
        } ON COLUMNS,
NON EMPTY { 
            [CAU7 TOP 10 COUNTRY]
        } ON ROWS
FROM [DB DATAWAREHOUSE]
WHERE [Arrival Time].[Arrival Date Year].&[2.016e3]
    
```

*Hình 3.117. Thay thế hàm vào câu lệnh truy vấn*

Country	Fact Hotel Booking Count
BRA	1065
NLD	1100
BEL	1133
IRL	1486
ITA	1905
ESP	3836
DEU	3865
FRA	5112
GBR	5576
PRT	22260

*Hình 3.118. Kết quả câu truy vấn trong Visual Studio 2019*

#### 9.4.2. Công cụ Pivot Excel

Kéo thả độ đo Fact Hotel Booking từ bảng Measures vào bảng Values, thuộc tính CAU7 TOP 10 COUNTRY thuộc bảng Arrival Time vào bảng Rows và thuộc tính thuộc tính CAU7 TOP 10 COUNTRY Arrival Date Year thuộc bảng Arrival Time lọc lấy các dòng dữ liệu có giá trị của thuộc tính năm nhận phòng là 2016 trong trường Filters ta được kết quả:

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

Arrival Date Year	Fact Hotel Booking Count
BRA	1065
NLD	1100
BEL	1133
IRL	1486
ITA	1905
ESP	3836
DEU	3865
FRA	5112
GBR	5576
PRT	22260

Hình 3.119. Kết quả khi dùng công cụ Pivot Excel

### 9.4.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

Fact Hotel Booking Count	
BRA	1065
NLD	1100
BEL	1133
IRL	1486
ITA	1905
ESP	3836
DEU	3865
FRA	5112
GBR	5576
PRT	22260

Hình 3.120. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

### ❖ Giải thích câu truy vấn MDX:

- Hàm Order dùng để sắp xếp thứ tự Country (quốc tịch) dựa trên số lượng giao dịch đặt phòng của từng quốc gia theo thứ tự tăng dần.

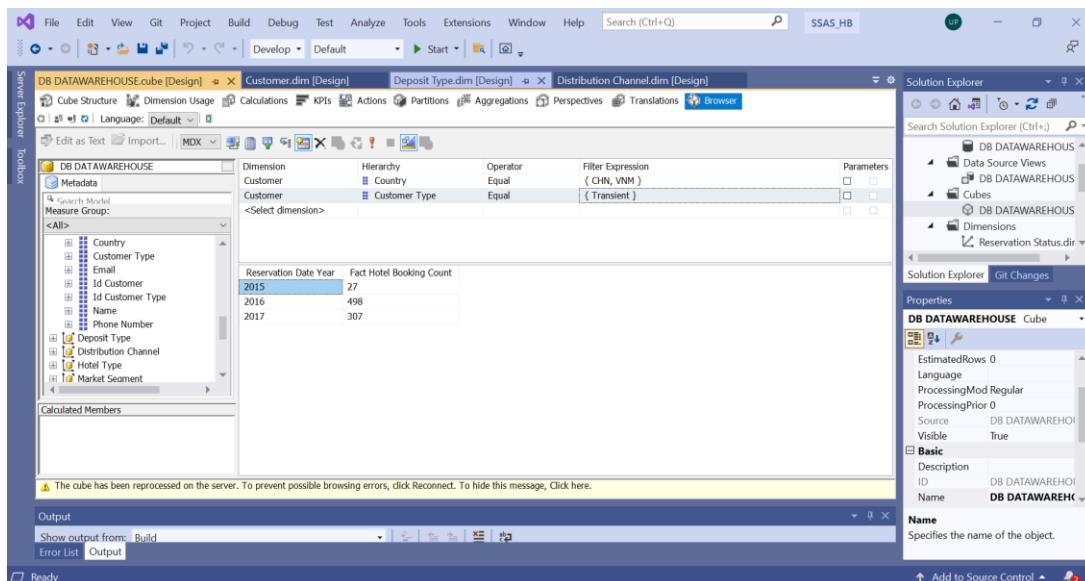
- Hàm Tail để lấy 10 dòng cuối của tập dữ liệu đã lọc trong Order.
- Hàm Non Empty để loại bỏ các giá trị null.

## 9.5. Liệt kê số lượng đơn đặt phòng ở ‘China’ và ‘Việt Nam’ theo từng năm và loại khách hàng là ‘Transient’

*Ý nghĩa câu truy vấn:* Cho biết tiềm năng của nhóm khách hàng là “Transient” ở hai quốc gia là Trung Quốc và Việt Nam.

### 9.5.1. Công cụ SSAS trên các khối Cube

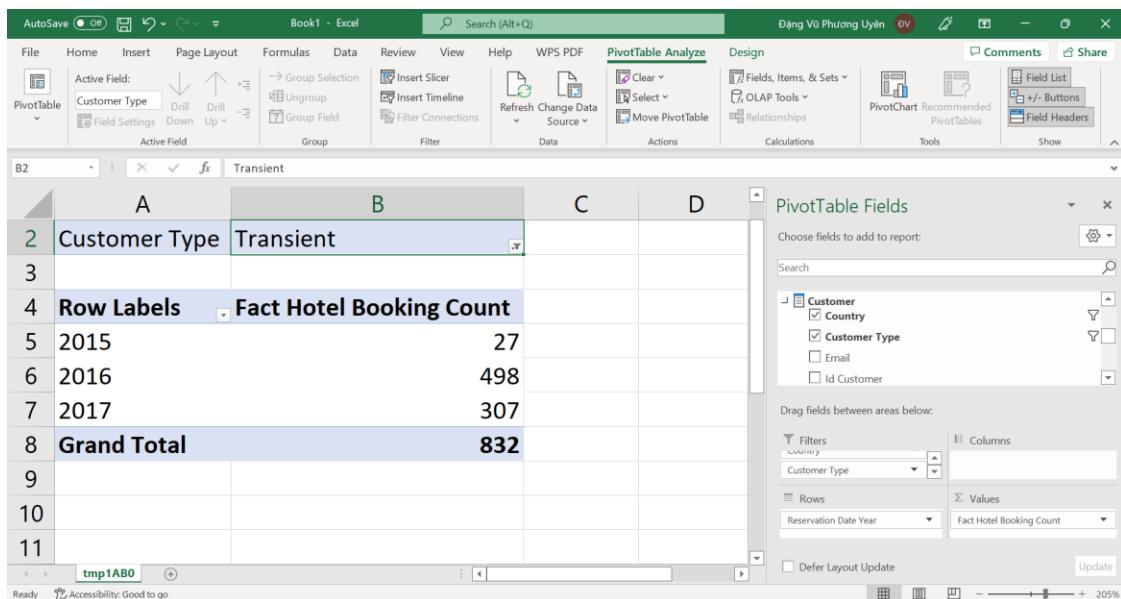
Kéo thuộc tính Hotel Type trong bảng Hotel type, thuộc tính Arrival Date year trong bảng Arrival Time, Adr và Fact Hotel Booking Count trong bảng Measures từ Measure Group sang cửa sổ lọc dữ liệu và cửa sổ thực thi. Ta thu được kết quả câu truy vấn



Hình 3.121. Kết quả câu truy vấn trong Visual Studio 2019

### 9.5.2. Công cụ Pivot Excel

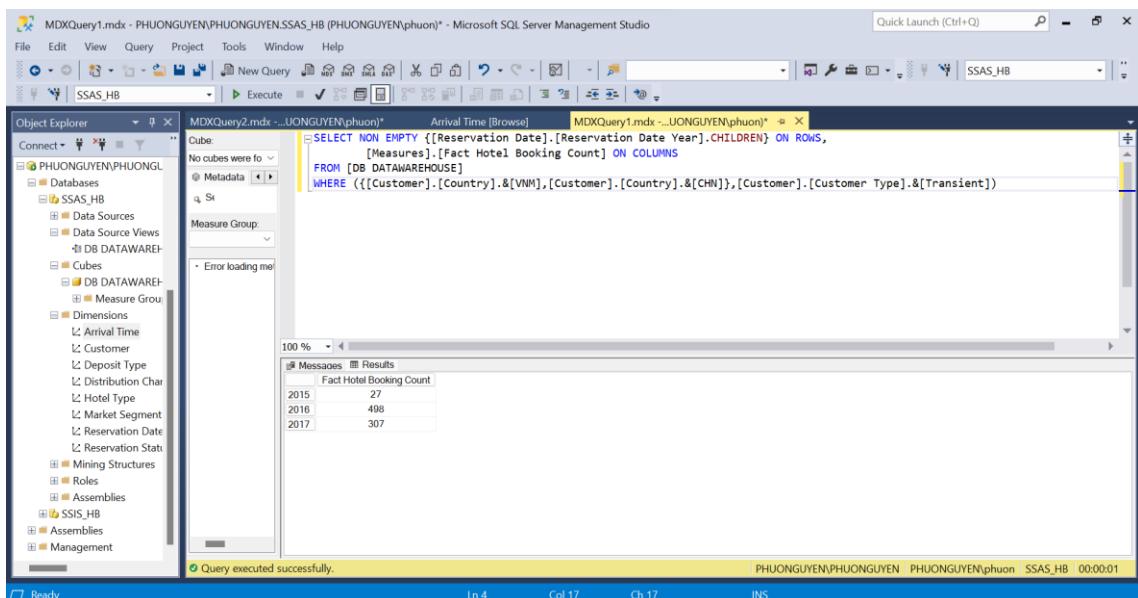
Kéo thả Fact Hotel Booking từ bảng Measures vào bảng Values, thuộc tính Reservation Date Year thuộc bảng Reservation Date vào bảng Rows. Sau đó, lần lượt lọc loại khách hàng là “Transient” và tên quốc gia là “CHN” và “VNM” thông qua Filters, ta được kết quả:



Hình 3.122. Kết quả khi dùng công cụ Pivot Excel

### 9.5.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:



Hình 3.123. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

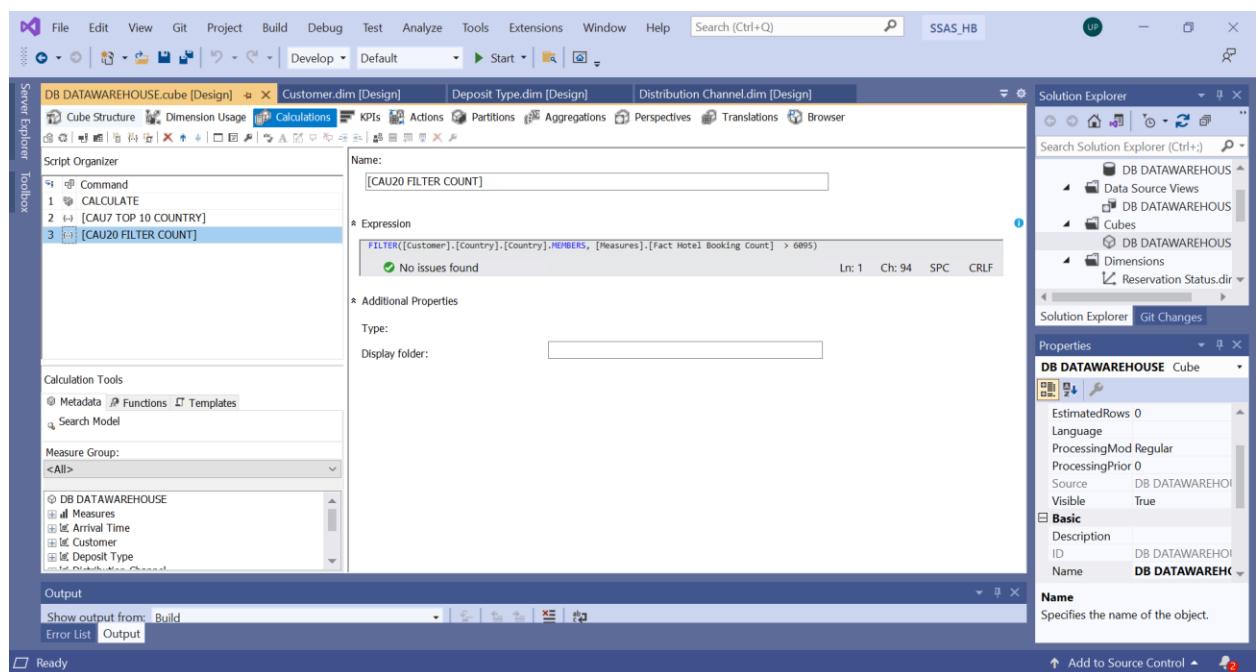
- Dùng Where để lọc các dòng dữ liệu có quốc tịch khách hàng là VietNam (VNM) hoặc China (CHN) thuộc loại khách hàng là Transient.

## 9.6. Tìm các quốc gia có số lượng đơn đặt phòng lớn hơn lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn.

*Ý nghĩa câu truy vấn:* Thống kê các quốc gia có số lượng đặt phòng lớn nhằm đánh giá tiềm năng kinh doanh của quốc gia đó.

### 9.6.1. Công cụ SSAS trên các khối Cube

Tạo Named Set có tên là [CAU20 FILTER COUNT] dùng để lấy thông các quốc gia có số lượng đơn đặt phòng lớn hơn lượng đặt phòng lớn hơn tổng số lượng đặt phòng trung bình của hệ thống khách sạn.



Hình 3.124.Tạo Named Set

Kéo thuộc tính Country trong bảng Customer và độ đo Fact Hotel Booking Count trong bảng Measures từ Measure Group sang cửa sổ thực thi và Named Set [CAU20 FILTER COUNT] từ bảng chiều Arrival Time sang cửa sổ lọc dữ liệu. Ta thu được kết quả câu truy vấn

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

The screenshot shows the Microsoft Visual Studio 2019 interface for Analysis Services (SSAS). The main area displays a pivot table titled "Fact Hotel Booking Count" with data grouped by "Country". The table shows the following data:

Country	Fact Hotel Booking Count
DEU	7286
ESP	8271
FRA	10324
GBR	12109
PRT	48370

The cube browser at the top shows dimensions like "Customer.dim [Design]" and "Deposit Type.dim [Design]". The Solution Explorer and Properties windows are also visible on the right.

Hình 3.125. Kết quả câu truy vấn trong Visual Studio 2019

### 9.6.2. Công cụ Pivot Excel

Kéo thả độ đo Fact Hotel Booking từ bảng Measures vào bảng Values, Named Set [CAU20 FILTER COUNT] thuộc bảng Arrival Time vào bảng Rows ta được kết quả:

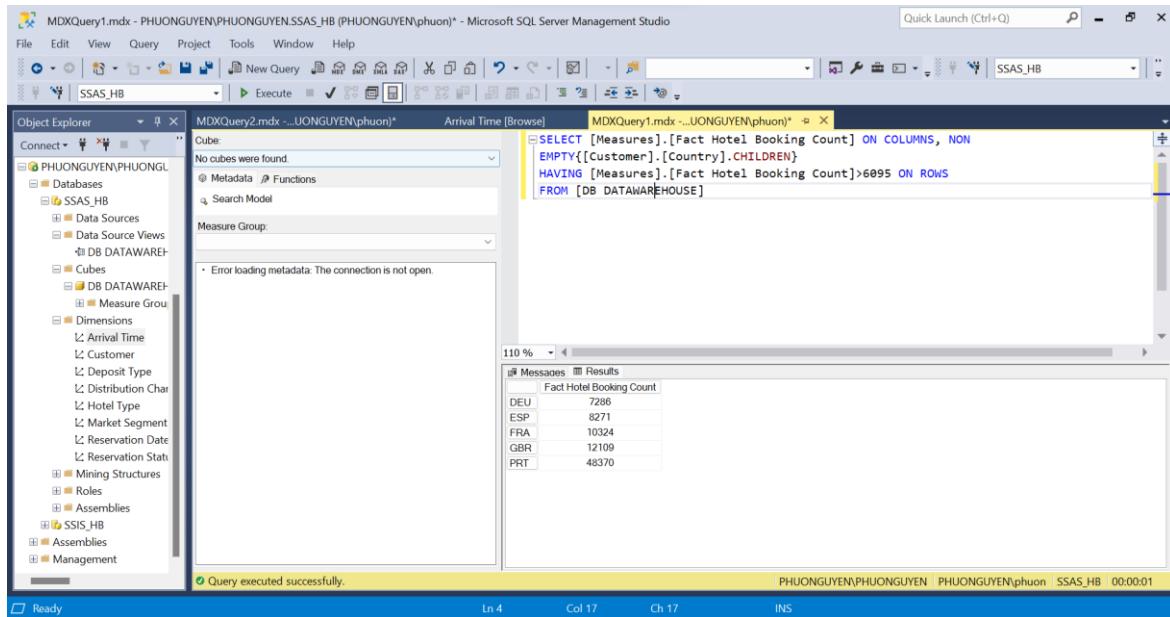
The screenshot shows a Microsoft Excel spreadsheet with a PivotTable report. The PivotTable Fields pane on the right shows the "Fact Hotel Booking" measure and its filters. The main table shows the following data:

Row Labels	Fact Hotel Booking Count
DEU	7286
ESP	8271
<b>FRA</b>	<b>10324</b>
GBR	12109
PRT	48370

Hình 3.126. Kết quả khi dùng công cụ Pivot Excel

### 9.6.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:



Hình 3.127. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích câu truy vấn MDX:

- Hàm Having có tác dụng như hàm Filter dùng để lọc tổng số giao dịch của hệ thống khách sạn theo từng quốc tịch của khách hàng trong điều kiện lớn hơn 6095 giao dịch.

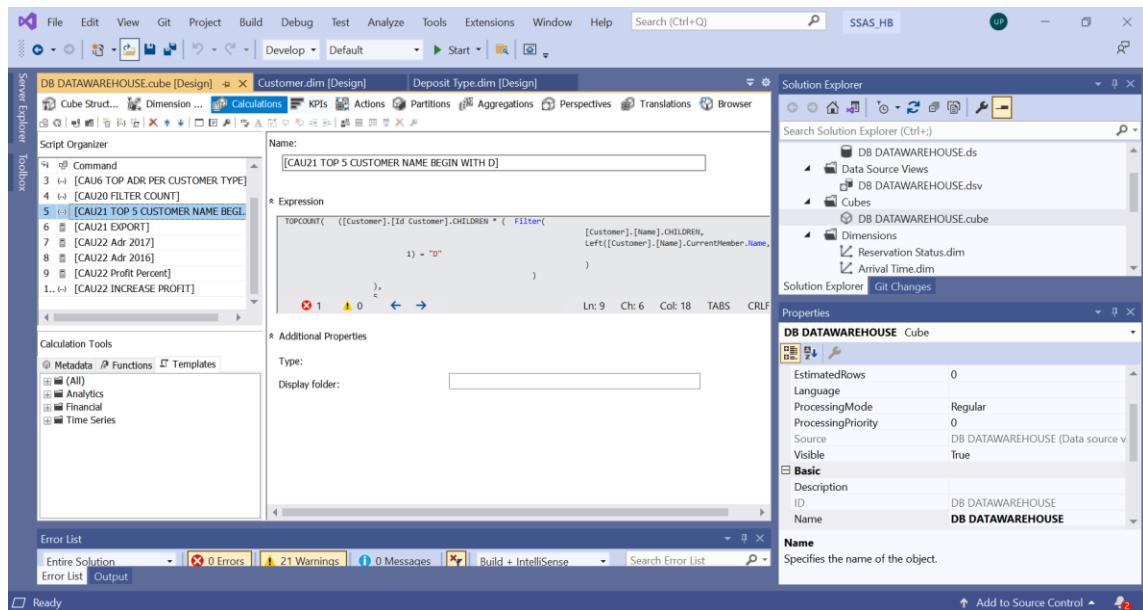
**9.7. Liệt kê 5 khách hàng có tên bắt đầu bằng chữ D ở Bồ Đào Nha (PRT), nếu khách hàng có tổng chi tiêu > 500 và là khách hàng quay lại thì xuất "Khách hàng thân thiết". Nếu tổng chi tiêu > 500 và là khách hàng lần đầu tiên tới thì xuất "Khách hàng tiềm năng". Ngược lại xuất "Khách hàng vãng lai."**

Ý nghĩa câu truy vấn: Giúp phân loại khách hàng nhằm đưa ra các khuyến mãi tri ân khách hàng phù hợp.

#### 9.7.1. Công cụ SSAS trên các khối Cube

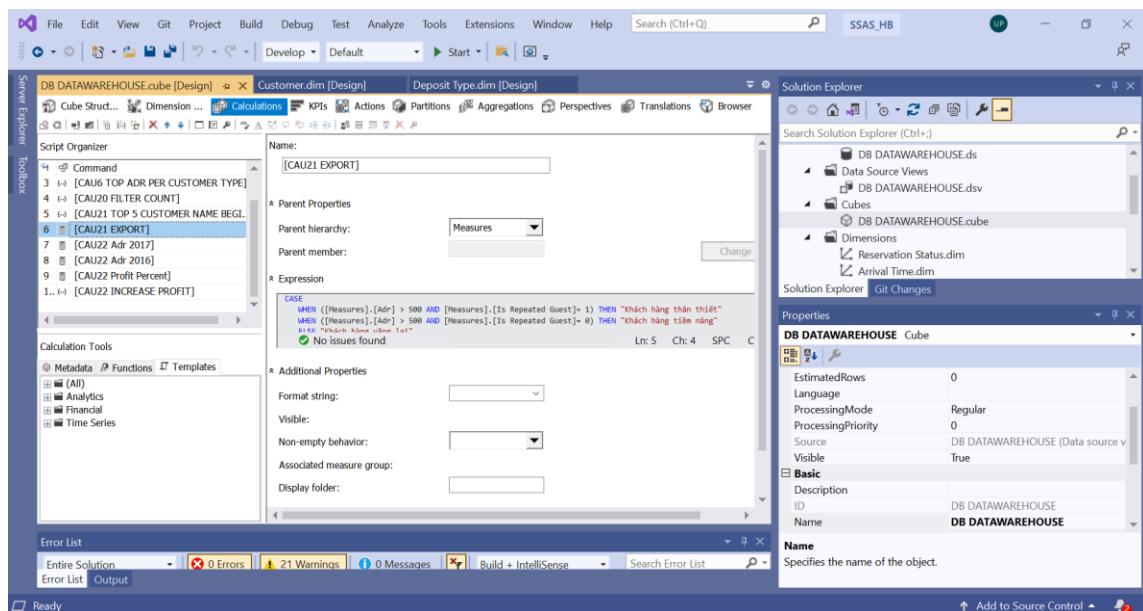
Tạo Named Set có tên là [CAU21 TOP 5 CUSTOMER NAME BEGIN WITH D] dùng để lấy thông 5 khách hàng có tên bắt đầu bằng chữ D ở Bồ Đào Nha (PRT)

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



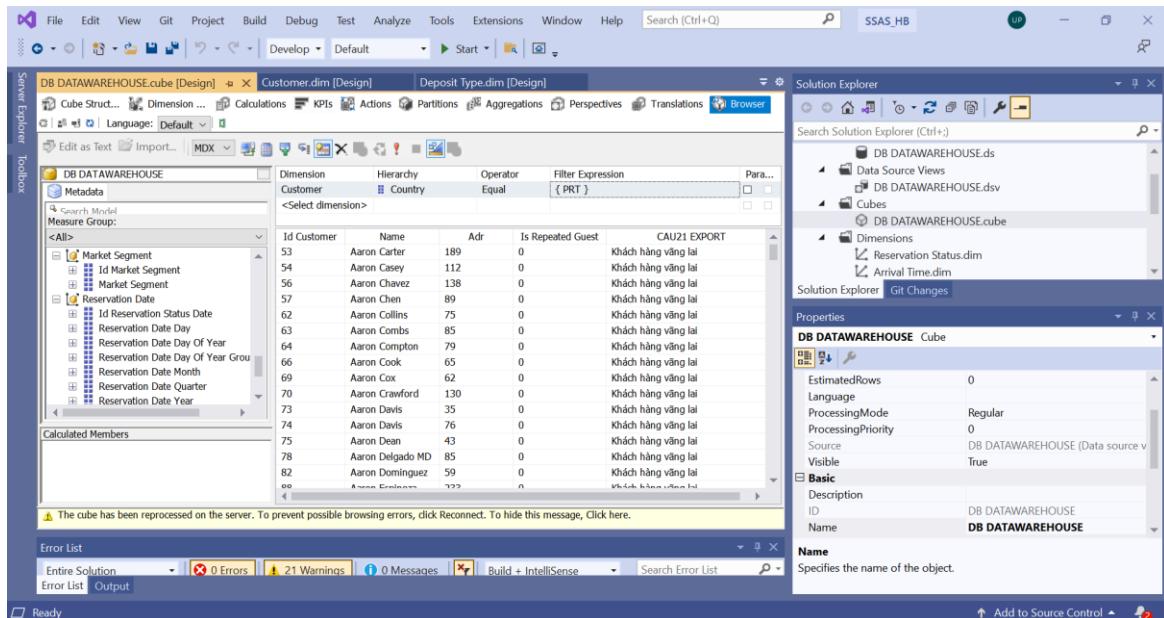
Hình 3.128. Tạo Named Set

Tạo Calculated Member có tên là [CAU21 EXPORT] dùng để xuất thông tin đánh giá loại khách hàng thông qua tổng chi tiêu



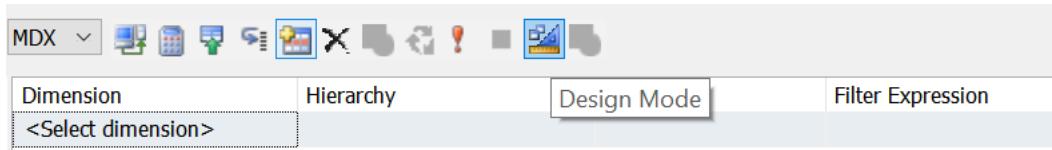
Hình 3.129. Tạo Calculated Member

Kéo lần lượt các thuộc tính Id Customer, Name trong bảng Customer và độ đo Adr, Is Repeated Guest, CAU21 EXPORT trong bảng Measures từ Measure Group sang cửa sổ thực thi và thuộc tính Country với giá trị lọc là PRT từ bảng chiêu Customer sang cửa sổ lọc dữ liệu. Ta thu được kết quả câu truy vấn



Hình 3.130. Kết quả câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng cách nhấp chuột vào biểu tượng Design Mode như hình minh họa:



Hình 3.130. Design Mode trên thanh công cụ

```

SELECT NON EMPTY {
    [Measures].[Adr],
    [Measures].[Is Repeated Guest],
    [Measures].[CAU21 EXPORT]
} ON COLUMNS,
NON EMPTY {
    ([Customer].[Id Customer].[Id Customer].ALLMEMBERS *
    [Customer].[Name].[Name].ALLMEMBERS )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM (
    SELECT (
        { [Customer].[Country].&[PRT] }
    ) ON COLUMNS
    FROM [DB DATAWAREHOUSE]
)
WHERE ( [Customer].[Country].&[PRT] ) CELL PROPERTIES VALUE,
      BACK_COLOR,
      FORE_COLOR,
      FORMATTED_VALUE,
      FORMAT_STRING,
      FONT_NAME,
      FONT_SIZE,
      FONT_FLAGS

```

*Hình 3.131. Câu lệnh truy vấn ban đầu*

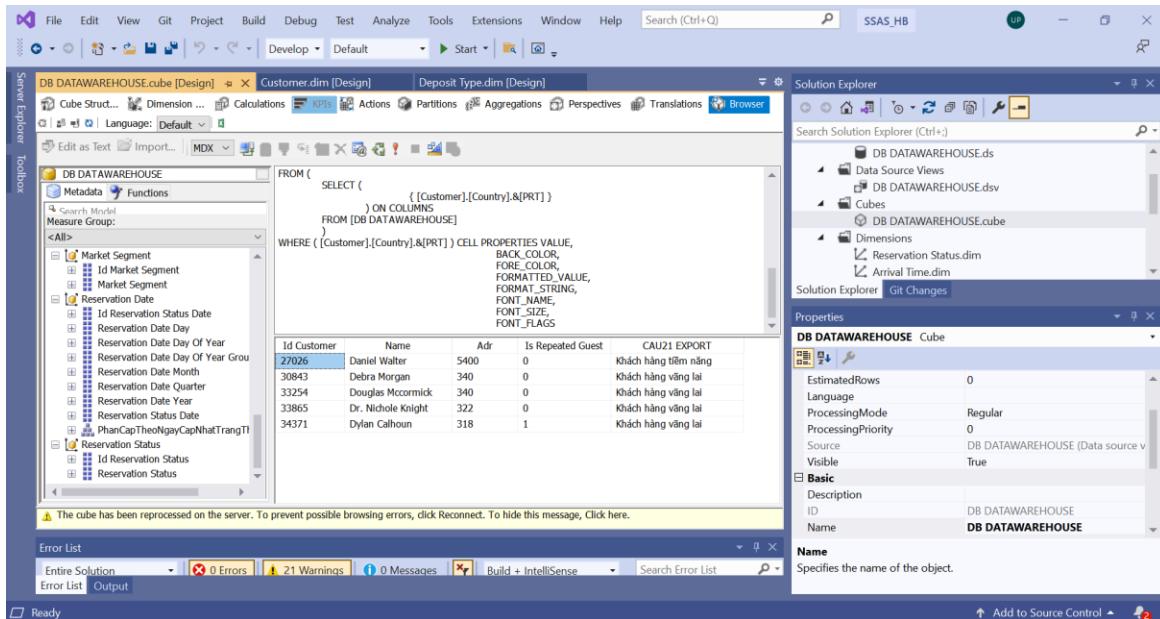
```

SELECT NON EMPTY {
    [Measures].[Adr],
    [Measures].[Is Repeated Guest],
    [Measures].[CAU21 EXPORT]
} ON COLUMNS,
NON EMPTY {
    TOPCOUNT(
        (
            [Customer].[Id Customer].CHILDREN *
            (
                Filter(
                    [Customer].[Name].CHILDREN,
                    Left([Customer].[Name].CurrentMember.Name, 1) = "D"
                )
            ),
            5,
            [Measures].[Adr]
        )
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM (
    SELECT (
        { [Customer].[Country].&[PRT] }
    ) ON COLUMNS
    FROM [DB DATAWAREHOUSE]
)
WHERE ( [Customer].[Country].&[PRT] ) CELL PROPERTIES VALUE,
      BACK_COLOR,
      FORE_COLOR,
      FORMATTED_VALUE,
      FORMAT_STRING,
      FONT_NAME,
      FONT_SIZE,
      FONT_FLAGS

```

*Hình 3.132. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.133. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.7.2. Công cụ Pivot Excel

Kéo thả Adr, Is Repeated Guest, CAU21 EXPORT từ bảng Measures vào bảng Values và Named Set CAU21 TOP 5 CUSTOMER NAME BEGIN WITH D thuộc bảng Measures vào bảng Rows. Tiếp đến, ta lọc dữ liệu qua trường Filters với điều kiện là quốc giá của khách hàng là Bồ Đào Nha (PRT) ta được kết quả:

	A	B	C	D	E	F
1	Country	PRT				
3	Row Labels	CAU21 EXPORT	Adr	Is Repeated Guest		
4	27026					
5	Daniel Walter	Khách hàng tiềm năng	5400	0		
6	30843					
7	Debra Morgan	Khách hàng vãng lai	340	0		
8	33254					
9	Douglas McCormick	Khách hàng vãng lai	340	0		
10	33865					
11	Dr. Nichole Knight	Khách hàng vãng lai	322	0		
12	34371					
13	Dylan Calhoun	Khách hàng vãng lai	318	1		
14						

Hình 3.134. Kết quả khi dùng công cụ Pivot Excel

### 9.7.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

The screenshot shows the Microsoft SQL Server Management Studio interface. In the center, there is a query editor window titled "MDXQuery1.mdx - PHUONGUYEN\phuong\* - Microsoft SQL Server Management Studio". The query itself is written in MDX and retrieves data from a cube named "DE DATAWAREHOUSE". The results are displayed in a table titled "Messages" with columns: Adr, Is Repeated Guest, CaseResult. The results are as follows:

Adr	Is Repeated Guest	CaseResult		
27026	Daniel Waller	5400	0	Khách hàng tiềm năng
30843	Debra Morgan	340	0	Khách hàng vãng lai
33254	Douglas McCormick	340	0	Khách hàng vãng lai
33885	Dr. Nichole Knight	322	0	Khách hàng vãng lai
34371	Dylan Calhoun	318	1	Khách hàng vãng lai

Below the table, a message says "Query executed successfully.".

Hình 3.135. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

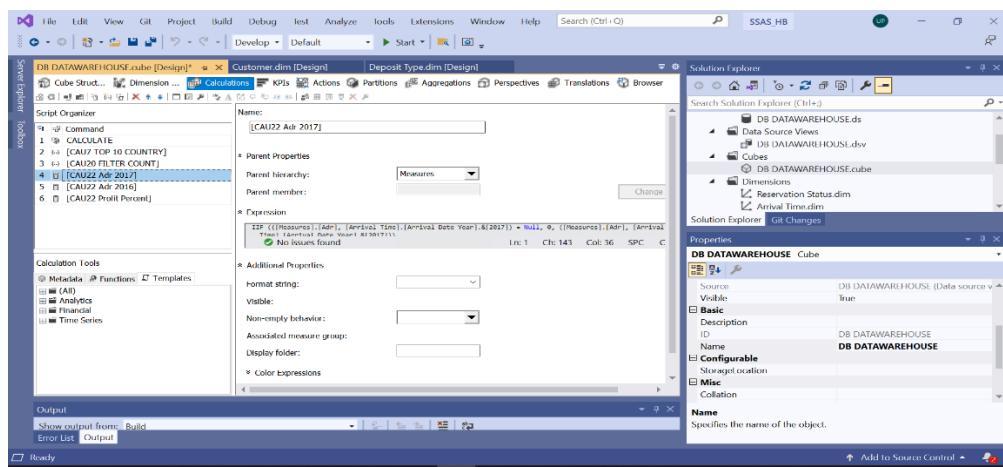
- Tạo độ đo CaseResult dùng để xuất “Khách hàng than thiết” đối với những khách hàng đã từng quay lại sử dụng giao dịch với hệ thống khách sạn và chi tiêu lớn hơn 500 \$. Và xuất “Khách hàng tiềm năng” đối với những khách hàng lần đầu sử dụng giao dịch đặt phòng nhưng có chi tiêu lớn hơn 500\$. Còn lại thì xuất “Khách hàng vãng lai”.
- Dùng hàm Filter để lọc ra khách hàng có tên bắt đầu bằng chữ “D”.
- Dùng hàm TopCount để sắp xếp tích chéo giữa mã khách hàng (Id Customer) và tên khách hàng đã được lọc thông qua hàm Filter và sắp xếp theo thứ tự giảm dần dựa trên độ đo chi tiêu khách hàng (Adr) và lấy 5 dòng đầu tiên của tập dữ liệu đó.
- Hàm Non-Empty dùng để loại bỏ các dữ liệu có giá trị null.

## 9.8. Liệt kê các quốc gia thu được lợi nhuận năm 2017 nhiều hơn năm 2016 và sắp xếp theo lợi nhuận giảm dần

*Ý nghĩa câu truy vấn:* Cung cấp góc nhìn tổng quan về độ thu hút khách hàng cũng như tiềm năng kinh doanh của các quốc gia.

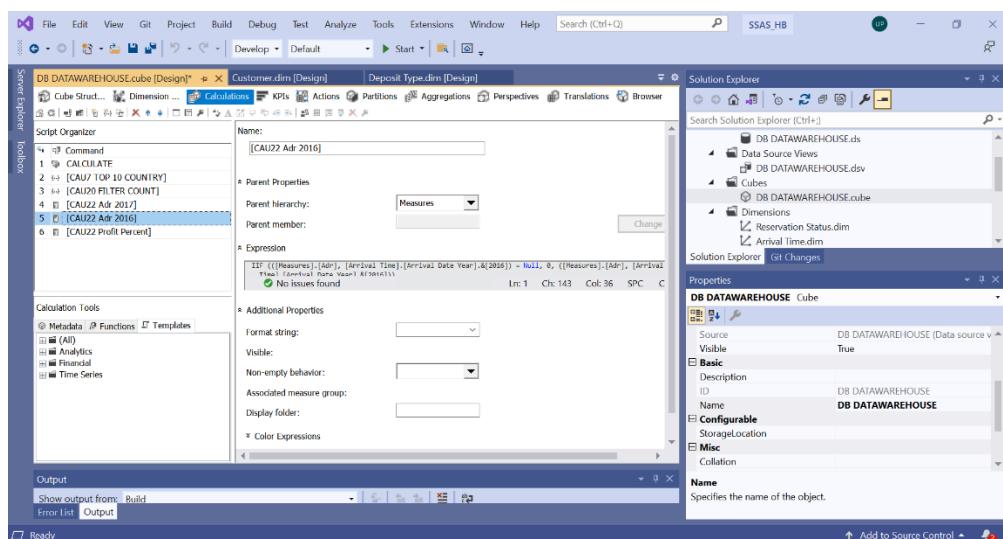
### 9.8.1. Công cụ SSAS trên các khối Cube

Tạo Calculated Member có tên là [CAU22 Adr 2017] dùng để chuyển các dữ liệu Null trong độ đo Adr năm 2017 thành 0.



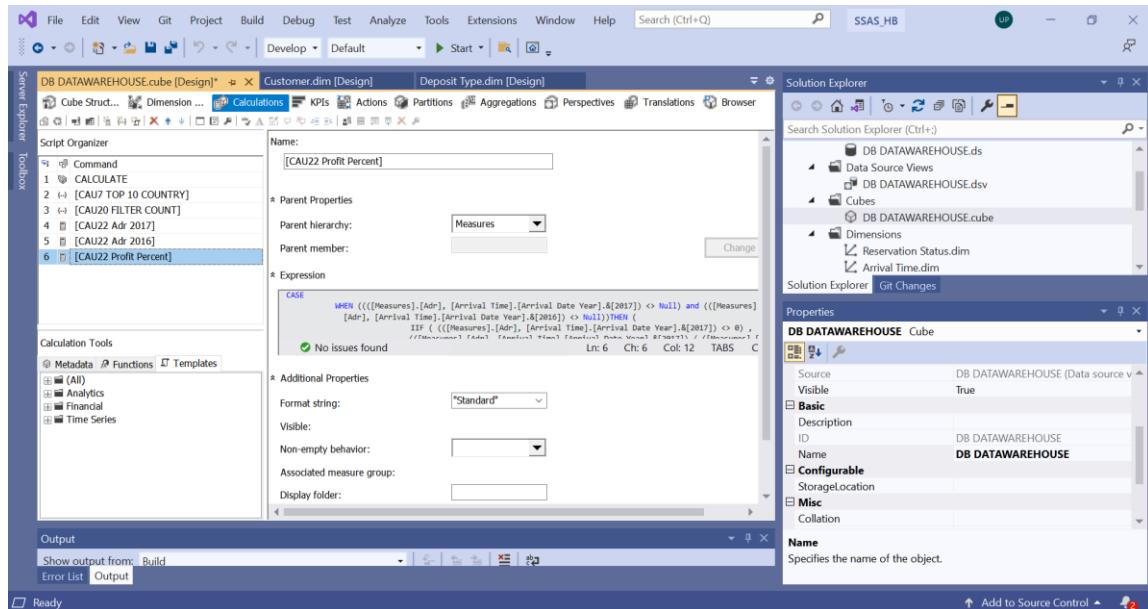
Hình 3.136. Tạo Calculated Member

Tạo Calculated Member có tên là [CAU22 Adr 2016] dùng để chuyển các dữ liệu Null trong độ đo Adr trong năm 2016 thành 0.



Hình 3.137. Tạo Calculated Member

Tạo Calculated Member tên là [CAU22 Profit Percent] dùng để tính toán tỷ lệ giữa độ đo [CAU22 Adr 2017] trên độ đo [CAU22 Adr 2016] trong bảng Measures.



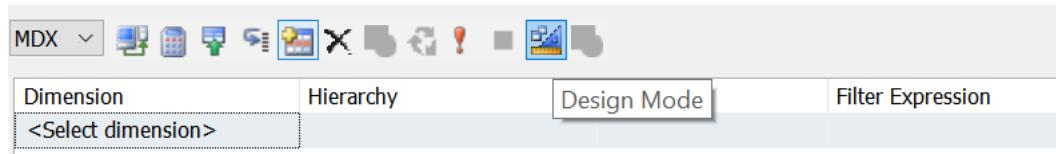
Hình 3.138. Tạo Caculated Member

Kéo lùn lượt các thuộc tính Country trong bảng Customer và các độ đo vừa tạo từ Measure Group sang cửa sổ thực thi để được kết quả câu truy vấn sau:

Country	CAU22 Adr 2016	CAU22 Adr 2017	CAU22 Profit Percent
ABW	158	99	0.6265823
AGO	21758	15295	0.7029598
AIA	265	0	0
ALB	486	353	0.7263374
AND	491	928	1.89002
ARE	4506	2409	0.5346205
ARG	12587	11312	0.898705
ARM	499	252	0.50501
ASM	0	72	0
ATA	176	93	0.5284091
ATF	0	126	0
AUS	19149	28137	1.469372
AUT	68535	57671	0.8414825
AZE	762	958	1.2577218
BDI	46	0	0
BEL	121077	124859	1.031236
BEN	125	244	1.952
BFA	80	0	0

Hình 3.139. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng cách nhấp chuột vào biểu tượng Design Mode như hình minh họa:



*Hình 3.140. Design Mode trên thanh công cụ*

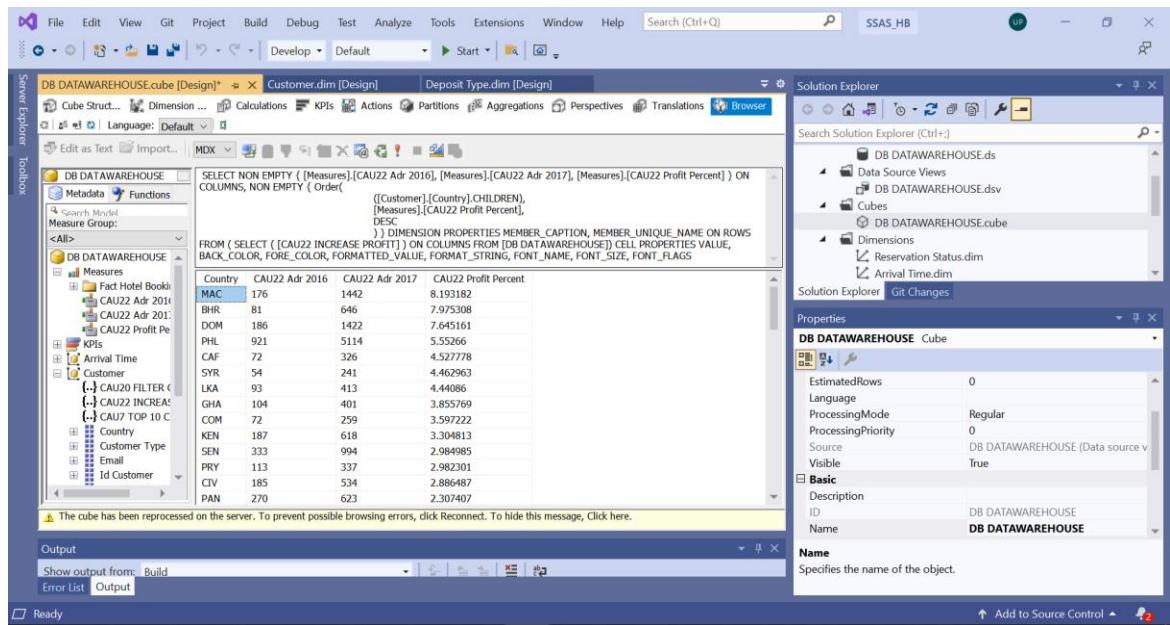
```
SELECT NON EMPTY {
    [Measures].[CAU22 Adr 2016],
    [Measures].[CAU22 Adr 2017],
    [Measures].[CAU22 Profit Percent]
} ON COLUMNS,
NON EMPTY {
    ([Customer].[Country].[Country].ALLMEMBERS )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS|
```

*Hình 3.141. Câu lệnh truy vấn ban đầu*

```
SELECT NON EMPTY {
    [Measures].[CAU22 Adr 2016],
    [Measures].[CAU22 Adr 2017],
    [Measures].[CAU22 Profit Percent]
} ON COLUMNS,
NON EMPTY {
    Order(
        ([Customer].[Country].CHILDREN),
        [Measures].[CAU22 Profit Percent],
        DESC
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS|
```

*Hình 3.141. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.142. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.8.2. Công cụ Pivot Excel

Kéo thả CAU22 Profit Percent, CAU22 Adr 2017, CAU22 Adr 2016 từ bảng Measures vào bảng Values, thuộc tính Country thuộc bảng Customer vào bảng Rows ta được kết quả:

	CAU22 Adr 2016	CAU22 Adr 2017	CAU22 Profit Percent
MAC	176	1442	819.32%
BHR	81	646	797.53%
DOM	186	1422	764.52%
PHL	921	5114	555.27%
CAF	72	326	452.78%
SYR	54	241	446.30%
LKA	93	413	444.09%
GHA	104	401	385.58%
COM	72	259	359.72%
KEN	187	618	330.48%
SEN	333	994	298.50%
PRY	113	337	298.23%
CIV	185	534	288.65%
PAN	270	623	230.74%

Hình 3.143. Kết quả khi dùng công cụ Pivot Excel

### 9.8.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

```

MDXQuery1.mdx - PHUONGUYEN\PHUONGUYEN.SSAS_HB (PHUONGUYEN\phuong)* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
SSAS_HB
MDXQuery1.mdx - UONGUyen(phuong)*
Cube: DB DATAWAREHOUSE
Measure Group: <All>
<All>
DB DATAWAREHOUSE
Measures
KPIs
Arrival Time
Arrival Date Year
Arrivals
Deposit Type
Distribution Channel
Hotel Type
Market Segment
Reservation Date
Reservation Status
with Member [Measures].[Adr 2017] as
IIF([Measures].[Adr], [Arrival Time].[Arrival Date Year].A[2017]) = Null, 0, ([Measures].[Adr], [Arrival Time].[Arrival Date Year].A[2017]))
Member [Measures].[Adr 2016] as
IIF([Measures].[Adr], [Arrival Time].[Arrival Date Year].A[2016]) = Null, 0, ([Measures].[Adr], [Arrival Time].[Arrival Date Year].A[2016]))
Member [Measures].[Profit Percent] as
CASE
    WHEN ([Measures].[Adr 2017]) > 0 AND [Measures].[Adr 2016] > 0 THEN [Measures].[Adr 2017] / [Measures].[Adr 2016]
    ELSE 0
    END +Percent
select [Measures].[Adr 2017],
[Measures].[Adr 2016],
[Measures].[Profit Percent]
on columns
non empty
filter(
    [Customer].[Country].children,
    [Measures].[Profit Percent] > 1
)
order by [Measures].[Profit Percent] desc
filter(
    [Customer].[Country].children,
    [Measures].[Profit Percent] > 1
)
Results
Adr 2017 Adr 2016 Profit Percent
MAC 1442 176 819.52%
BHR 646 81 797.53%
DOM 1422 186 764.52%
PHL 5114 921 555.27%
CAF 326 72 452.78%
SYR 241 54 446.30%
LKA 413 93 444.09%
GHA 401 104 385.58%
COM 259 72 359.72%
KEN 618 187 330.49%

```

Hình 3.144. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

- Lần lượt tạo các độ đo:

+ Adr 2017: để chuyển đổi các giá trị Adr khi có năm nhận phòng vào năm 2017 là null thành số 0.

+ Adr 2016: để chuyển đổi các giá trị Adr khi có năm nhận phòng vào năm 2016 là null thành số 0.

+ Profit Percent: Dùng để tính toán tỷ lệ giữa doanh thu thu được của từng nước trong năm 2017 trên chi tiêu thu được của từng nước trong năm 2016 với điều kiện tử số và mẫu số của tỷ lệ đó khác 0. Ngược lại trả về giá trị 0. Và Định dạng kiểu dữ liệu thành phần trăm (Percent).

- Dùng hàm Order dùng để sắp xếp tên các quốc gia của các khách hàng theo thứ tự giảm dần dựa trên độ đo Profit Percent vừa tạo.

- Hàm Filter dùng để lọc ra tập dữ liệu được sắp xếp từ hàm Order với điều kiện tỷ lệ doanh thu lớn hơn 1.

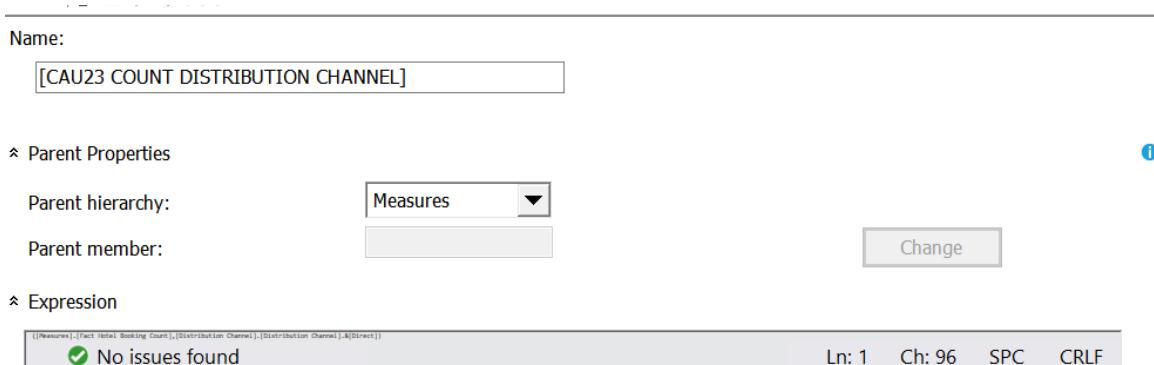
- Hàm Non Empty để loại bỏ các dữ liệu có giá trị null.

## 9.9. Liệt kê các phân khúc thị trường mà tổng doanh thu của của kênh phân phối "Corporate" đó cộng lại tối thiểu 500 (lấy theo thứ tự doanh thu tăng dần) hoặc những phân khúc thị trường có số lượng lượt đặt phòng của kênh phân phối "Direct" tối thiểu 500.

Ý nghĩa câu truy vấn: Thể hiện các phân khúc thị trường có hiệu quả kinh doanh cao ở các kênh phân phối "Corporate" và "Direct".

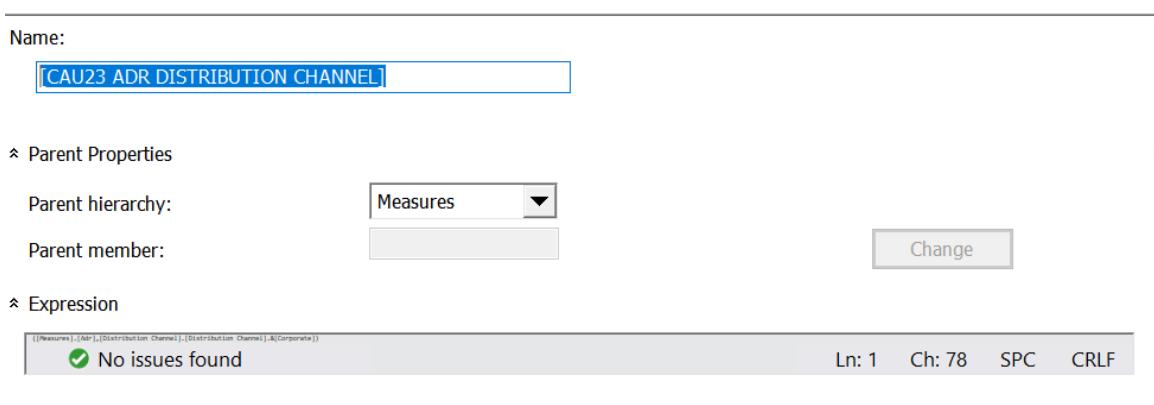
### 9.9.1. Công cụ SSAS trên các khối Cube

Tạo Calculated Member có tên là [CAU23 ADR DISTRIBUTION CHANNEL] dùng để lấy thông tin số lượng lượt đặt phòng của kênh phân phối "Direct".



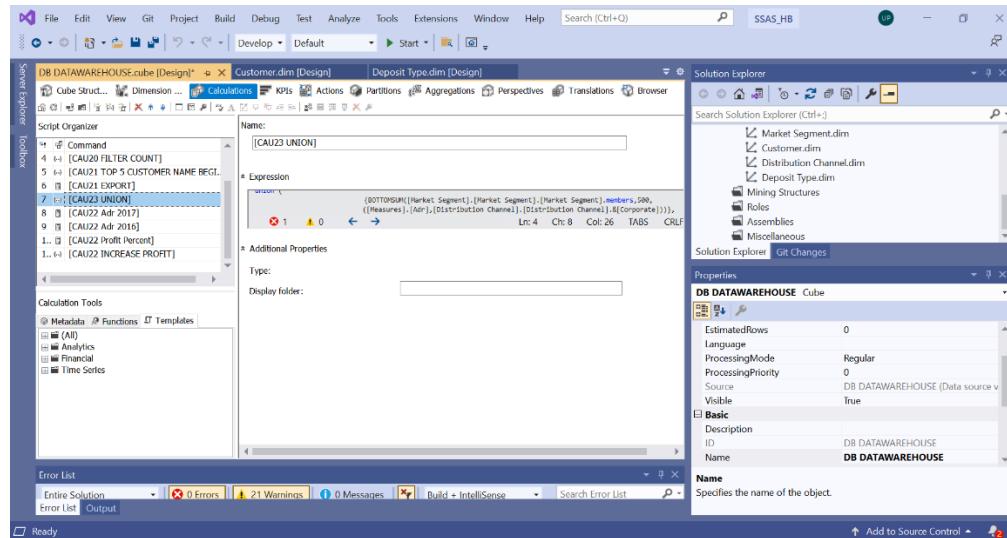
Hình 3.145. Tạo Calculated Member

Tạo Calculated Member có tên là [CAU23 COUNT DISTRIBUTION CHANNEL] dùng để lấy thông doanh thu của của kênh phân phối "Corporate".



Hình 3.146. Tạo Calculated Member

Tạo Named Set có tên là [CAU23 UNION] dùng để lấy thông tin về phân khúc thị trường của tổng doanh thu của của kênh phân phối "Corporate" đó cộng lại tối thiểu 500 (lấy theo thứ tự doanh thu tăng dần) hoặc những phân khúc thị trường có số lượng lượt đặt phòng của kênh phân phối "Direct" tối thiểu 500.



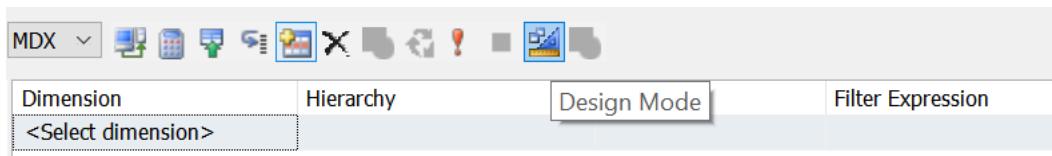
Hình 3.147. Tạo Named Set

Kéo lẩn lượt các thuộc tính Market Segment trong bảng Market Segment, Distribution Channel của bảng Distribution Channel và các độ đo vừa tạo từ Measure Group sang cửa sổ thực thi và Named Set vừa tạo qua cửa sổ lọc được kết quả câu truy vấn sau:

Market Segment	Distribution Channel	CAU23 ADR DISTRIBUTION CHANNEL	CAU23 COUNT DISTRIBUTION CHANNEL
Complementary	Corporate	54	574
Complementary	Direct	54	574
Complementary	GDS	54	574
Complementary	TA/TO	54	574
Direct	Corporate	5529	12204
Direct	Direct	5529	12204
Direct	GDS	5529	12204
Direct	TA/TO	5529	12204
Online TA	Corporate	4417	133
Online TA	Direct	4417	133
Online TA	GDS	4417	133
Online TA	TA/TO	4417	133

Hình 3.148. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng các nhấn chuột vào biểu tượng Design Mode như hình minh họa:



*Hình 3.149. Design Mode trên thanh công cụ*

```
SELECT NON EMPTY {  
    [Measures].[CAU23 ADR DISTRIBUTION CHANNEL],  
    [Measures].[CAU23 COUNT DISTRIBUTION CHANNEL]  
} ON COLUMNS,  
NON EMPTY {  
    (  
        [Market Segment].[Market Segment].[Market Segment].ALLMEMBERS *  
        [Distribution Channel].[Distribution Channel].[Distribution Channel].ALLMEMBERS  
    )  
}  
DIMENSION PROPERTIES MEMBER_CAPTION,  
MEMBER_UNIQUE_NAME ON ROWS  
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,  
    BACK_COLOR,  
    FORE_COLOR,  
    FORMATTED_VALUE,  
    FORMAT_STRING,  
    FONT_NAME,  
    FONT_SIZE,  
    FONT_FLAGS
```

*Hình 3.150. Câu lệnh truy vấn ban đầu*

```
SELECT NON EMPTY {  
    [Measures].[CAU23 ADR DISTRIBUTION CHANNEL],  
    [Measures].[CAU23 COUNT DISTRIBUTION CHANNEL]  
} ON COLUMNS,  
NON EMPTY {  
    [CAU23 UNION]  
} ON ROWS  
FROM [DB DATAWAREHOUSE]
```

*Hình 3.151. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

The screenshot shows the Microsoft Visual Studio 2019 interface with the SSAS\_HB project open. The central area displays an MDX query results grid. The query is:

```
SELECT NON EMPTY {
    [Measures].[CAU23 ADR DISTRIBUTION CHANNEL],
    [Measures].[CAU23 COUNT DISTRIBUTION CHANNEL]
} ON COLUMNS,
NON EMPTY {
    [CAU23 UNION]
} ON ROWS
FROM [DB DATAWAREHOUSE]
```

The grid shows the following data:

Market Segment	CAU23 ADR DISTRIBUTION CHANNEL	CAU23 COUNT DISTRIBUTION CHANNEL
Complementary	54	574
Online TA	4417	133
Direct	5529	12204

The Solution Explorer on the right shows the project structure for 'SSAS\_HB'.

Hình 3.152. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.9.2. Công cụ Pivot Excel

Kéo thả CAU23 COUNT DISTRIBUTION CHANNEL và CAU23 ADR DISTRIBUTION CHANNEL từ bảng Measures vào bảng Values, Named Set CAU23 UNION của bảng Market Segment vào bảng Rows ta được kết quả:

The screenshot shows a Microsoft Excel spreadsheet titled 'Book1 - Excel'. It contains a PivotTable with the following data:

	B	C
Row Labels	CAU23 ADR DISTRIBUTION CHANNEL	CAU23 COUNT DISTRIBUTION CHANNEL
Complementary	54	574
Direct	5529	12204
Online TA	4417	133

The PivotTable Fields pane on the right shows the configuration: Row Labels set to 'CAU23 UNION', Columns set to 'Values', and Values set to 'CAU23 ADR DISTRIBUTION CHANNEL' and 'CAU23 COUNT DISTRIBUTION CHANNEL'.

Hình 3.153. Kết quả khi dùng công cụ Pivot Excel

### 9.9.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

```

SELECT {
    {[Measures].[Adr],[Distribution Channel].[Distribution Channel].&[Corporate]},
    {[Measures].[Fact Hotel Booking Count],[Distribution Channel].[Distribution Channel].&[Direct]}
} on columns,
NON EMPTY (
    {TOPSUM([Market Segment].[Market Segment].[Market Segment].members,500,([Measures].[Adr],[Distribution Channel].[Distribution Channel].&[Corporate]))},
    {BOTTOMSUM([Market Segment].[Market Segment].[Market Segment].members,500,([Measures].[Fact Hotel Booking Count],[Distribution Channel].[Distribution Channel].&[Direct]))}
) on rows
from [DB DATAWAREHOUSE];

```

Results	Adr	Fact Hotel Booking Count
Corporate	54	574
Online TA	4417	133
Direct	5529	12204

Hình 3.154. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

- Hàm BottomSum dùng để sắp xếp các dữ liệu theo thứ tự tăng dần và từ dòng đầu tiên, lấy ra các dòng dữ liệu sao cho xuất hiện kết quả đầu tiên của tổng giá trị Adr của kênh kinh doanh là Corporate là 500.

```

SELECT {
    {[Measures].[Adr],[Distribution Channel].[Distribution Channel].&[Corporate]},
    {[Measures].[Fact Hotel Booking Count],[Distribution Channel].[Distribution Channel].&[Direct]}
} on columns,
NON EMPTY (
    {BOTTOMSUM([Market Segment].[Market Segment].[Market Segment].members,500,([Measures].[Adr],[Distribution Channel].[Distribution Channel].&[Corporate]))}
) on rows
from [DB DATAWAREHOUSE];

```

Results	Adr	Fact Hotel Booking Count
Corporate	54	574
Online TA	4417	133

Hình 3.155. Hàm BottomSum

- Hàm TopSum dùng để sắp xếp các dữ liệu theo thứ tự giảm dần và từ dòng đầu tiên, lấy ra các dòng dữ liệu sao cho xuất hiện kết quả đầu tiên của tổng giá trị Fact Hotel Booking Count của kênh kinh doanh là Direct là 500.

```

SELECT {
    ([Measures].[Adr],[Distribution Channel].[Distribution Channel].&[Corporate]),
    ([Measures].[Fact Hotel Booking Count],[Distribution Channel].[Distribution Channel].&[Direct])
} on columns,
NON EMPTY (
    [TOPSUM({[Market Segment].[Market Segment].members},500,{[Measures].[Fact Hotel Booking Count]},{[Distribution Channel].[Distribution Channel].&[Direct]})]
) on rows
from [DB DATAWAREHOUSE];

```

	Adr	Fact Hotel Booking Count
Corporate		
Direct	5529	12204

*Hình 3.156. Hàm TopSum*

- Hàm Union dùng để lấy kết quả của cả hai tập dữ liệu.
- Hàm Non Empty dùng để loại bỏ các dữ liệu có giá trị null.

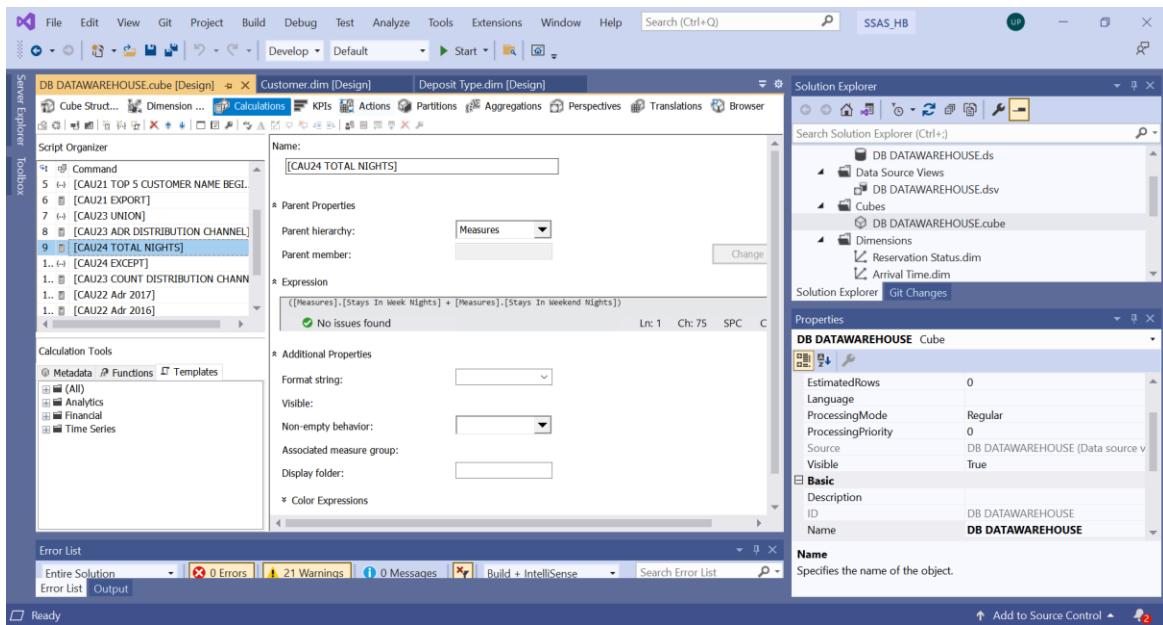
## **9.10. Liệt kê tổng số đêm thuê của các quốc gia trong năm 2015 trừ (ngày nhận phòng vào thứ 7 và không đặt tiền cọc trước của trong năm 2015)**

*Ý nghĩa câu truy vấn:* Thể hiện xu hướng đặt phòng của khách hàng ở các quốc gia để có chiến lược kinh doanh phù hợp.

### *9.10.1. Công cụ SSAS trên các khối Cube*

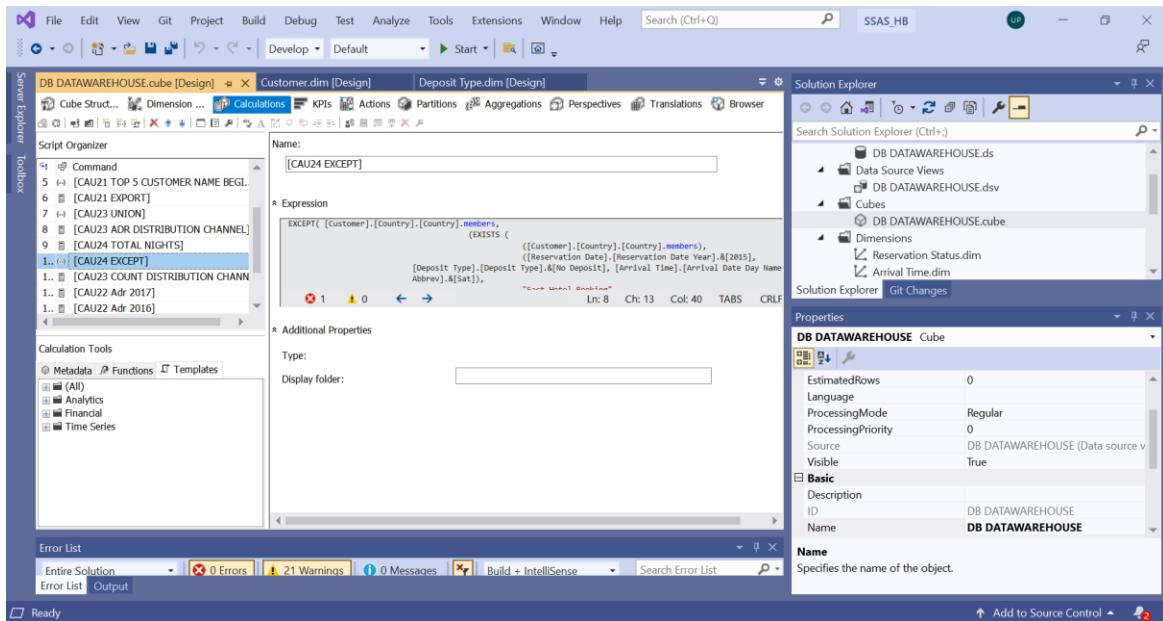
Tạo Calculated Member có tên là [CAU24 TOTAL NIGHTS] dùng để lấy thông tin tổng số đêm thuê của từng dòng dữ liệu.

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.157. Tạo Calculated Member

Tạo Named Set có tên là [CAU24 EXCEPT] dùng để lấy thông tin các quốc gia trong năm 2015 trừ (ngày nhận phòng vào thứ 7 và không đặt tiền cọc trước của trong năm 2015)



Hình 3.158. Tạo Calculated Member

Kéo thuộc tính Country trong bảng Customer và độ đo vừa tạo từ Measure Group sang cửa sổ thực thi và Named Set vừa tạo và Arrival Date Year qua cửa sổ lọc với năm được chọn là 2015 được kết quả câu truy vấn sau:

Hình 3.159. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

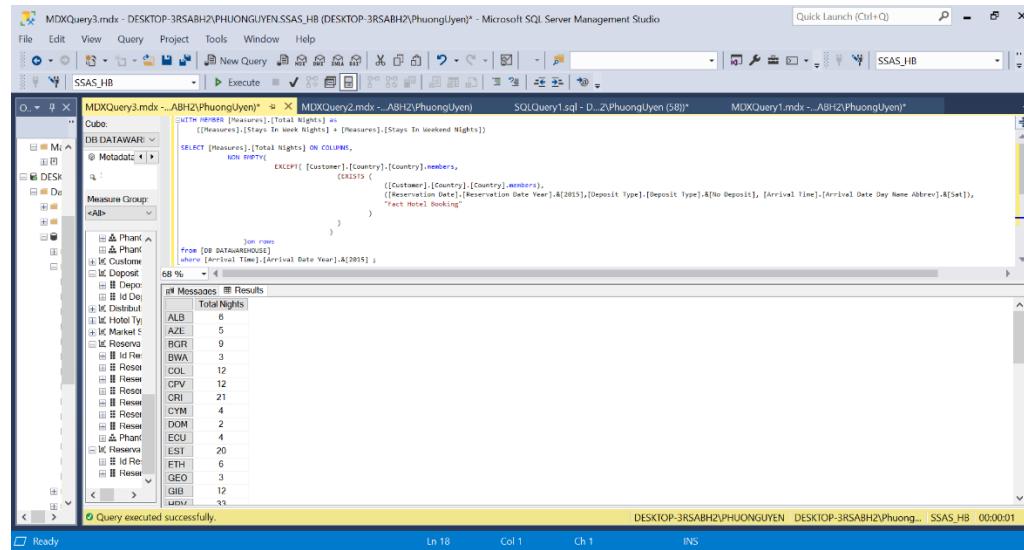
#### 9.10.2. Công cụ Pivot Excel

Kéo thả Calculated Member CAU24 TOTAL NIGHTS từ bảng Measures vào bảng Values, Named Set CAU24 EXCEPT thuộc bảng Customer vào bảng Rows. Tiếp đến, ta lọc dữ liệu với điều kiện Arrival Date Year là 2015 thông qua trường Filters được kết quả:

Hình 3.160 Kết quả khi dùng công cụ Pivot Excel

### 9.10.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:



Hình 3.160. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

- Hàm Exists để tìm những giá trị trong thuộc tính Country có dòng dữ liệu thỏa mãn ba điều kiện là năm thay đổi trạng thái cuối cùng là 2015, có loại hình thanh toán là No Deposit và ngày nhận phòng vào thứ bảy trong bảng Fact Hotel Booking.
- Hàm Except để làm phép toán trừ, ta lấy tất cả các giá trị trong thuộc tính Country trừ đi các giá trị thu được từ kết quả hàm Exists.
- Hàm NotEmpty dung để loại bỏ các dữ liệu có giá trị null.
- Dùng Where để lấy các dữ liệu có năm nhận phòng trong năm 2015.

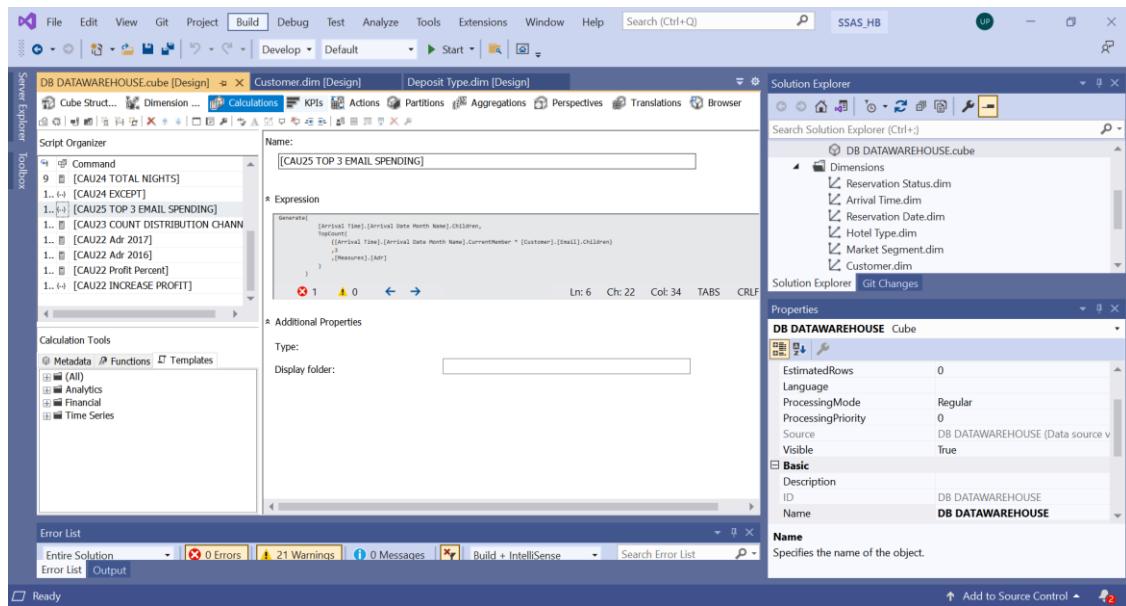
### 9.11. Với mỗi tháng đưa ra 3 email của khách hàng có chi tiêu cao nhất.

Ý nghĩa câu truy vấn: Cho biết các khách hàng tiềm năng của khách sạn từ đó đưa ra các chương trình khuyến mãi đặc biệt nhằm tri ân khách hàng.

#### 9.11.1. Công cụ SSAS trên các khôi Cube

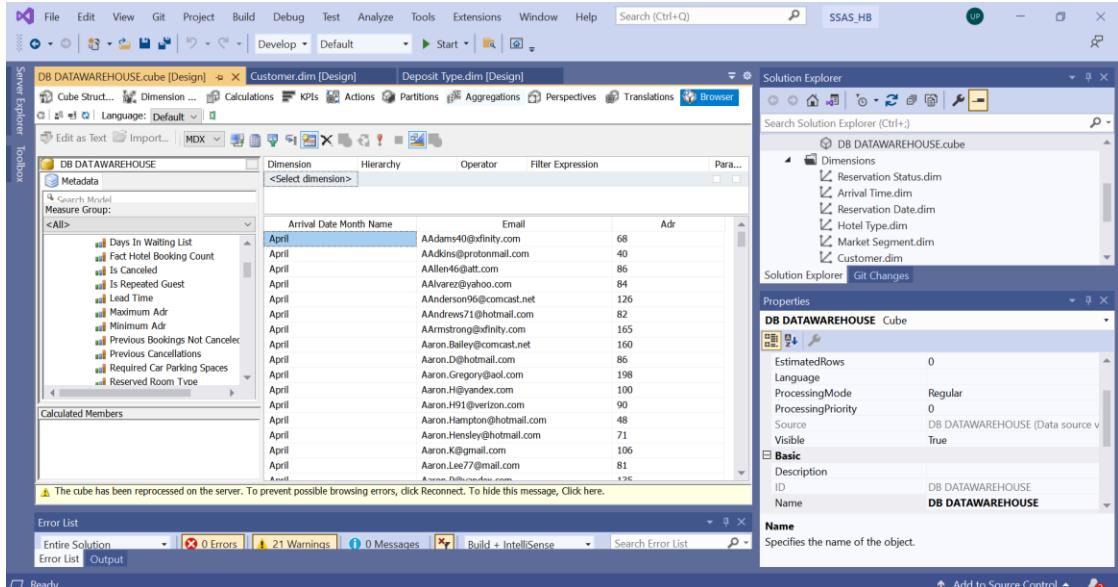
Tạo Named Set có tên là [CAU25 TOP 3 EMAIL SPENDING] dùng để lấy thông tin tháng và 3 email của khách hàng có chi tiêu cao nhất trong tháng đó.

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.161. Tạo Named Set

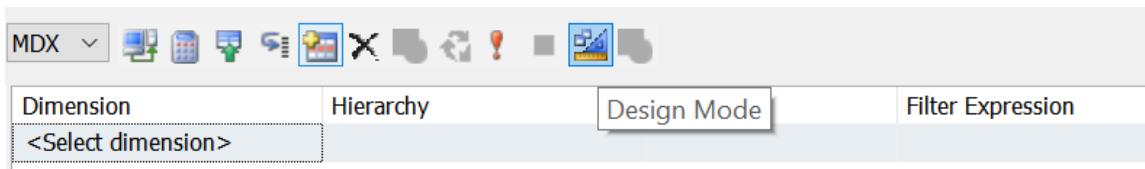
Kéo thuộc tính Email trong bảng Customer, thuộc tính Arrival Date Month Name trong bảng Arrival Time và độ đo Adr từ Measure Group sang cửa sổ thực thi ta được kết quả câu truy vấn sau:



Hình 3.162. Kết quả câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng các nhấn chuột vào biểu tượng

Design Mode như hình minh họa:



*Hình 3.163. Design Mode trên thanh công cụ*

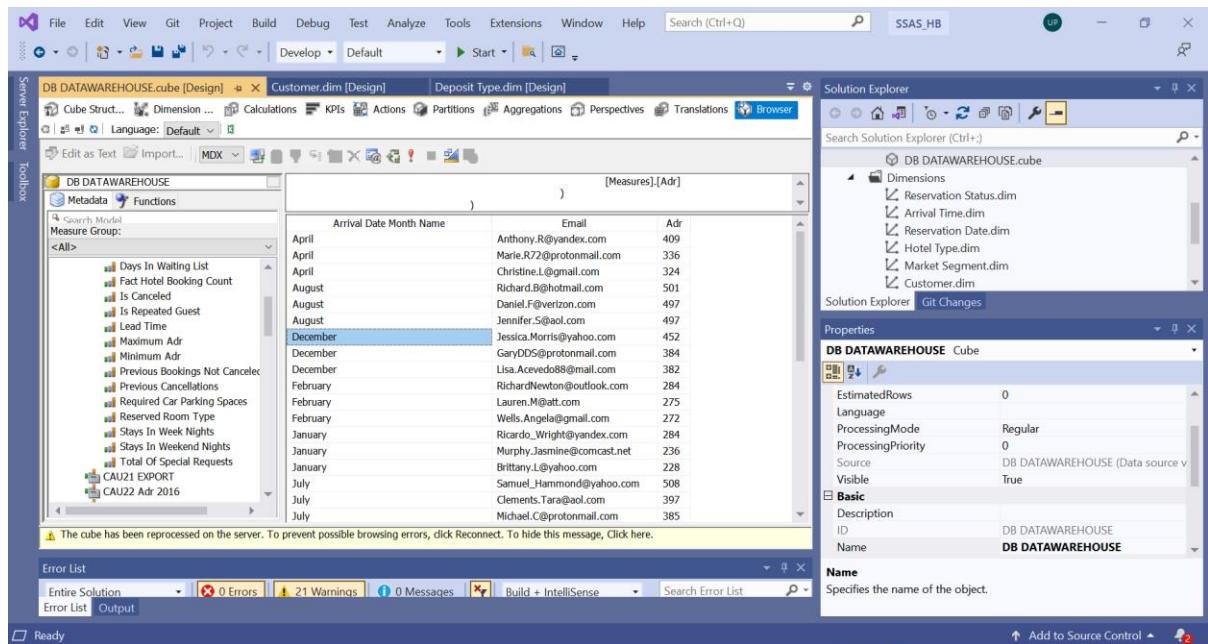
```
SELECT NON EMPTY {
    [Measures].[Adr]
} ON COLUMNS,
NON EMPTY {
    (
        [Arrival Time].[Arrival Date Month Name].[Arrival Date Month Name].ALLMEMBERS *
        [Customer].[Email].[Email].ALLMEMBERS
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS
```

*Hình 3.164. Câu lệnh truy vấn ban đầu*

```
SELECT NON EMPTY {
    [Measures].[Adr]
} ON COLUMNS,
NON EMPTY {
    Generate(
        [Arrival Time].[Arrival Date Month Name].Children,
        TopCount(
            (
                [Arrival Time].[Arrival Date Month Name].CurrentMember *
                [Customer].[Email].Children
            ),
            3,
            [Measures].[Adr]
        )
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS
```

*Hình 3.165. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.166. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.11.2. Công cụ Pivot Excel

Kéo thả Adr từ bảng Measures vào bảng Values, và Named Set CAU25 TOP 3 EMAIL SPENDING của bảng Measures vào bảng Rows ta được kết quả:

	Row Labels	Adr
1	April	
3	Anthony.R@yandex.com	409
4	Marie.R72@protonmail.com	336
5	Christine.L@gmail.com	324
6	August	
7	Richard.B@hotmail.com	501
8	Daniel.F@verizon.com	497
9	Jennifer.S@aol.com	497
10	December	
11	Jessica.Morris@yahoo.com	452
12	GaryDDS@protonmail.com	384
13	Lisa.Acevedo88@mail.com	382
14	February	
15	RichardNewton@outlook.com	284
16	Lauren.M@att.com	275

Hình 3.167. Kết quả khi dùng công cụ Pivot Excel

### 9.11.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

```

MDXQuery3.mdx - DESKTOP-3RSABH2\PHUONGUYEN.SSAS_HB (DESKTOP-3RSABH2\PhuongUyen)* - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
SSAS_HB
MDXQuery3.mdx - .\BHZ\PhuongUyen*
MDXQuery1.mdx - .\BHZ\PhuongUyen*
SQLQuery1.sql - D:\PhuongUyen (58)*
MDXQuery1.mdx - .\BHZ\PhuongUyen*
Select {[Measures].[Adr]} columns,
NonEmpty ({Generate
[Arrival Time].[Arrival Date Month Name].Children,
TopCount(
[Arrival Time].[Arrival Date Month Name].CurrentMember * {[customer].[Email].Children}
,3
),[Measures].[Adr]
)})
) } On Rows
From [DB DATAWAREHOUSE];

```

Month	Email	Adr
April	Anthony.R@yandex.com	409
April	Marie.R72@protonmail.com	336
April	Christine.L@gmail.com	324
August	Richard.B@hotmail.com	501
August	Daniel.F@verizon.com	497
August	Jennifer.S@aid.com	497
December	Jessica.Morris@yahoo.com	452
December	Gary.DOS@protonmail.com	384
December	Lisa.Arendt88@gmail.com	382
February	Richard.Wenlow@outlook.com	284
February	Lauren.M@att.com	275
February	Wells.Angele@gmail.com	272
January	Ricardo.Wright@yandex.com	264
January	Murphy.Jasmine@comcast.net	236
January	Brittany.L@yahoo.com	228

Hình 3.168. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

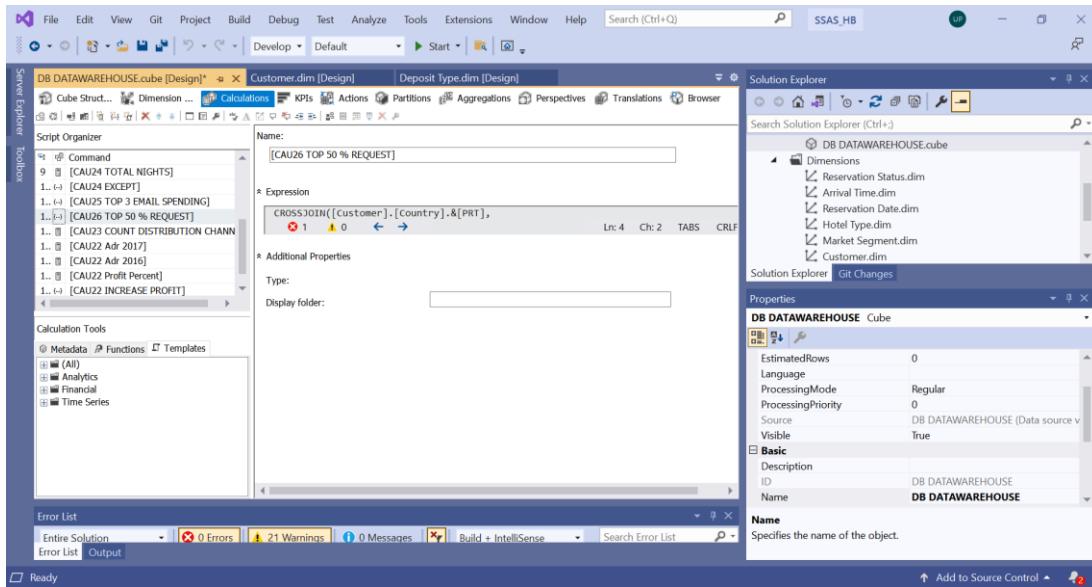
- Hàm TopCount dùng để sắp xếp tập dữ liệu tích chọn giữa tháng trong ngày nhận phòng (Arrival Date Month Name) và email trong bảng chiều Customer theo thứ tự giảm dần và lấy 3 dòng dữ liệu đầu tiên.
- Hàm Generate dùng để duyệt từng giá trị trong tháng dạng chữ của ngày nhận phòng ở đối số thứ nhất. Tại đối số thứ hai, hàm Generate lấy ra tập dữ liệu đã tạo trong TopCount đã trình bày trước đó.
- Hàm NonEmpty dùng để loại bỏ các dữ liệu có giá trị null.

### 9.12. Liệt kê các quý theo từng năm của quốc gia Bồ Đào Nha, thuộc top 50% quý có tổng yêu cầu đặc biệt trong tổng số nhu cầu đặc biệt thu được.

Ý nghĩa câu truy vấn: Thể hiện các quý có các yêu cầu đặc biệt nhiều nhất để đưa ra chiến lược kinh doanh phù hợp cho khách sạn.

### 9.12.1. Công cụ SSAS trên các khối Cube

Tạo Named Set có tên là [CAU26 TOP 50 % REQUEST] dùng để lấy thông tin các các quý theo từng năm của quốc gia Bồ Đào Nha, thuộc top 50% quý có tổng yêu cầu đặc biệt nhiều nhất



Hình 3.169. Tạo Named Set

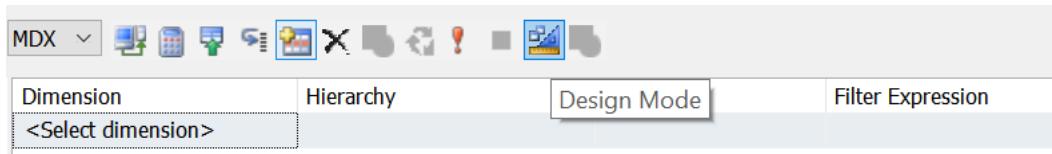
Kéo thuộc tính Country trong bảng Customer, thuộc tính Arrival Date Year trong bảng Arrival Time và các độ đo Adr, Total Of Special Requests từ Measure Group sang cửa sổ thực thi ta được kết quả truy vấn sau:

Country	Arrival Date Year	Arrival Date Quarter	Adr	Total Of Special Re...
ABW	2016	2	158	2
ABW	2017	2	99	1
AGO	2015	3	4038	19
AGO	2015	4	1307	9
AGO	2016	1	6419	19
AGO	2016	2	4841	6
AGO	2016	3	5466	23
AGO	2016	4	5032	29
AGO	2017	1	7973	40
AGO	2017	2	5116	22
AGO	2017	3	2206	16
AIA	2016	3	265	0
ALB	2015	3	107	0
ALB	2015	4	66	0
ALB	2016	1	35	2
ALB	2016	2	222	0

Hình 3.170. Kết quả câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng các nhấn chuột vào biểu tượng

Design Mode như hình minh họa:



*Hình 3.171. Design Mode trên thanh công cụ*

```
SELECT NON EMPTY {
    [Measures].[Adr],
    [Measures].[Total Of Special Requests]
} ON COLUMNS,
NON EMPTY {
    (
        [Customer].[Country].[Country].ALLMEMBERS *
        [Arrival Time].[PhanCapNgayNhanPhong(YQMD)].[Arrival Date Quarter].ALLMEMBERS
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
    BACK_COLOR,
    FORE_COLOR,
    FORMATTED_VALUE,
    FORMAT_STRING,
    FONT_NAME,
    FONT_SIZE,
    FONT_FLAGS
```

*Hình 3.172. Câu lệnh truy vấn ban đầu*

```
SELECT NON EMPTY {
    [Measures].[Adr],
    [Measures].[Total Of Special Requests]
} ON COLUMNS,
NON EMPTY {
    CROSSJOIN(
        [Customer].[Country].&[PRT],
        DRILLDOWNLEVEL(
            [Arrival Time].[Arrival Date Year].[Arrival Date Year]*TOPPERCENT(
                [Arrival Time].[Arrival Date Quarter].[Arrival Date Quarter],
                50,
                [Measures].[Total Of Special Requests]
            ),,
            0
        )
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
    BACK_COLOR,
    FORE_COLOR,
    FORMATTED_VALUE,
    FORMAT_STRING,
    FONT_NAME,
    FONT_SIZE,
    FONT_FLAGS
```

*Hình 3.173. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

The screenshot shows the Visual Studio 2019 interface for an SSAS project named "SSAS\_HB". The central area displays an MDX query:

```

SELECT NON EMPTY {
    [Measures].[Adr],
    [Measures].[Total Of Special Requests]
} ON COLUMNS,
NON EMPTY (
    CROSSJOIN(
        [Customer].[Country].&[PRT],
        DRILLDOWNLEVEL(
            Year).[Arrival Date Year]*[Arrival Time].[Arrival Date]
    )
    TOPPERCENT(
        [Arrival Time].[Arrival Date Quarter].[Arrival Date Quarter],
        50,
        [Measures].[Total Of Special Requests]
    )
),

```

The results pane shows a table with the following data:

Country	Arrival Date Year	Arrival Date Quarter	Adr	Total Of Special Requests
PRT	2016	3	585359	2299
PRT	2016	4	405193	1763
PRT	2017	2	606235	1754
PRT	2017	3	388845	1869

The Solution Explorer shows the cube's dimensions and measures. The Properties pane displays basic information about the cube.

Hình 3.174. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.12.2. Công cụ Pivot Excel

Kéo thả Adr và Total Of Special Requests từ bảng Measures vào bảng Values, Named Set [CAU26 TOP 50 % REQUEST] của bảng Measures vào bảng Rows ta được kết quả:

The screenshot shows an Excel spreadsheet with a PivotTable. The PivotTable Fields pane on the right shows fields from the 'Customer' dimension and a named set 'CAU26 TOP 50 % REQUEST'. The main table displays data for Portugal (PRT) grouped by year (2016, 2017) and quarter (3, 4).

	A	B	C	D	E
1	Row Labels	Adr	Total Of Special Requests		
2	PRT				
3	2016				
4	3	585359	2299		
5	4	405193	1763		
6	2017				
7	2	606235	1754		
8	3	388845	1869		

Hình 3.175. Kết quả khi dùng công cụ Pivot Excel

### 9.12.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

The screenshot shows the Microsoft SQL Server Management Studio interface with the following details:

- Query Editor:** Displays an MDX query:
 

```
SELECT {[Measures].[Adr]}, {[Measures].[Total Of Special Requests]} ON COLUMNS,
NON EMPTY(
    CROSSJOIN([Customer].[Country].&[PRT],
    DRILLDOWNLEVEL([Arrival Time].[Arrival Date Year].[Arrival Date Year])
    )TOPPERCENT([Arrival Time].[Arrival Date Quarter].[Arrival Date Quarter], 50, [Measures].[Total Of Special Requests]),0
)ON ROWS
FROM [DB DATAWAREHOUSE]
```
- Object Explorer:** Shows the cube structure under the DB DATAWAREHOUSE cube, including measures like Fact Hotel Booking, Adr, Adults, and Total Of Special Requests.
- Results Grid:** Shows the output of the query, grouped by Adr (Address) and year (Arrival Date Year). The data includes:
 

Adr	Arrival Date Year	Total Of Special Requests
PRT	2016	585359
PRT	2016	405193
PRT	2017	606235
PRT	2017	388845
		2299
		1763
		1754
		1869
- Status Bar:** Shows "Query executed successfully." and other system information.

Hình 3.176. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích ý nghĩa câu truy vấn MDX:

- Hàm TopPercent để sắp xếp quý của ngày nhận phòng khách sạn theo thứ tự giảm dần dựa trên độ đo tổng yêu cầu đặc biệt của từng quý và từ dòng dữ liệu đầu tiên có quý chứa tổng yêu cầu đặc biệt cao nhất cho tới khi tổng tất cả các yêu cầu chiếm ít nhất 50% giá trị của Total Of Special Request thì dừng lại.

- Trong hàm DrillDownLevel, ở đối số thứ nhất truyền vào tích chéo giữa thuộc tính năm nhận phòng (Arrival Date Year) .với từng quý (Arrival Date Quarter) đã được lọc ở hàm TopPercent và mức chỉ mục Index với giá trị 0 để truy cập đến mức thấp nhất.

- Hàm CrossJoin dùng để tích chéo thuộc tính quốc tịch là nước Bồ Đào Nha (PRT) với các thuộc tính của tập dữ liệu đã được xử lý bên DrillDownLevel.

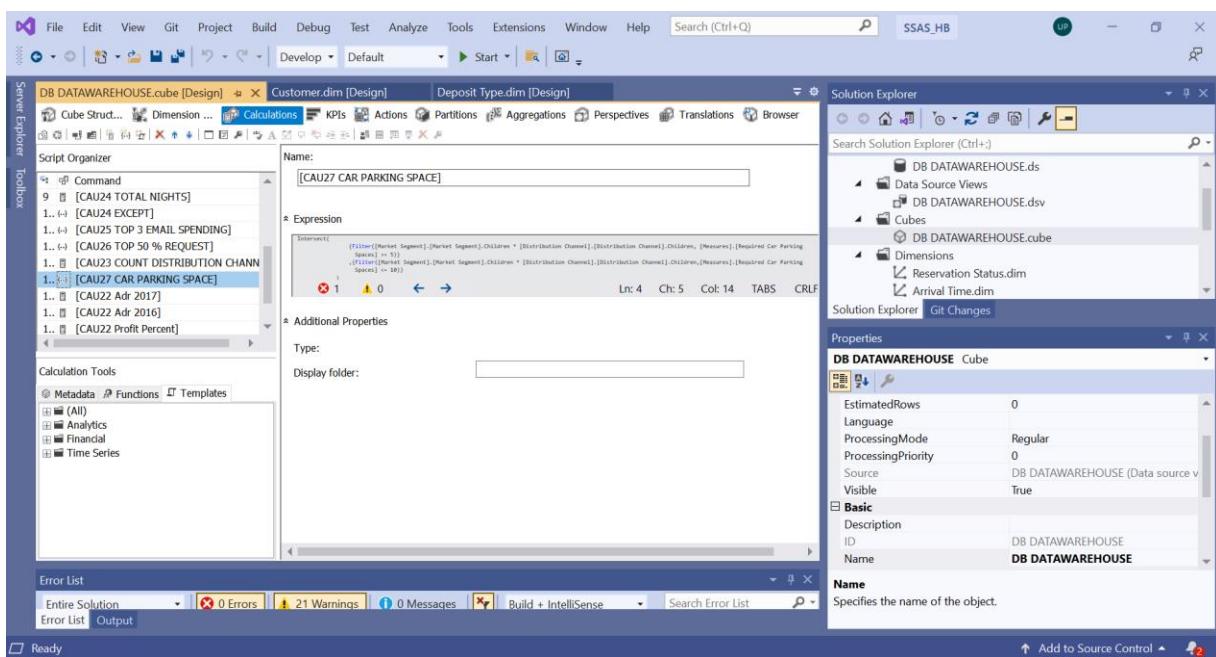
-Hàm NonEmpty dùng để loại bỏ các dữ liệu có giá trị null.

### 9.13. Liệt kê những phân khúc thị trường và kênh phân phối có yêu cầu bãi đỗ cho từ 5 đến 10 xe.

*Ý nghĩa câu truy vấn:* Cho biết các nhu cầu bãi đỗ của các phân khúc thị trường và kênh phân phối để khách sạn có những sắp xếp hoặc mở rộng bãi đỗ để đáp ứng nhu cầu cần thiết.

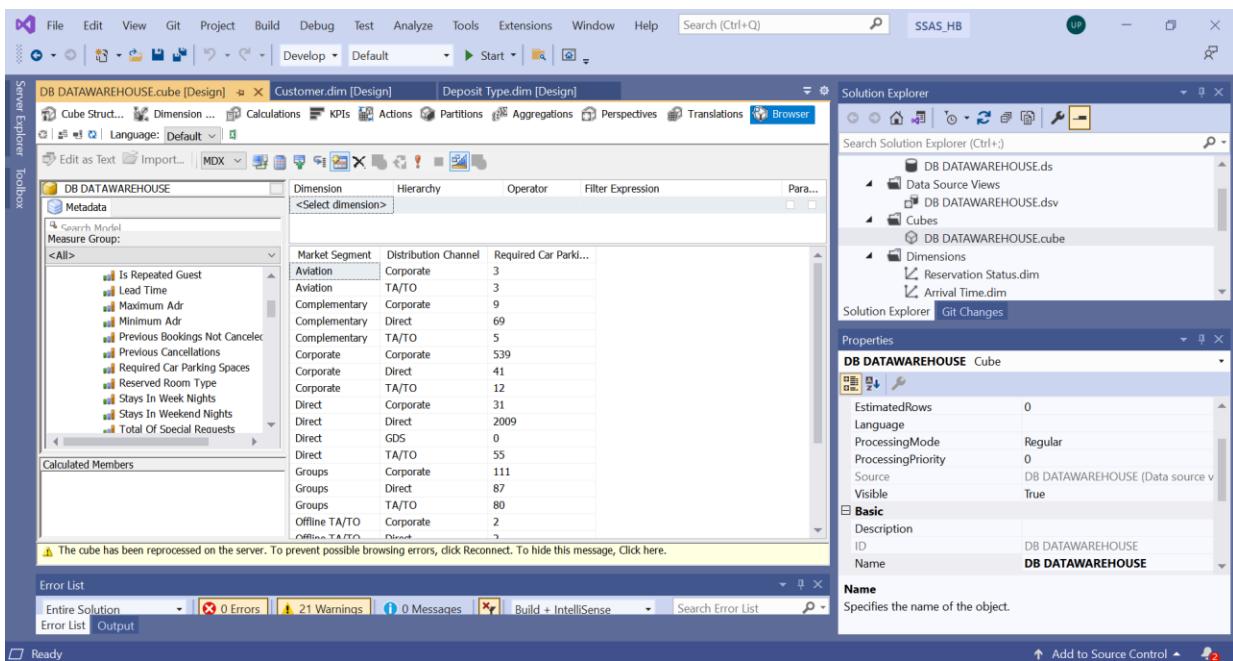
#### 9.13.1. Công cụ SSAS trên các khối Cube

Tạo Named Set có tên là [CAU26 TOP 50 % REQUEST] dùng để lấy thông tin những phân khúc thị trường và kênh phân phối có yêu cầu bãi đỗ cho từ 5 đến 10 xe



Hình 3.177. Tạo Named Set

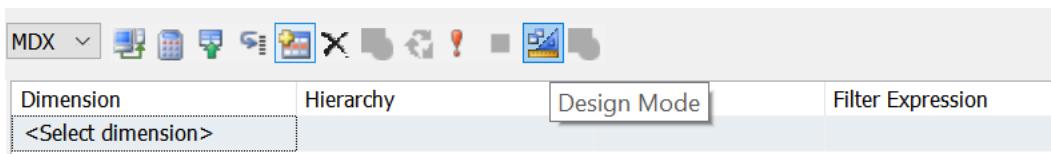
Kéo thuộc tính Market Segment trong bảng Market Segment và thuộc tính Distribution Channel trong bảng Distribution Channel và độ đo Required Car Parking Spaces từ Measure Group sang cửa sổ thực thi ta được kết quả câu truy vấn sau:



*Hình 3.177. Kết quả câu truy vấn trong Visual Studio 2019*

Chuyển từ Design Mode sang Script Mode bằng các nhấn chuột vào biểu tượng

Design Mode như hình minh họa:



*Hình 3.178. Design Mode trên thanh công cụ*

```

SELECT NON EMPTY {
    [Measures].[Required Car Parking Spaces]
} ON COLUMNS,
NON EMPTY {
    (
        [Market Segment].[Market Segment].[Market Segment].ALLMEMBERS *
        [Distribution Channel].[Distribution Channel].[Distribution Channel].ALLMEMBERS
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS

```

*Hình 3.179. Câu lệnh truy vấn ban đầu*

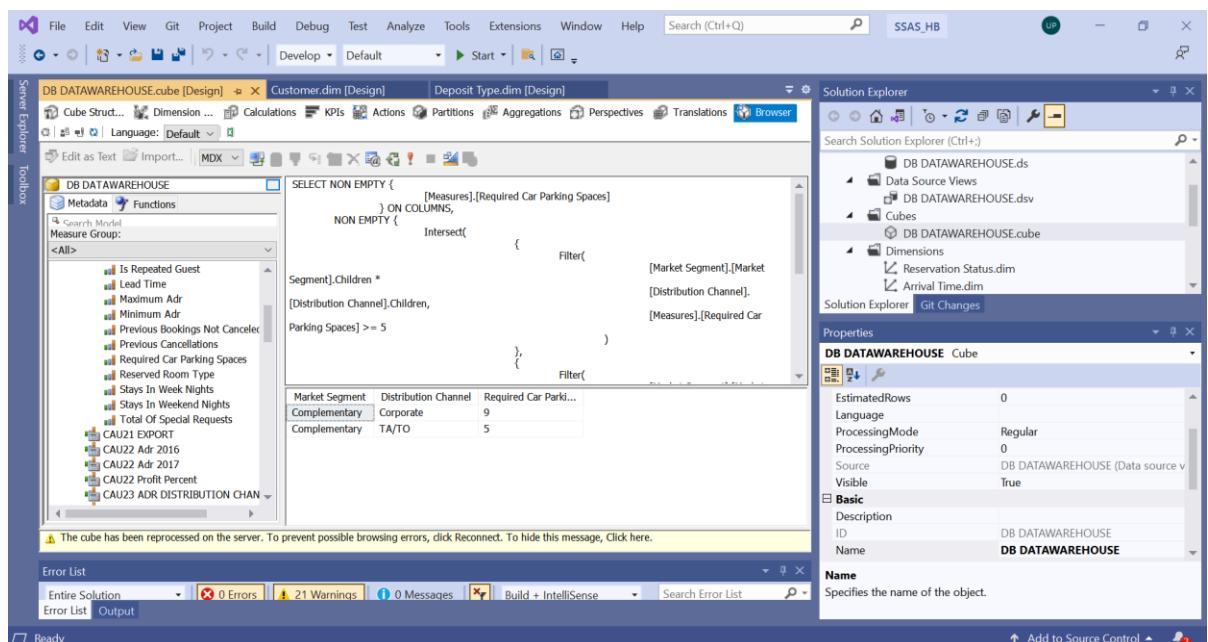
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

```

SELECT NON EMPTY {
    [Measures].[Required Car Parking Spaces]
} ON COLUMNS,
NON EMPTY {
    Intersect(
        {
            Filter(
                [Market Segment].[Market Segment].Children *
                [Distribution Channel].[Distribution Channel].Children,
                [Measures].[Required Car Parking Spaces] >= 5
            )
        },
        {
            Filter(
                [Market Segment].[Market Segment].Children *
                [Distribution Channel].[Distribution Channel].Children,
                [Measures].[Required Car Parking Spaces] <= 10
            )
        }
    )
},
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS

```

Hình 3.180. Thay thế hàm vào câu lệnh truy vấn



Hình 3.181. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.13.2. Công cụ Pivot Excel

Kéo thả Required Car Parking Spaces từ bảng Measures vào bảng Values, Named Set [CAU27 CAR PARKING SPACE] thuộc bảng Measures vào bảng Rows ta được kết quả:

A	B	C	D	E
1 Row Labels	Required Car Parking Spaces			
2 Complementary				
3 Corporate	9			
4 TA/TO	5			
5				
6				
7				
8				
9				
10				
11				

Hình 3.182. Kết quả khi dùng công cụ Pivot Excel

### 9.13.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

```

Select {[Measures].[Required Car Parking Spaces]} on Columns,
       Intersect(
           {Filter([Market Segment].[Market Segment].Children * [Distribution Channel].[Distribution Channel].[Measures].[Required Car Parking Spaces] >= 5)},
           {Filter([Market Segment].[Market Segment].Children * [Distribution Channel].[Distribution Channel].[Measures].[Required Car Parking Spaces] <= 10)}
       )
    on Rows
  From [DB DATAWAREHOUSE]
  
```

	Required Car Parking Spaces
Complementary   Corporate	9
Complementary   TA/TO	5

Hình 3.182. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

### ❖ Giải thích ý nghĩa câu truy vấn MDX:

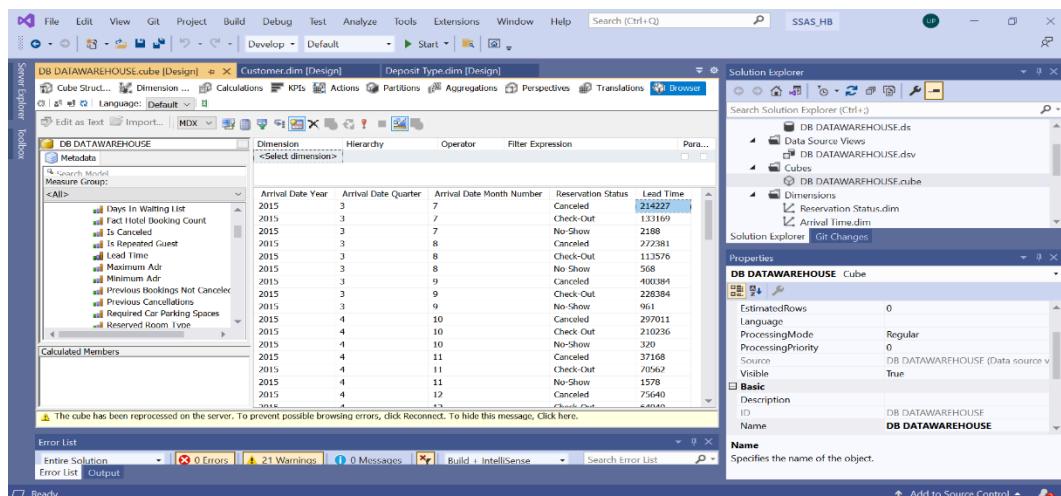
- Operator “\*” dùng để tích chéo thuộc tính Market Segment với thuộc tính Distribution Channel.
- Hàm Filter đầu tiên dùng để lọc các giá trị tích chéo trong operator vừa tạo với điều kiện tổng số lượng bãі đỗ xe hơi được yêu cầu ít nhất 5.
- Hàm Filter thứ hai dùng để lọc các giá trị tích chéo trong operator vừa tạo với điều kiện tổng số lượng bãі đỗ xe hơi được yêu cầu nhiều nhất 10.
- Hàm Intersect dùng để giao hai kết quả của hai lần lọc với nhau để lấy giá trị những tổng số bãі đỗ xe nằm trong khoảng từ 5 đến 10 của từng phân khúc thị trường trong từng kênh phân phối.

### 9.14. Thống kê tổng thời gian chờ của từng hình thức thanh toán đặt cọc mà khách hàng đặt phòng theo từng tháng, quý, năm

*Ý nghĩa câu truy vấn:* Dựa vào thống kê thời gian chờ này khách sạn có thể có những dự trù, sắp xếp các hoạt động đặt phòng một cách tối ưu nhất.

#### 9.14.1. Công cụ SSAS trên các khối Cube

Kéo các thuộc tính Arrival Date Year, Arrival Date Quarter, Arrival Date Month trong bảng Arrival Time, thuộc tính Reservation Status trong bảng Reservation Status và độ đo Lead Time trong bảng Measures từ Measure Group sang cửa sổ lọc dữ liệu và cửa sổ thực thi. Ta thu được kết quả câu truy vấn



Hình 3.183. Kết quả câu truy vấn trong Visual Studio 2019

### 9.14.2. Công cụ Pivot Excel

Kéo thả Lead Time từ bảng Measures vào bảng Values, thuộc tính Arrival Date Year, Arrival Date Quarter, Arrival Date Month từ PhanCapNgayNhanPhong(YQMDD) thuộc bảng Arrival Time và thuộc tính Reservation Status của bảng Reservation Status vào bảng Rows ta được kết quả:

Lead Time		
2015	3	7
Canceled	214227	
Check-Out	133169	
No-Show	2188	
		8
Canceled	272381	
Check-Out	113576	
No-Show	568	
		9
Canceled	400384	
Check-Out	228384	

Hình 3.184. Kết quả khi dùng công cụ Pivot Excel

### 9.14.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:

	Lead Time	Lead Time	Lead Time
All	630899	593751	68959
2015	1206811	819967	6765
3	88992	475129	3717
7	214227	133169	2188
8	272381	113576	568
9	400384	228384	961
4	409819	344838	3048
10	297011	210236	320
11	37168	70562	1578
12	75640	64040	1150
Total	2652877	208107	145145

Hình 3.185. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

#### ❖ Giải thích câu tuy cần MDX:

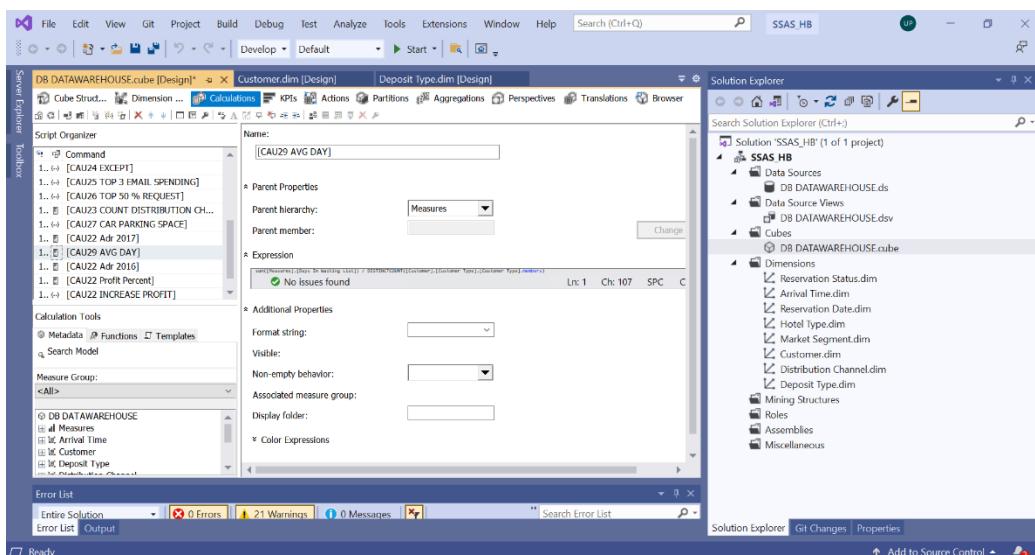
- Hàm DrillDownLevel sẽ truy vấn xuống từng phân cấp trong PhanCapNgayNhanPhong(YQMDD) đã tạo trước đó và bắt đầu từ thuộc tính năm (Arrival Date Year) xuống quý (Arrival Date Quarter) và dừng lại ở tháng (Arrival Date Month Number).
  - Dùng operator “\*” để tính tích chéo của thuộc tính Reservation Status trong bảng chiều Reservation Status với độ đo Lead Time trong Measures.
  - Hàm NonEmpty dùng để loại bỏ các dữ liệu có giá trị null.

**9.15. Tính trung bình ngày chờ trên mỗi loại khách hàng của loại hình phân khúc thị trường của mỗi quốc gia và được thực hiện với năm thay đổi trạng thái bằng năm nhân phòng trừ loại hình phân khúc là Complementary**

*Ý nghĩa câu truy vấn:* Từ số liệu trung bình ngày chờ của từng loại khách hàng của từng phân khúc ( ngoại trừ phân khúc “Complementary”) khách sạn có thể có những dự trù, sắp xếp các hoạt động đặt phòng một cách tối ưu nhất.

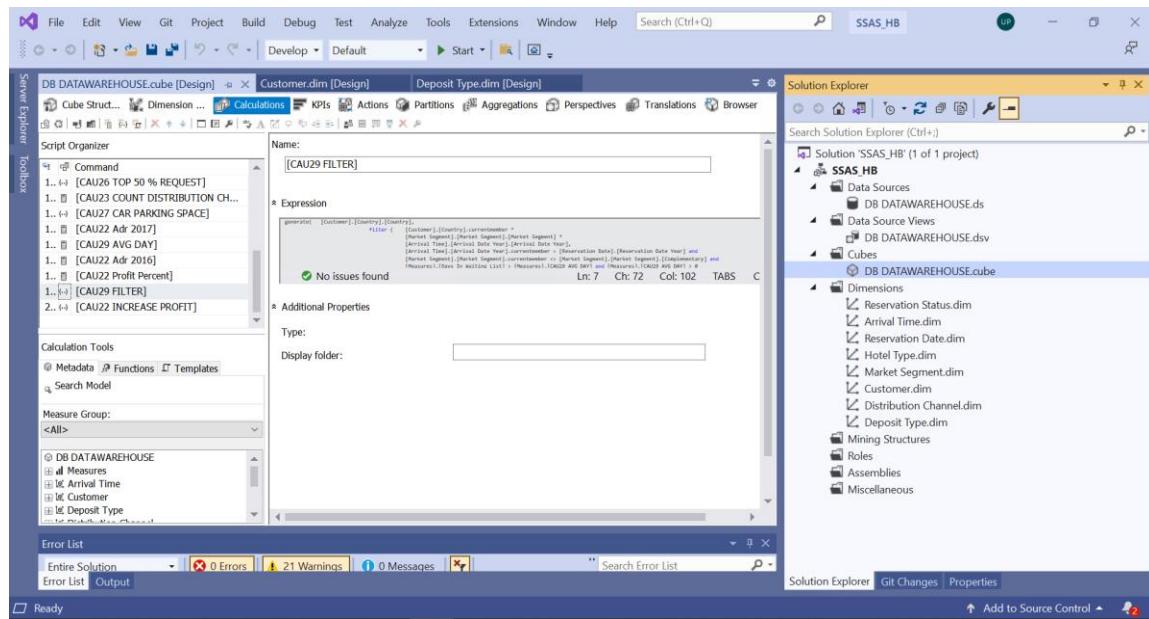
#### 9.15.1. Công cụ SSAS trên các khối Cube

Tạo Calculated Member có tên là [CAU29 AVG DAY] dùng để tính trung bình ngày chờ trên mỗi loại khách hàng



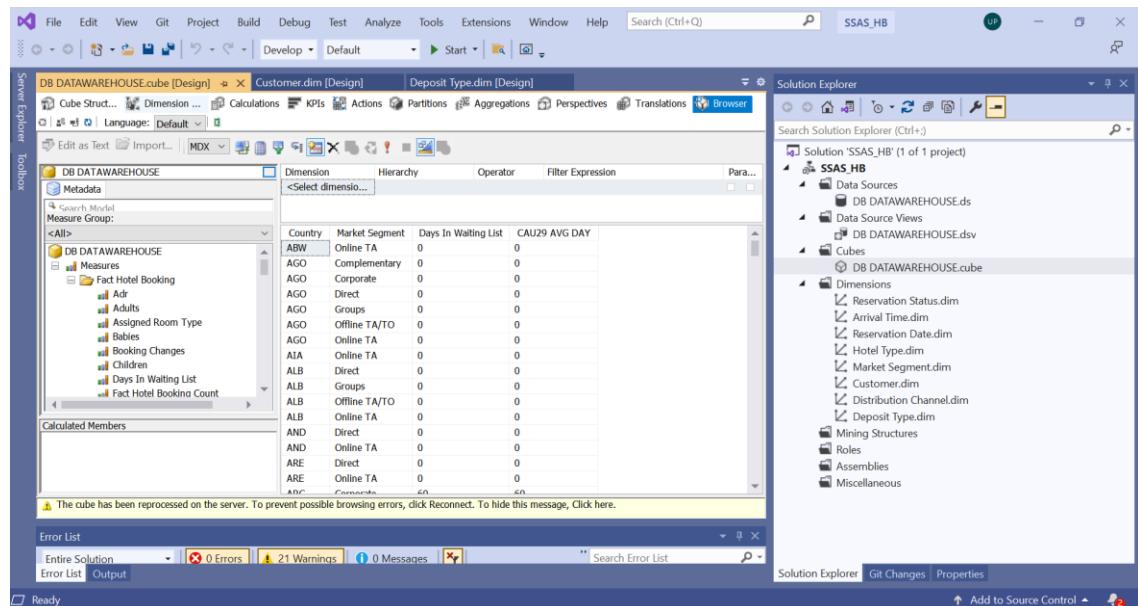
Hình 3.186. Tao Calculated Member

Tạo Named Set có tên là [CAU29 FILTER] dùng để lấy thông tin tên quốc gia, phân khúc thị trường và năm nhận phòng trong điều kiện năm thay đổi trạng thái bằng năm nhận phòng trừ loại hình phân khúc là Complementary



Hình 3.187. Tạo Named Set

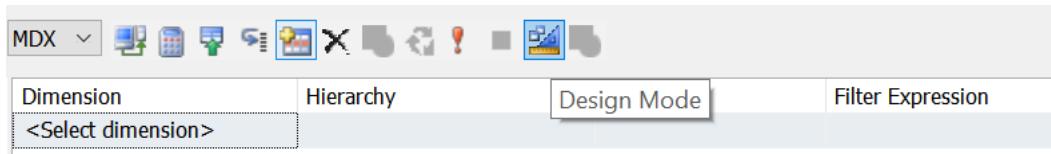
Kéo lần lượt các thuộc tính Market Segment trong bảng Market Segment, Country của bảng Customer và độ đo Days In Waiting List, CAU29 AVG DAY tạo từ Measure Group sang cửa sổ thực thi ta được kết quả câu truy vấn sau:



Hình 3.188. Kết quả câu truy vấn trong Visual Studio 2019

Chuyển từ Design Mode sang Script Mode bằng các nhấn chuột vào biểu tượng

Design Mode như hình minh họa:



*Hình 3.189. Design Mode trên thanh công cụ*

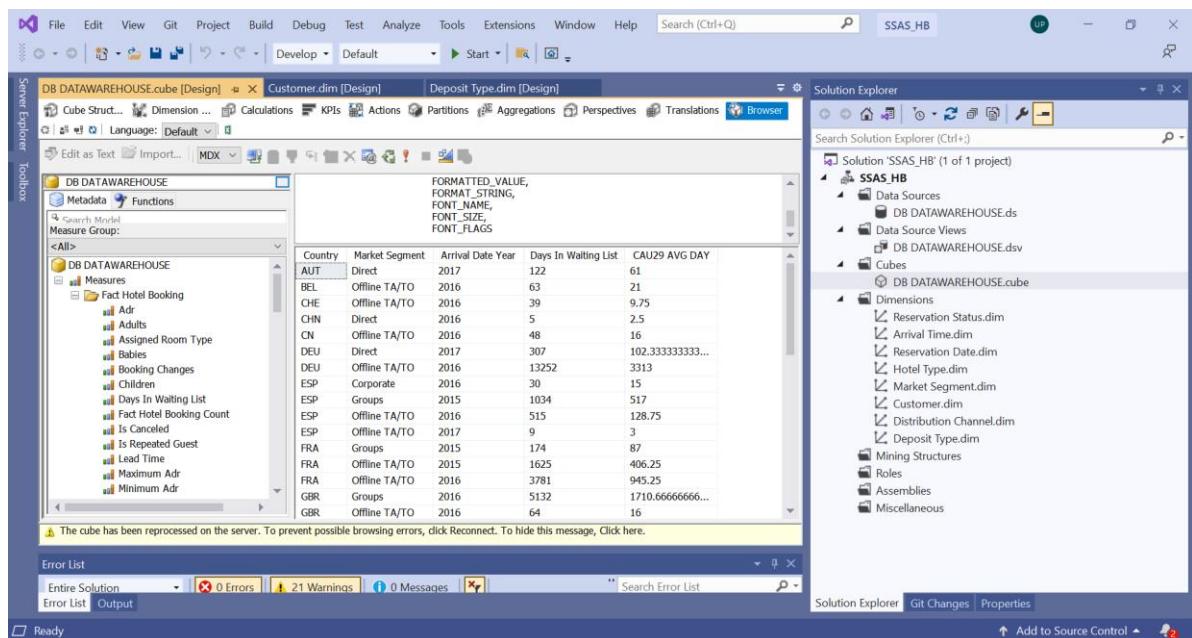
```
SELECT NON EMPTY {
    [Measures].[Days In Waiting List],
    [Measures].[CAU29 AVG DAY]
} ON COLUMNS,
NON EMPTY {
    (
        [Customer].[Country].[Country].ALLMEMBERS *
        [Market Segment].[Market Segment].[Market Segment].ALLMEMBERS
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS
```

*Hình 3.190. Câu lệnh truy vấn ban đầu*

```
SELECT NON EMPTY {
    [Measures].[Days In Waiting List],
    [Measures].[CAU29 AVG DAY]
} ON COLUMNS,
NON EMPTY {
    generate(
        [Customer].[Country].[Country],
        filter (
            [Customer].[Country].currentmember *
            [Market Segment].[Market Segment].[Market Segment] *
            [Arrival Time].[Arrival Date Year].[Arrival Date Year],
            [Arrival Time].[Arrival Date Year].currentmember = [Reservation Date].[Reservation Date Year] and
            [Market Segment].[Market Segment].currentmember <> [Market Segment].[Market Segment].[Complementary] and
            [Measures].[Days In Waiting List] > [Measures].[CAU29 AVG DAY] and
            [Measures].[CAU29 AVG DAY] > 0
        )
    )
}
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [DB DATAWAREHOUSE] CELL PROPERTIES VALUE,
BACK_COLOR,
FORE_COLOR,
FORMATTED_VALUE,
FORMAT_STRING,
FONT_NAME,
FONT_SIZE,
FONT_FLAGS
```

*Hình 3.191. Thay thế hàm vào câu lệnh truy vấn*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 3.191. Kết quả cuối cùng của câu truy vấn trong Visual Studio 2019

### 9.15.2. Công cụ Pivot Excel

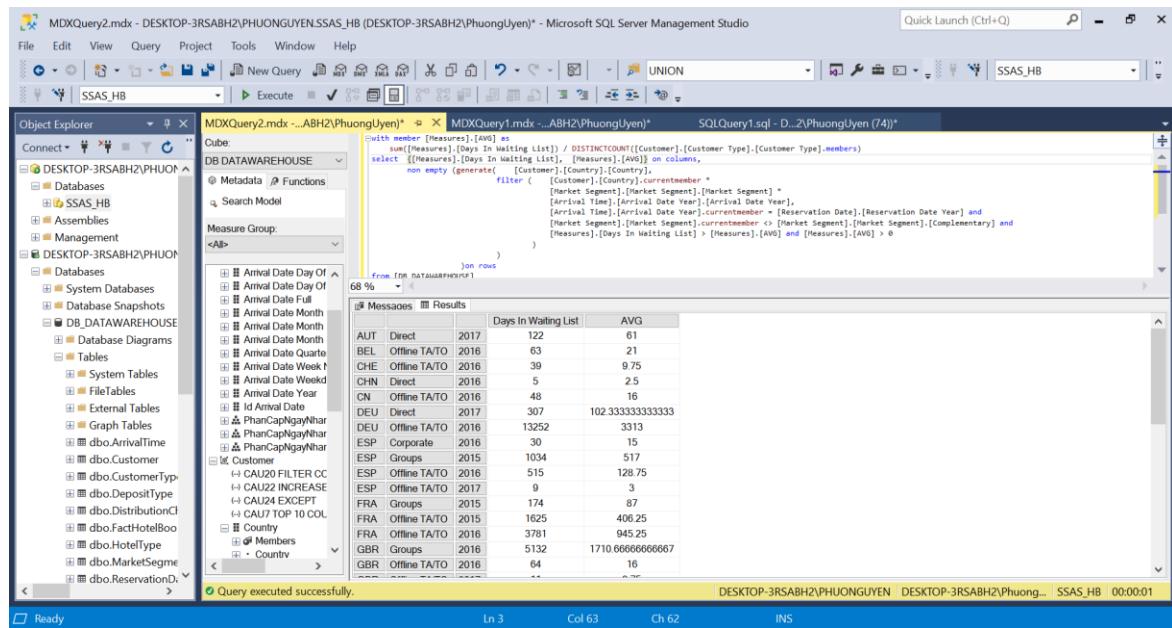
Kéo thả Days In Waiting List, CAU29 AVG DAY từ bảng Measures vào bảng Values, thuộc tính CAU29 FILTER thuộc bảng Measures vào bảng Rows ta được kết quả:

	Days In Waiting List	CAU29 AVG DAY
1 Row Labels	Days In Waiting List	CAU29 AVG DAY
2 AUT		
3 Direct		
4 2017	122	61
5 BEL		
6 Offline TA/TO		
7 2016	63	21
8 CHE		
9 Offline TA/TO		
10 2016	39	9.75
11 CHN		
12 Direct		
13 2016	5	2.5
14 CN		
15 Offline TA/TO		
16 2016	48	16
17 DEU		
18 Direct		
19 2017	307	102.333333
20 Offline TA/TO		
21 2016	13252	3313

Hình 3.192. Kết quả khi dùng công cụ Pivot Excel

### 9.15.3. Ngôn ngữ MDX

Sử dụng câu lệnh truy vấn MDX trong Microsoft Sql Server Management Studio 18 ta được:



Hình 3.193. Kết quả câu truy vấn khi dùng ngôn ngữ MDX

❖ Giải thích câu truy vấn MDX:

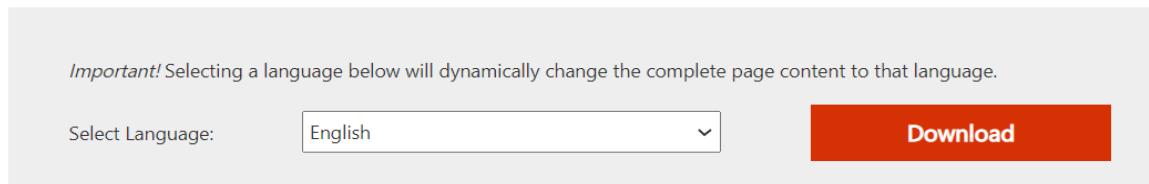
- Tạo độ đo có tên AVG dung để tính trung bình ngày chờ dựa trên từng loại khách hàng.
  - Dùng operator “\*” để tích chéo các giá trị trong thuộc tính phân khúc thị trường (Market Segment) với lần lượt thuộc tính Arrival Date Year (năm khách hàng nhận phòng) trong điều kiện năm nhận phòng cùng năm khách hàng thực hiện giao dịch lần cuối, phân khúc thị trường khác “Complement”, số ngày chờ đến khi khách sạn liên lạc lại để trao đổi về thông tin đặt phòng phải lớn hơn số ngày trung bình và số ngày trung bình phải lớn hơn 0.
  - Hàm Generate dùng để duyệt từng giá trị trong Country ở đối số thứ nhất. Tại đối số thứ hai, hàm Generate lấy ra từng quốc tịch khách hàng hiện tại thông qua CurrentMember trong hàm Filter.

## **CHƯƠNG IV. QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)**

### **1. Cài đặt Microsoft Sql Server 2019 Reporting Services**

- Download công cụ Microsoft SQL Server 2019 Reporting Services.

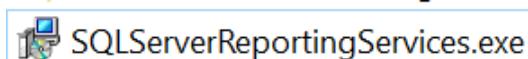
Microsoft SQL Server 2019 Reporting Services



*Hình 4.1: Giao diện cài đặt công cụ Microsoft SQL Server 2019 Reporting Services.*

- Cài đặt công cụ Microsoft SQL Server 2019 Reporting Services.

- Sau khi tải thành công ta có được file exe như sau:



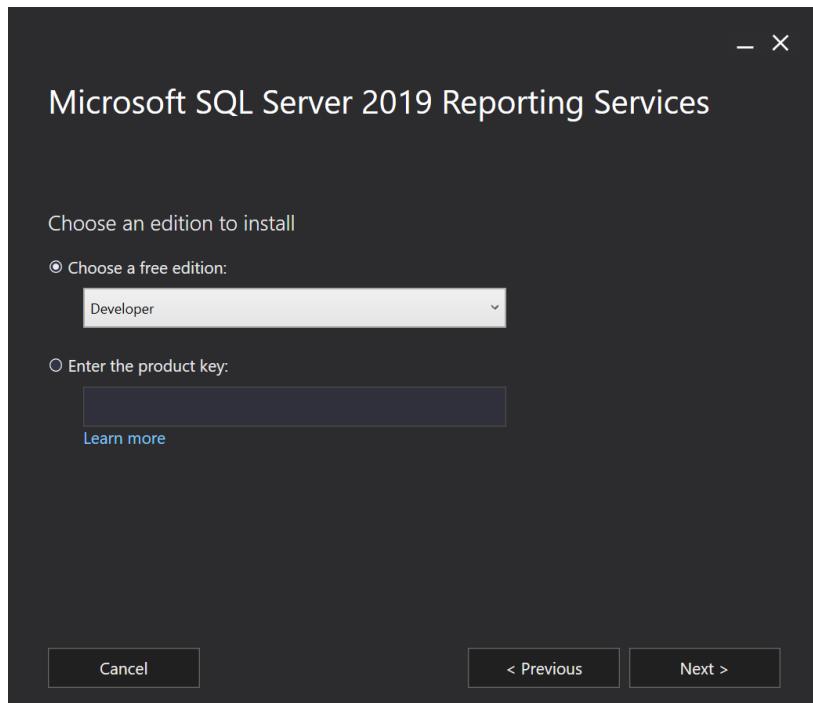
*Hình 4.2 : File cài đặt SQLServerReportingServices.*

- Bước 1: Click chuột phải chọn Install ta được giao diện cài đặt Microsoft SQL Server 2019 Reporting Services.



*Hình 4.3: Giao diện cài đặt Microsoft SQL Server 2019 Reporting Services.*

- Bước 2: Tại trang Choose an edition to install, chọn choose a free edition và tiếp tục nhấn Developer và nhấn Next.



*Hình 4.4: Giao diện màn hình Choose a free edition.*

- Bước 3: Tại trang giao diện Review the license terms, chọn I accept the license terms và nhấn Next:



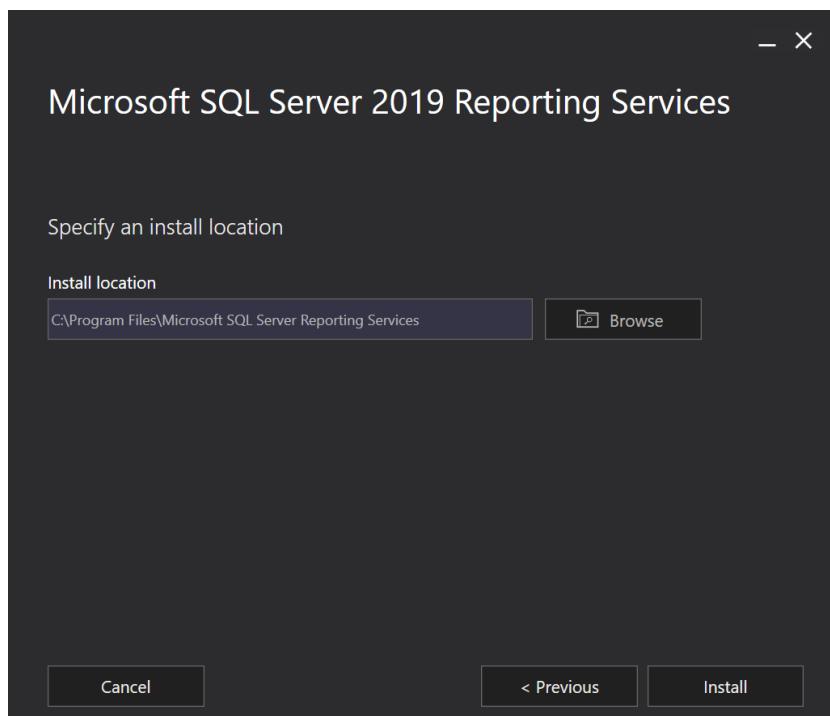
*Hình 4.5: Giao diện màn hình Review the license terms.*

- Bước 4: Tiếp tục nhấp Next:



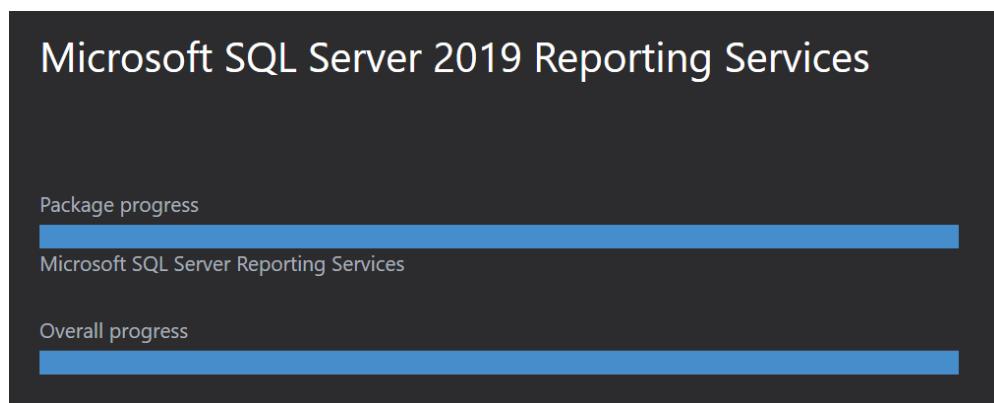
*Hình 4.6 : Giao diện màn hình Install Database Engine.*

- Bước 5: Nhấn Next để qua trang kế tiếp.



*Hình 4.7: Giao diện màn hình Specify an install location.*

- Bước 6: Chọn install để tiến hành cài đặt công cụ.



*Hình 4.8: Giao diện khi tiến hành cài đặt của Microsoft.*

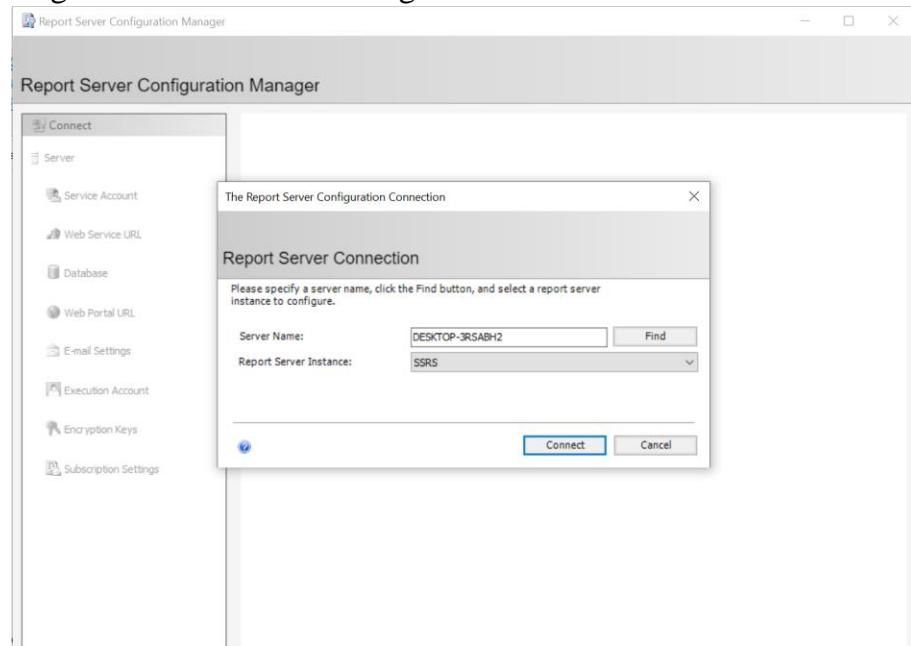
- Bước 7: Sau khi cài đặt thành công, click chọn Configure Report Server để mở giao diện Setup Completed.



*Hình 4. 9: Kết quả cài đặt thành công.*

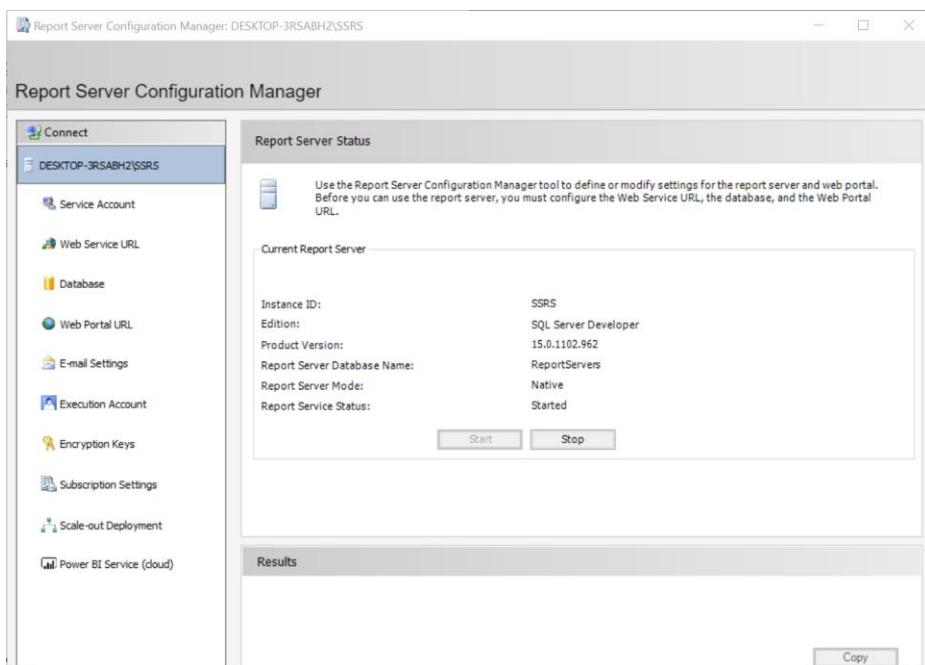
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 8: Các bước cấu hình Report Server Configuration Manager.
  - Bước 8.1: Chọn Configure report server hoặc mở công cụ Report Server Configuration Connection trong thanh tìm kiếm.



*Hình 4.10: Giao diện sau khi cài đặt thành công Microsoft SQL Server 2019 Reporting Services.*

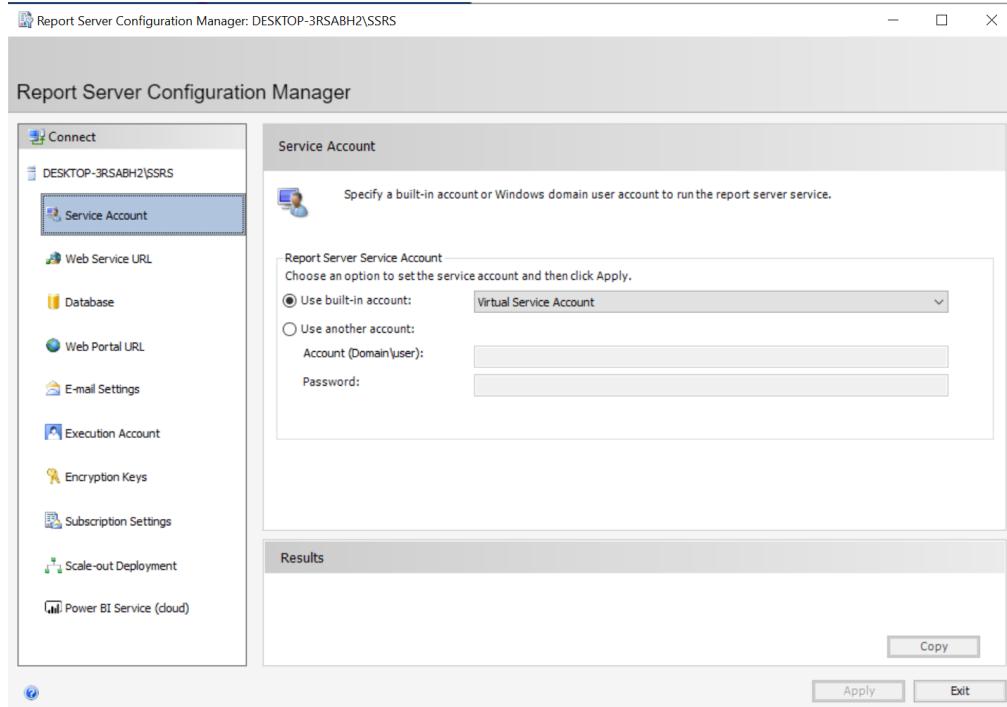
- Bước 8.2: Giao diện tổng thể của công cụ Report Server Configuration Manager.



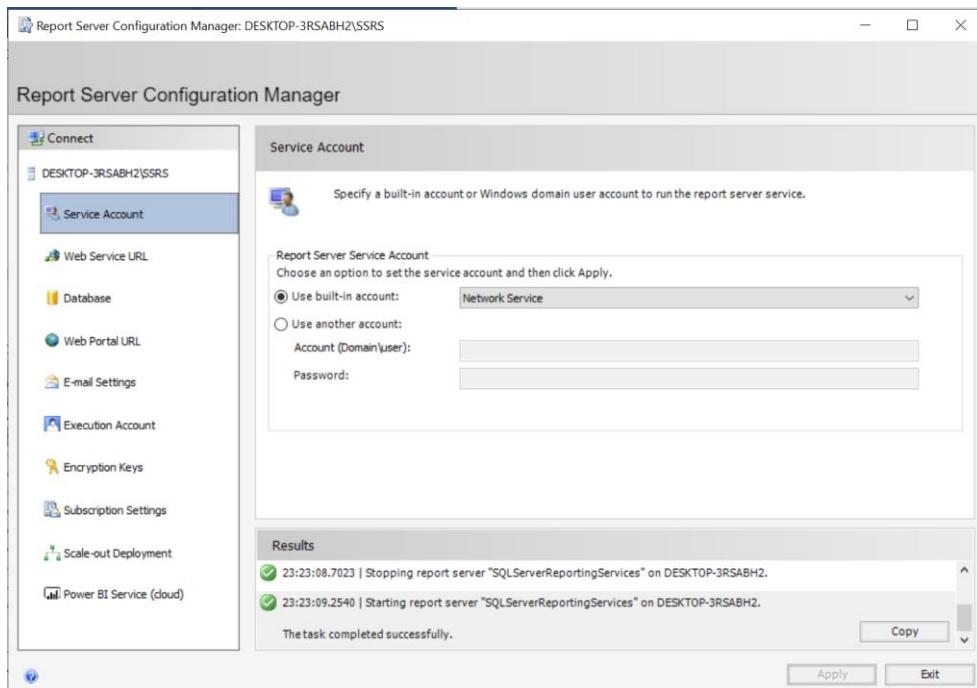
*Hình 4.11: Giao diện Report Server Configuration Manager sau khi kết nối thành công.*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 8.3:Tại tab Service Account, chuyển đổi giá trị trong Use a built – in account, chọn Apply, ta thu được kết quả như sau:

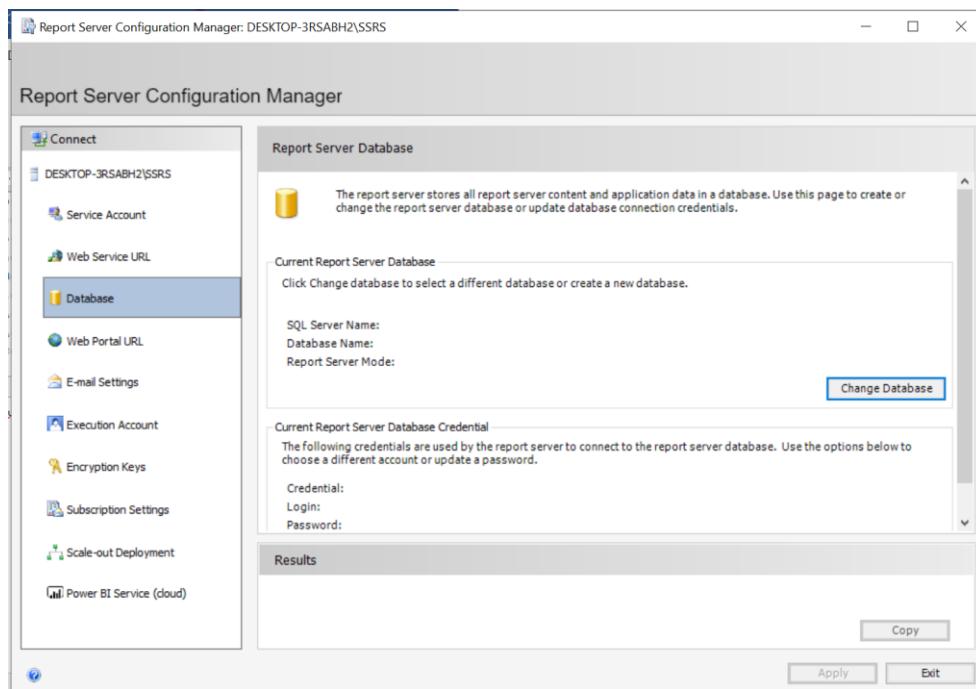


*Hình 4.12: Giao diện tab Service Account ban đầu.*



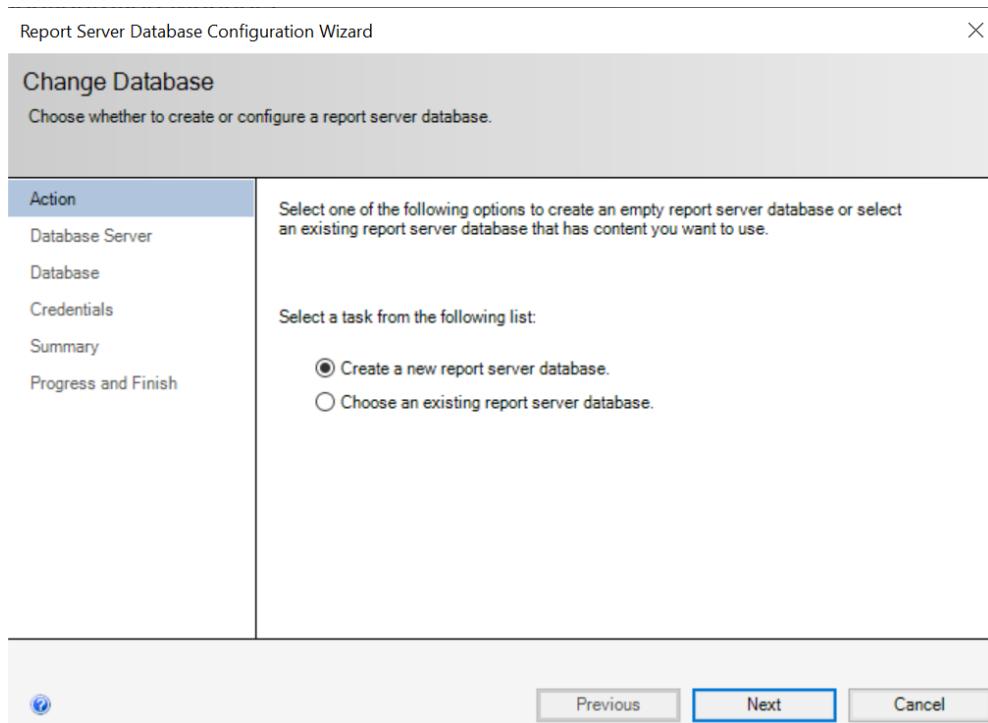
*Hình 4.13: Thay đổi tài khoản ban đầu thành Network Service và nhấn Apply để hoàn tất.*

- **Bước 8.4:** Tại tab Database, tạo mới một dataset bằng cách chọn Change Database.



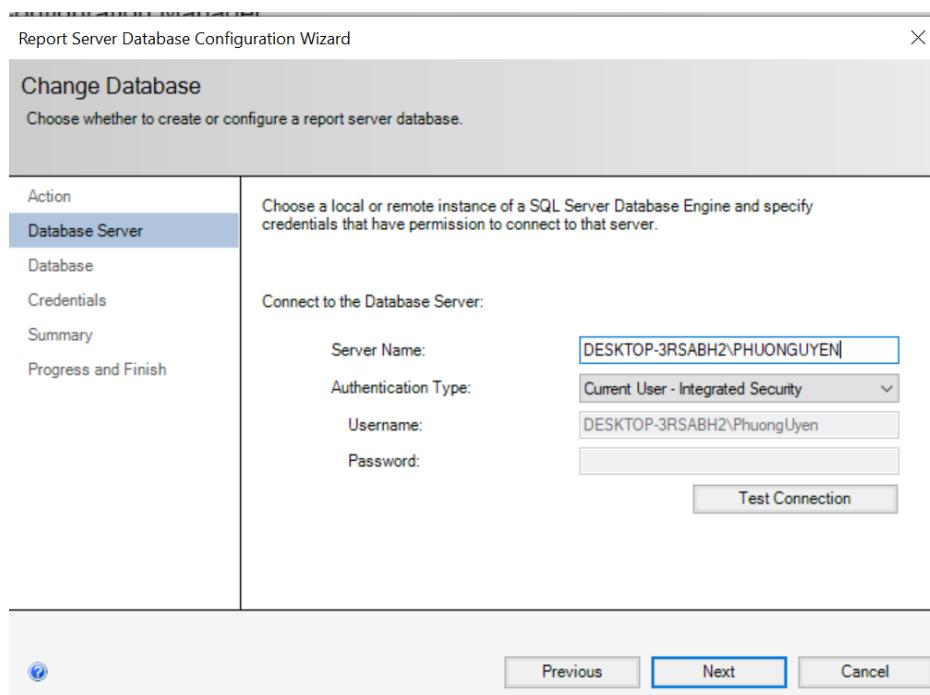
*Hình 4.14: Giao diện tab Database của Report Server Configuration Manager.*

- **Bước 8.5:** Chọn Create a new report server database và nhấn Next.



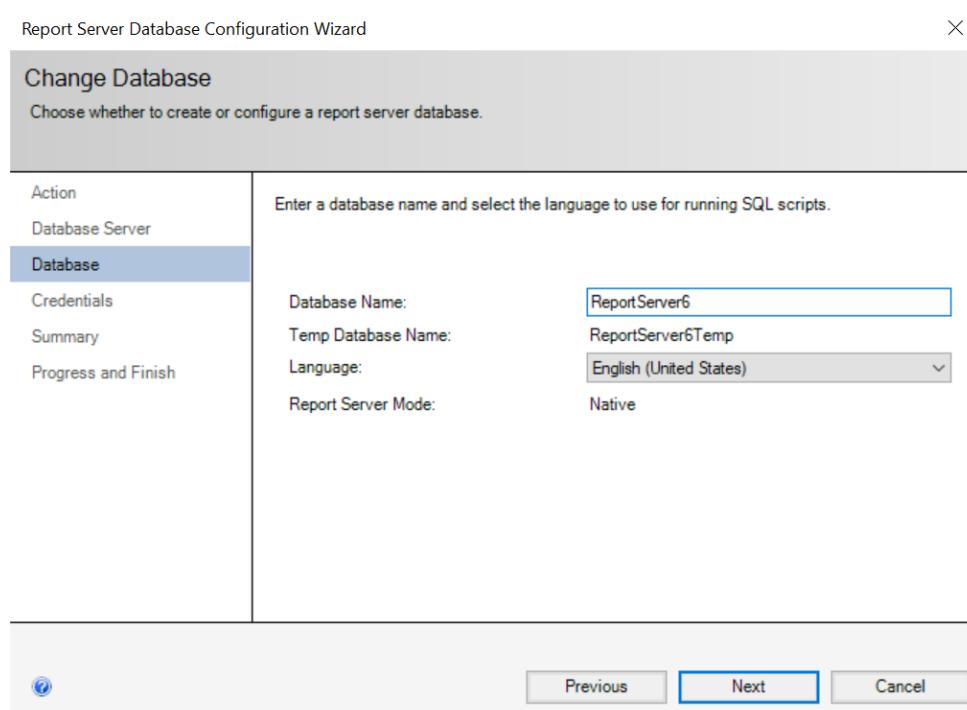
*Hình 4.15: Chọn Create a new report server database và nhấn Next.*

- **Bước 8.6:** Điền vào Server Name phù hợp và nhấn Next.



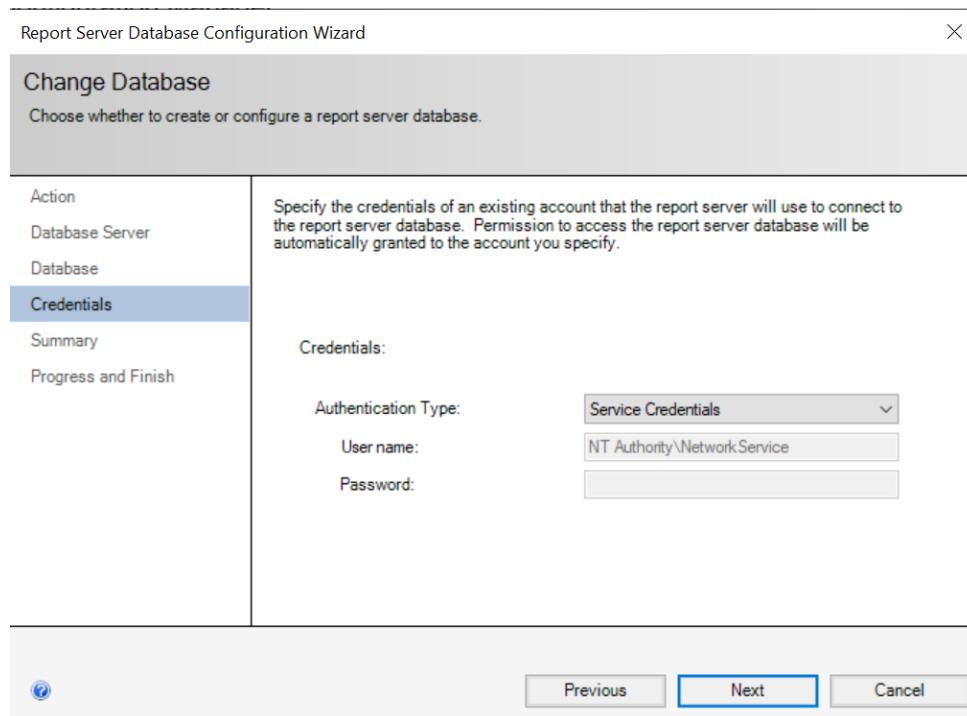
*Hình 4.16: Điền Server Name phù hợp và nhấn Next.*

- **Bước 8.7:** Hình: Đặt tên cho Database và nhấn Next.



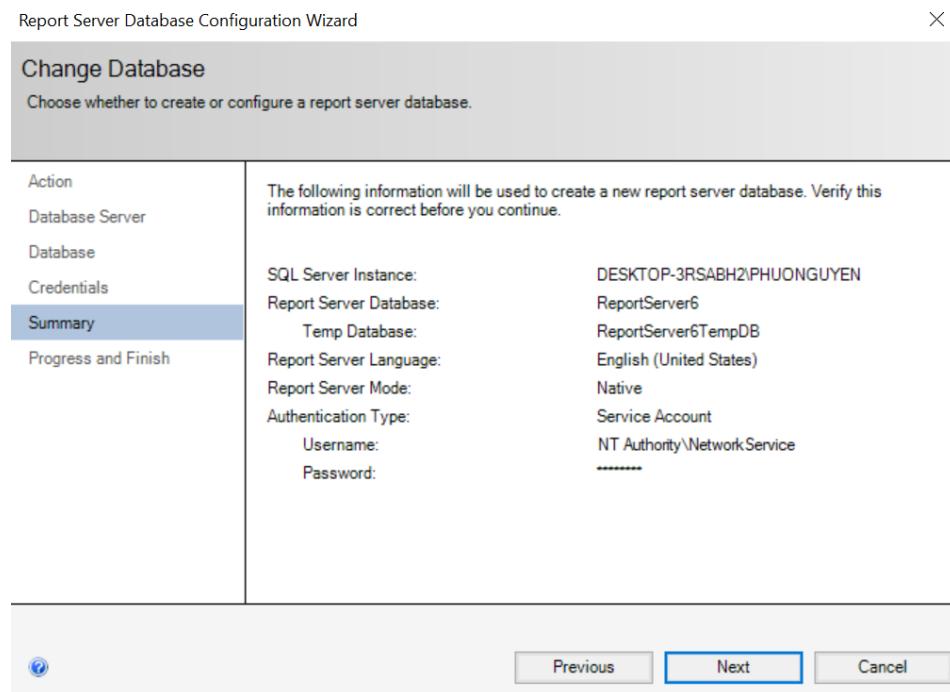
*Hình 4.17: Đặt tên cho Database và nhấn Next.*

- Bước 8.8: Trong tab Credentials tiếp tục nhấn Next.



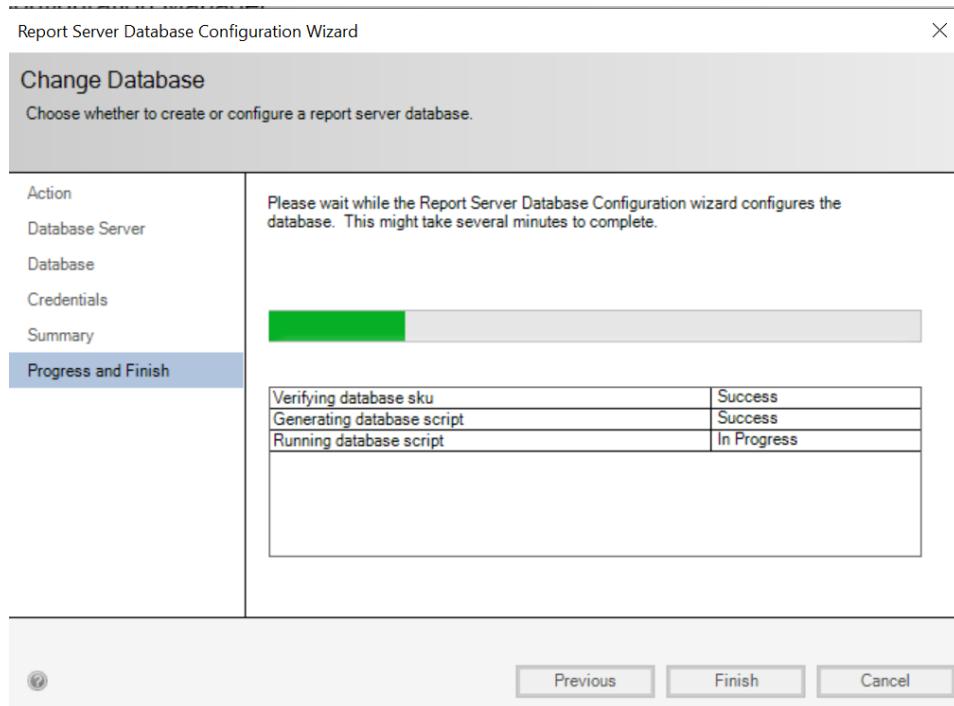
*Hình 4.18: Trong tab Credentials tiếp tục nhấn Next.*

- Bước 8.9: Trong tab Summary, tiếp tục nhấn Next.



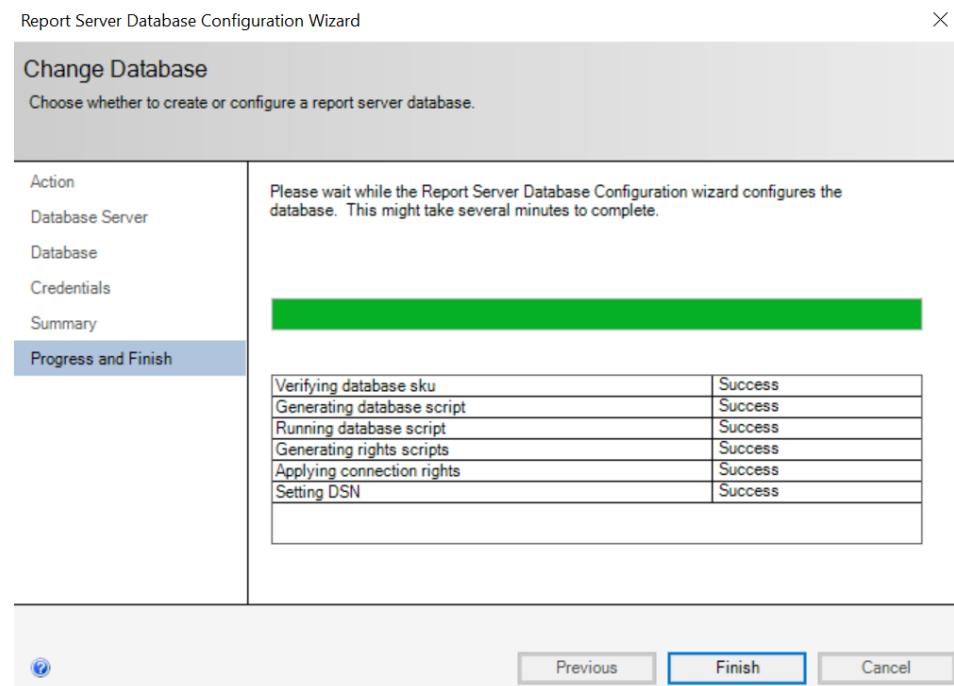
*Hình 4.19: Trong tab Summary, tiếp tục nhấn Next.*

▪ Bước 8.10: Hệ thống tiến hành tạo Database.



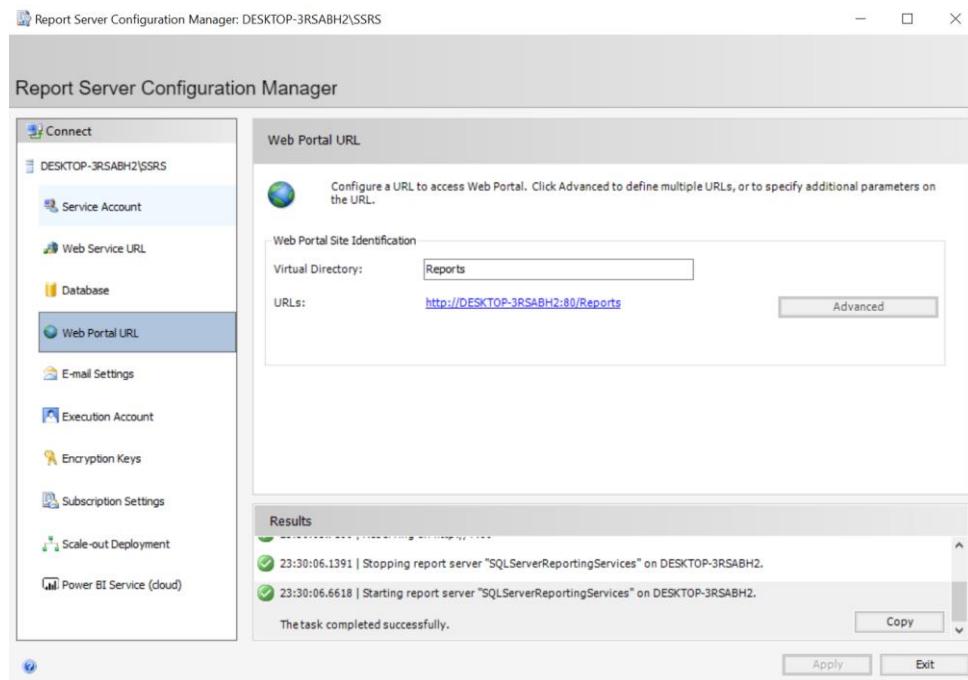
*Hình 4.20: Tiến hành tạo Database.*

▪ Bước 8.11: Tạo database thành công.



*Hình 4.21: Tạo database thành công.*

- Bước 8.12: Tại giao diện tab Web Portal URL, chọn Apply để xây dựng trang web.

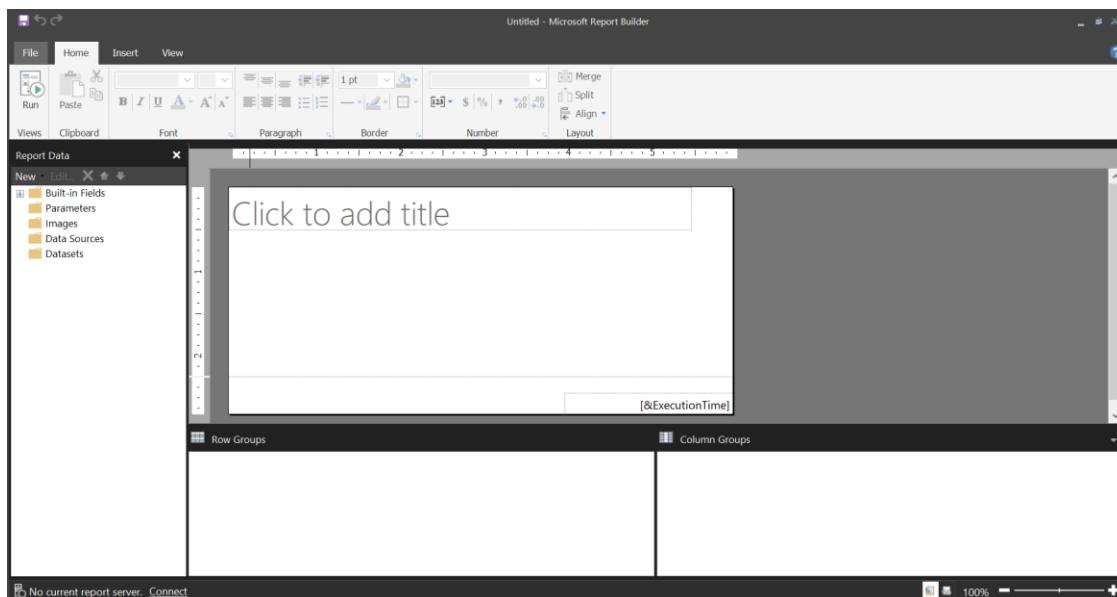


*Hình 4.22: Giao diện tab Web Portal URL.*

## 2. Quá trình lập báo biểu bằng công cụ Report Builder.

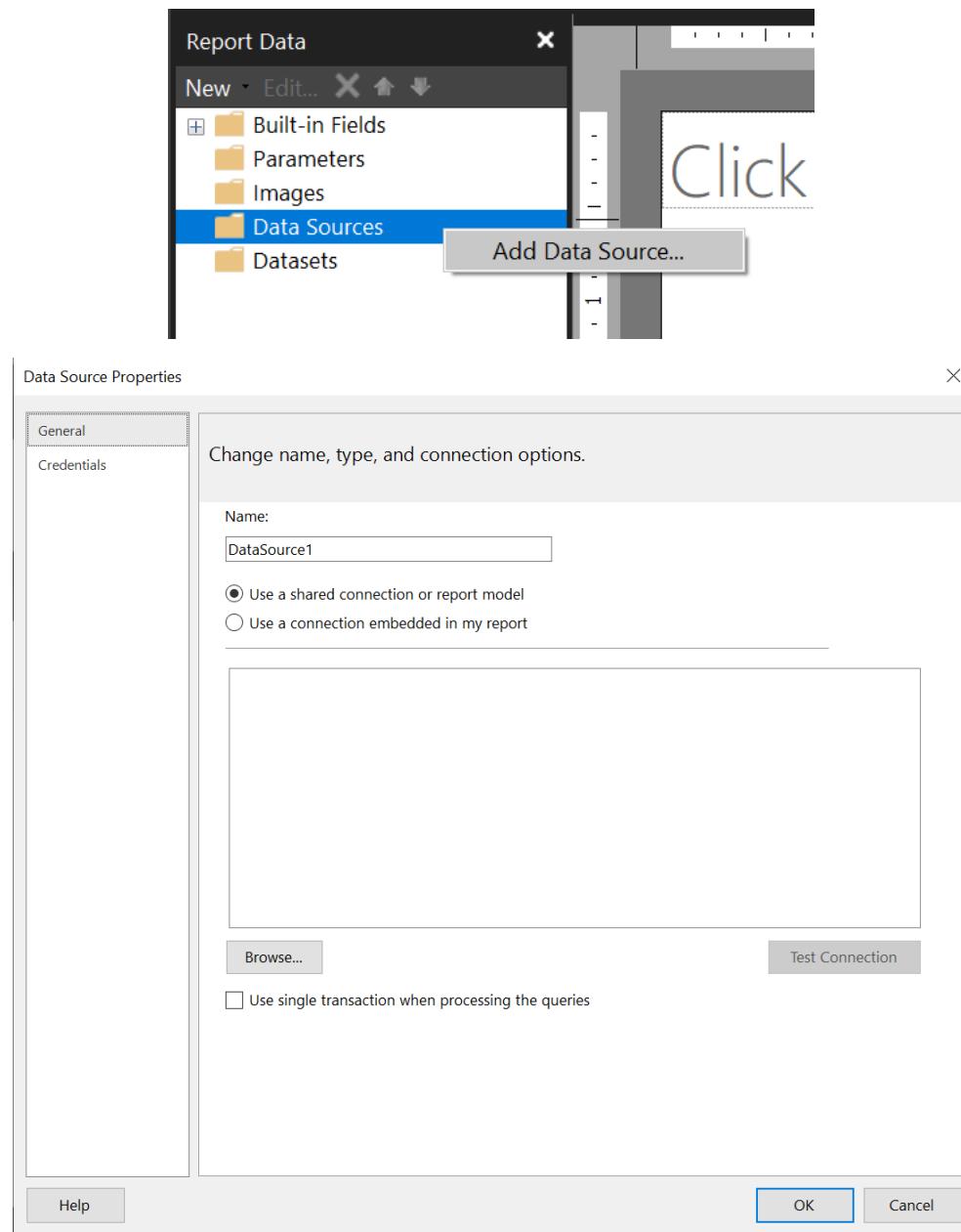
### 2.1. Tạo kết nối với kho dữ liệu.

- Giao diện công cụ Report Builder khi cài đặt thành công:



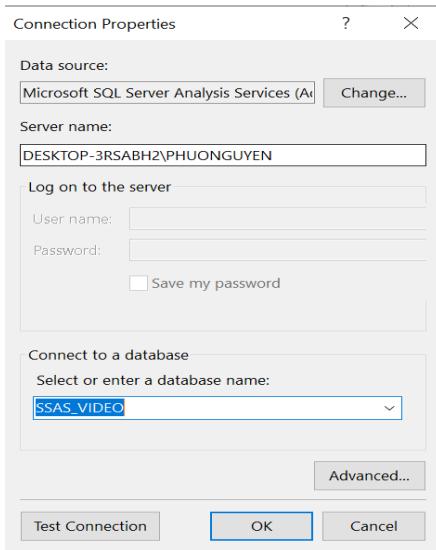
*Hình 4.23: Giao diện công cụ Report Builder.*

- Bước 1: Tại thanh Report Data → Chọn Data Source → chọn add data source như hình minh họa:



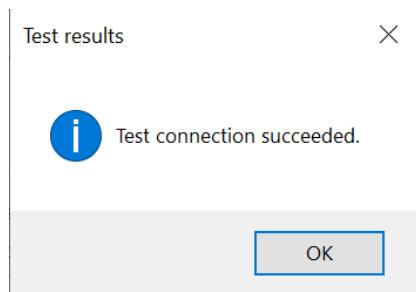
*Hình 4.24,25: Thêm Data Source*

- Bước 2: Trường Type chọn Microsoft SQL Server Analysis Service → Chọn Type và cấu hình như hình minh họa (tùy tên) → rồi nhấn Test Connection. Nếu thành công nhấn OK → nhấn OK.



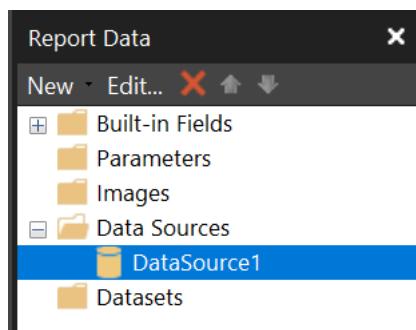
*Hình 4.26: Chọn kết nối đến project SSAS đã tạo*

- Bước 3: Click Test Connection để kiểm tra kết nối.



*Hình 4.27 : Kết nối Project SSAS thành công*

- Bước 4: Nhấn Ok để hoàn tất quá trình kết nối Data Source.

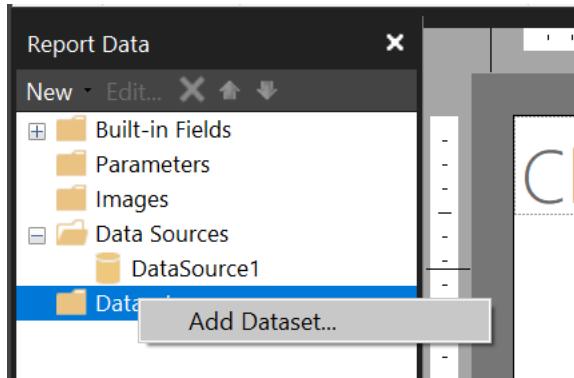


*Hình 4.28: Tạo thành công kết nối với Data Sources*

## 2.2. Thực hiện báo biểu bảng công cụ Report Builder.

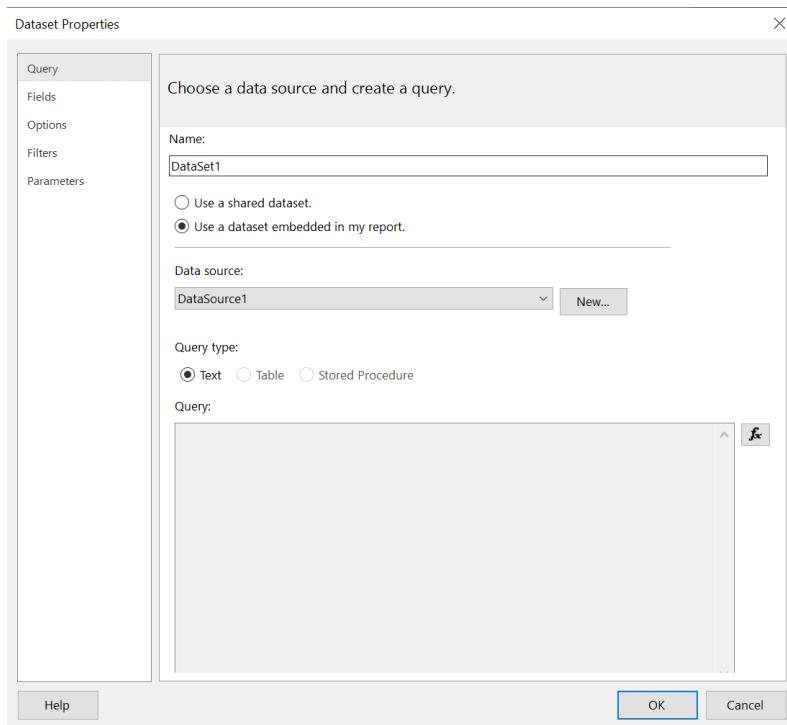
### 2.2.1. Báo biểu 1: Xây dựng báo cáo doanh thu khách sạn từ năm 2015 đến năm 2017.

- Bước 1: Thêm datasets cần thiết cho báo biểu 1 đặt tên là datasets1, bằng cách nhấp chuột phải vào Datasets → Add new dataset...



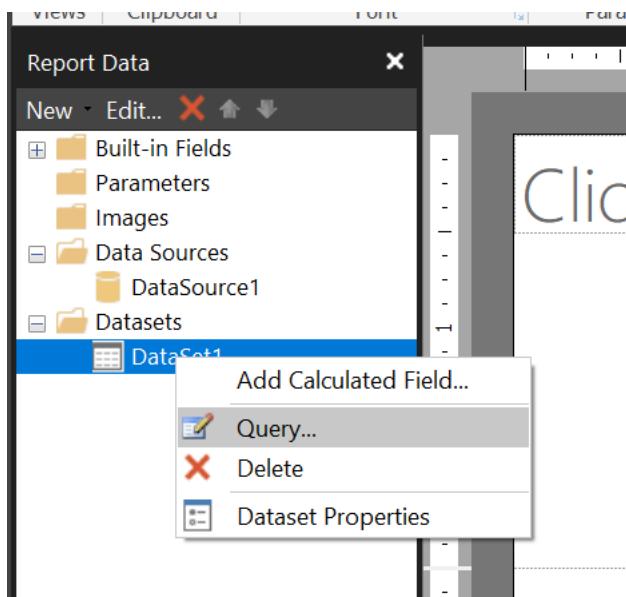
*Hình 4. 29: Thêm datasets cho câu báo biểu 1*

- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường Name đặt tên là Dataset1 chọn chế độ Use a dataset embedded in my report. Trường Data Source chọn data source tại phần tạo kết nối data source.

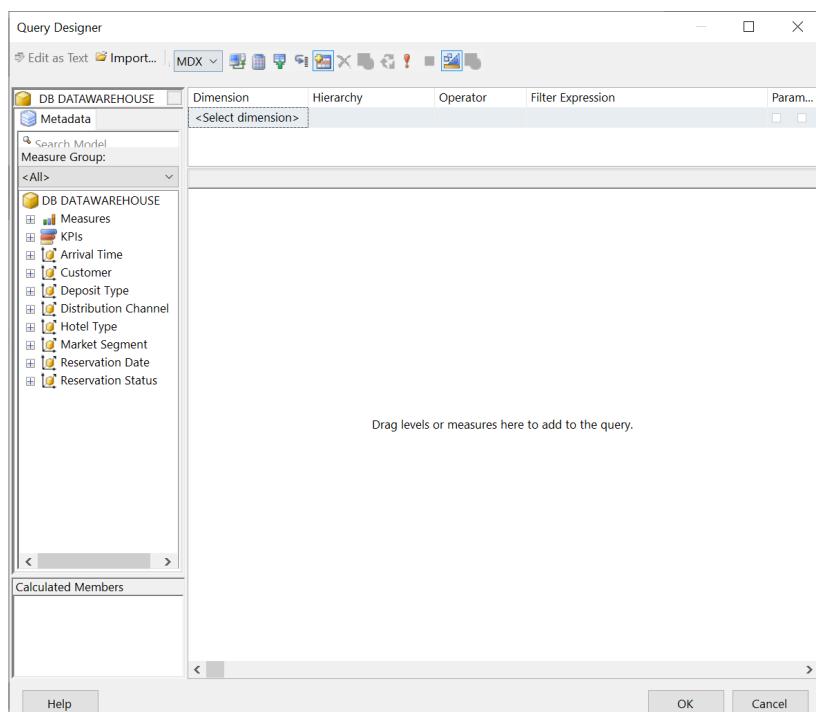


*Hình 4.30: Điều chỉnh tên, datasource cho báo biểu 1*

- Bước 3: Click chuột phải và chọn Query hộp thoại sẽ hiện ra:

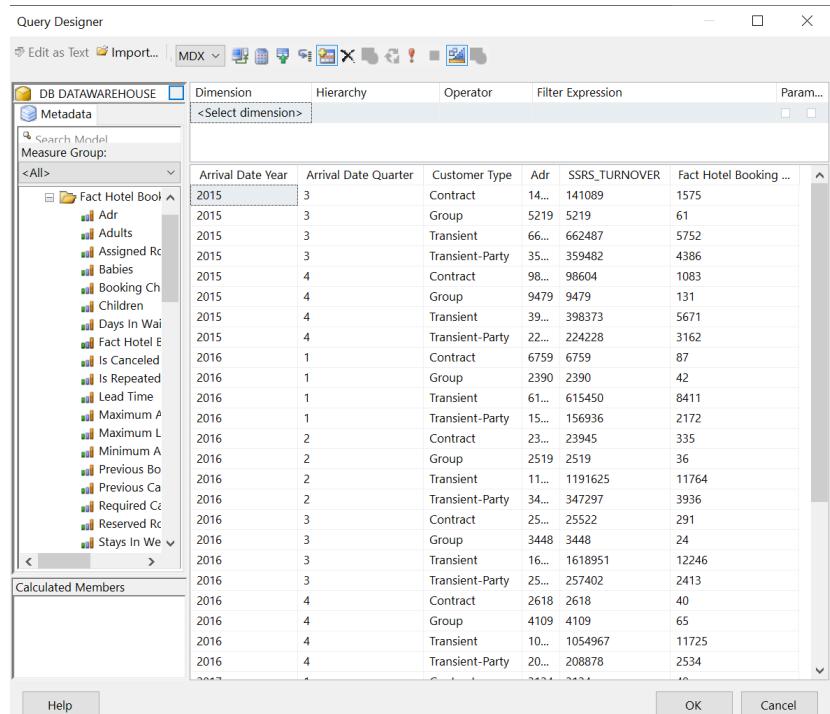


*Hình 4.31: Thêm query cho tập dataset.*



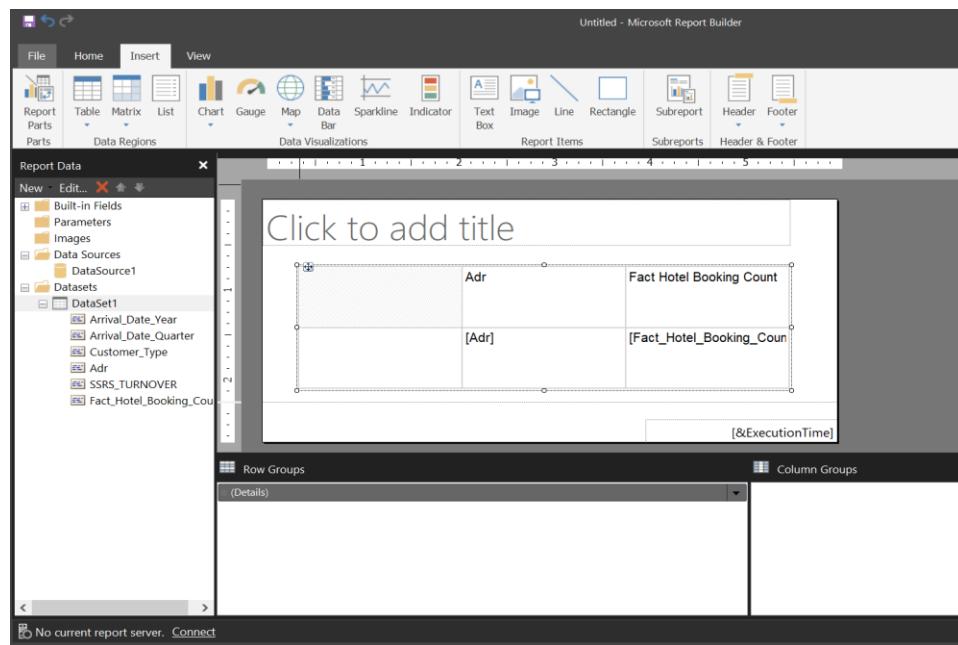
*Hình 4.32: Cửa sổ làm việc Query Designer*

- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.



*Hình 4.33: Chọn những thuộc tính và độ đo cần thiết cho báo biểu 1*

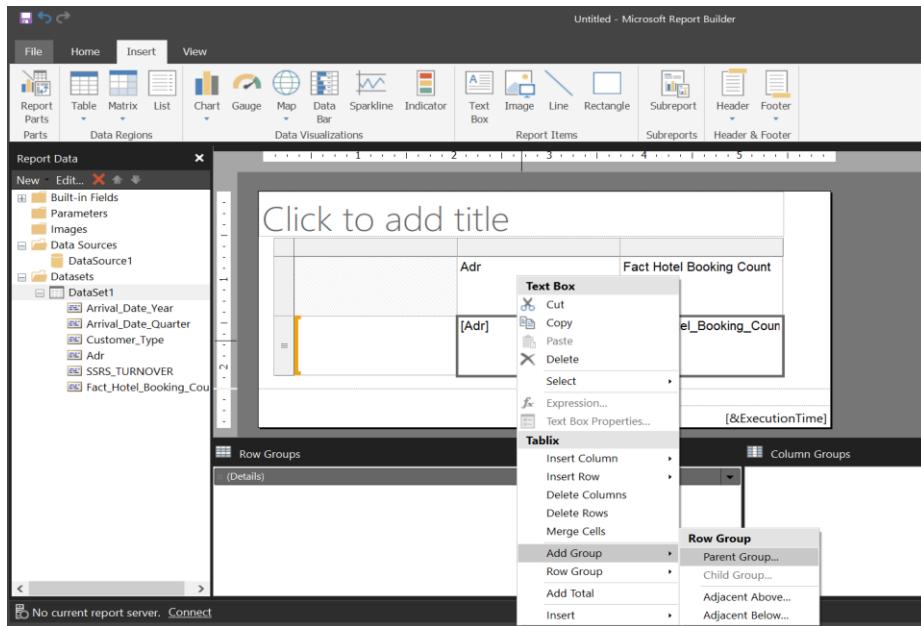
- Bước 5: Insert một Table vào với tùy chọn là Insert Table và kéo table qua vùng thiết kế (Design), kéo những thuộc tính vào bảng table như hình minh họa:



*Hình 4.34: Kéo thả những thuộc tính cho câu báo biểu 1 vào bảng design*

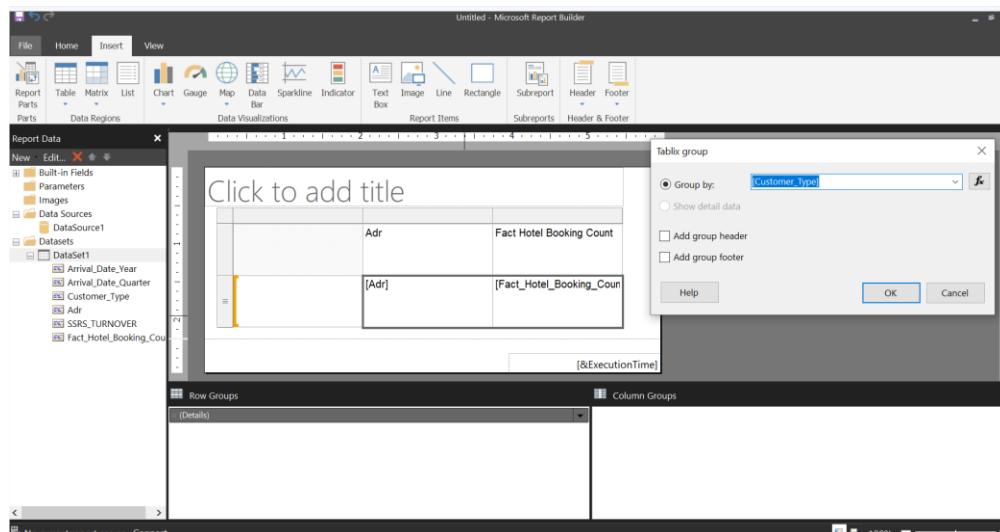
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 6: Tạo Grouping bằng cách tô đen thuộc các thuộc tính muốn được gom nhóm, chọn Add Group để gom nhóm dữ liệu.
- Bước 6.1: Gom nhóm dữ liệu Adr và Fact Hotel Booking Count.



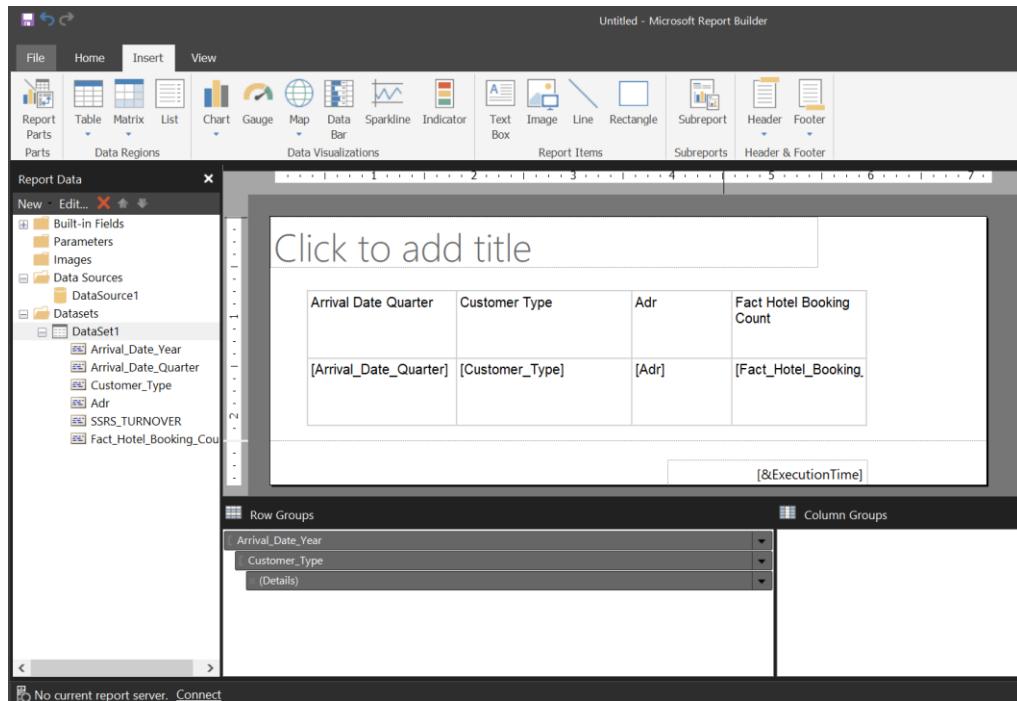
Hình 4.35: Gom nhóm dữ liệu Adr và Fact Hotel Booking Count.

- Bước 6.2: Gom nhóm theo các độ đo theo loại khách hàng (Customer Type).



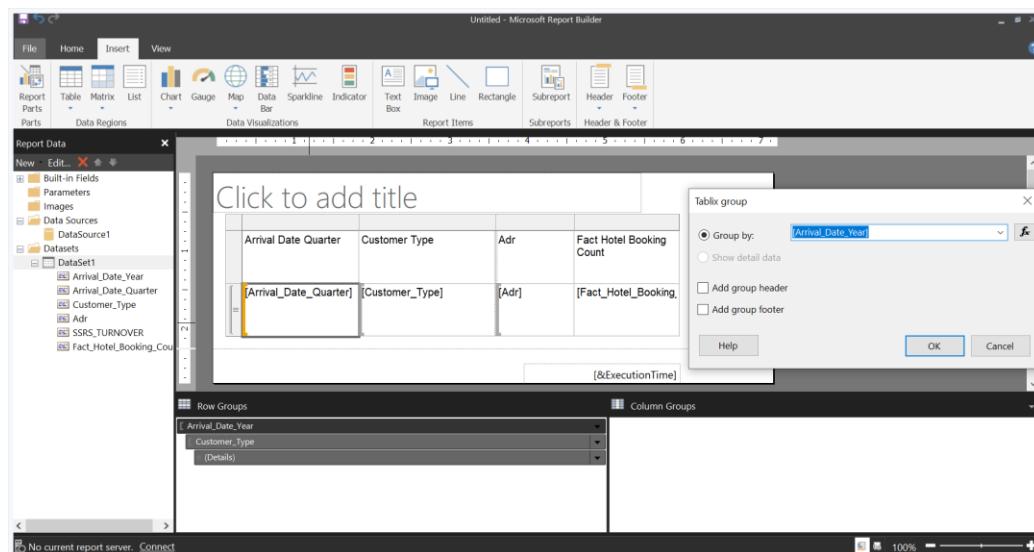
Hình 4.36: Gom nhóm theo các độ đo theo loại khách hàng (Customer Type)

▪ Bước 6.3: Gom nhóm Customer Type theo quý (Arrival\_Date\_Quarter).



Hình 4.37: Gom nhóm Customer Type theo quý (Arrival\_Date\_Quarter)

▪ Bước 6.4: Gom nhóm quý nhận phòng (Arrival\_Date\_Quarter) theo năm nhận phòng (Arrival\_Date\_Year).



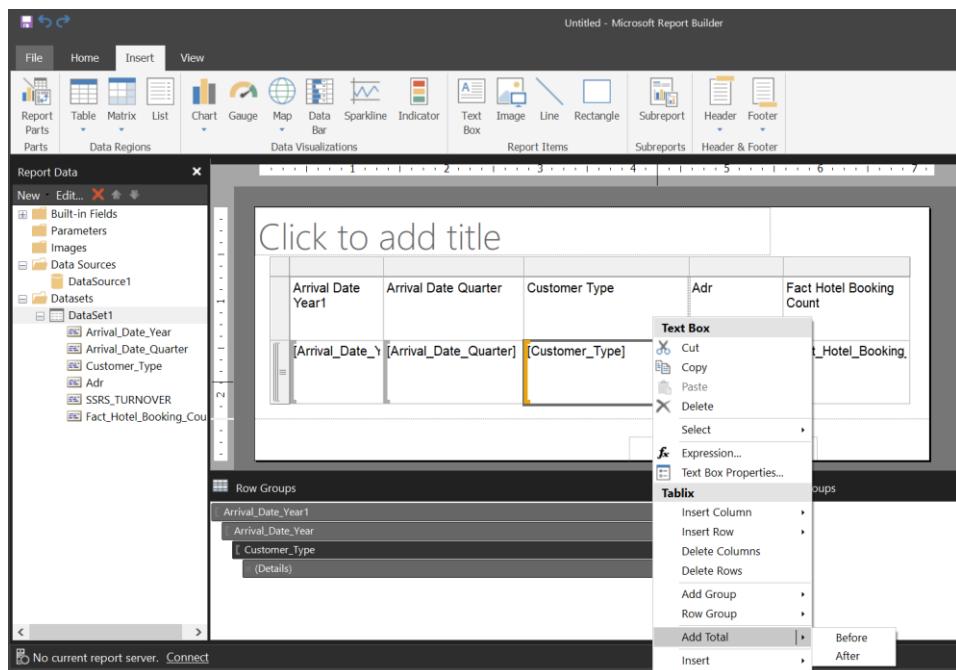
Hình 4.38: Gom nhóm quý nhận phòng (Arrival\_Date\_Quarter) theo năm nhận phòng (Arrival\_Date\_Year).

▪ Bước 6.5: Hoàn tất quá trình gom nhóm.

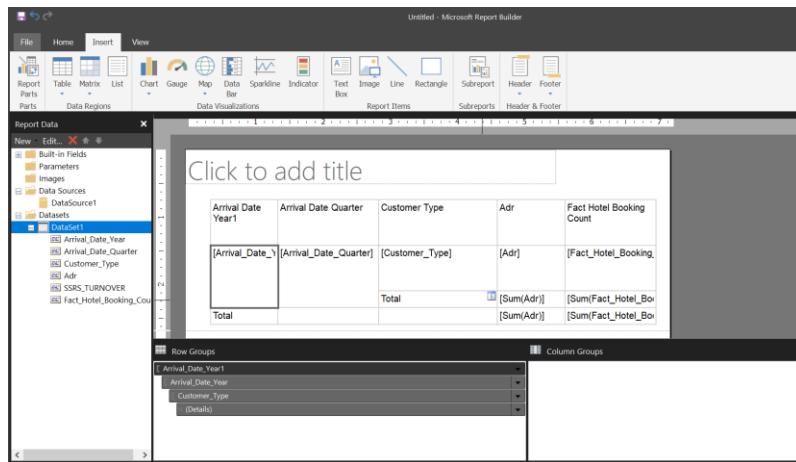
The screenshot shows a Microsoft Report Builder interface with a table titled "Click to add title". The table has five columns: "Arrival Date Year1", "Arrival Date Quarter", "Customer Type", "Adr", and "Fact Hotel Booking Count". A summary row is added at the bottom, containing fields: "[Arrival\_Date\_Year1]", "[Arrival\_Date\_Quarter]", "[Customer\_Type]", "[Adr]", and "[Fact\_Hotel\_Booking\_Count]". The "Customer Type" column is highlighted with a yellow border. The report also includes a header section with the text "[&ExecutionTime]".

Hình 4.39: Kết quả của quá trình gom nhóm.

- Bước 7: Tạo Total bằng cách chọn vào thuộc tính mà ta muốn tính tổng các độ đo theo thuộc tính đó, chọn Add Total để tính tổng các giá trị.



Hình 4.40: Tính tổng giá trị dữ liệu Adr và Fact Hotel Booking Count.

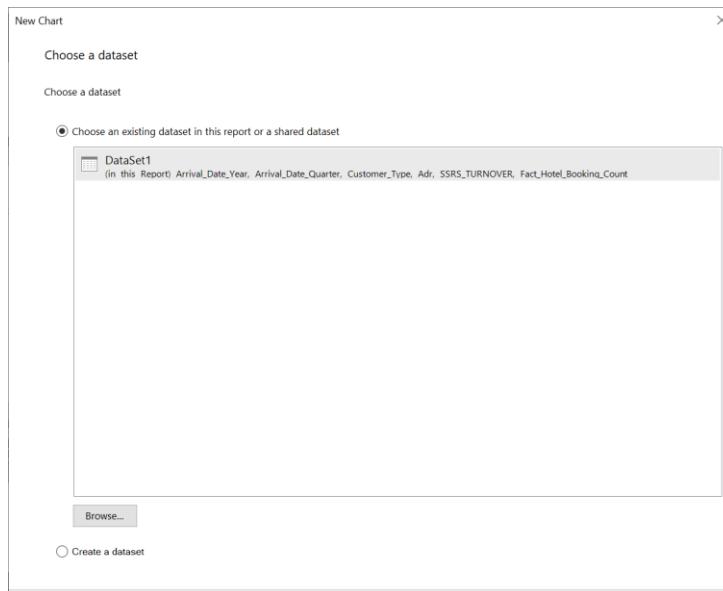


*Hình 4.41: Tính tổng quý nhận phòng (Arrival\_Date\_Quarter) theo năm nhận phòng (Arrival\_Date\_Year)*

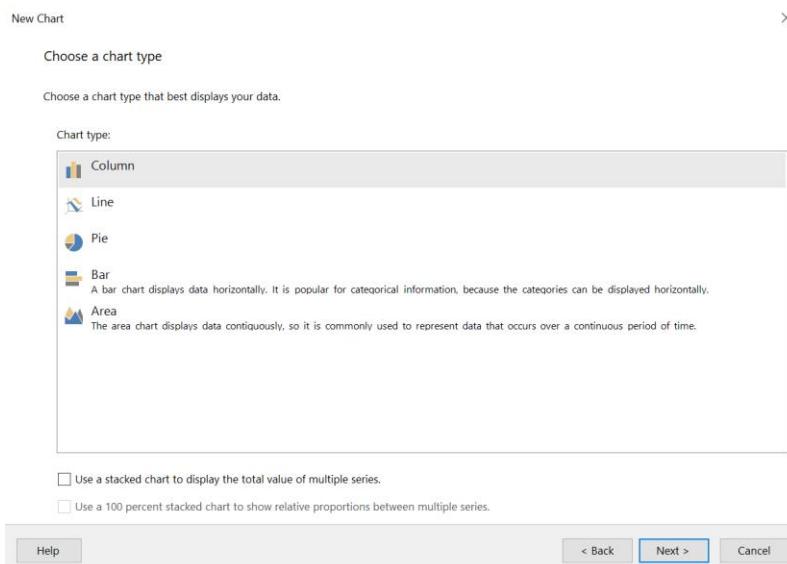
Arrival Date Year1	Arrival Date Quarter	Customer Type	Adr	Fact Hotel Booking Count
[Arrival_Date_Y]	[Arrival_Date_Quarter]	[Customer_Type]	[Adr]	[Fact_Hotel_Booking_
		Total	[Sum(Adr)]	[Sum(Fact_Hotel_Boi]

*Hình 4.42: Kết quả của quá trình gom nhóm.*

- Bước 8: Để thống kê doanh thu của từng khách hàng theo từng năm, ta chọn Insert một Chart → Chọn Chart Wizard → Chọn lược đồ dạng Column.

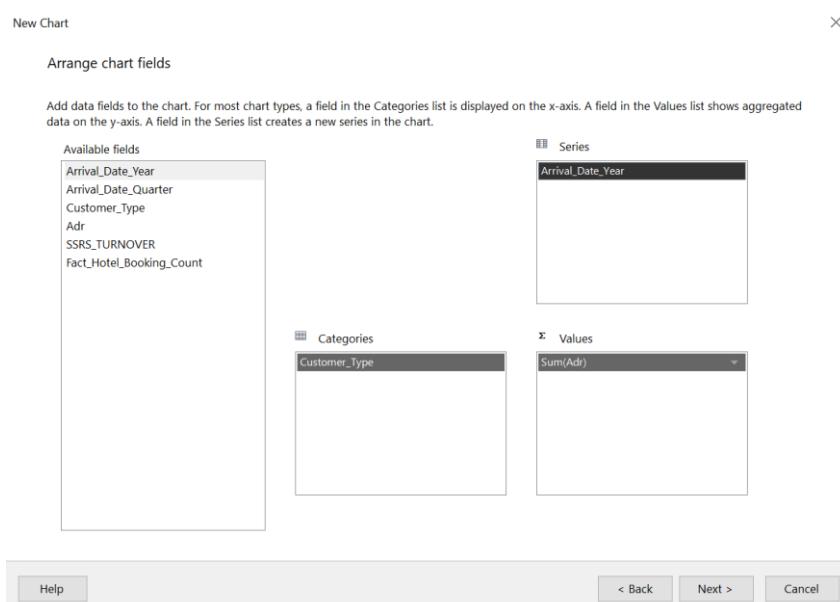


*Hình 4.43 : Màn hình lựa chọn Dataset.*

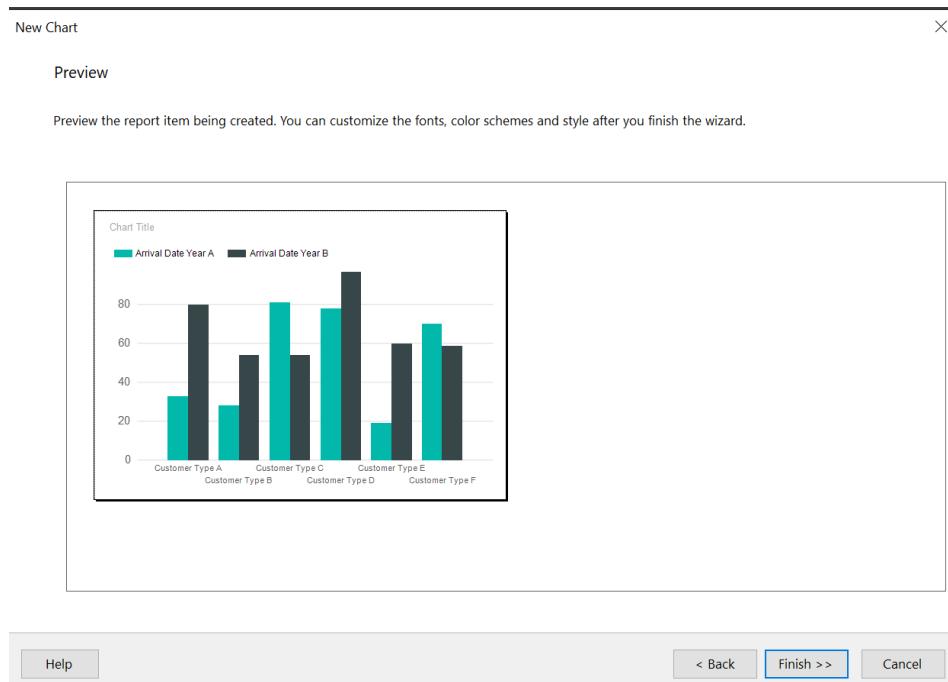


*Hình 4.44: Màn hình lựa chọn Chart*

- Bước 9: Kéo thả các thuộc tính và độ đo vào trường phù hợp.

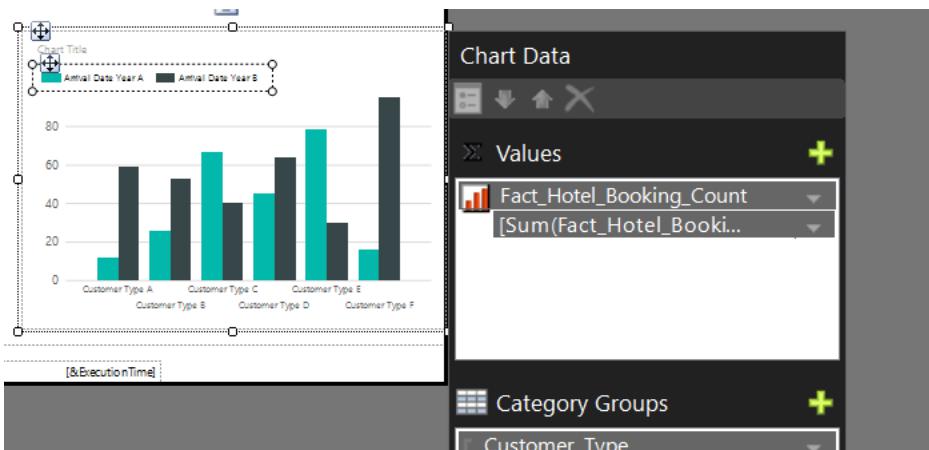


*Hình 4.45: Kéo thả Arrival\_Date\_Year vào ô Series, độ đo Adr vào ô Values và thuộc tính Customer\_Type vào ô Categories.*



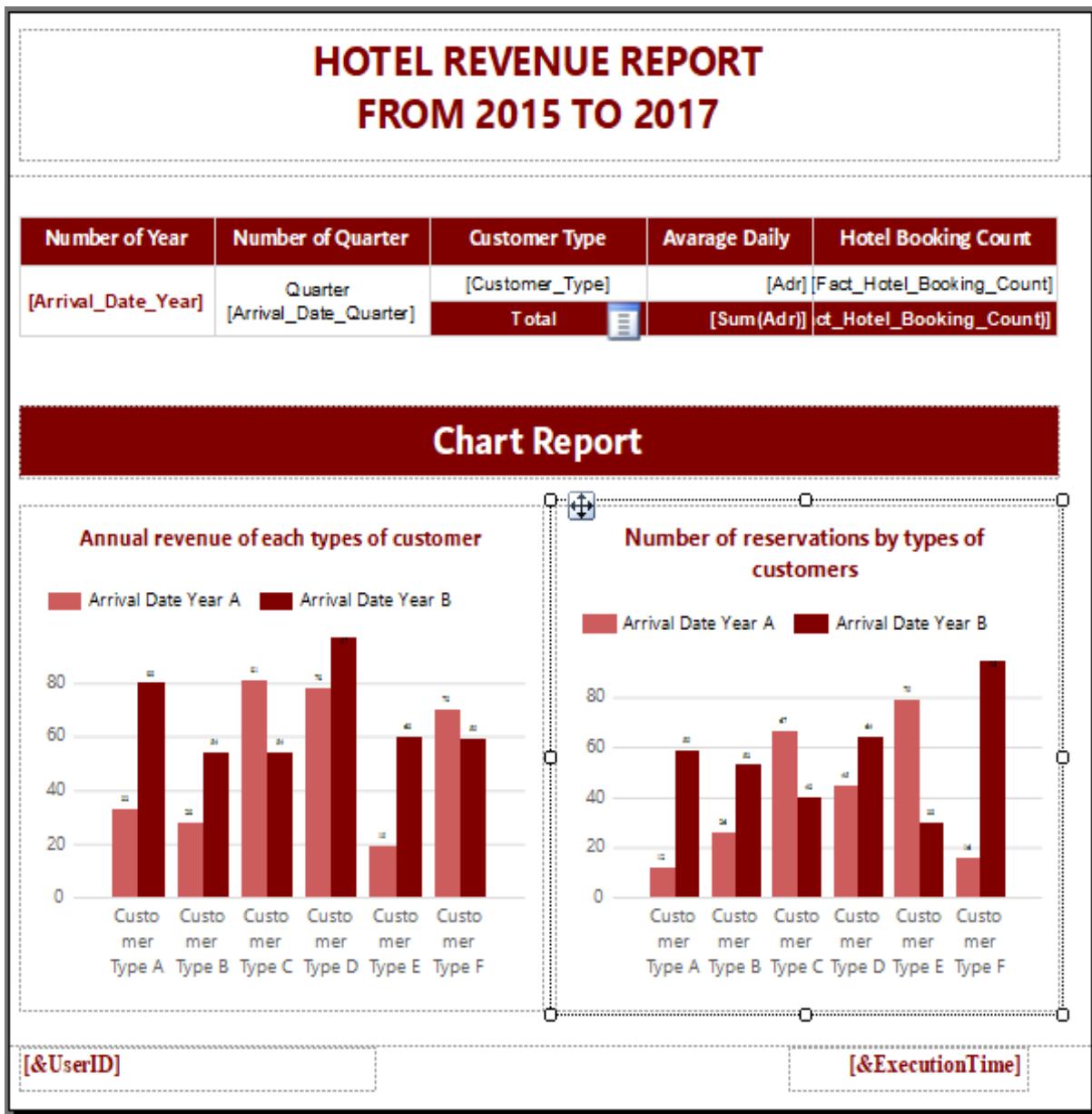
*Hình 4.46: Kết quả thu được*

- Bước 10: Thực hiện tương tự với biểu đồ thể hiện số lượng đăng ký đặt phòng của từng loại khách hàng theo từng năm.



*Hình 4.47: Biểu đồ thể hiện số lượng đăng ký đặt phòng của từng loại khách hàng theo từng năm.*

- Bước 11: Định dạng report



Hình 4.48: Định dạng Report 1

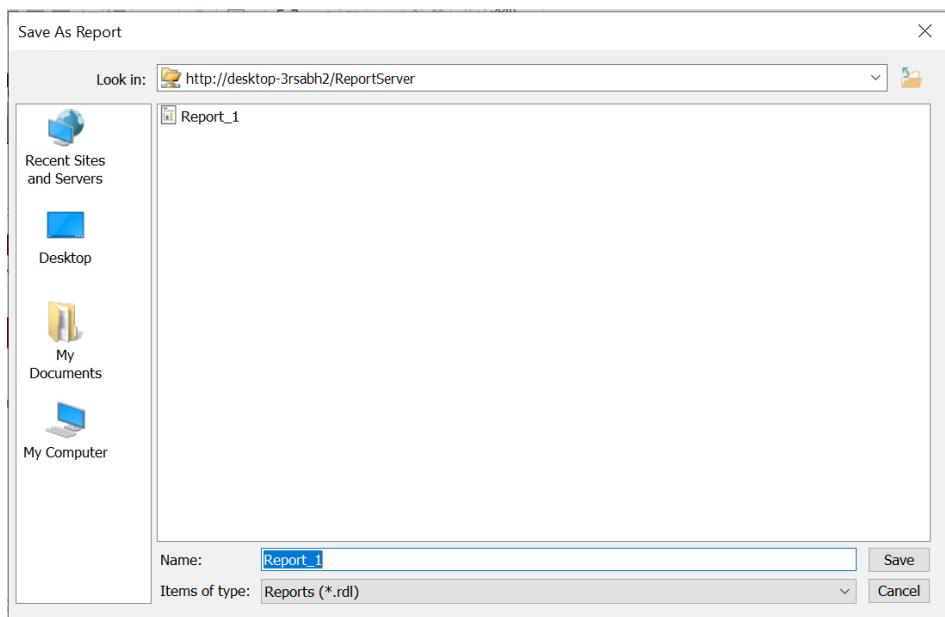
- Bước 12: In báo cáo ra file excel bằng cách qua tab review → Save file with excel.

HOTEL REVENUE REPORT FROM 2015 TO 2017				
Number of Year	Number of Quarter	Customer Type	Average Daily Rate	Hotel Booking Count
2015	Quarter 3	Contract	4111.000.00	1.075.00
		Group	45.219.00	51.00
		Transient	4552.487.00	5.752.00
		Transient-Party	4353.482.00	4.306.00
	Quarter 4	<b>Total</b>	<b>\$1,158,272.00</b>	<b>11,224.00</b>
		Contract	438.584.00	1.003.00
		Group	45.473.00	55.00
		Transient	4338.373.00	5.674.00
2016	Quarter 1	Transient-Party	4224.228.00	3.162.00
		<b>Total</b>	<b>\$738,684.00</b>	<b>10,842.00</b>
		Contract	46.755.00	87.00
		Group	42.558.00	42.00
	Quarter 2	Transient	4515.458.00	8.411.00
		Transient-Party	4156.355.00	2.472.00
		<b>Total</b>	<b>\$289,535.00</b>	<b>10,242.00</b>
		Contract	423.345.00	935.00
2017	Quarter 3	Group	42.519.00	35.00
		Transient	41.191.625.00	11.754.00
		Transient-Party	4347.237.00	3.336.00
		<b>Total</b>	<b>\$1,565,386.00</b>	<b>15,074.00</b>
	Quarter 4	Contract	425.522.00	291.00
		Group	45.449.00	24.00
		Transient	41.618.351.00	12.246.00
		Transient-Party	4257.482.00	2.415.00
		<b>Total</b>	<b>\$1,278,572.00</b>	<b>14,364.00</b>
2017	Quarter 1	Contract	45.154.00	43.00
		Group	45.587.00	37.00
		Transient	4358.354.00	10.545.00
		Transient-Party	4155.384.00	2.866.00
	Quarter 2	<b>Total</b>	<b>\$289,533.00</b>	<b>12,545.00</b>
		Contract	427.724.00	403.00
		Group	47.525.00	31.00
		Transient	41.888.578.00	14.512.00
		Transient-Party	4247.521.00	2.427.00
		<b>Total</b>	<b>\$2,003,459.00</b>	<b>17,527.00</b>
	Quarter 3	Contract	425.585.00	158.00
		Group	47.365.00	43.00
		Transient	41.384.349.00	8.637.00
		Transient-Party	4162.525.00	1.258.00
		<b>Total</b>	<b>\$1,568,510.00</b>	<b>10,282.00</b>



Hình 4.49: Kết quả câu Report 1

- Bước 13: Nhấn Save dẫn đến đường dẫn → Recent Sites and Servers → Save



*Hình 4.50: Lưu đến đường dẫn Report Server hiện hành*

- Bước 14: Vào trang <http://desktop-3rsabh2/ReportServer>  
➔ Tìm đến Report\_1.

## **desktop-3rsabh2/ReportServer - /**

---

Saturday, May 14, 2022 12:56 AM      78004 [Report\\_1](#)

---

Microsoft SQL Server Reporting Services Version 15.0.1102.962

*Hình 4.51: Kết nối thành công.*

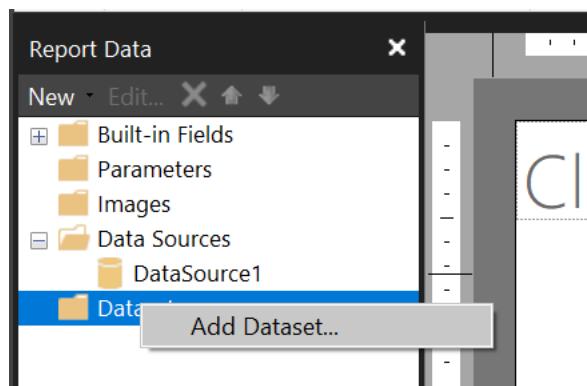
## HOTEL REVENUE REPORT FROM 2015 TO 2017

Number of Year	Number of Quarter	Customer Type	Average Daily Rate	Hotel Booking Count
2015	Quarter 3	Contract	\$141,089.00	1,575.00
		Group	\$5,219.00	61.00
		Transient	\$662,487.00	5,752.00
		Transient-Party	\$359,482.00	4,386.00
		<b>Total</b>	<b>\$1,168,277.00</b>	<b>11,774.00</b>
	Quarter 4	Contract	\$98,604.00	1,083.00
		Group	\$9,479.00	131.00
		Transient	\$398,373.00	5,671.00
		Transient-Party	\$224,228.00	3,162.00
		<b>Total</b>	<b>\$730,684.00</b>	<b>10,047.00</b>
2016	Quarter 1	Contract	\$6,759.00	87.00
		Group	\$2,390.00	42.00
		Transient	\$615,450.00	8,411.00
		Transient-Party	\$156,936.00	2,172.00
		<b>Total</b>	<b>\$781,535.00</b>	<b>10,712.00</b>
	Quarter 2	Contract	\$23,945.00	335.00
		<b>Total</b>	<b>\$23,945.00</b>	<b>335.00</b>
		Contract	\$141,089.00	1,575.00
		Group	\$5,219.00	61.00
		Transient	\$662,487.00	5,752.00
2017	Quarter 3	Transient-Party	\$359,482.00	4,386.00
		<b>Total</b>	<b>\$1,168,277.00</b>	<b>11,774.00</b>
		Contract	\$98,604.00	1,083.00
		Group	\$9,479.00	131.00
		Transient	\$398,373.00	5,671.00
	Quarter 4	Transient-Party	\$224,228.00	3,162.00
		<b>Total</b>	<b>\$730,684.00</b>	<b>10,047.00</b>
		Contract	\$6,759.00	87.00
		Group	\$2,390.00	42.00
		Transient	\$615,450.00	8,411.00
2018	Quarter 1	Transient-Party	\$156,936.00	2,172.00
		<b>Total</b>	<b>\$781,535.00</b>	<b>10,712.00</b>
		Contract	\$23,945.00	335.00
		<b>Total</b>	<b>\$23,945.00</b>	<b>335.00</b>
		Contract	\$141,089.00	1,575.00
	Quarter 2	Group	\$5,219.00	61.00
		Transient	\$662,487.00	5,752.00
		Transient-Party	\$359,482.00	4,386.00
		<b>Total</b>	<b>\$1,168,277.00</b>	<b>11,774.00</b>
		Contract	\$98,604.00	1,083.00
2019	Quarter 3	Group	\$9,479.00	131.00
		Transient	\$398,373.00	5,671.00
		Transient-Party	\$224,228.00	3,162.00
		<b>Total</b>	<b>\$730,684.00</b>	<b>10,047.00</b>
		Contract	\$6,759.00	87.00
	Quarter 4	Group	\$2,390.00	42.00
		Transient	\$615,450.00	8,411.00
		Transient-Party	\$156,936.00	2,172.00
		<b>Total</b>	<b>\$781,535.00</b>	<b>10,712.00</b>
		Contract	\$23,945.00	335.00

Hình 4.52: In report 1 trên Report Server

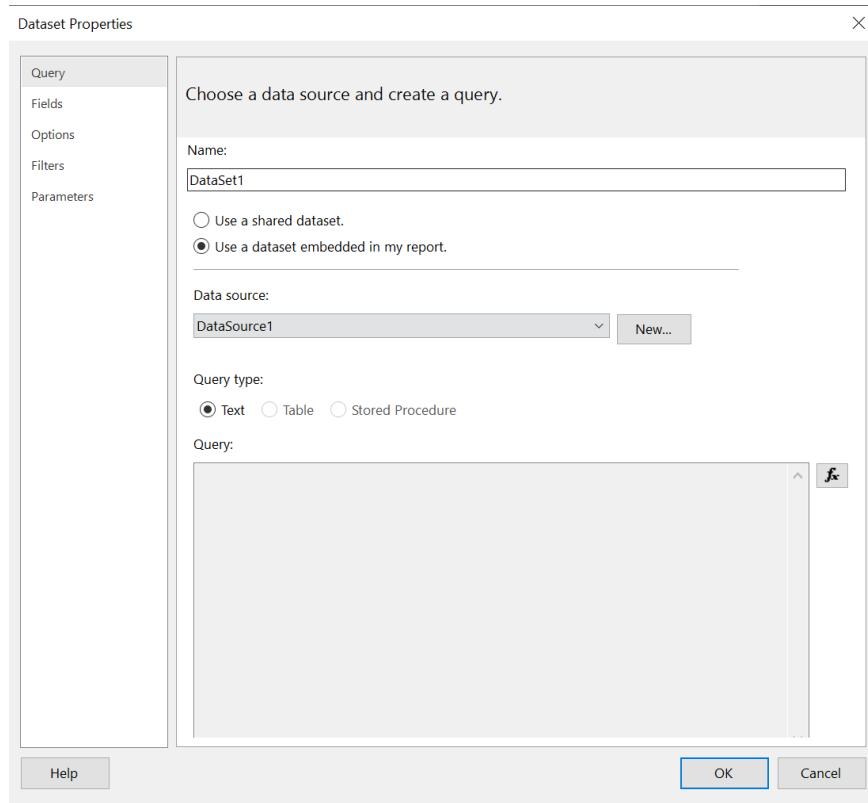
### 2.2.2. Báo biểu 2: Xây dựng báo cáo đặt phòng của khách hàng tại khách sạn.

- Bước 1: Thêm datasets cần thiết cho báo biểu 1 đặt tên là dataset1, bằng cách nhấn chuột phải vào Datasets → Add new dataset...



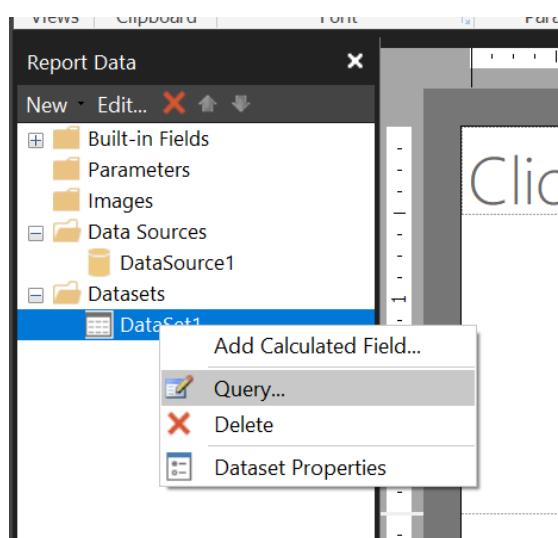
Hình 4.53: Thêm datasets cho câu báo biểu

- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường Name đặt tên là DataSet1 chọn chế độ Use a dataset embedded in my report. Trường Data Source chọn data source tại phần tạo kết nối data source.



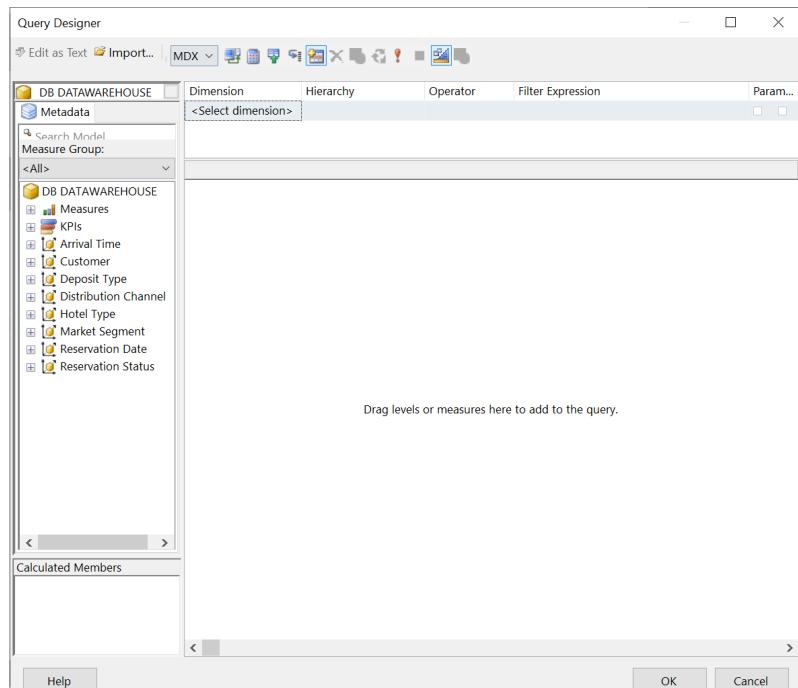
*Hình 4.54: Điều chỉnh tên, datasource cho báo biểu 2*

- Bước 3: Click chuột phải và chọn Query hộp thoại sẽ hiện ra:



*Hình 4.55: Thêm query cho tập dataset.*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



*Hình 4.56: Cửa sổ làm việc Query Designer*

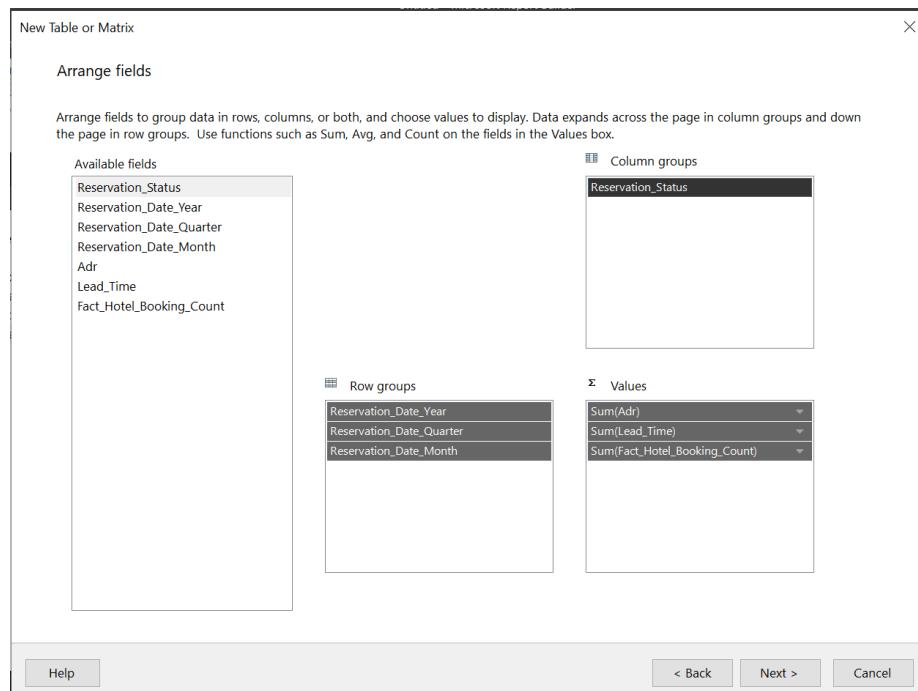
- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.

Reservation Status	Reservation Date ...	Reservation Date ...	Reservation Date ...	Adr	Lead Tim
Canceled	2014	4	10	11...	52040
Canceled	2014	4	11	0	297
Canceled	2015	1	1	55...	293370
Canceled	2015	1	2	4528	11173
Canceled	2015	1	3	6832	19049
Canceled	2015	2	4	14...	26787
Canceled	2015	2	5	24...	48646
Canceled	2015	2	6	59...	107618
Canceled	2015	3	7	18...	481638
Canceled	2015	3	8	10...	139045
Canceled	2015	3	9	96...	87299
Canceled	2015	4	10	16...	732664
Canceled	2015	4	11	85...	112311
Canceled	2015	4	12	10...	202607
Canceled	2016	1	1	21...	331657
Canceled	2016	1	2	19...	291595
Canceled	2016	1	3	20...	302751
Canceled	2016	2	4	20...	255400
Canceled	2016	2	5	15...	152111
Canceled	2016	2	6	16...	172744
Canceled	2016	3	7	16...	181512
Canceled	2016	3	8	15...	168922
Canceled	2016	3	9	17...	238484
Canceled	2016	4	10	14	105786

*Hình 4.57: Chọn những thuộc tính và độ đo cần thiết cho báo biểu 2*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 5: Vào tab Insert, chọn Matrix vào với tùy chọn là Matrix Wizard và kéo những thuộc tính vào các ô Values, Column groups và Row groups như hình minh họa:

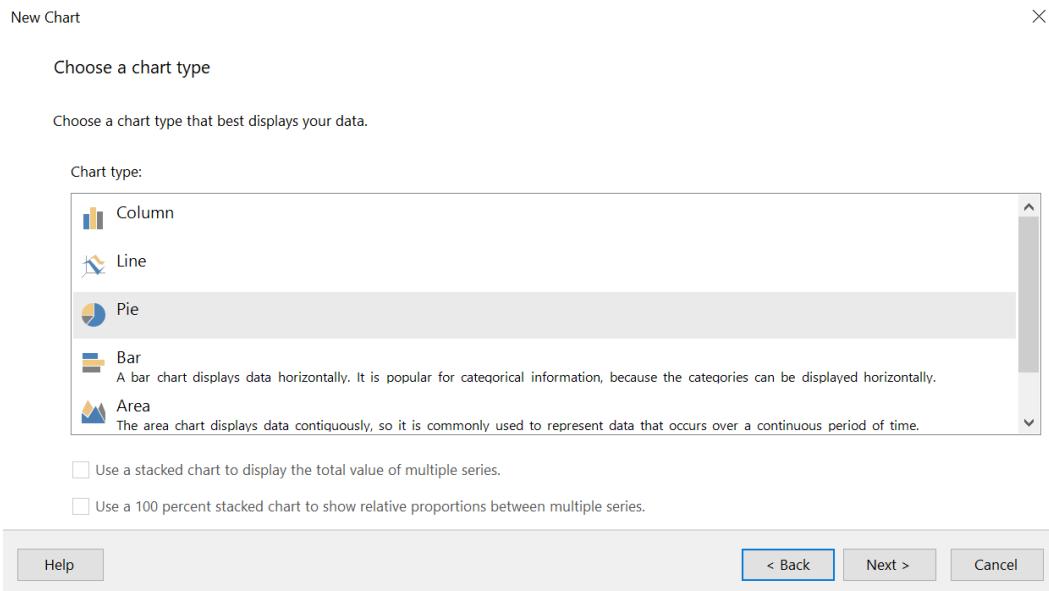


*Hình 4.58: Thêm các thuộc tính và độ đo vào các trường dữ liệu phù hợp.*

*Hình 4.59: Kết quả của bảng báo cáo 2.*

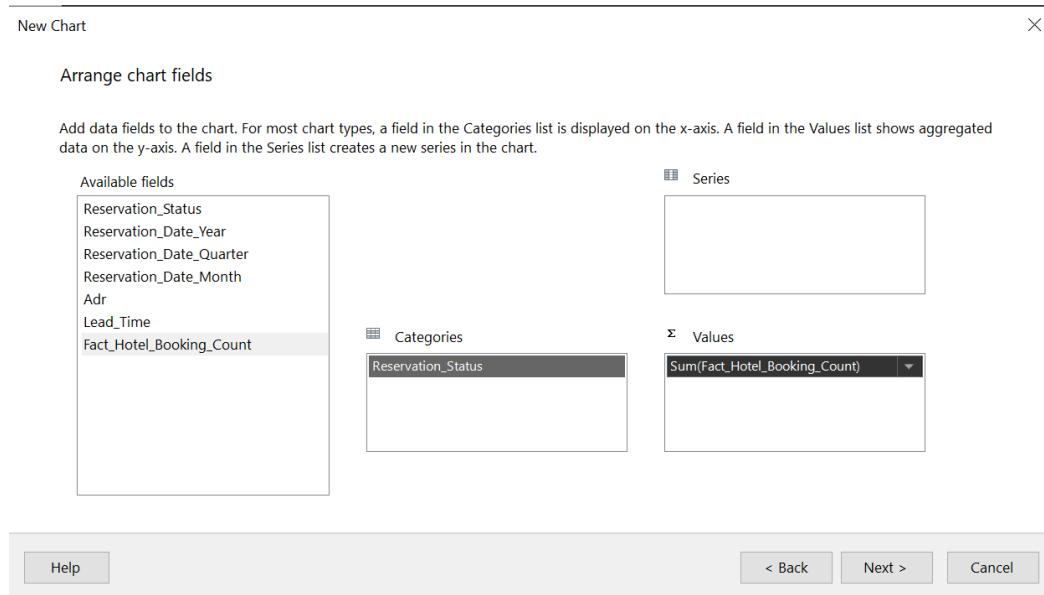
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 6: Để thống kê doanh thu của từng khách hàng theo từng năm, ta chọn Insert một Chart → Chọn Chart Wizard → Chọn lược đồ các dạng hình tròn, cột và đường.



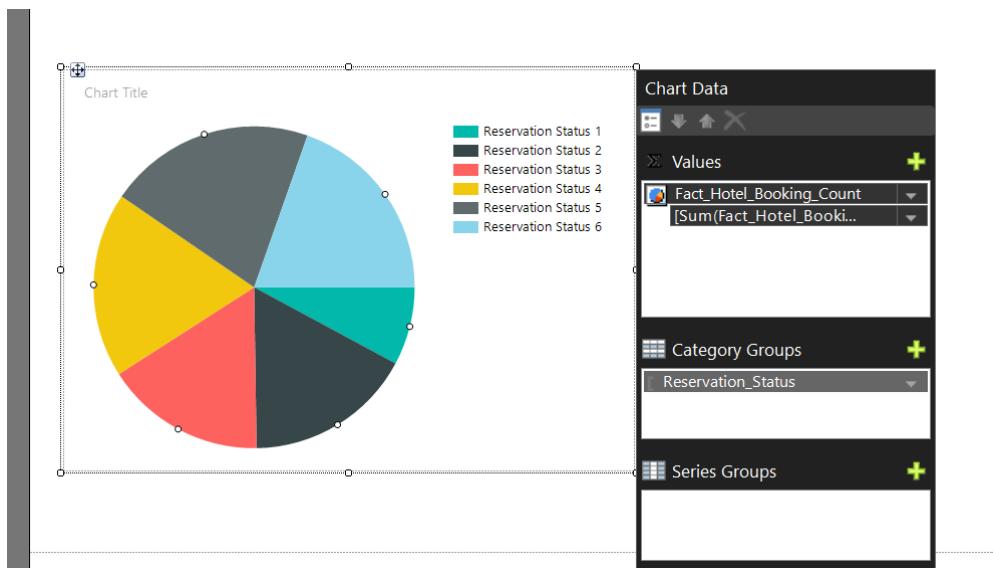
*Hình 4.60: Màn hình lựa chọn biểu đồ tròn.*

- Bước 7: Kéo thả độ đo Fact\_Hotel\_Booking vào ô Values và thuộc tính Reservation\_Status vào ô Categories.



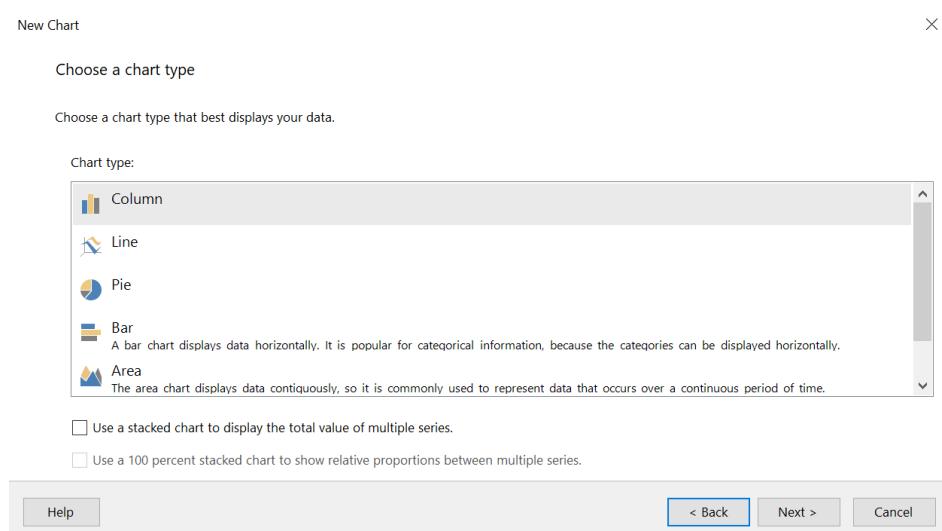
*Hình 4.61: Kéo thả độ đo Fact\_Hotel\_Booking vào ô Values và thuộc tính Reservation\_Status vào ô Categories.*

- Bước 8: Kết quả thu được biểu đồ:



*Hình 4.62: Kết quả trả về của biểu đồ tròn.*

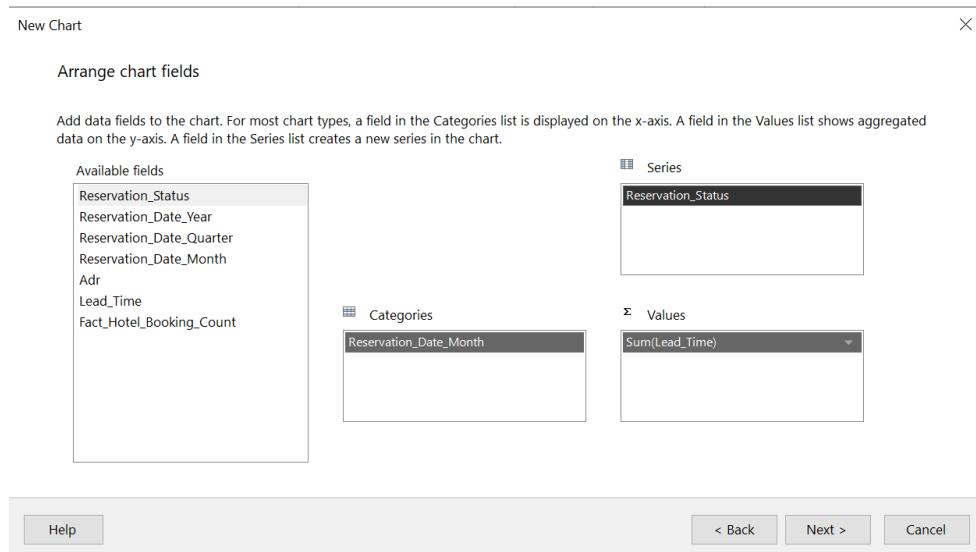
- Bước 9: Thực hiện tương tự, tạo biểu đồ cột để thống kê số ngày chờ của mỗi trạng thái trong từng tháng.



*Hình 4.63: Màn hình lựa chọn biểu đồ cột.*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: Kéo thả độ đo Lead\_Time vào ô Values, thuộc tính Reservation\_Status vào ô Series và thuộc tính Reservation\_Date\_Month vào ô Categories.



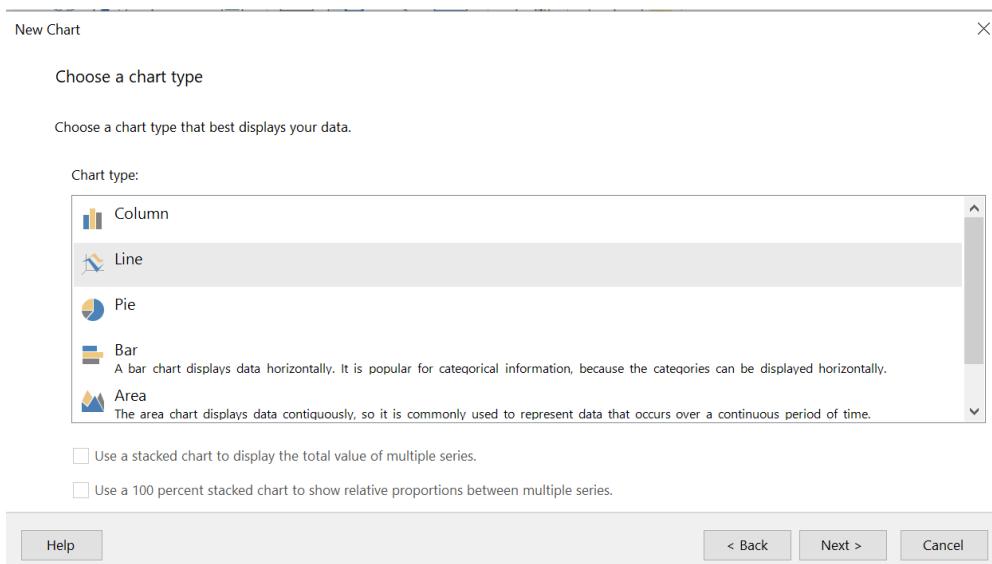
*Hình 4.64: Kéo thả độ đo Lead\_Time vào ô Values, thuộc tính Reservation\_Status vào ô Series và thuộc tính Reservation\_Date\_Month vào ô Categories.*

- Bước 11: Thu được biểu đồ cột:



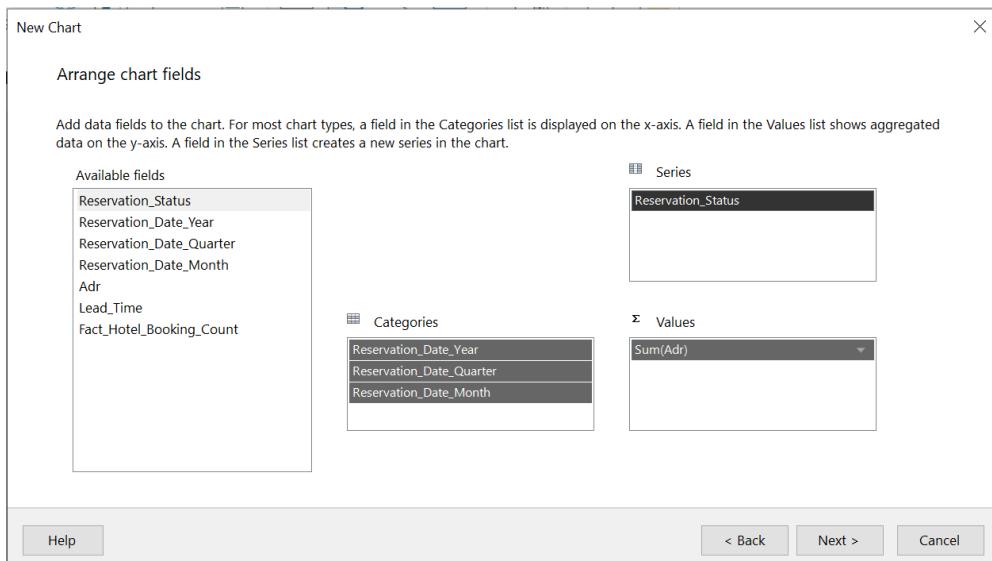
*Hình 65: Kết quả trả về của biểu đồ cột.*

- Bước 12: Thực hiện tương tự, tạo biểu đồ đường để thống kê tổng doanh thu của mỗi trạng thái trong từng tháng của từng năm.



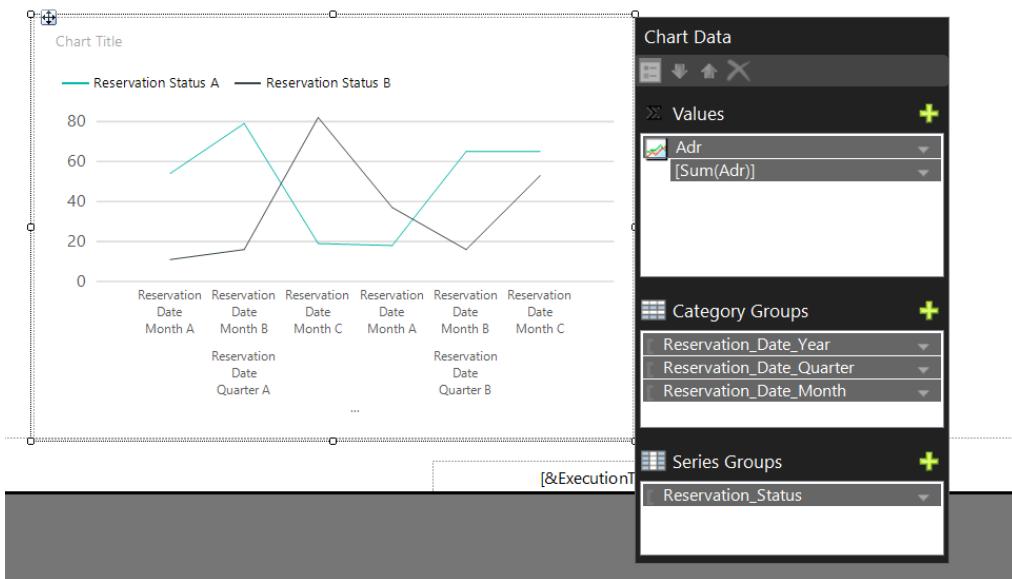
*Hình 4.66: Màn hình lựa chọn biểu đồ đường.*

- Bước 13: Kéo thả độ đo Adr vào ô Values, thuộc tính Reservation\_Status vào ô Series và các thuộc tính Reservation\_Date\_Year, Reservation\_Date\_Quarte và Reservation\_Date\_Month vào ô Categories.



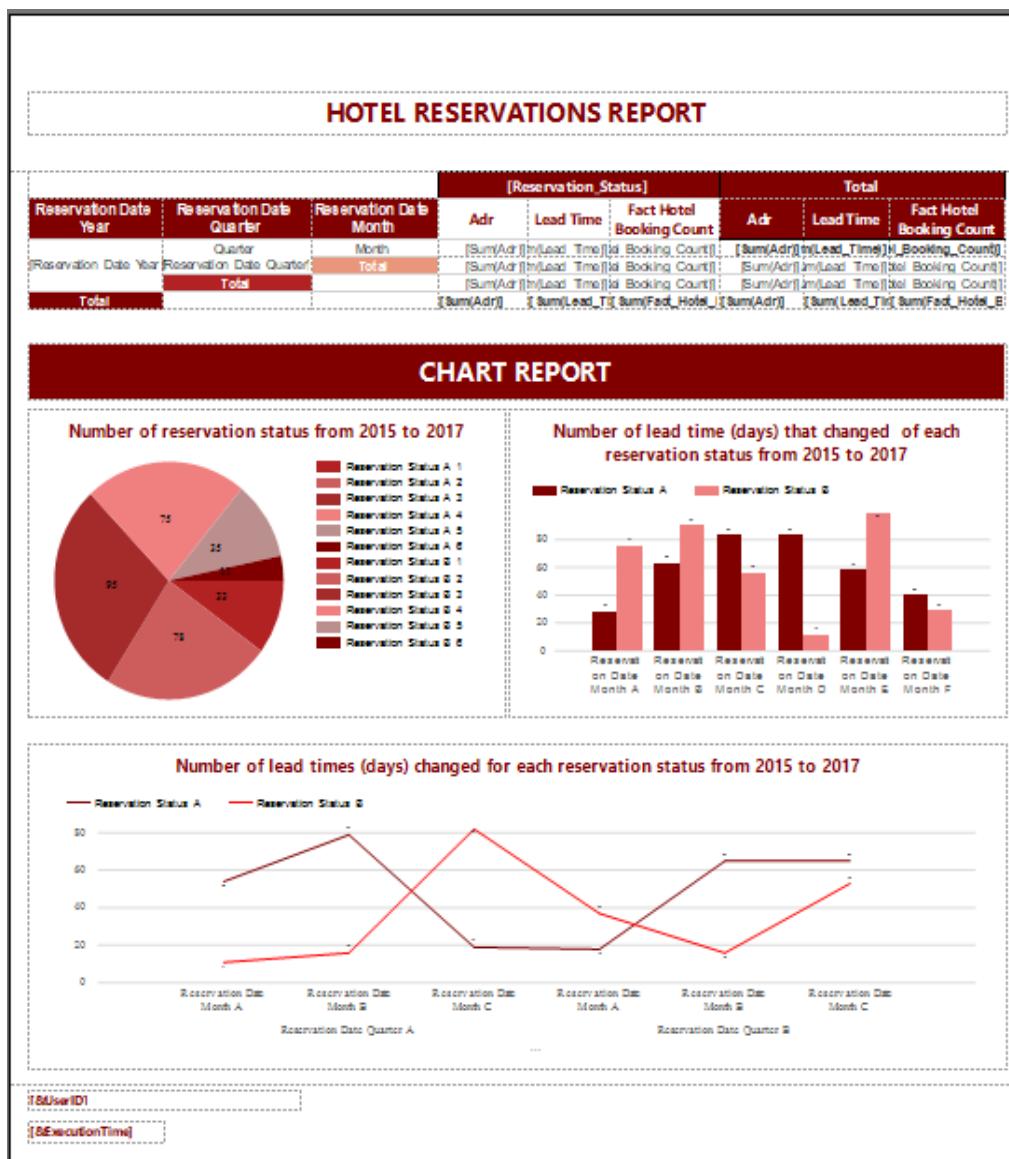
*Hình 4.67: Kéo thả độ đo Adr vào ô Values, thuộc tính Reservation\_Status vào ô Series và các thuộc tính Reservation\_Date\_Year, Reservation\_Date\_Quarte và Reservation\_Date\_Month vào ô Categories.*

- Bước 14: Kết quả trả về của biểu đồ đường:



*Hình 4.68: Kết quả trả về của biểu đồ đường.*

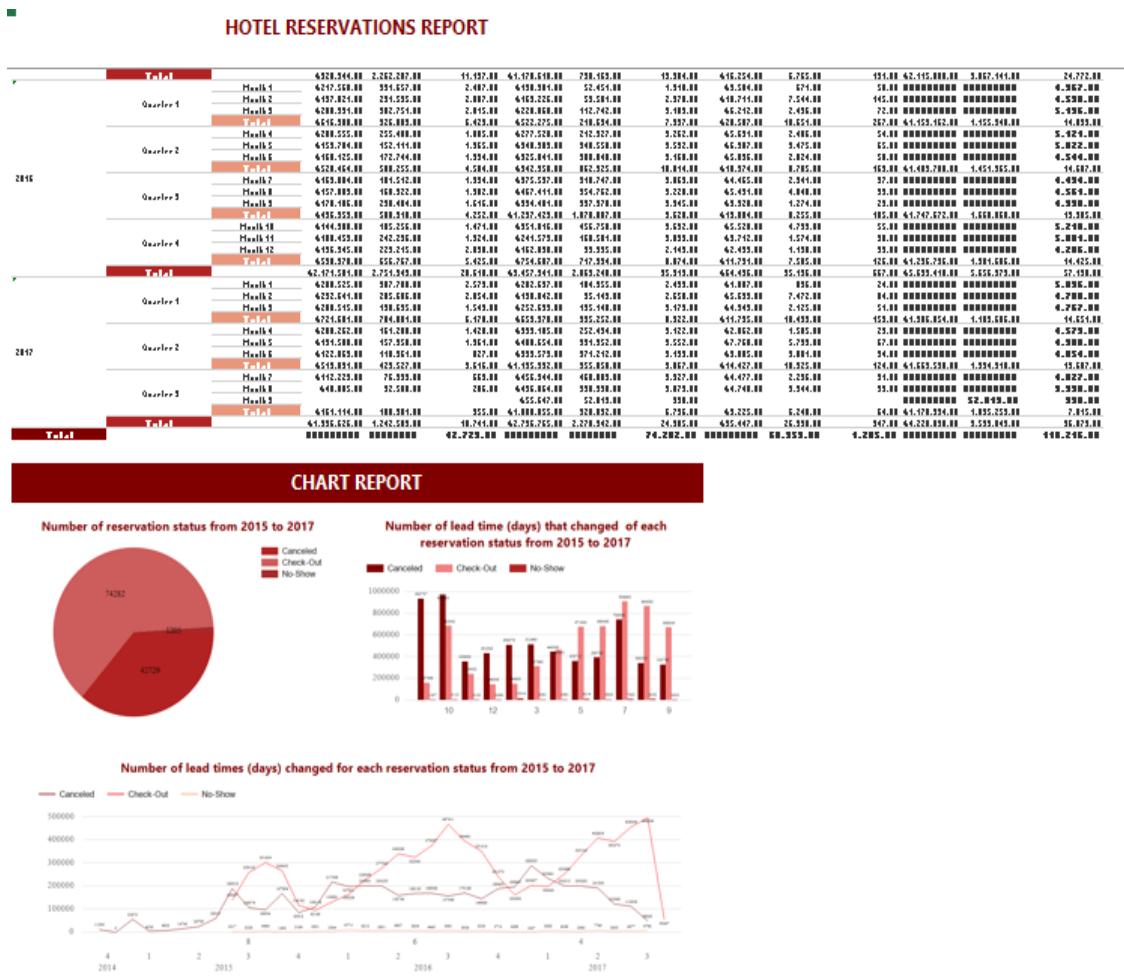
- Bước 15: Định dạng report



*Hình 4.69: Định dạng Report 2*

## Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 16: In báo cáo ra file excel bằng cách qua Tab Review → Save file with excel.



Giao viên hướng dẫn: ThS. Đỗ Thị Minh Phụng

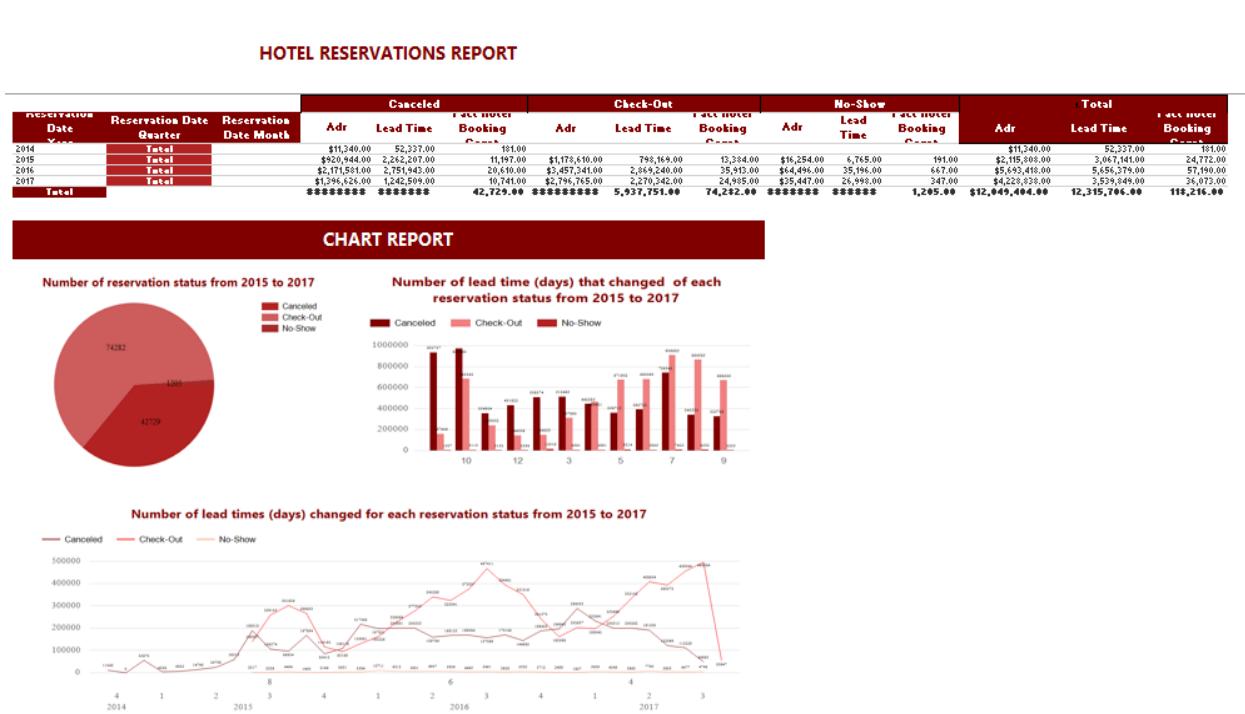
- Bước 17: Vào trang <http://desktop-3rsabh2/ReportServer>
  - ➔ Tìm đến Report\_2.

## desktop-3rsabh2/ReportServer - /

Saturday, May 14, 2022 12:56 AM      78004 Report\_1  
 Saturday, May 14, 2022 1:53 AM      172936 Report\_2

Microsoft SQL Server Reporting Services Version 15.0.1102.962

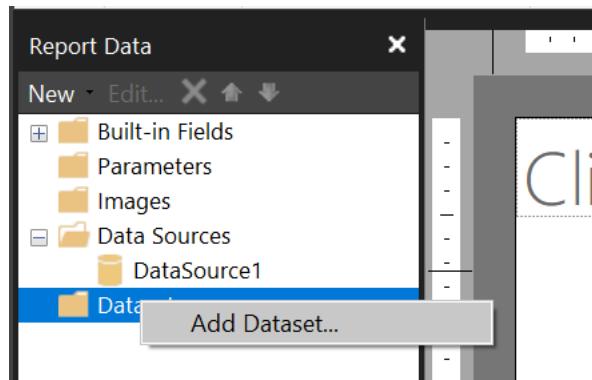
*Hình 4.71: Kết nối thành công.*



*Hình 4.72: In report 2 trên Report Server.*

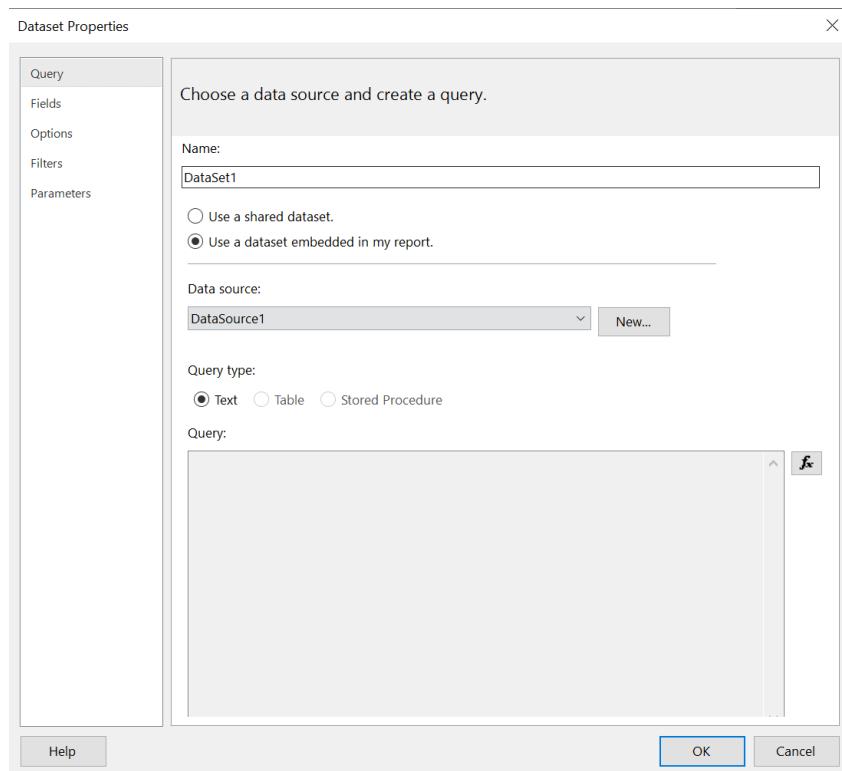
**2.2.3. Báo biểu 3: Thống kê lượng khách hàng từ năm 2015 đến năm 2017 của mỗi loại khách sạn.**

- Bước 1: Thêm datasets cần thiết cho báo biểu 3 đặt tên là datasets1, bằng cách nhấp chuột phải vào Datasets → Add new dataset...



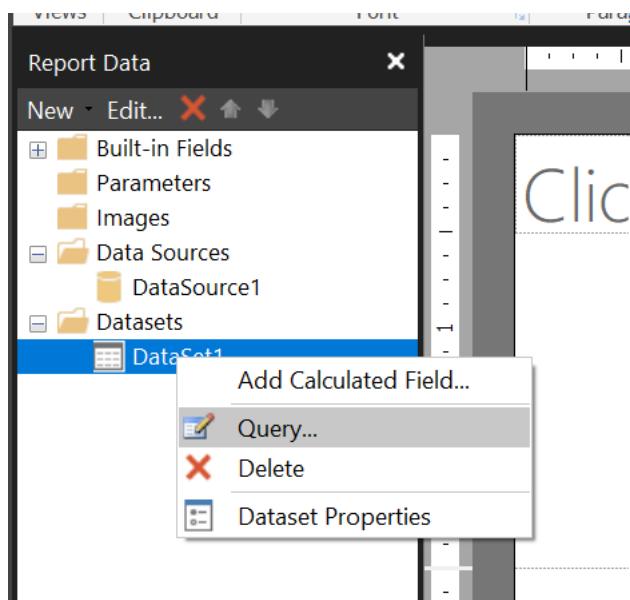
*Hình 4.73: Thêm datasets cho câu báo biểu 3*

- Bước 2: Trong tab Datasets Properties → Trường Query. Tại trường Name đặt tên là Dataset\_Query1 chọn chế độ Use a dataset embedded in my report. Trường Data Source chọn data source tại phần tạo kết nối data source.

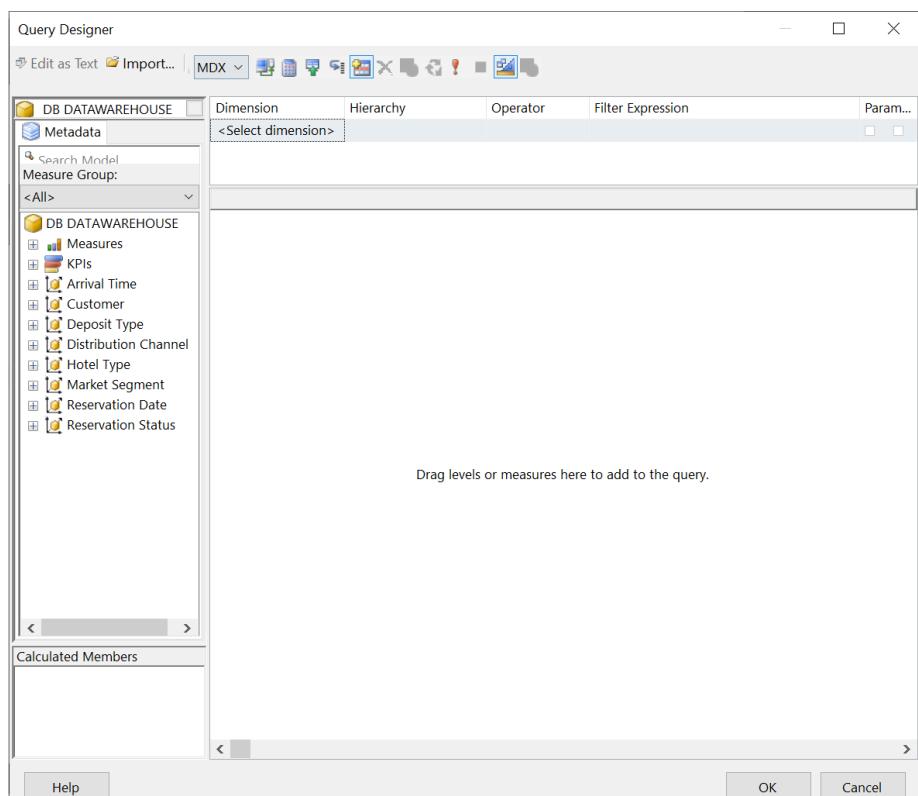


*Hình 4.74: Điều chỉnh tên, datasource cho báo biểu 3.*

- Bước 3: Click chuột phải và chọn Query hộp thoại sẽ hiện ra:

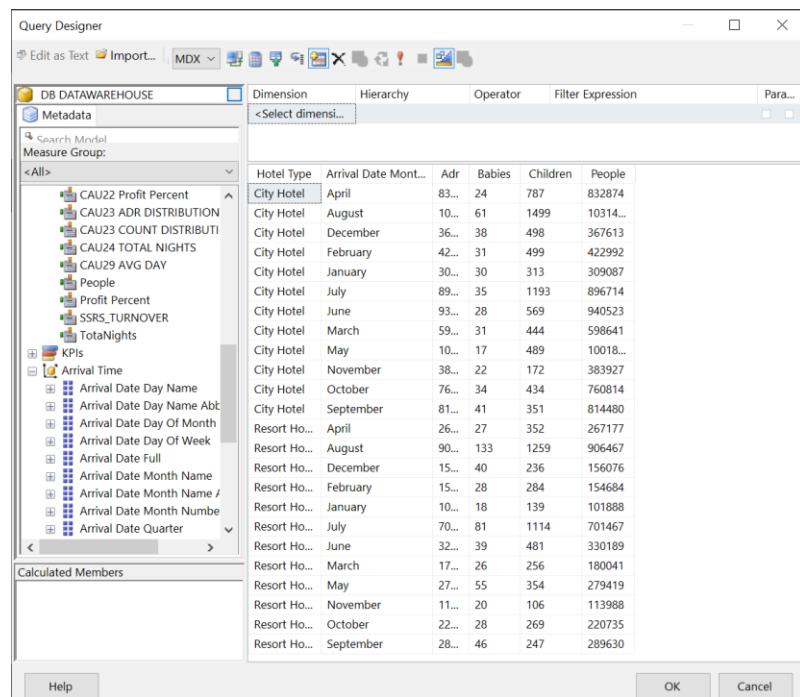


*Hình 4.75: Thêm query cho tập dataset.*



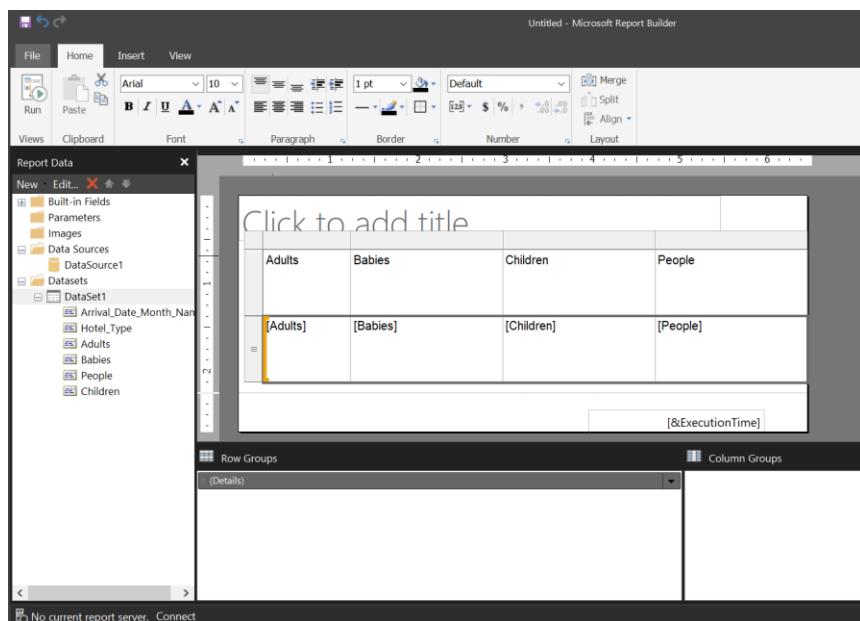
*Hình 4.76: Cửa sổ làm việc Query Designer*

- Bước 4: Chọn những thuộc tính cần thiết cho báo biểu giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn OK.



Hình 4.77: Chọn những thuộc tính và độ đo cần thiết cho báo biểu 3

- Bước 5: Insert một Table vào với tùy chọn là Insert Table và kéo table qua vùng thiết kế (Design), kéo những thuộc tính vào bảng table như hình minh họa:



Hình 4.78: Kéo thả những thuộc tính cho câu báo biểu 1 vào bảng design

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 6: Tạo Grouping bằng cách tô đen thuộc các thuộc tính muốn được gom nhóm, chọn Add Group để gom nhóm dữ liệu.

Hotel Type	Arrival Date Month Name	Adults	Babies	Children	People
[Hotel_Type]	[Arrival_Date_M]	[Adults]	[Babies]	[Children]	[People]

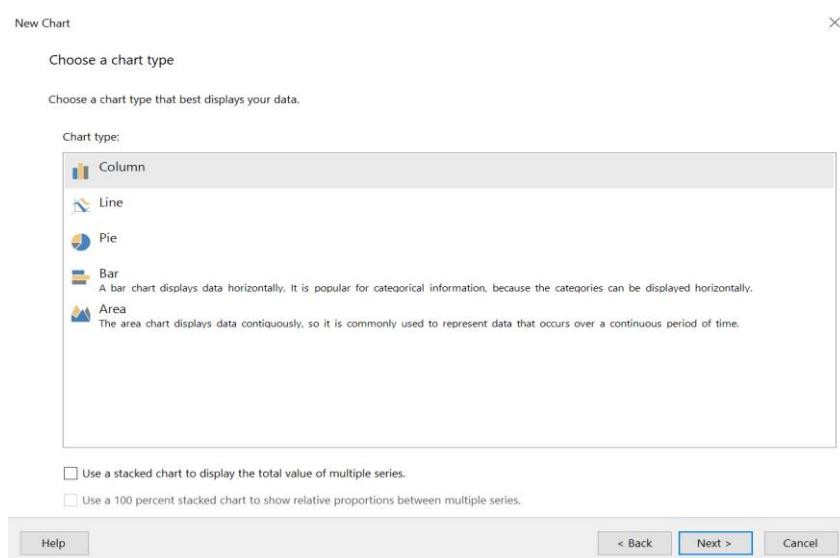
Hình 4.79: Kết quả của quá trình gom nhóm số lượng khách hàng theo tháng và theo loại khách sạn.

- Bước 7: Tạo Total bằng cách chọn vào thuộc tính mà ta muốn tính tổng các độ đo theo thuộc tính đó, chọn Add Total để tính tổng các giá trị.

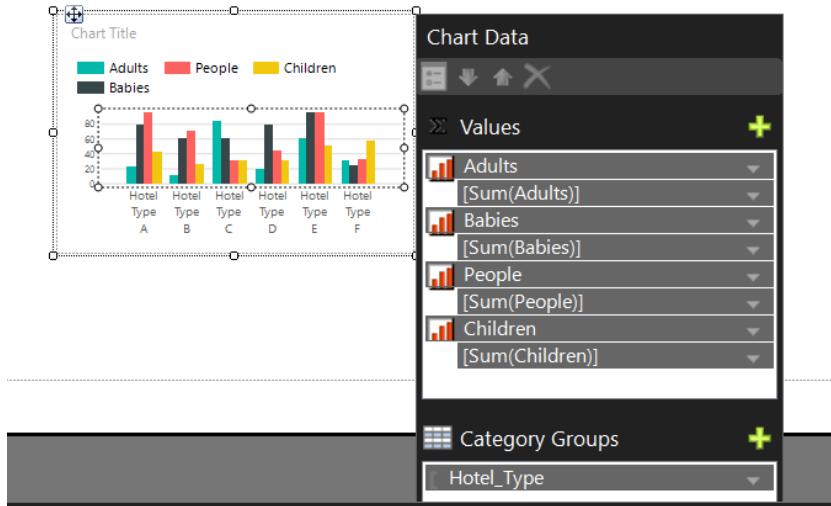
Hotel Type	Arrival Date Month Name	Adults	Babies	Children	People
[Hotel_Type]	[Arrival_Date_M]	[Adults]	[Babies]	[Children]	[People]
Total		[Sum(Adults)]	[Sum(Babies)]	[Sum(Children)]	[Sum(People)]

Hình 4.80: Tính tổng giá trị dữ liệu Adults, Babies, Children, People theo loại khách hàng (Hotel\_Type).

- Bước 8: Để thống kê lượng khách của từng khách sạn theo từng tháng, ta chọn Insert một Chart → Chọn Chart Wizard → Chọn lược đồ dạng cột.

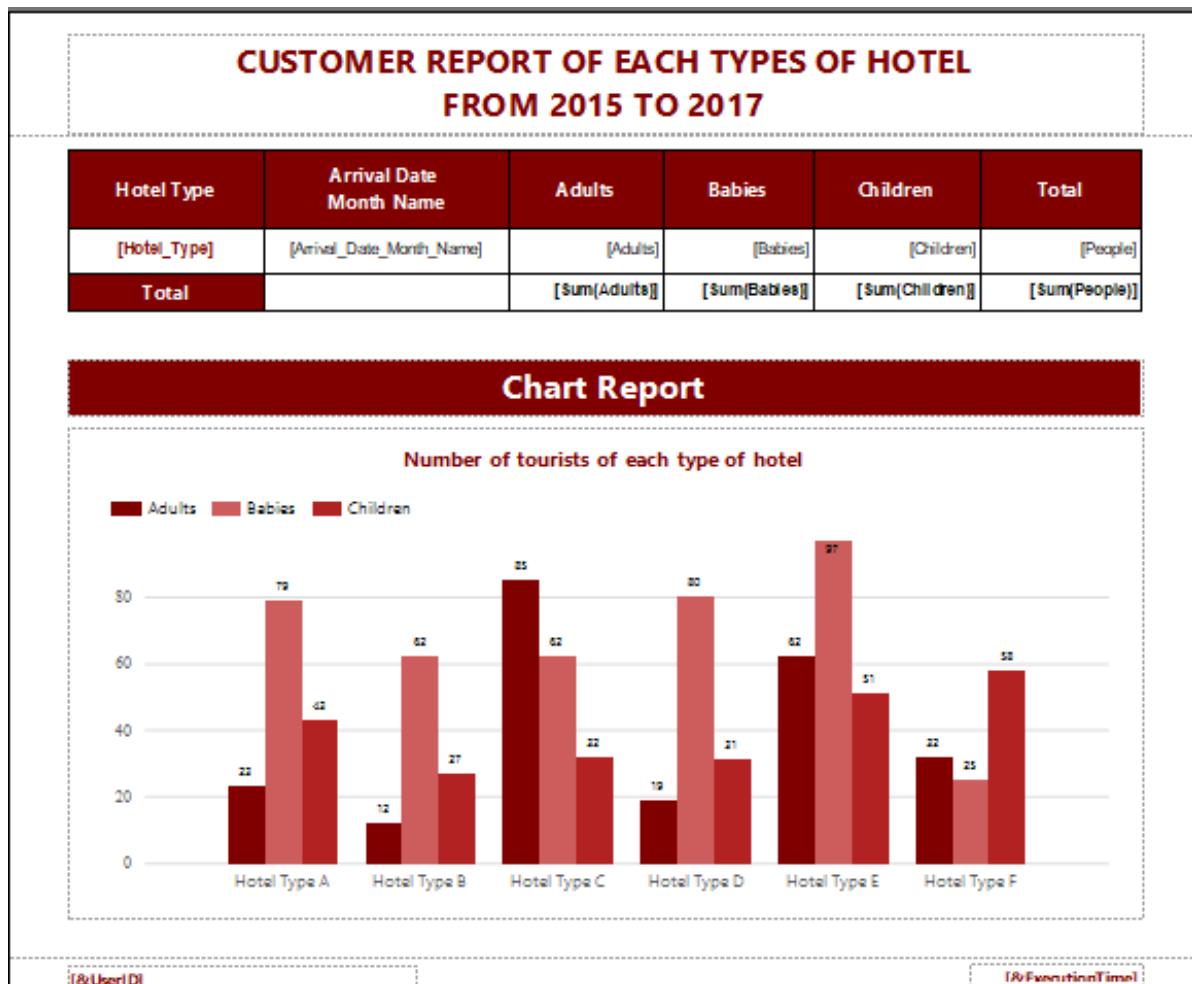


Hình 4.81: Màn hình lựa chọn biểu đồ cột.



Hình 4.82: Kết quả trả về của biểu đồ cột.

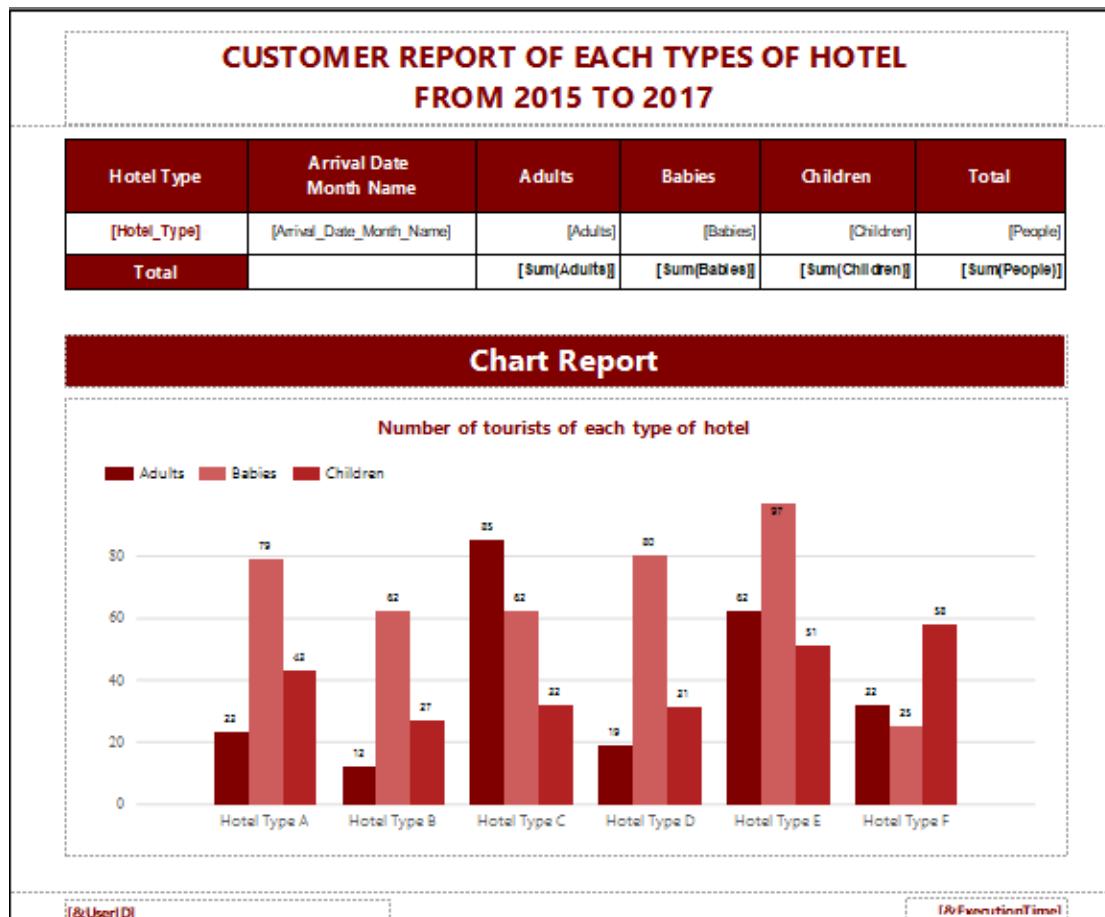
- Bước 9: Định dạng report



Hình 4.83: Định dạng Report 3

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 10: In báo cáo ra file excel bằng cách qua tab review  save file with excel



Hình 4.84: Kết quả câu Report 3

- Bước 11: Vào trang <http://desktop-3rsabh2/ReportServer>  
→ Tìm đến Report\_1.

## **desktop-3rsabh2/ReportServer - /**

Saturday, May 14, 2022 12:56 AM      78004 [Report\\_1](#)  
Saturday, May 14, 2022 1:53 AM      172936 [Report\\_2](#)  
Saturday, May 14, 2022 1:58 AM      88777 [report\\_3](#)

Microsoft SQL Server Reporting Services Version 15.0.1102.962

Hình 4.85: Kết nối thành công.

## CUSTOMER REPORT OF EACH TYPES OF HOTEL FROM 2015 TO 2017

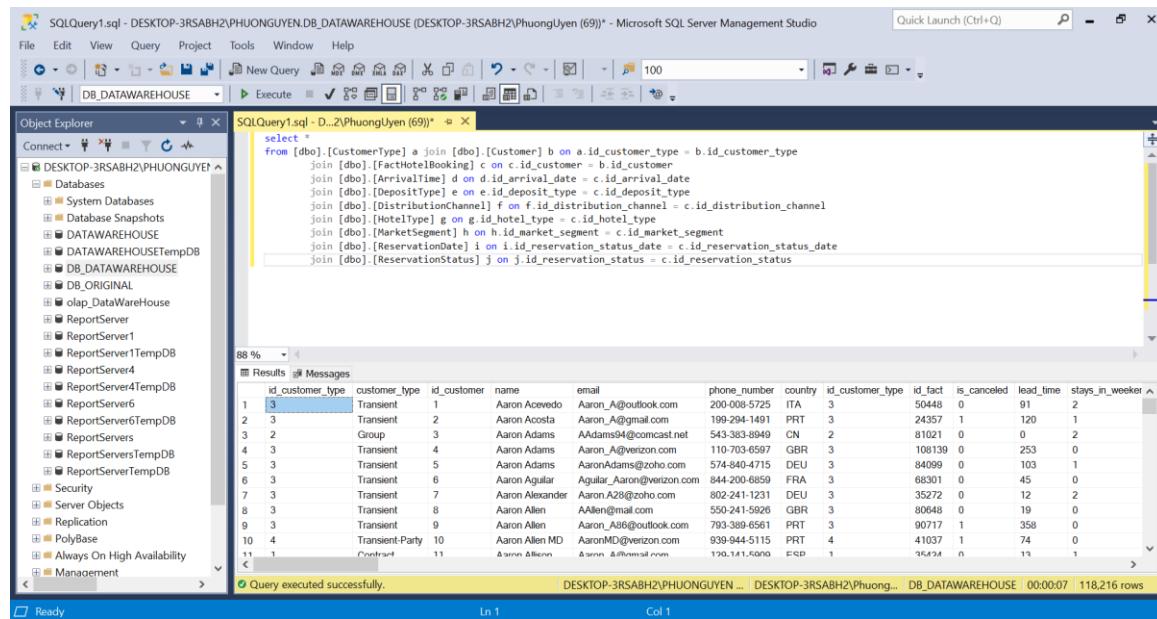
<b>Hotel Type</b>	<b>Arrival Date Month Name</b>	<b>Adults</b>	<b>Babies</b>	<b>Children</b>	<b>Total</b>
<b>City Hotel</b>	April	14,290.00	24.00	787.00	832,874.00
	August	17,832.00	61.00	1,499.00	1,031,420.00
	December	7,609.00	38.00	498.00	367,613.00
	February	8,846.00	31.00	499.00	422,992.00
	January	6,378.00	30.00	313.00	309,087.00
	July	15,975.00	35.00	1,193.00	896,714.00
	June	14,519.00	28.00	569.00	940,523.00
	March	11,858.00	31.00	444.00	598,641.00
	May	14,958.00	17.00	489.00	1,001,836.00
	November	7,330.00	22.00	172.00	383,927.00
	October	13,665.00	34.00	434.00	760,814.00
	September	13,569.00	41.00	351.00	814,480.00

Hình 4.86: In report 3 trên Report Server

### 3. Quá trình lập báo biểu bằng công cụ Google Data Studio

#### 3.1. Tạo file CSV từ dữ liệu đã lọc từ quá trình SSIS.

- Bước 1: Vào Microsoft SQL Server → Chọn Database Engine đã tạo ở quá trình SSIS → Nhấn Connect để kết nối với database.
- Bước 2: Thực thi câu lệnh truy vấn trong cửa sổ SQL



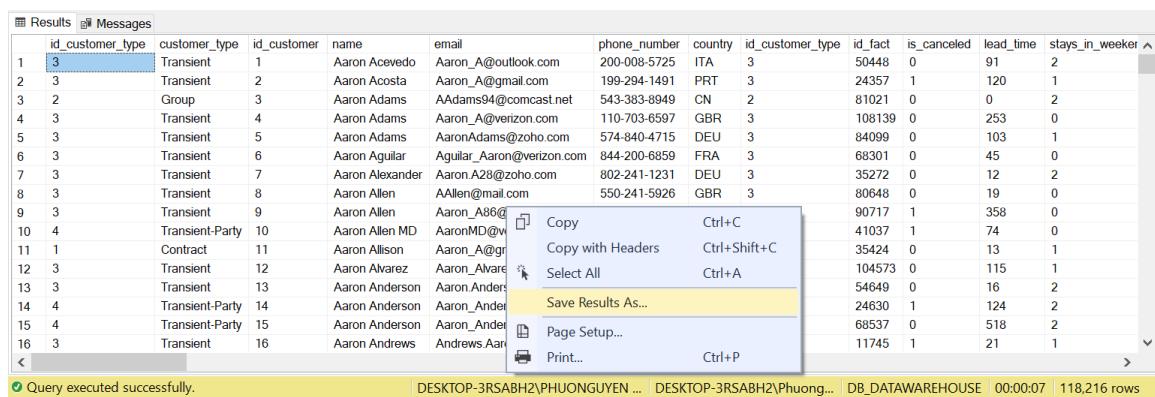
```

SQLQuery1.sql - DESKTOP-3RSABH2\PHUONGUYEN.DB_DATABRICKS (DESKTOP-3RSABH2\PhuongUyen (69)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query Execute
Object Explorer
Connect Databases System Databases Database Snapshots DB_DATABRICKS DB_DATABRICKSTempDB DB_DB_DATABRICKSE DB_DB_DATABRICKSETempDB DB_DB_DATABRICKSDB DB_DB_DATABRICKSDBTempDB DB_DB_DATABRICKSDBOriginal DB_DB_DATABRICKSDataWarehouse ReportServer ReportServer1 ReportServerTempDB ReportServer4 ReportServer4TempDB ReportServer6 ReportServer6TempDB ReportServers ReportServersTempDB ReportServerTempDB Security Server Objects Replication PolyBase Always On High Availability Management
SQLQuery1.sql - D...2\PhuongUyen (69) - + X
select *
from [dbo].[CustomerType] a join [dbo].[Customer] b on a.id_customer_type = b.id_customer_type
join [dbo].[FactHotelBooking] c on c.id_customer = b.id_customer
join [dbo].[ArrivalTime] d on d.id_arrival_date = c.id_arrival_date
join [dbo].[DepositType] e on e.id_deposit_type = c.id_deposit_type
join [dbo].[DistributionChannel] f on f.id_distribution_channel = c.id_distribution_channel
join [dbo].[HotelType] g on g.id_hotel_type = c.id_hotel_type
join [dbo].[MarketSegment] h on h.id_market_segment = c.id_market_segment
join [dbo].[ReservationDate] i on i.id_reservation_status_date = c.id_reservation_status_date
join [dbo].[ReservationStatus] j on j.id_reservation_status = c.id_reservation_status
Results Messages
id_customer_type customer_type id_customer name email phone_number country id_customer_type id_fact is_canceled lead_time stays_in_weeker
1 3 Transient 1 Aaron Acevedo Aaron_A@outlook.com 200-008-5725 ITA 3 50448 0 91 2
2 3 Transient 2 Aaron Acosta Aaron_A@gmail.com 199-294-1491 PRT 3 24357 1 120 1
3 2 Group 3 Aaron Adams AAAdams94@comcast.net 543-383-8949 CN 2 81021 0 0 2
4 3 Transient 4 Aaron Adams Aaron_A@verizon.com 110-703-6597 GBR 3 108139 0 253 0
5 3 Transient 5 Aaron Adams AaronAdams@zoho.com 574-840-4715 DEU 3 84099 0 103 1
6 3 Transient 6 Aaron Aguilar Aguilar_Aaron@verizon.com 844-200-6859 FRA 3 68301 0 45 0
7 3 Transient 7 Aaron Alexander Aaron_A28@zoho.com 802-241-1231 DEU 3 35272 0 12 2
8 3 Transient 8 Aaron Allen AAAllen@mail.com 550-241-5926 GBR 3 80648 0 19 0
9 3 Transient 9 Aaron Allen Aaron_A86@outlook.com 793-389-6561 PRT 3 90717 1 358 0
10 4 Transient-Party 10 Aaron Allen MD AaronMD@verizon.com 939-944-5115 PRT 4 41037 1 74 0
11 1 Contract 11 Aaron Allison Aaron_A@gr... 170.1.1.1:49000 EGP 1 11745 0 21 1
< 11 1 Contract 11 Aaron Allison Aaron_A@gr... 170.1.1.1:49000 EGP 1 11745 0 21 1
Query executed successfully. DESKTOP-3RSABH2\PHUONGUYEN ... DESKTOP-3RSABH2\Phuong... DB_DATABRICKS 00:00:07 118,216 rows
Ready

```

Hình 4.87: Truy vấn tất cả các thuộc tính có trong bảng cơ sở dữ liệu.

- Bước 3: Tại vùng kết quả câu truy vấn nhấn chuột phải chọn Save Result As ... → Chọn đường dẫn rồi nhấn lưu file CSV.



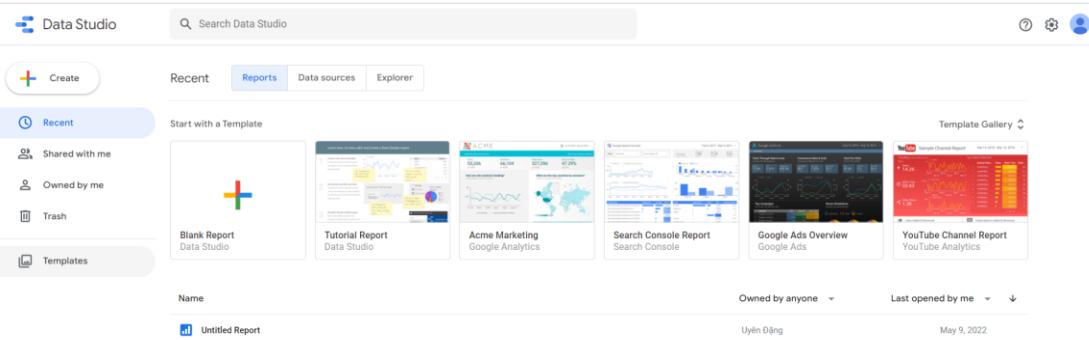
id_customer_type	customer_type	id_customer	name	email	phone_number	country	id_customer_type	id_fact	is_canceled	lead_time	stays_in_weeker
1	3	Transient	1	Aaron Acevedo	Aaron_A@outlook.com	200-008-5725	ITA	3	50448	0	91
2	3	Transient	2	Aaron Acosta	Aaron_A@gmail.com	199-294-1491	PRT	3	24357	1	120
3	2	Group	3	Aaron Adams	AAAdams94@comcast.net	543-383-8949	CN	2	81021	0	0
4	3	Transient	4	Aaron Adams	Aaron_A@verizon.com	110-703-6597	GBR	3	108139	0	253
5	3	Transient	5	Aaron Adams	AaronAdams@zoho.com	574-840-4715	DEU	3	84099	0	103
6	3	Transient	6	Aaron Aguilar	Aguilar_Aaron@verizon.com	844-200-6859	FRA	3	68301	0	45
7	3	Transient	7	Aaron Alexander	Aaron_A28@zoho.com	802-241-1231	DEU	3	35272	0	12
8	3	Transient	8	Aaron Allen	AAAllen@mail.com	550-241-5926	GBR	3	80648	0	19
9	3	Transient	9	Aaron Allen	Aaron_A86@outlook.com	793-389-6561	PRT	3	90717	1	358
10	4	Transient-Party	10	Aaron Allen MD	AaronMD@verizon.com	939-944-5115	PRT	4	41037	1	74
11	1	Contract	11	Aaron Allison	Aaron_A@gr...	170.1.1.1:49000	EGP	1	11745	0	21
12	3	Transient	12	Aaron Alvarez	Aaron_Alvare...						
13	3	Transient	13	Aaron Anderson	Aaron_Ander...						
14	4	Transient-Party	14	Aaron Anderson	Aaron_Ander...						
15	4	Transient-Party	15	Aaron Anderson	Aaron_Ander...						
16	3	Transient	16	Aaron Andrews	Andrews.Aar...						

Query executed successfully. DESKTOP-3RSABH2\PHUONGUYEN ... DESKTOP-3RSABH2\Phuong... DB\_DATABRICKS 00:00:07 118,216 rows

Hình 4.88: Màn hình xuất file CSV từ câu lệnh SQL

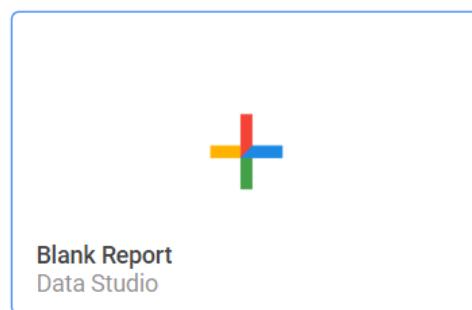
### 3.2. Dựa dữ liệu và lọc dữ liệu từ công cụ Google Data Studio.

- Bước 1: Chuẩn bị file CSV hoặc một file excel đã chứa dữ liệu cần phân tích.
- Bước 2: Địa chỉ của Google Data Studio: <https://datastudio.google.com/>



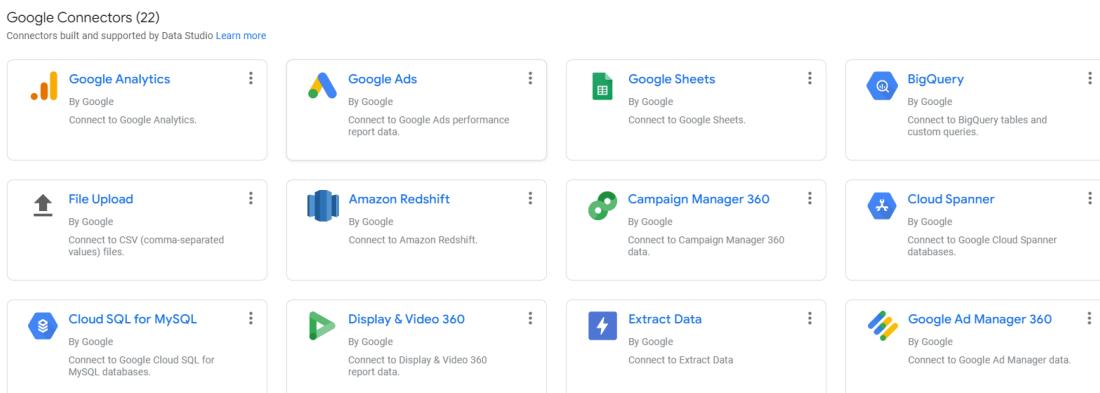
*Hình 4.89: Cửa sổ quản lý của Google Data Studio*

- Bước 3: Đầu tiên hành tạo mới một Report ta nhấn vào Blank Report



*Hình 4.90: Tạo Blank Report*

- Bước 4: Tiếp tục sau đó thêm nguồn dữ liệu bằng file Excel hoặc file csv hoặc có thể lấy dữ liệu từ câu truy vấn Pivot Excel trong công cụ SSAS xuất ra file csv → Chọn File Upload.



*Hình 4.91: Chọn File Upload.*

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Chọn file csv vừa tạo upload lên

The screenshot shows the Google Data Studio 'File Upload' interface. At the top, there's a 'File Upload' button and a note: 'By Google. You can bring data into Data Studio from almost any source by uploading CSV (comma-separated values) files. File upload lets you report on data not supported by a specific connector.' Below this are 'LEARN MORE' and 'REPORT AN ISSUE' links. On the left, a sidebar lists 'Data Sets' with 'olap\_utm.csv' selected. The main area shows the uploaded file 'olap\_utm.csv' with details: 'TOTAL FILE SIZE 27 MB (28% of 100MB used)', 'NUMBER OF FILES 1', 'CREATION DATE 5/14/22 10:29 AM', and 'LAST MODIFIED DATE 5/14/22 10:29 AM'. A 'VIEW FILES IN CLOUD' link is at the bottom right. Below the file details is a table with one row for 'olap\_utm.csv', showing 'File name', 'Uploaded at', 'Size', and 'Status' (Uploaded). There's also a 'DELETE DATA SET' link. At the bottom left is a 'CREATE NEW DATA SET' button.

Hình 4.92: Màn hình Upload từ file CSV

- Cửa sổ định nghĩa lại dữ liệu các trường và thêm các trường cần thiết. → Nhấn Connect.

The screenshot shows the Google Data Studio 'olap\_utm.csv' data source editor. At the top, it displays the file name and various settings: 'shness: 12 hours', 'Community visualizations access: On', 'Field editing in reports: On', 'CREATE REPORT', and 'EXPLORE'. Below these are 'EDIT CONNECTION' and 'FILTER BY EMAIL' buttons, and 'ADD A FIELD' and 'ADD A PARAMETER' buttons. A message box says: 'Data source editors can now refresh fields, edit connections, and edit custom SQL.' with a 'Dismiss' button. The main area is titled 'DIMENSIONS (47)' and lists 47 dimensions in a table. The columns are 'Field', 'Type', and 'Default'. Each dimension entry includes a three-dot menu icon. The table shows various types like Number, Text, ISO Week, Date, Quarter, and Year. At the bottom, there's a 'REFRESH FIELDS' button and a status bar indicating '48 / 48 Fields'.

Hình 4.93 : Màn hình lọc dữ liệu, chỉnh sửa kiểu dữ liệu

### 3.3. Quá trình thực hiện báo cáo bằng công cụ Google Data Studio.

3.3.1. Báo cáo 1: Thống kê số lượng khách hàng ở từng quốc gia từ năm 2015 đến năm 2017 đi du lịch.

- Bước 1: Tạo Header cho Report

The screenshot shows the Google Data Studio interface. At the top left is the Data Studio logo. To its right is the title 'HOTEL CUSTOMER COUNTS REPORT' in a large, bold, blue font. Below the title is a search bar with the placeholder 'Type to search'.

Hình 4. 94: Thêm Header cho Report 4 (Google Data Studio).

- Bước 2: Chọn dạng bảng tổng hợp là Table.
  - Các Dimension chọn các country.
  - Các Metric lọc theo adults, babies, children.

The screenshot shows a table report in Google Data Studio. The table has columns: country, adults, babies, and children. The data is as follows:

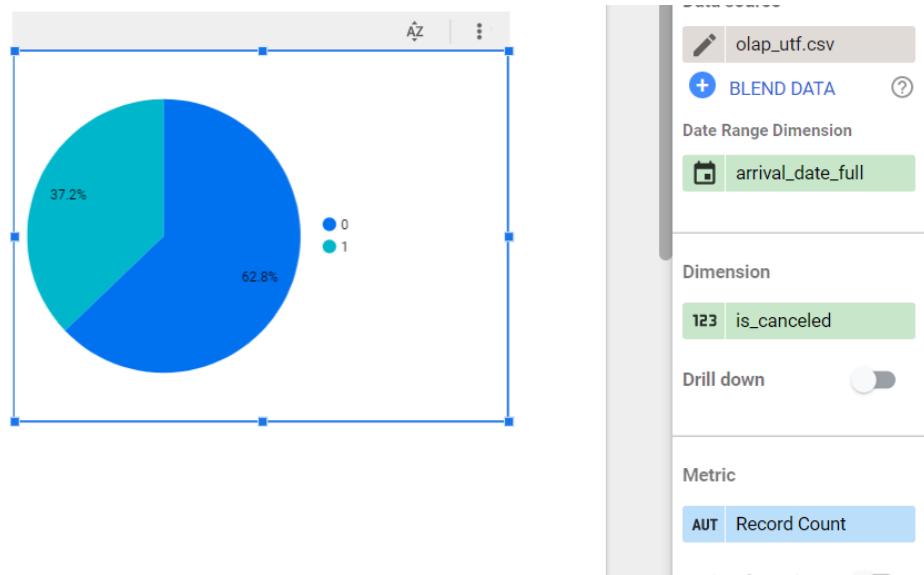
	country	adults	babies	children
1.	PRT	85,524	430	3,436
2.	GBR	23,184	92	1,253
3.	FRA	20,136	75	1,207
4.	ESP	16,038	125	1,405
5.	DEU	13,701	18	477
6.	ITA	7,380	16	456
7.	IRL	6,568	24	315
8.	BEL	4,586	12	311
9.	PER	4,447	24	292

To the right of the table is a configuration sidebar:

- DATA**:
  - Date Range Dimension: arrival\_date\_full
  - Dimension: country
    - RBC country
    - + Add dimension
  - Drill down:
- STYLE**
- Metric**:
  - SUM adults
  - SUM babies
  - SUM children
  - + Add metric
- Optional metrics**:
- Available Fields**: A list of fields including adr, adults, arrival\_date\_day\_name, arrival\_date\_month\_name, arrival\_date\_quarter, arrival\_date\_week\_number, arrival\_date\_weekend, arrival\_date\_year, arrival\_date\_full, arrival\_date\_month, arrival\_date\_quarter, arrival\_date\_weekend, arrival\_date\_year, and arrival\_date\_full.
- Actions**: ADD A FIELD, ADD A PARAMETER.

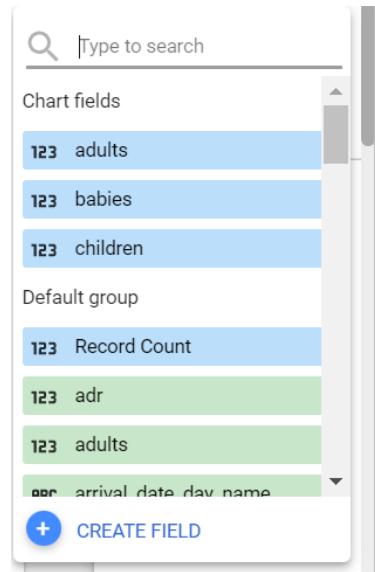
Hình 4.95: Bảng truy vấn.

- Bước 3: Chọn Pie Chart trong tab menu Insert để tiến hành tạo biểu đồ tròn:
  - Các Dimension chọn các is\_canceled.
  - Các Metric lọc theo Record Count.



Hình 4.96: Biểu đồ tròn thể hiện số lượng khách hàng hủy phòng và nhận phòng.

- Bước 4: Chọn Column Chart trong tab menu Insert để tiến hành tạo biểu đồ cột:
  - Các Dimension chọn các arrival\_date\_month\_name.
  - Các Metric lọc theo adults, babies, childrens.
  - Độ đo total được tạo tại Metric:



Hình 4.97: Giao diện lựa chọn các độ đo.

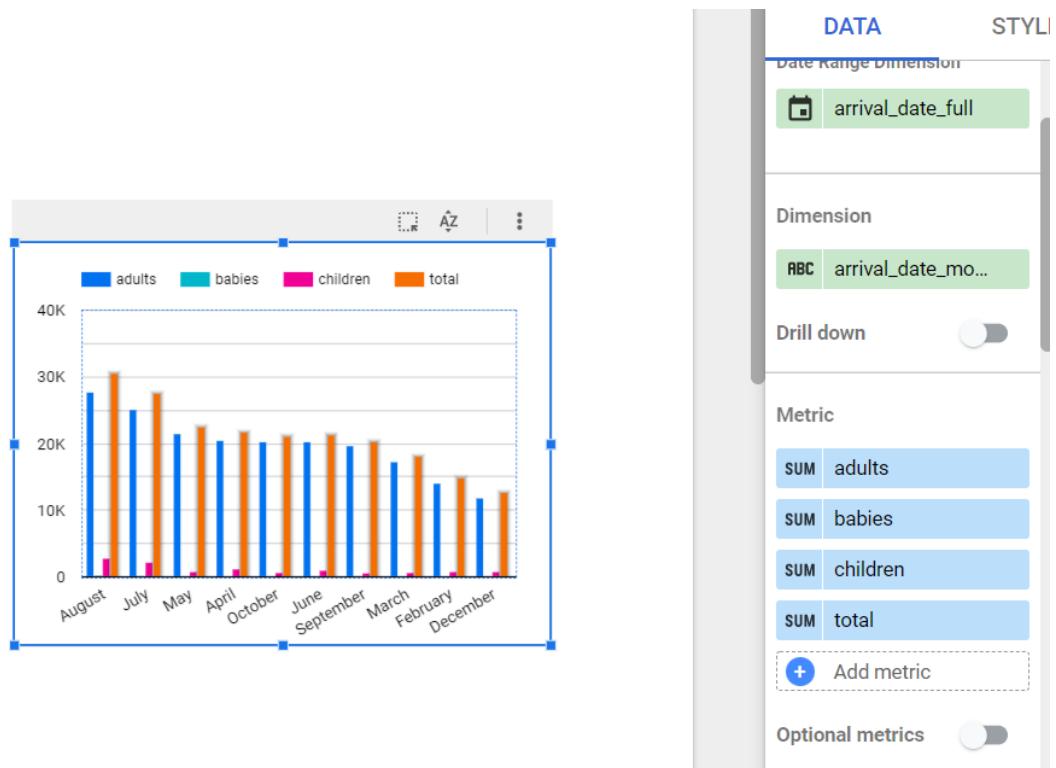
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

The screenshot shows the 'Create Field' dialog box. At the top, it says 'AUT New Field'. Below that is a 'Name' input field containing 'total'. Under 'Formula', there is a large text area with a single digit '1' on the left. To the right of the formula area are dropdown menus for 'Type' (set to 'Number'), 'Comparison calculation' (set to 'None'), and 'Running calculation' (set to 'None'). A blue 'APPLY' button is located at the bottom right.

*Hình 4.98: Giao diện sau khi chọn Create Field*

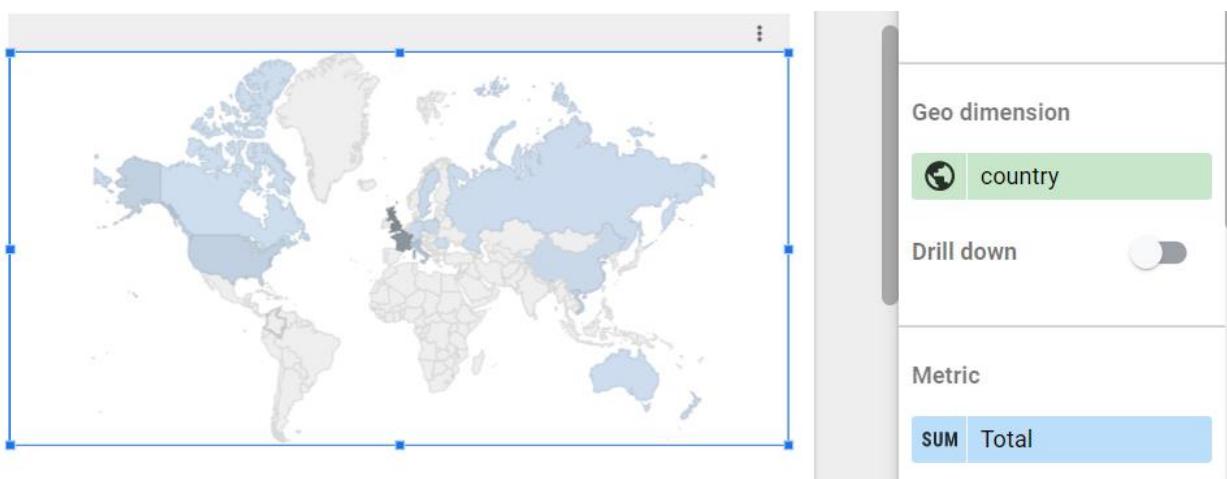
This screenshot shows the same 'Create Field' dialog box as above, but with a more complex formula entered in the formula field: '1 + adults + babies + children'. The formula is highlighted with green boxes around each term. The rest of the dialog box elements are identical to the previous screenshot.

*Hình 4.99: Điền vào Formula các phép tính để tạo ra một độ đo mới. Sau đó nhấn Apply.*



*Hình 4.100: Biểu đồ cột thống kê số lượng khách hàng theo người lớn, trẻ em, và em bé khi đặt phòng tại khách sạn.*

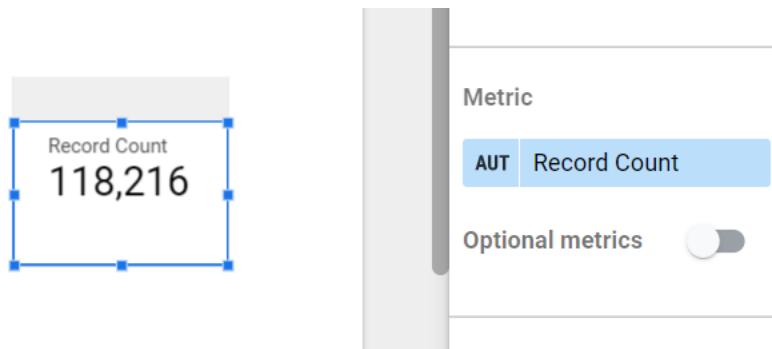
- Bước 5: Chọn Geo Chart trong tab menu Insert để tiến hành tạo bản đồ:
  - Các Geo Dimension chọn các country.
  - Các Metric lọc theo total.



*Hình 4.101: Bản đồ thế giới thể hiện sự phân bố của lượng người thuê phòng tại khách sạn.*

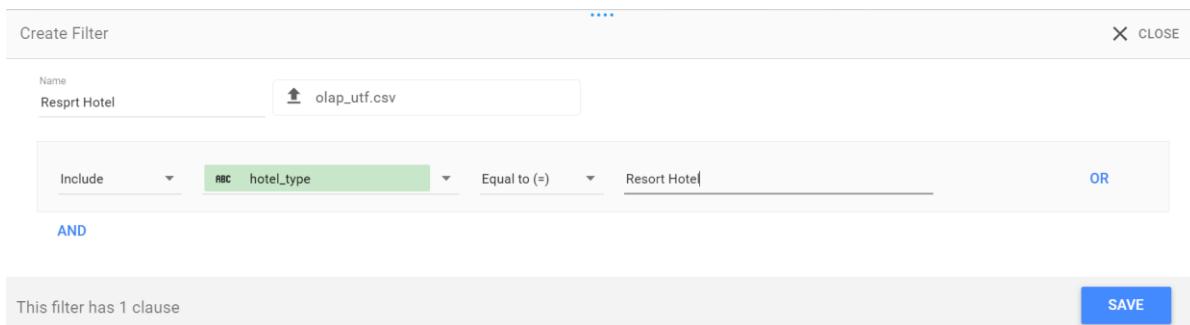
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 6: Chọn ScoreCard trong tab menu Insert để tiến hành tạo bảng thống kê:
  - Các Geo Dimension chọn các country.
  - Các Metric lọc theo total.

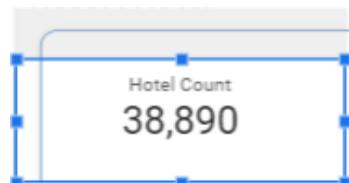


*Hình 4.102: Bảng thống kê số lượng đặt phòng.*

- Trường Filter với mỗi ScoreCard sẽ lọc với loại khách sạn khác nhau.



*Hình 4.103 Lọc dữ liệu theo loại khách sạn là Resort Hotel.*



*Hình 4.104: Kết quả ScoreCard thống kê số lượng khách hàng đặt phòng tại loại khách sạn là Resort Hotel.*

- Tương tự, ta lần lượt tạo các ScoreCard thống kê số lượng khách hàng tại City Hotel và doanh thu của từng loại khách sạn.

Hotel Count  
79,326

Hình 4.105: Kết quả ScoreCard thống kê số lượng khách hàng đặt phòng tại loại khách sạn là City Hotel.

Hotel Revenue  
8,353,684.74

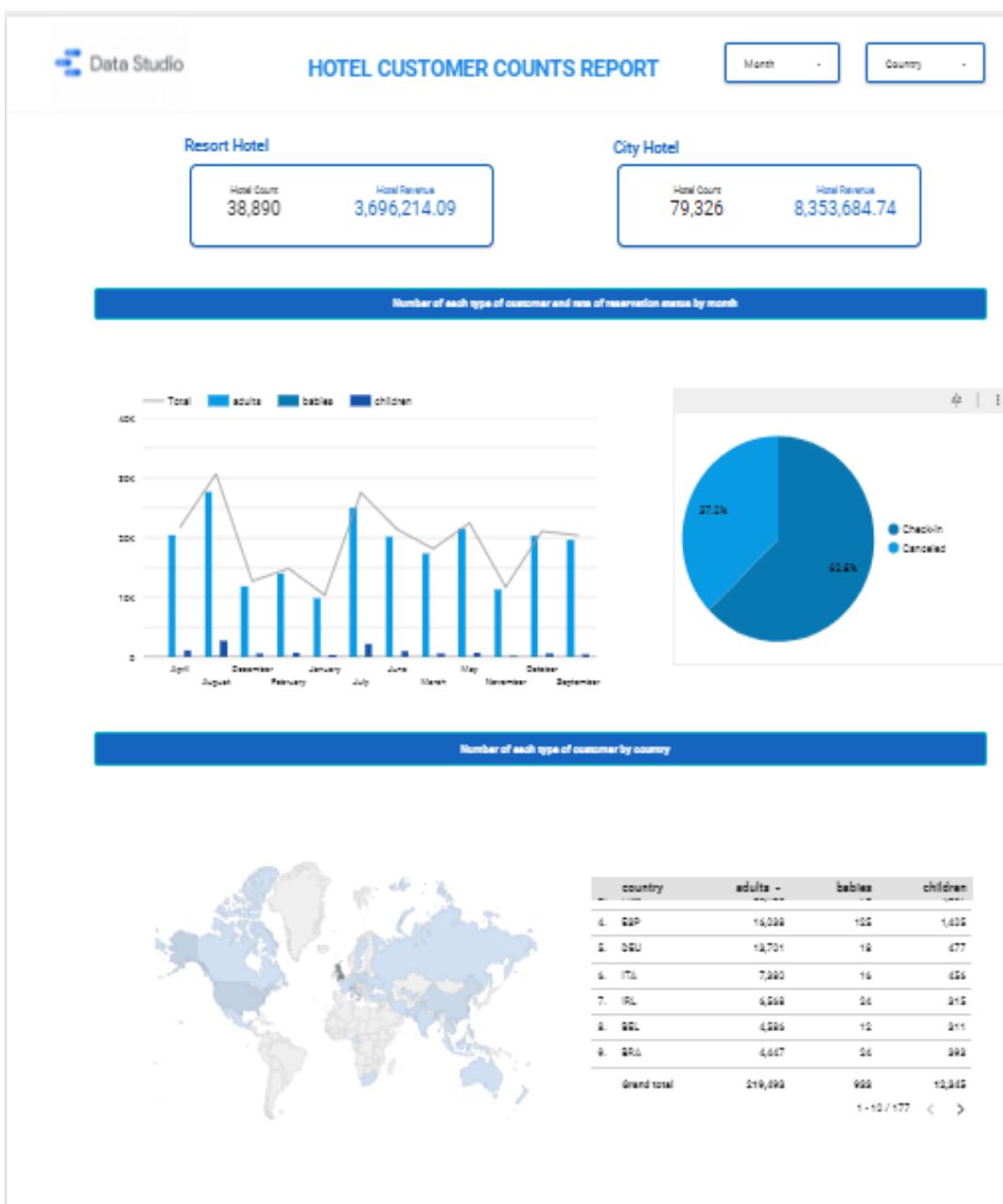
Hình 4.106: Kết quả ScoreCard thống kê doanh thu khách hàng đặt phòng tại loại khách sạn là City Hotel.

Hotel Revenue  
3,696,214.09

Hình 4.107: Kết quả ScoreCard thống kê doanh thu khách hàng đặt phòng tại loại khách sạn là Resort Hotel.

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

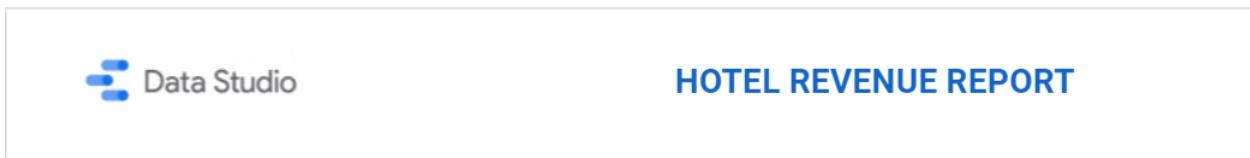
- Bước 7: Điều chỉnh các thông số cho Report hoàn thành ta được kết quả như hình dưới:



*Hình 4.108: Bảng thống kê doanh thu, số lượng khách hàng của từng quốc gia đăng ký đặt phòng tại hai loại khách sạn.*

3.3.2. Báo biểu 2: Thống kê doanh thu theo từng tháng và từng năm của từng kênh phân phối.

- Bước 1: Tạo Header cho Report



Hình 4.109: Thêm Header cho Report 5 (Google Data Studio).

- Bước 2: Chọn dạng bảng tổng hợp là Heat Map.
- Trong tab Data,
  - Các Row Dimension chọn arrival\_date\_month\_name.
  - Các Column Dimension chọn distribution\_channel.
  - Các Metric lọc theo adr.
  - Click chọn Show grand total theo cột và theo dòng để tính tổng các giá trị.
- Trong tab Style, thay đổi Metric #1 từ dạng Bar sang HeatMap.

Distribution Channel / Average Daily Rate			
Number of year	TA/TO	GDS	Grand total
2017			4,572,152.3
2016			5,522,993.1
2015			1,721,594.8
Grand total	9,830,053.4	23,267.0	11,816,740.2

Hình 4.110: Bảng truy vấn.

## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

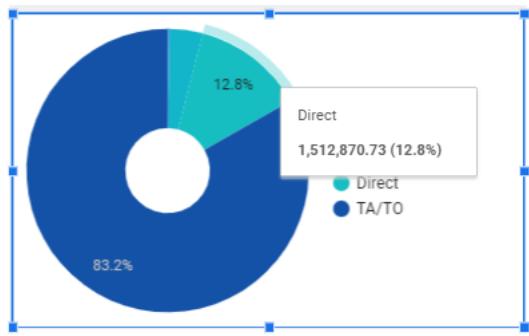
The screenshot shows the 'DATA' tab interface. On the left, there are three main sections: 'Row dimension', 'Column dimension', and 'Metric'. Under 'Row dimension', 'reservation\_stat...' is selected. Under 'Column dimension', 'Distribution Cha...' is selected. Under 'Metric', 'Avarage Daily Ra...' is selected. To the right is a sidebar titled 'Available Fields' containing a list of fields such as adr, adults, agent, arrival\_date\_day\_of\_m.., arrival\_date\_month, arrival\_date\_week\_nu..., arrival\_date\_year, assigned\_room\_type, babies, booking\_changes, children, company, country, credit\_card, customer\_type, days\_in\_waiting\_list, ADD A FIELD, and ADD A PARAMETER. There are also buttons for 'Add dimension' and 'Add metric'.

*Hình 4.111: Tab data.*

The screenshot shows the 'STYLE' tab interface. It includes sections for 'Table Labels' (with font size set to 11px and font family set to Open Sans) and 'Missing Data'. The 'Missing Data' section features a heatmap grid where 'Heatmap' is currently selected. Below the grid are options for 'Show number' (unchecked) and 'Bar scaling' (radio button set to 'Relative to data').

*Hình 4.112: Tab Style.*

- Bước 3: Chọn Pie Chart trong tab menu Insert để tiến hành tạo biểu đồ tròn:
  - Các Dimension chọn các distribution\_channel.
  - Các Metric lọc theo adr.

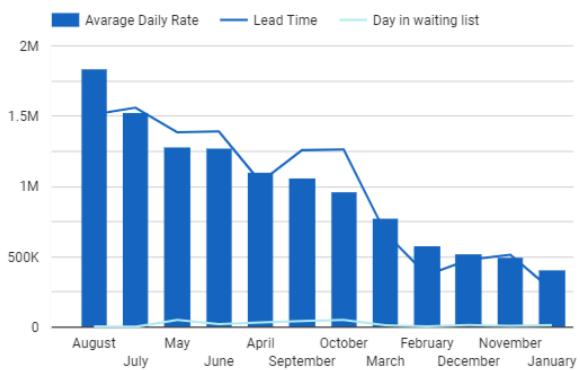


*Hình 4.113: Biểu đồ tròn thể hiện doanh thu của từng kênh phân phối.*

*Hình 4.114: Tab Data.*

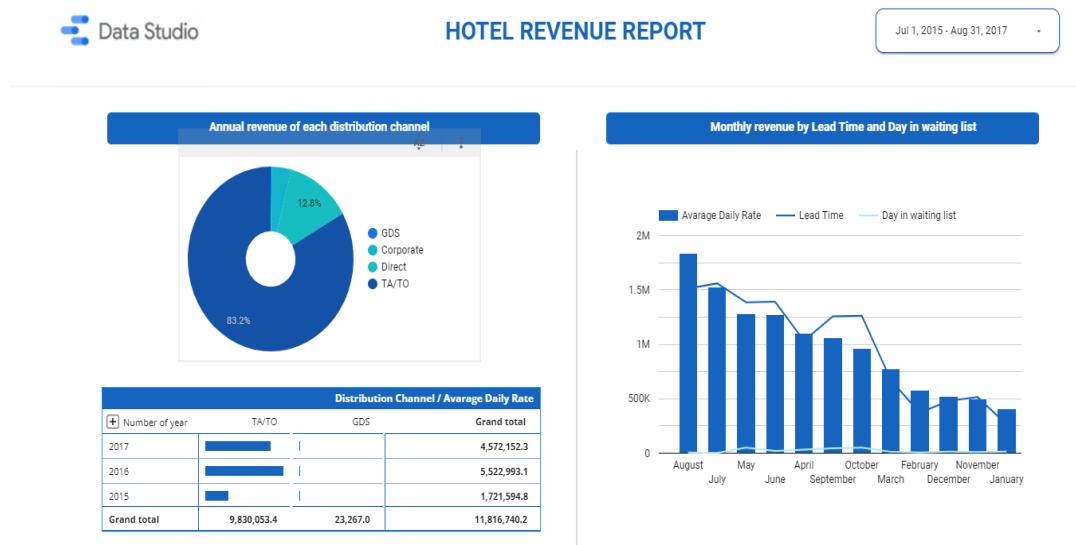
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 4: Chọn Combo Chart trong tab menu Insert để tiến hành tạo biểu đồ cột và đường:
  - Các Dimension chọn các arrival\_date\_month\_name.
  - Các Metric lọc theo adr, lead\_time, days\_in\_waiting\_list.



*Hình 4.115: Biểu đồ thể hiện tổng doanh thu, tổng lượng thời gian chờ và tổng thời gian xác nhận đặt phòng theo từng năm.*

- Bước 6: Điều chỉnh các thông số cho Report hoàn thành ta được kết quả như hình dưới:



*Hình 4.116: Bảng thống kê doanh thu, tổng thời gian chờ của từng kênh phân phối trong năm 2015 và 2017.*

3.3.3. Báo biểu 3: Thông kê số lượng yêu cầu đặc biệt của từng phân khúc thị trường trong năm 2015, 2016 và 2017.

- Bước 1: Tạo Header cho Report



### HOTEL MARKET SEGMENT'S REQUEST REPORT

Hình 4.117 Thêm Header cho Report 5 (Google Data Studio).

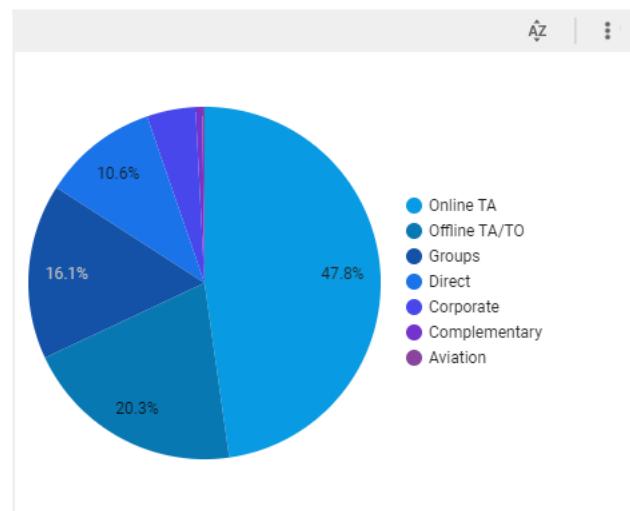
- Bước 2: Chọn dạng bảng tổng hợp là Pivot Table.
- Trong tab Data,
  - Các Row Dimension chọn market\_segment, assigned\_room\_type, reserved\_room\_type.
  - Các Column Dimension chọn meal.
  - Các Metric lọc theo total\_of\_special\_requests và required\_parking\_spaces.
  - Click chọn Show grand total theo cột và theo dòng để tính tổng các giá trị của độ đo.

		Meal / Total of special requests / Required parking car spaces									
market_segment	assigned_room_ty...	reserved_room...	Total of specia...	Required parki...	Total of specia...	Required parki...	Total of specia...	Required parki...	Total of ...	Required...	Grand total
Online TA	P	P	0	0	-	-	-	-	-	0	0
			<b>Total</b>	<b>0</b>	<b>0</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>0</b>	<b>0</b>
	K	F	-	-	-	-	-	-	-	1	0
		E	0	0	-	-	-	-	-	1	0
		D	5	0	2	0	-	-	-	30	0
		A	58	3	2	0	-	-	-	114	5
		<b>Total</b>	<b>63</b>	<b>3</b>	<b>4</b>	<b>0</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>146</b>	<b>5</b>
	I	H	-	-	-	-	-	-	-	3	2
		G	-	-	2	2	-	-	-	4	3
<b>Grand total</b>			<b>7,952</b>	<b>241</b>	<b>7,690</b>	<b>987</b>	<b>186</b>	<b>55</b>	<b>68,002</b>	<b>7,419</b>	

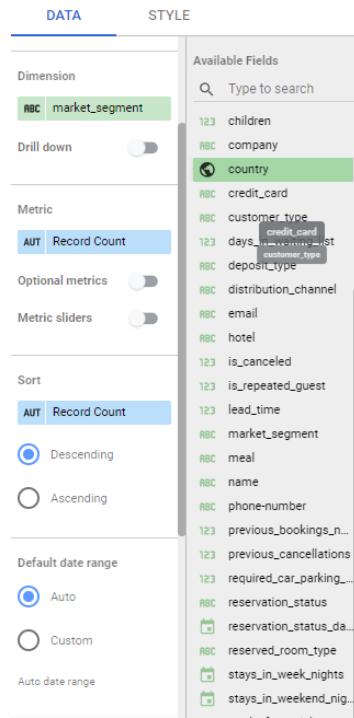
Hình 4.118: Bảng truy vấn.

*Hình 4.119: Tab data.*

- Bước 3: Chọn Pie Chart trong tab menu Insert để tiến hành tạo biểu đồ tròn:
  - Các Dimension chọn các market\_segment.
  - Các Metric lọc theo Record Count.

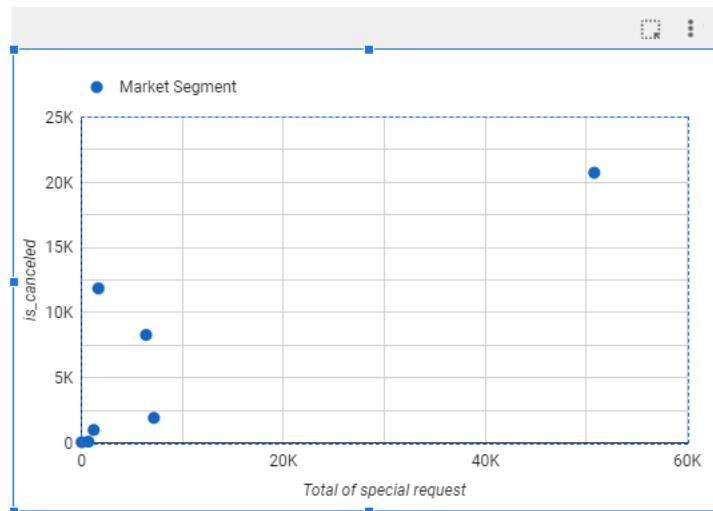


*Hình 4.120: Biểu đồ tròn thể hiện tổng số lượt đặt phòng của từng phân khúc thị trường.*



Hình 4.121: Tab Data.

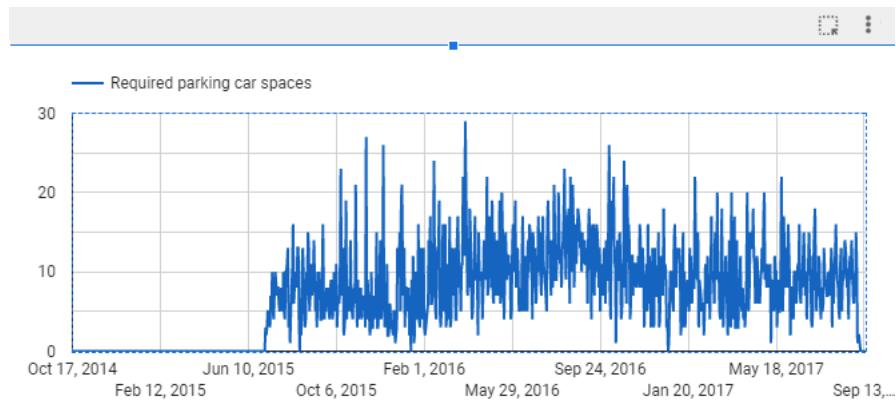
- Bước 4: Chọn Scatter Chart trong tab menu Insert để tiến hành tạo biểu đồ thể hiện sự phân bố của dữ liệu:
  - Các Dimension chọn các market\_segment.
  - Các Metric lọc theo total\_of\_special\_requests (theo cột X\_ và is\_canceled (theo cột Y)



Hình 4.122: Biểu đồ thể hiện sự phân bố của các phân khúc thị trường và cho biết lượng yêu cầu đặc biệt và lượng khách hàng trong từng phân khúc đó.

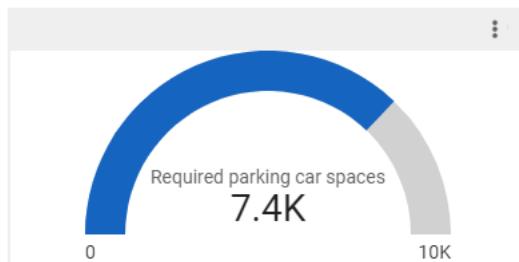
## *Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*

- Bước 5: Chọn Time Series trong tab menu Insert để tiến hành tạo biểu đồ:
  - Các Dimension chọn các reservation\_status\_date.
  - Các Metric lọc theo required\_paking\_car\_spaces.



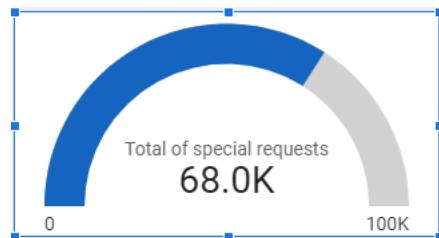
*Hình 4.123: Biểu đồ thể hiện sự thay đổi của số chỗ bãi đậu xe yêu cầu của khách hàng.*

- Bước 6: Chọn Gauge trong tab menu Insert để tiến hành tạo biểu đồ:
  - Các Metric lọc theo required\_paking\_car\_spaces.



*Hình 4.124: Biểu đồ thể hiện tổng số yêu cầu bãi đỗ xe của khách hàng và so sánh với giá trị mục tiêu.*

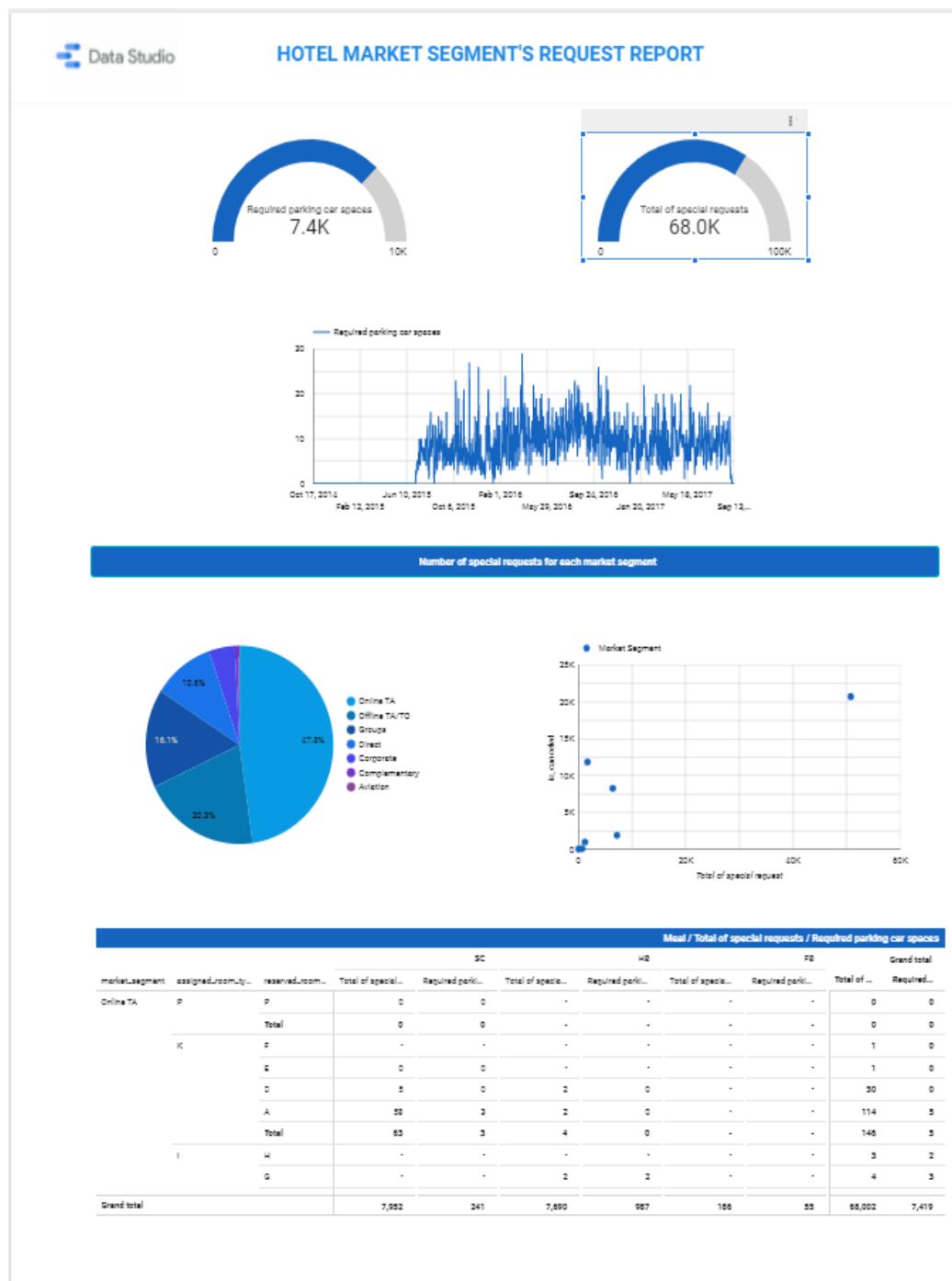
- Các Metric lọc theo total\_of\_special\_requests.



*Hình 4.125 : Biểu đồ thể hiện tổng số yêu cầu đặc biệt của khách hàng và so sánh với giá trị mục tiêu.*

## Đồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 7: Điều chỉnh các thông số cho Report hoàn thành ta được kết quả như hình dưới:

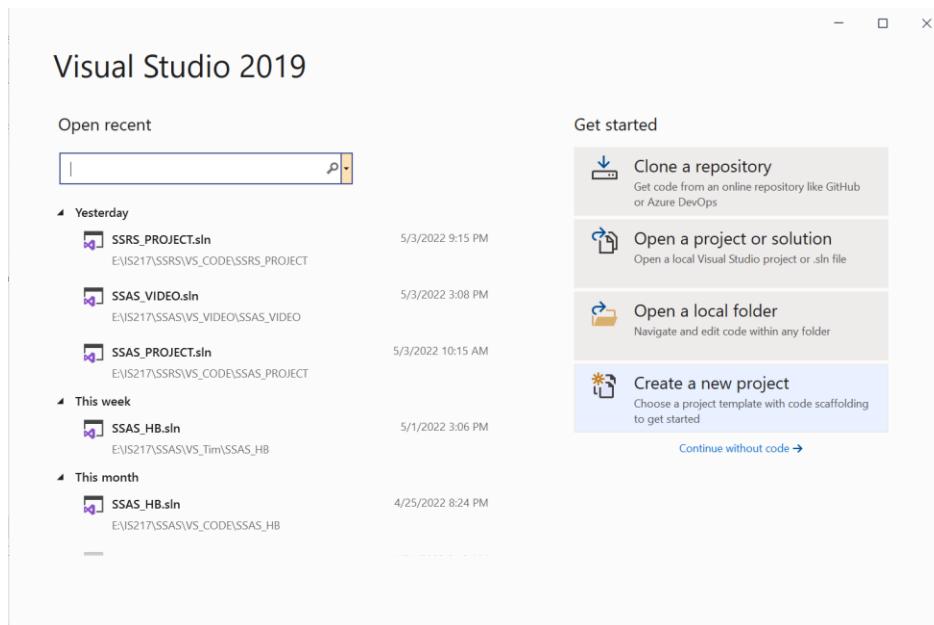


Hình 4.126: Bảng thống kê tổng yêu cầu đặc biệt và yêu cầu bãi đỗ xe của từng phân khúc thị trường.

## CHƯƠNG V. QUÁ TRÌNH DATA MINING

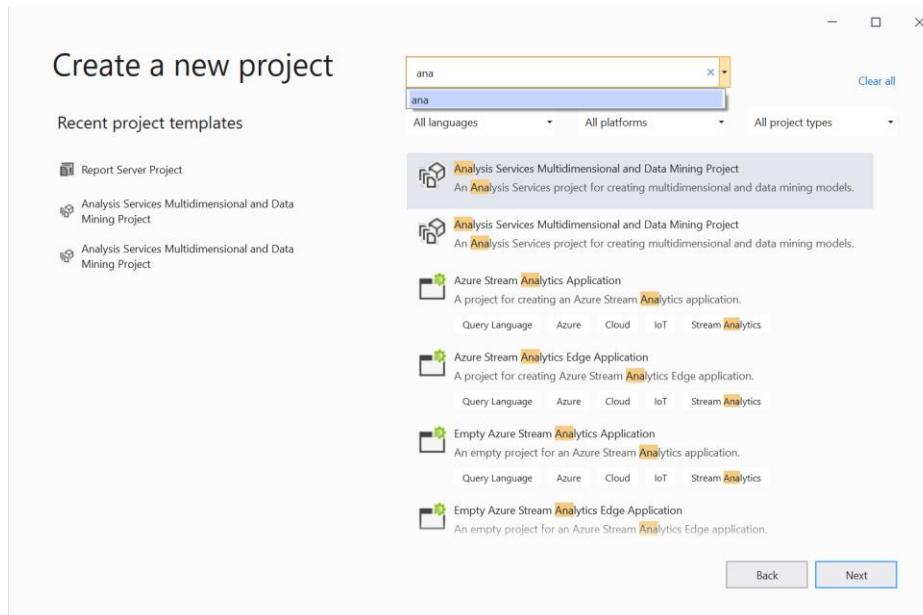
### 1. Tạo project Data Mining trong Visual Studio 2019

- Bước 1: Chọn phần mềm Visual Studio 2019 → Create a new project như hình:



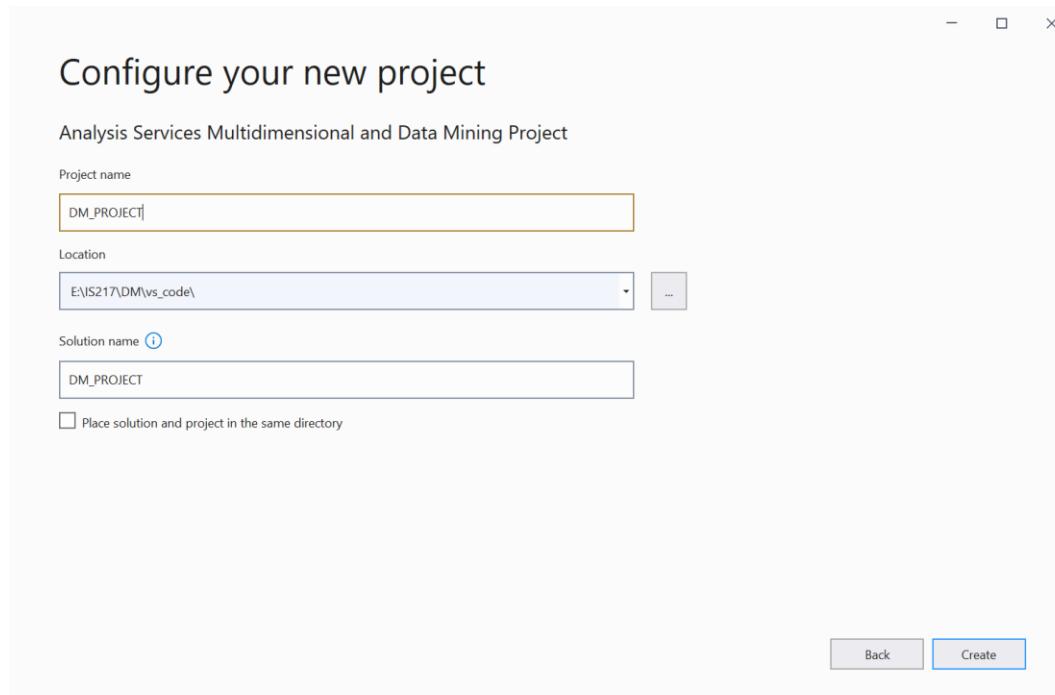
*Hình 5.1: Giao diện khi sử dụng Visual Studio 2019*

- Bước 2: Tìm project có công cụ Analysis Services Multidimensional and Data Mining project → Next như hình:



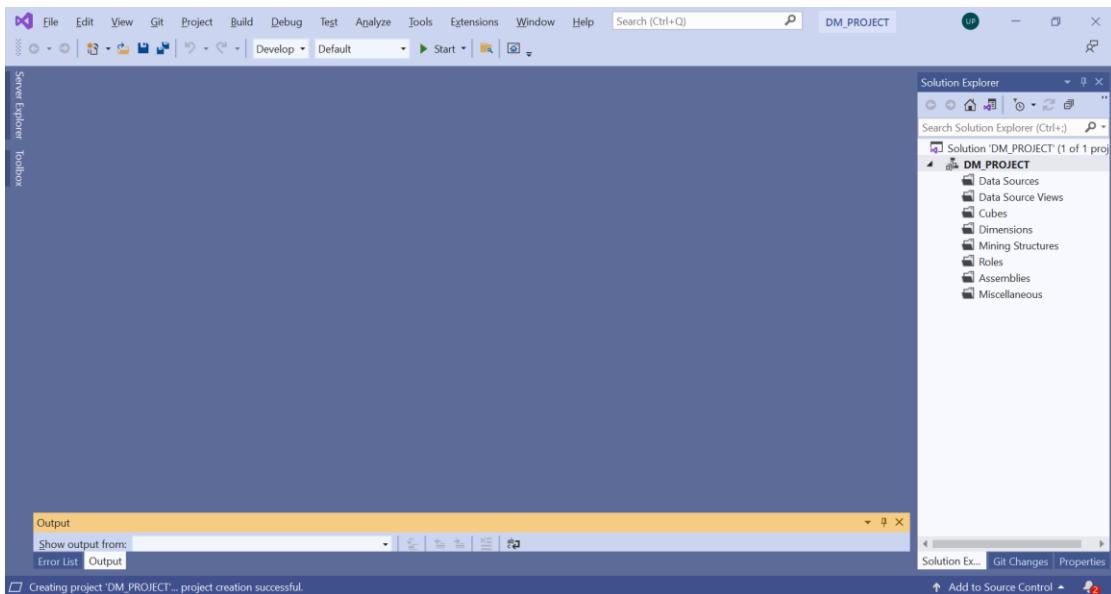
*Hình 5.2 Tìm kiếm một Project Data Mining của Visual Studio*

- Bước 3: Chọn đường dẫn để lưu project, đặt tên và nhấn Create.



*Hình 5.3 Giao diện cấu hình tên file của Data Mining*

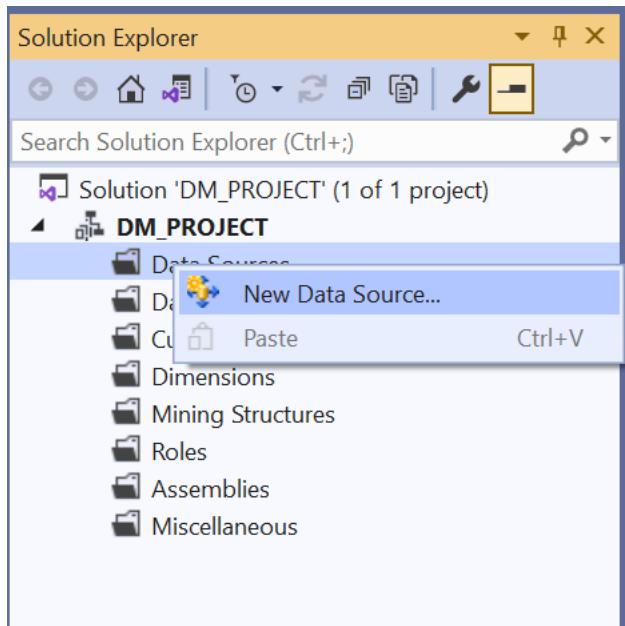
- Bước 4: Giao diện quá trình Data Mining của công cụ BI hiện ra.



*Hình 5.4. Giao diện của công cụ BI quá trình Data Mining*

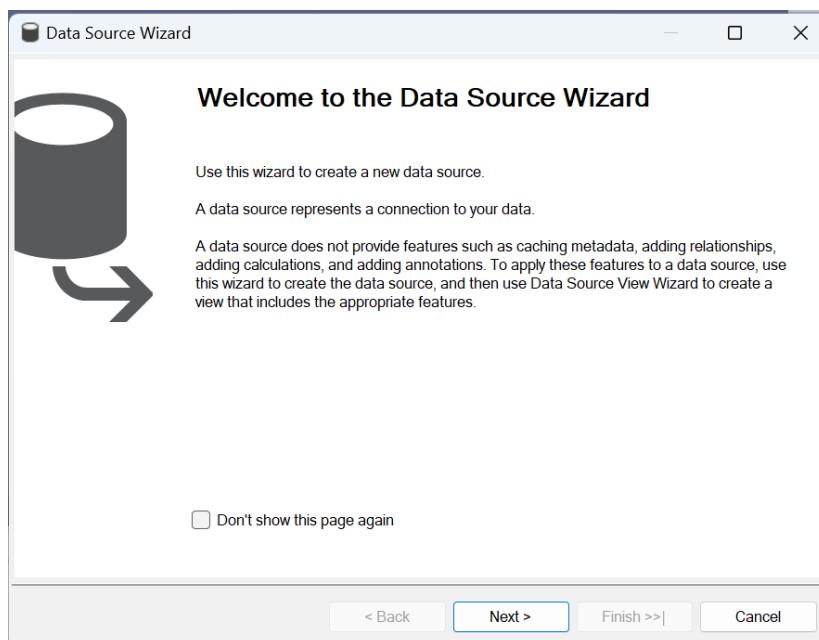
## **2. Xác định dữ liệu nguồn (DEFINE DATA SOURCE)**

- Bước 1: Bên gốc phải phần Solution Explorer nhấn chuột phải vào Data Source và chọn New Data Source.



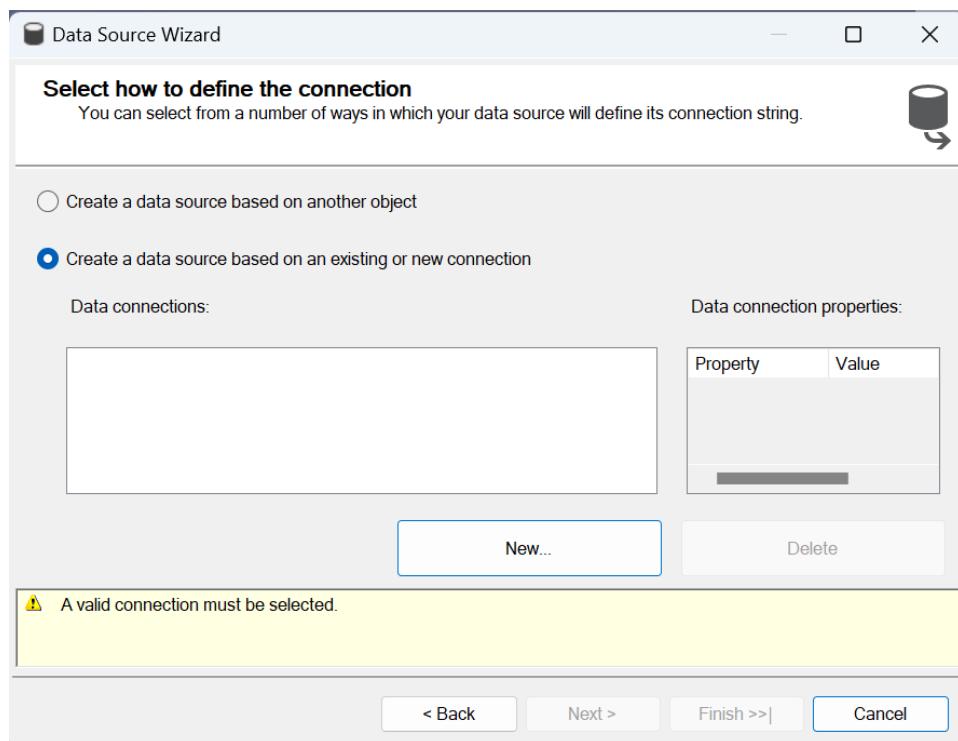
*Hình 5.5. Tạo một Data Source cho quá trình Data Mining*

- Bước 2: Một cửa sổ hiện ra tiếp tục nhấn Next để tiếp tục.

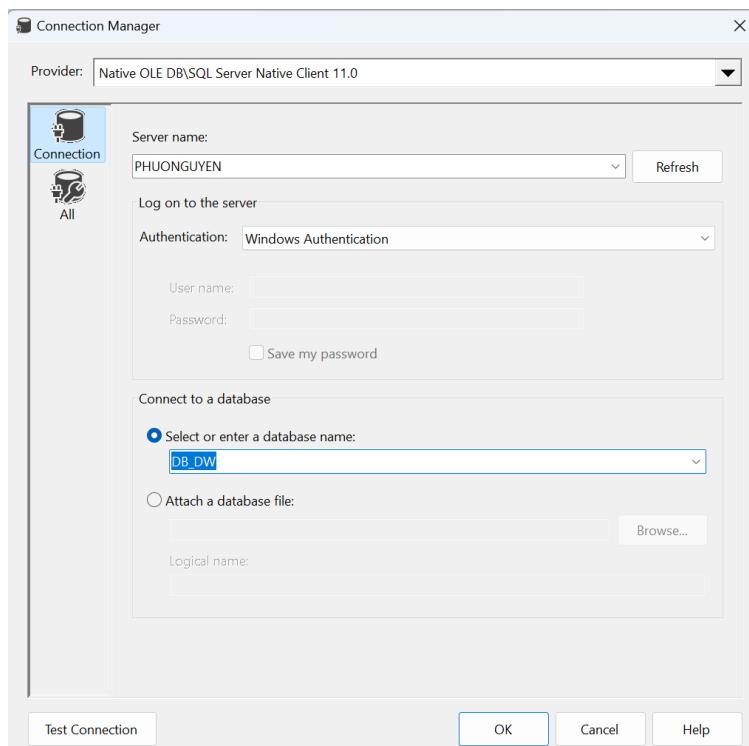


*Hình 5.6. Phần giới thiệu Data Source Wizard*

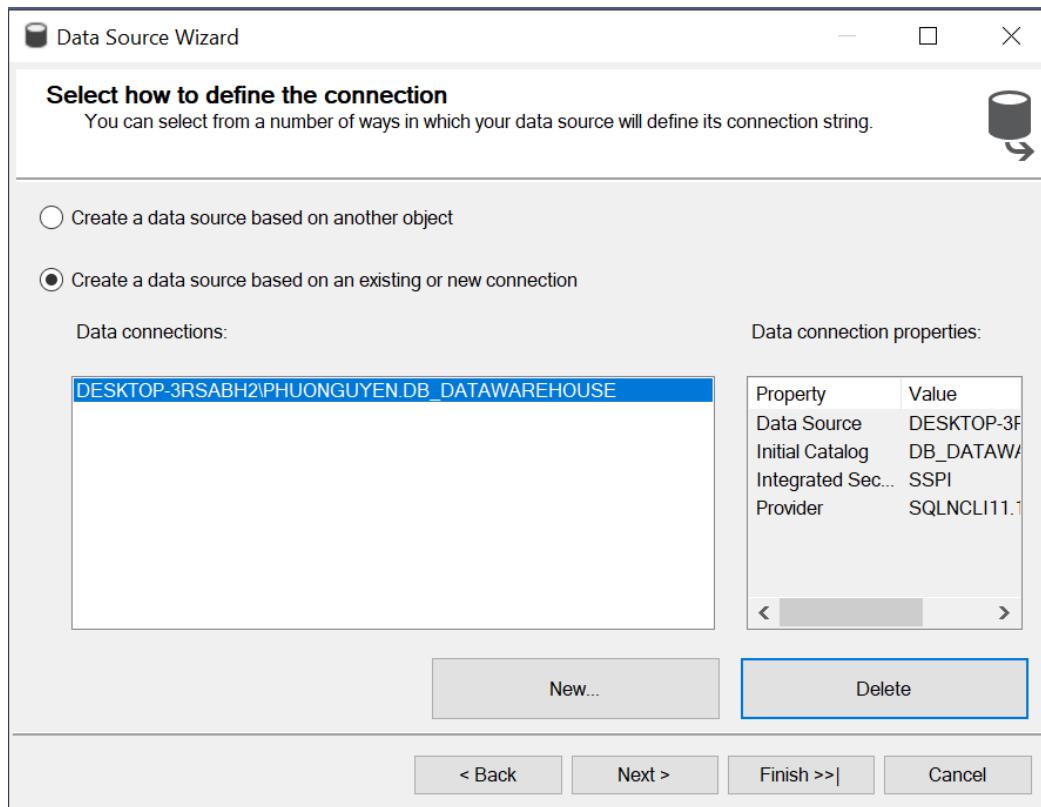
- Bước 3: Chọn kết nối đã tạo từ quá trình SSIS rồi tiếp tục nhấn Next.



*Hình 5.7. Chọn kết nối với Data Source đã có sẵn hoặc tạo kết nối mới*

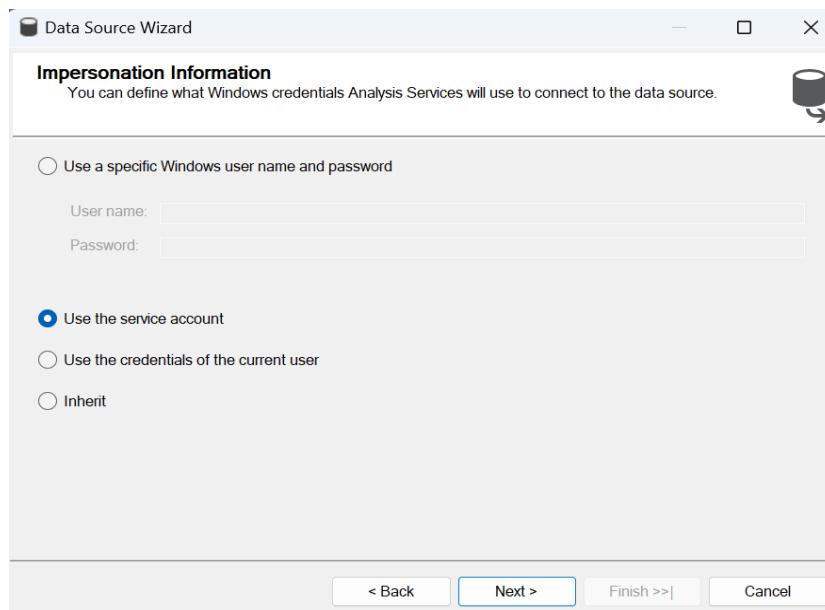


*Hình 5.8. Tao kết nối mới khi project chạy lần đầu*



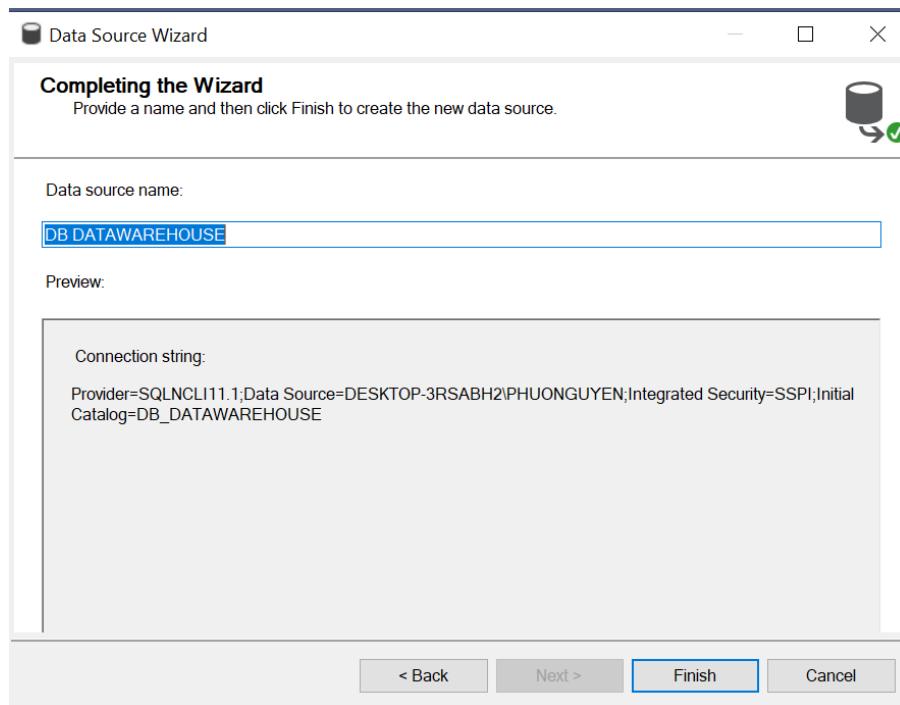
*Hình 5.9. Chọn kết nối với Data Source*

- Bước 4: Chọn Use the service account nhấn Next.



*Hình 5.10. Lựa chọn tài khoản phù hợp*

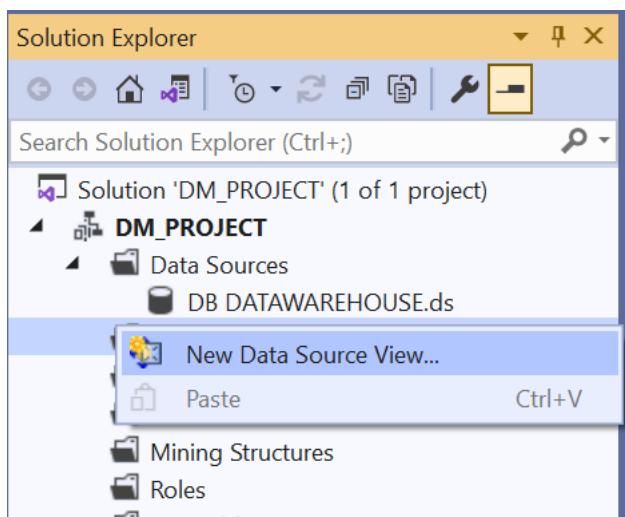
- Bước 5: Nhấn Finish để kết thúc quá trình xác định dữ liệu nguồn.



*Hình 5.11. Kết thúc quá trình Define Data Source*

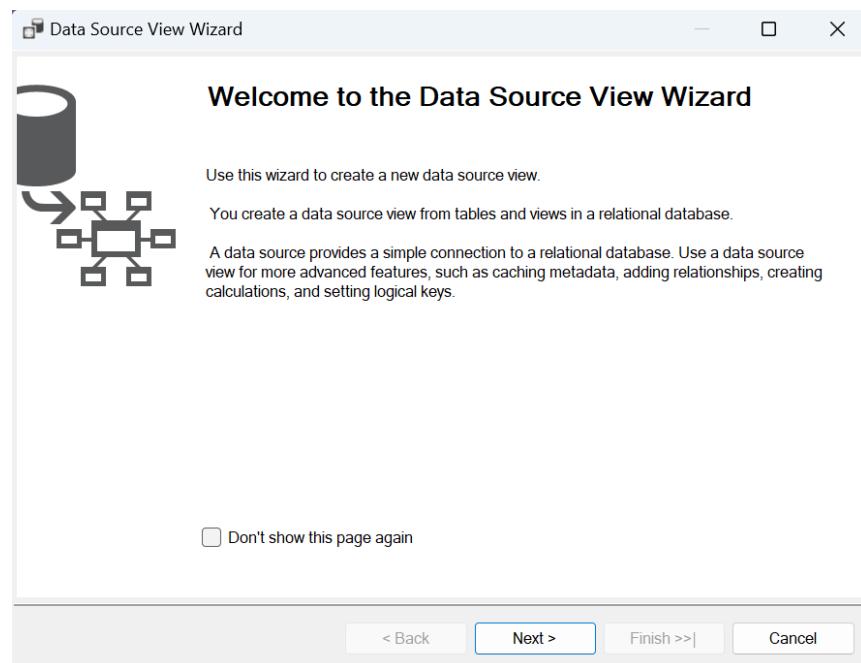
### **3. Xác định khung nhìn dữ liệu nguồn (DEFINE DATA SOURCE VIEW)**

- Bước 1: Bên gốc phải phần Solution Explorer nhấn chuột phải vào Data Source View và chọn New Data Source View.



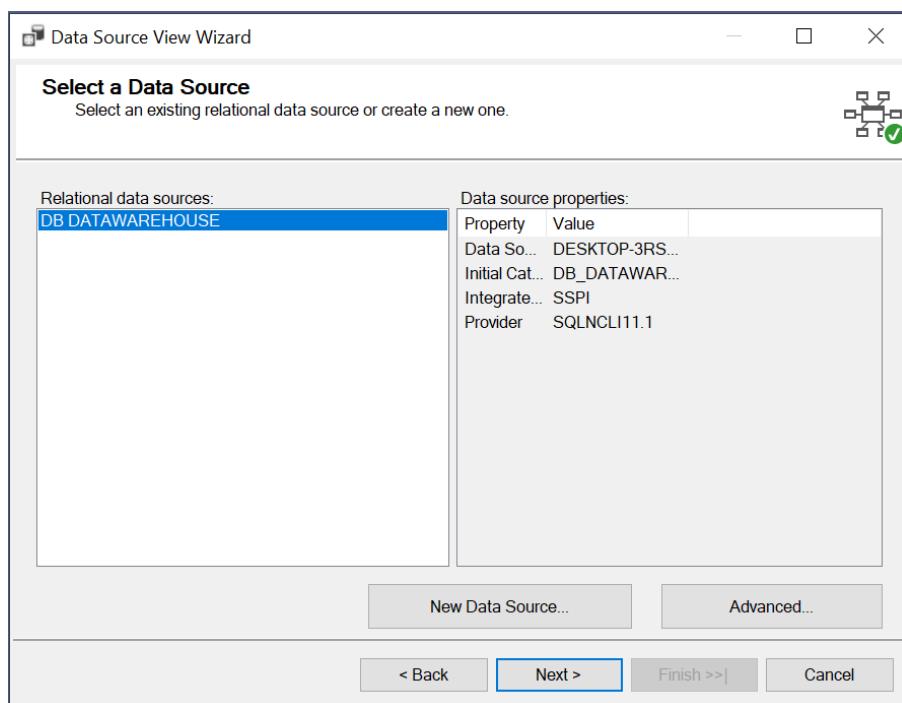
*Hình 5.12. Tạo Data Source View cho quá trình Data Mining*

- Bước 2: Một cửa sổ hiện ra tiếp tục nhấn Next để tiếp tục.



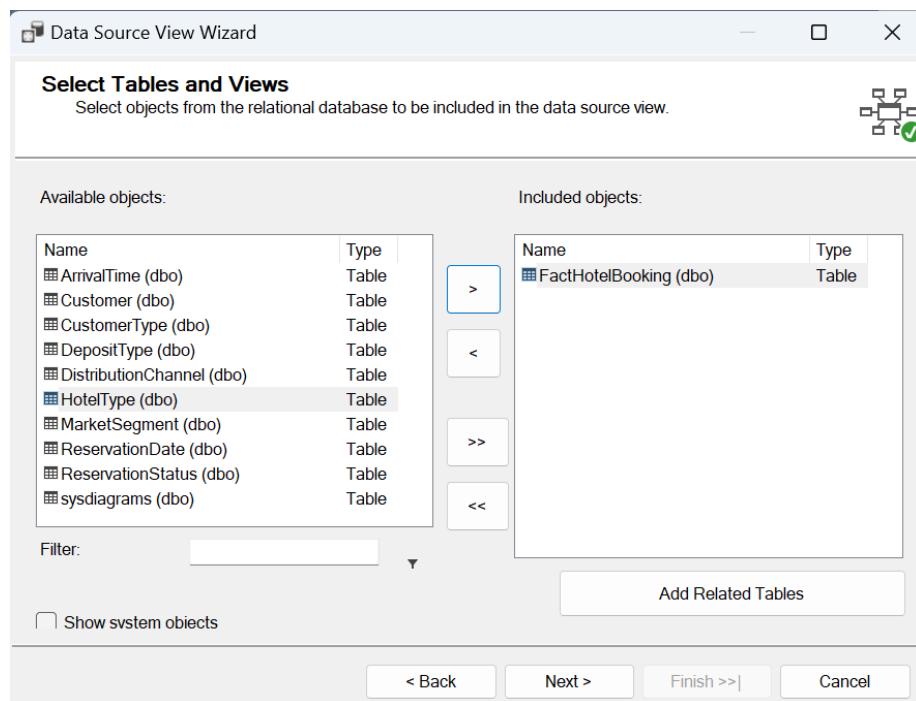
*Hình 5.13. Phân giới thiệu Data Source View Wizard*

- Bước 3: Cho kho dữ liệu DB WAREHOUSE rồi nhấn Next để tiếp tục

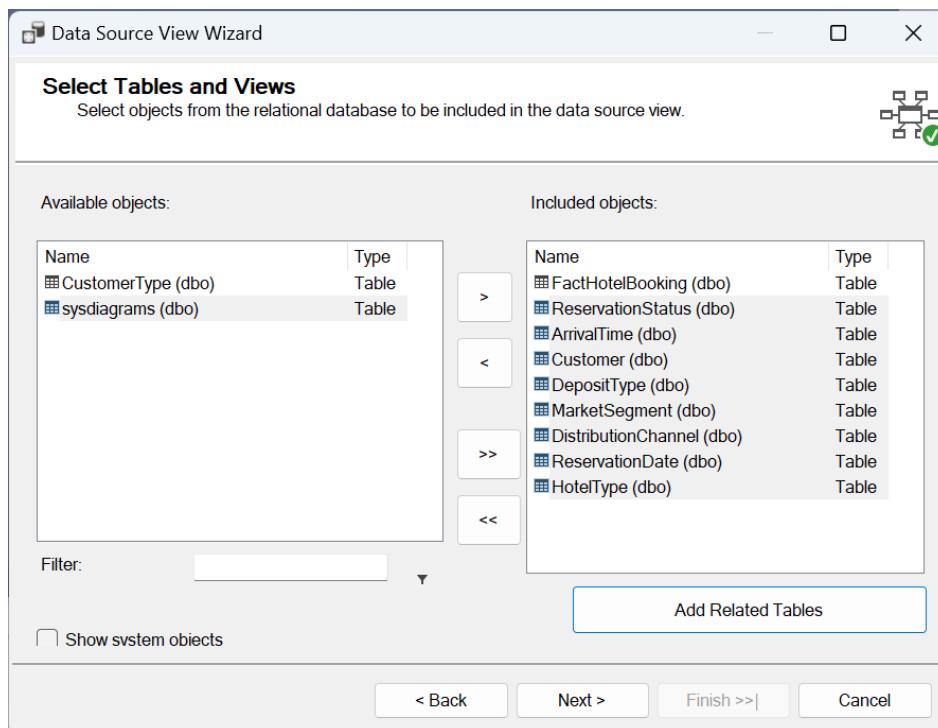


*Hình 5.14. Chọn kho dữ liệu cho quá trình Data Mining*

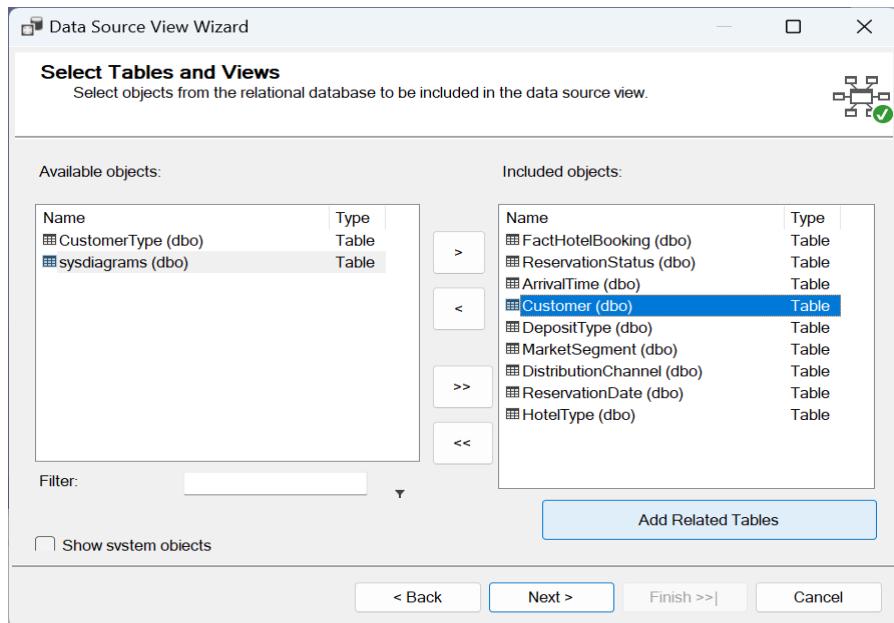
- Bước 4: Chọn bảng FACT và DIM cho quá trình phân tích sau đó nhấn Next.



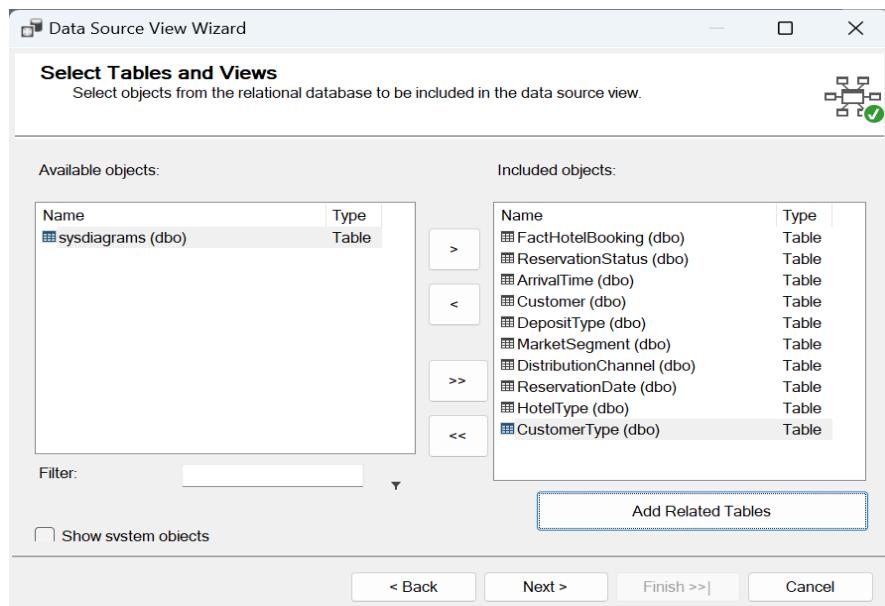
*Hình 5.15. Chọn bảng dữ liệu Fact cho quá trình phân tích*



*Hình 5.16. Chọn bảng Dimension có liên quan đến Fact*

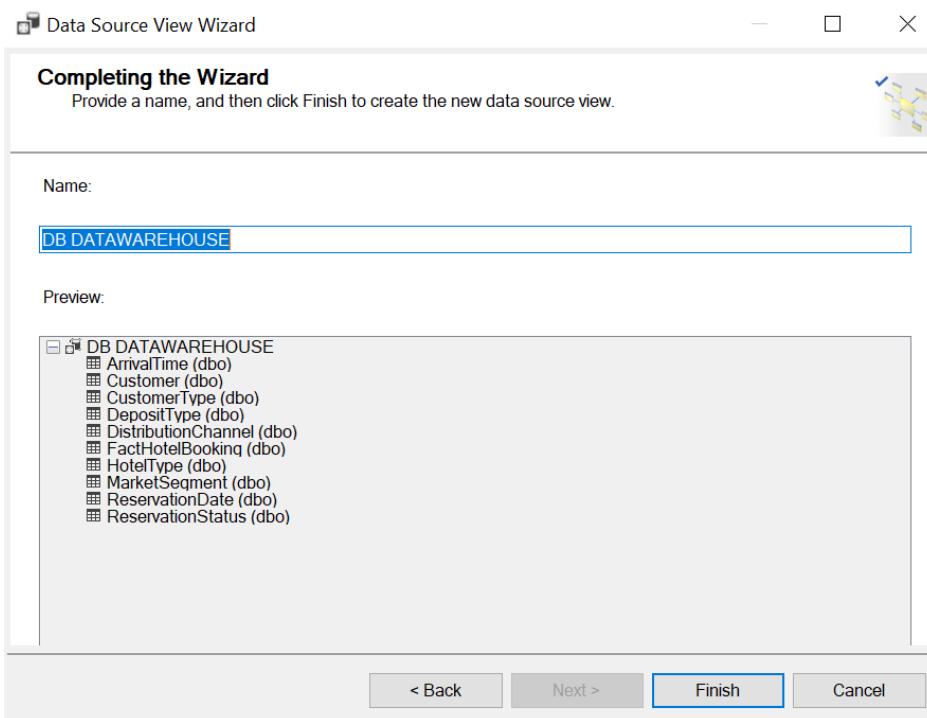


*Hình 5.17. Chọn bảng Customer và tiếp tục chọn Add related Table để lấy các bảng Dimension có liên quan với Customer*



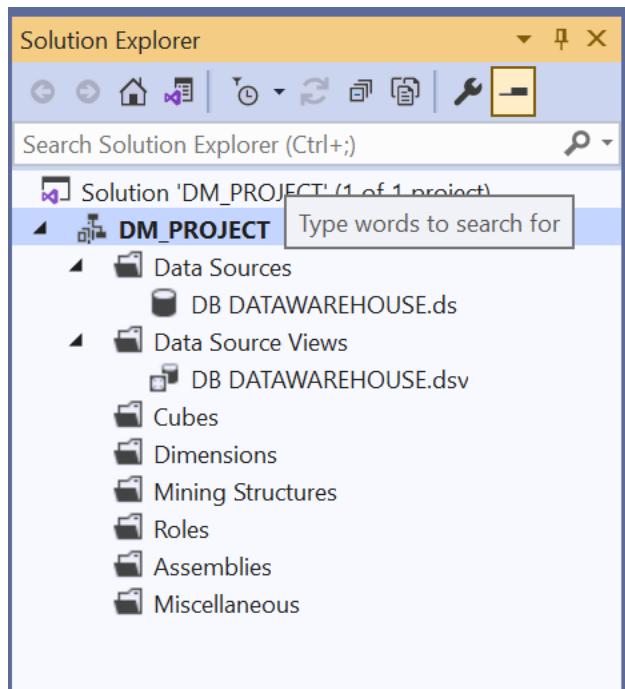
*Hình 5.18. Các bảng dữ liệu cho quá trình phân tích*

- Bước 6: Nhấn Next:



*Hình 5.19. Các bảng dữ liệu được tạo*

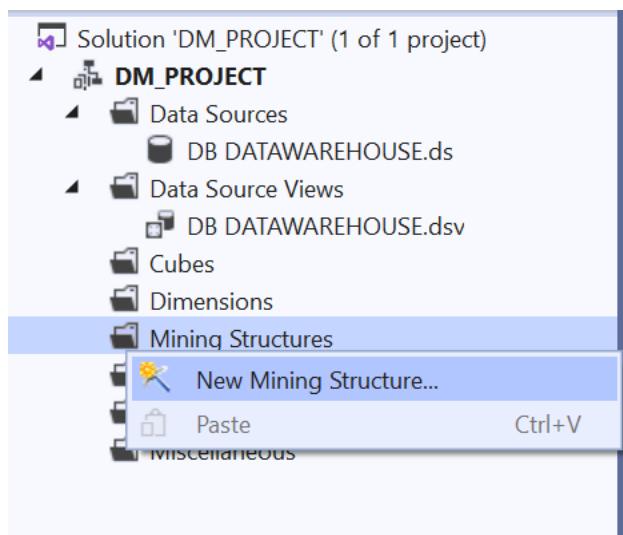
- Bước 7: Nhấn Finish để kết thúc quá trình tạo Data Source View và kết quả sẽ được hiển thị như hình:



*Hình 5.20. Hoàn thành quá trình tạo Data Source View*

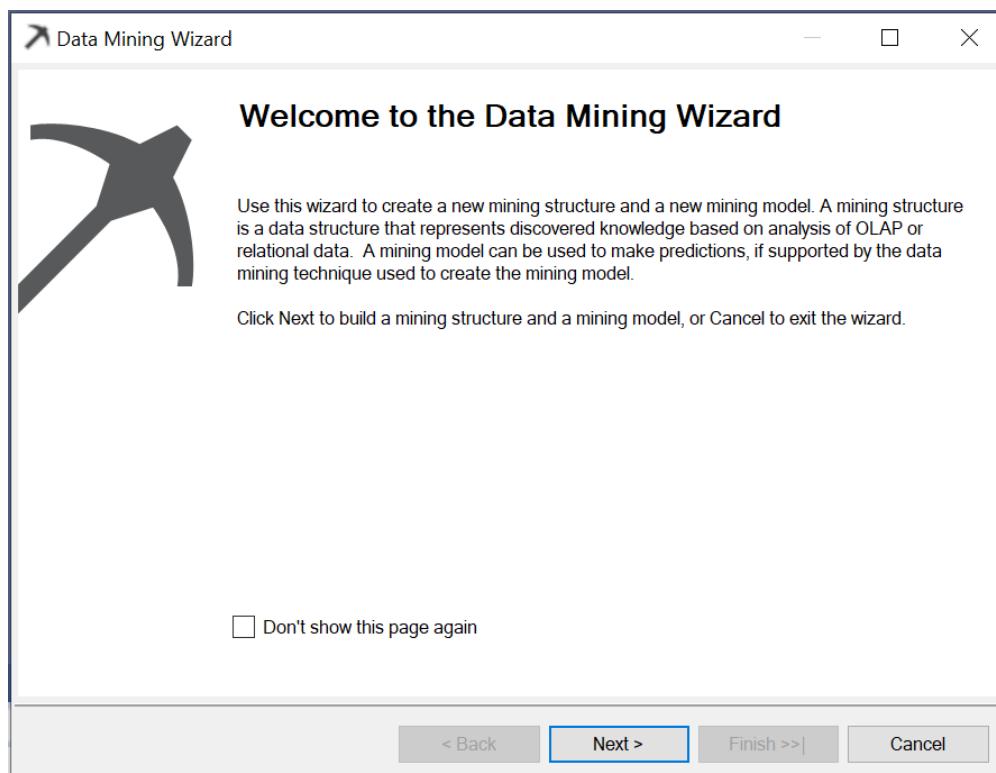
#### 4. Thiết lập cấu trúc Data Mining

- Bước 1: Click phải chuột vào Mining Structures và chọn New Mining Structure để mở Data Mining Wizard.



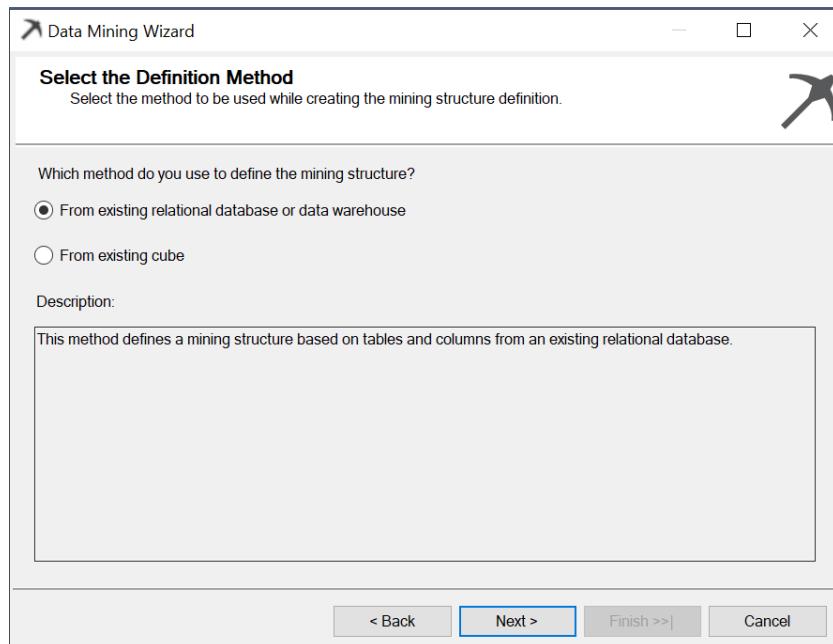
Hình 5.12. Tạo mới một cấu trúc Data Mining

- Bước 2: Trên trang Welcome to the Data Mining Wizard chọn Next.



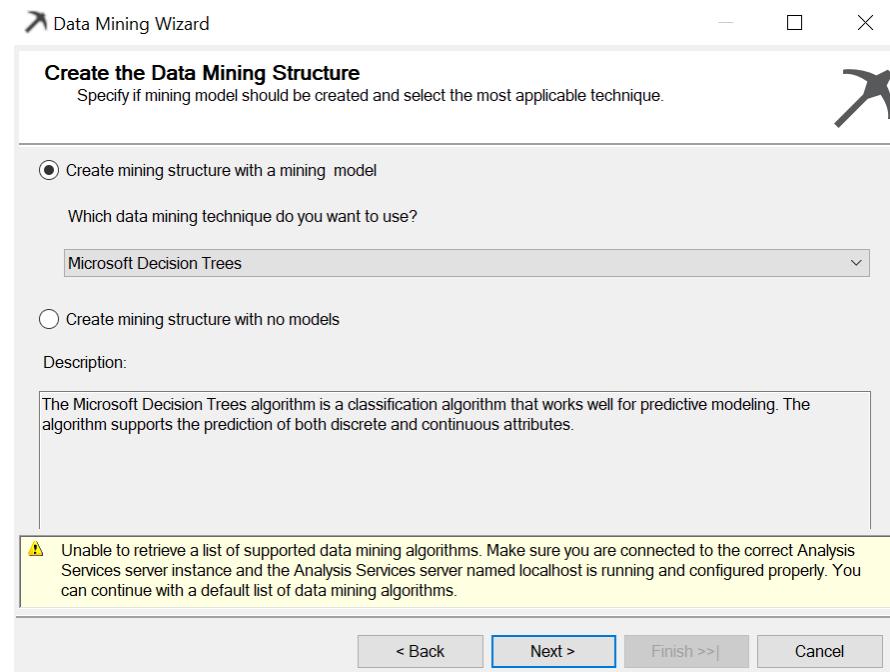
Hình 5.13. Phản giói thiệu Data Mining Wizard

- Bước 3: Trên trang Select the Definition Method, chọn From existing relational database or data warehouse → Next.



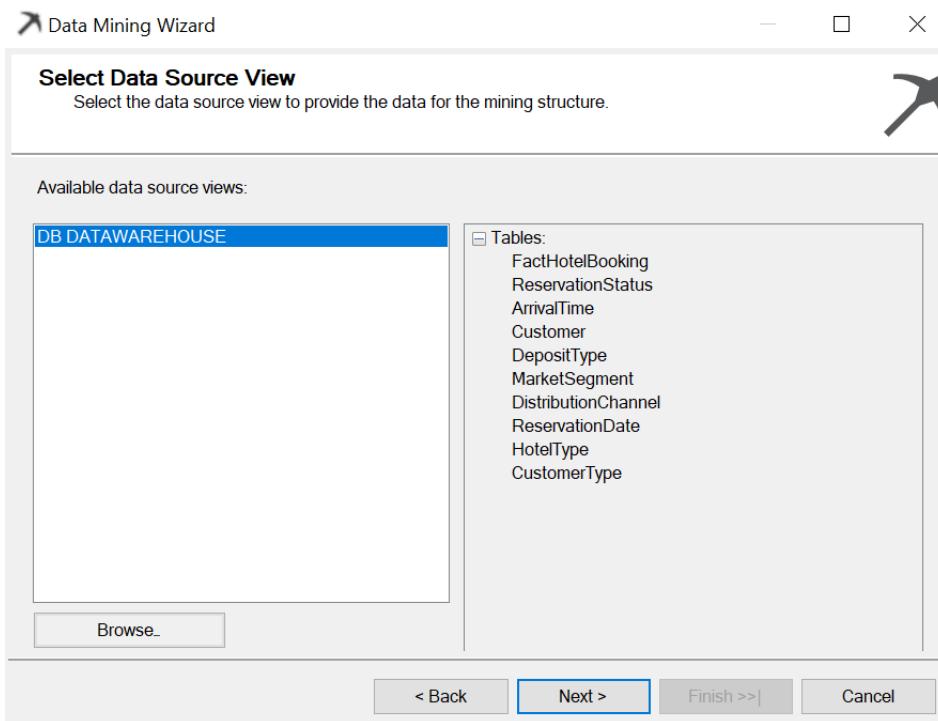
*Hình 5.14. Màn hình chọn phương thức để xác định cấu trúc Data Mining*

- Bước 4: Trên trang Create the Data Mining Structure, dưới Which data mining technique do you want to use? chọn Microsoft Decision Trees → Next.



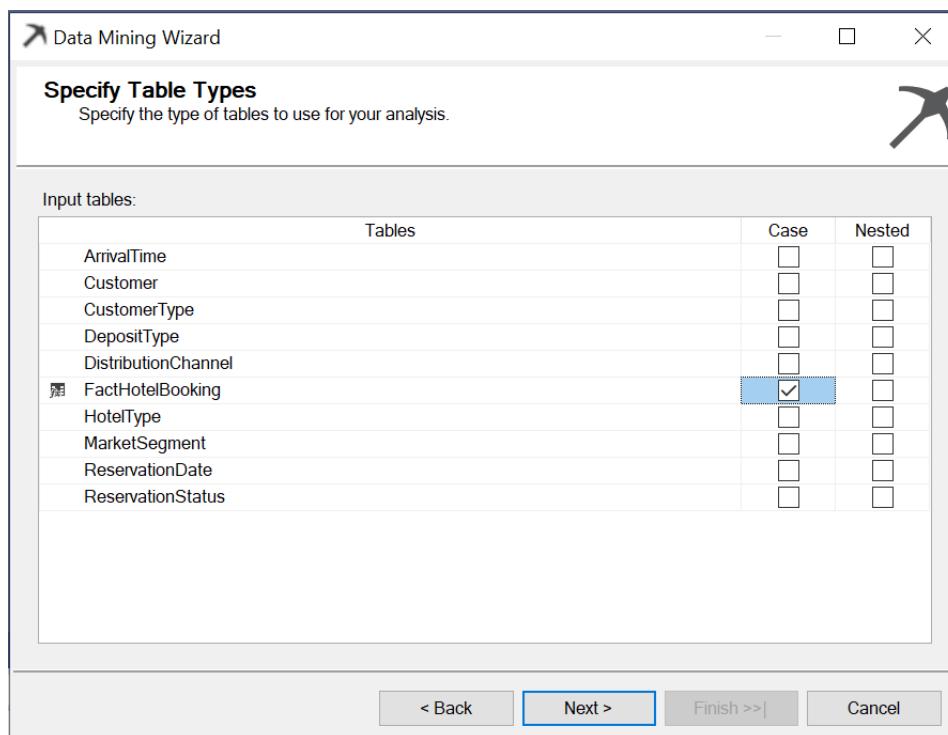
*Hình 5.15. Màn hình chọn thuật toán để sử dụng Data Mining.*

- Bước 5: Trên trang Select Data Source View, chọn Next.



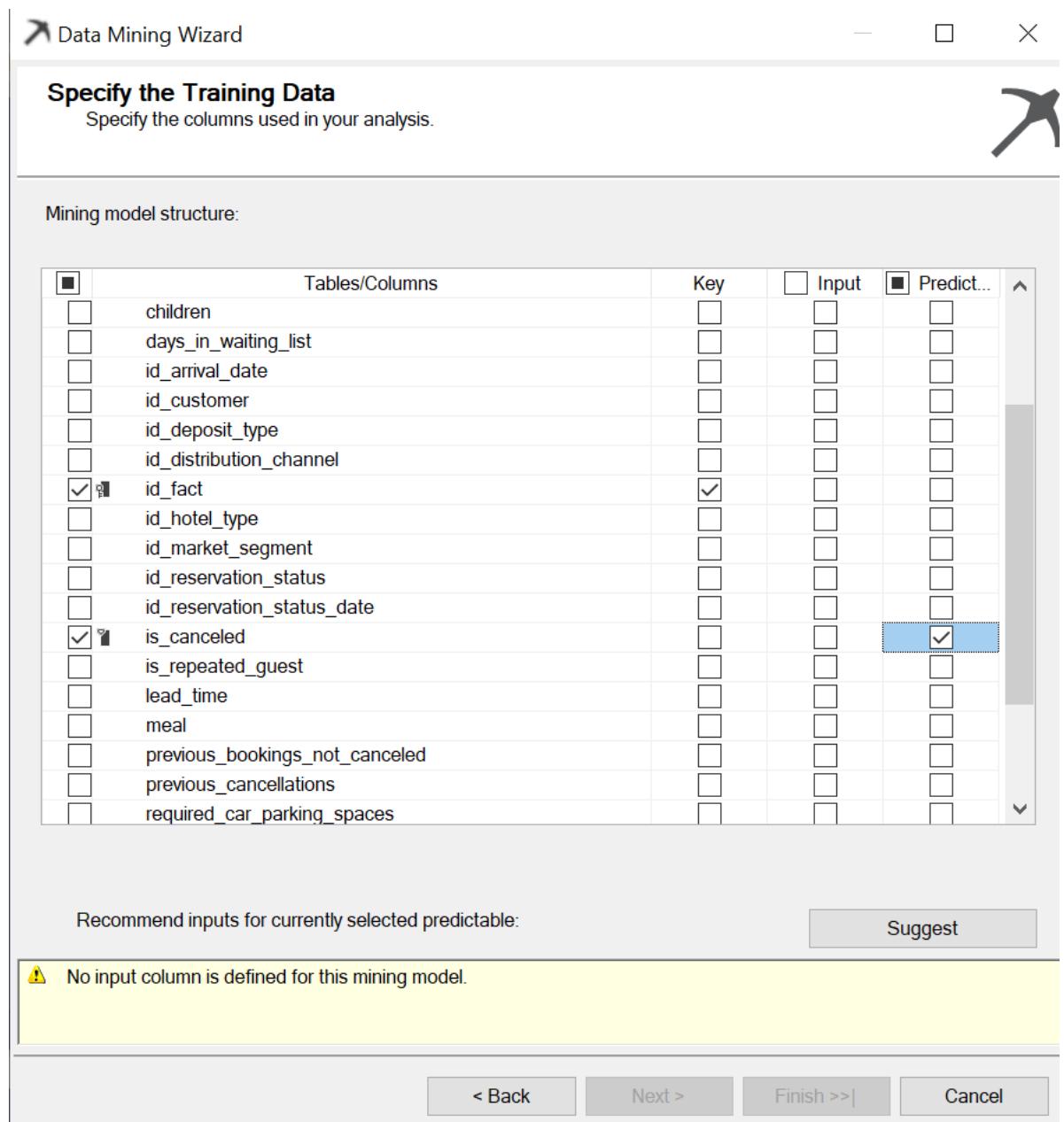
*Hình 5.19. Chọn nguồn dữ liệu Data Mining*

- Bước 6: Trên trang Specify Table Types, chọn bảng Fact → Next.



*Hình 5.20. Chọn bảng dữ liệu Data Mining*

- Bước 7: Trên trang Specify the Training Data, cột key để mặc định (mặc định sẽ là mã cột mã báo cáo), chọn is\_canceled làm trường predictable để dự đoán khả năng hủy nhận phòng của các giao dịch đặt phòng, chọn Suggest.



Hình 5.21: Quá trình chọn Key, Input, Predict

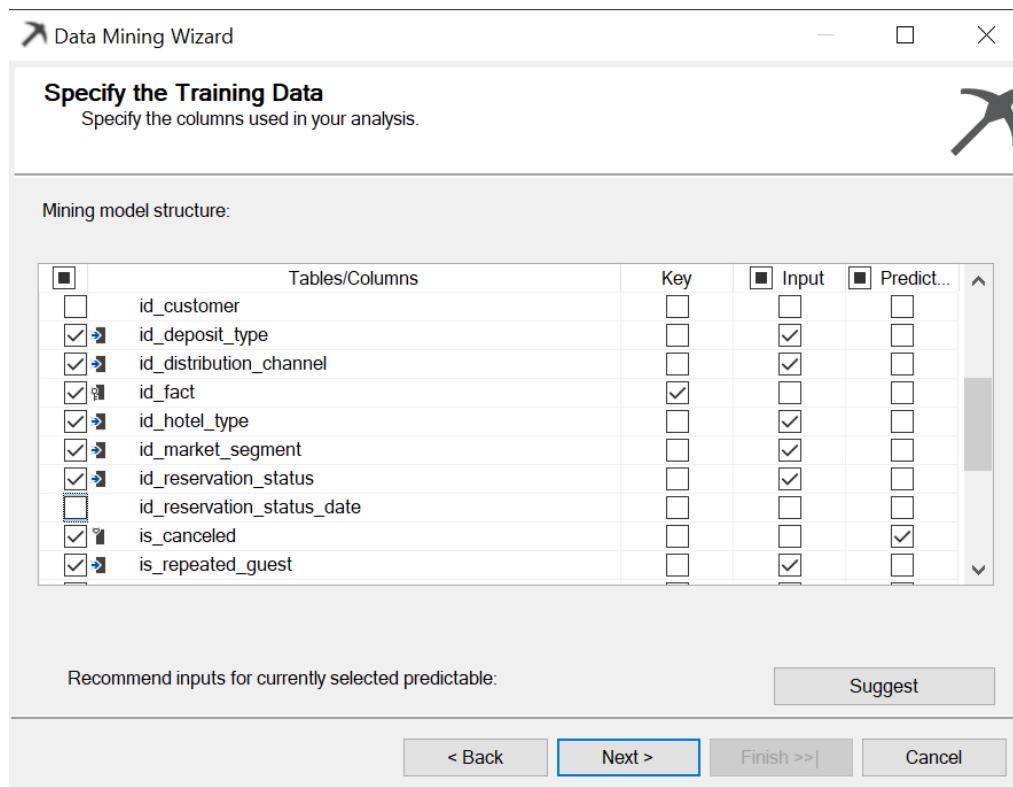
- Bước 8: Suggest đưa ra các input gợi ý dựa trên predictable đã chọn, ở đây ta không muốn dùng các input được gợi ý sẵn thì chọn cancel.

Column Name	Score	Input
id_reservation_status	1.000	x
id_deposit_type	0.218	x
previous_cancellations	0.059	x
total_of_special_requests	0.059	x
id_market_segment	0.055	x
required_car_parking_spaces	0.044	
id_distribution_channel	0.035	
assigned_room_type	0.035	
stays_in_week_nights	0.029	
days_in_waiting_list	0.026	
booking_changes	0.024	
previous_bookings_not_canc...	0.018	
id_hotel_type	0.012	
stays_in_weekend_nights	0.009	
is_repeated_guest	0.008	
reserved_room_type	0.008	
adults	0.003	
children	0.003	
babies	0.002	
meal	0.001	
id_customer		

OK Cancel

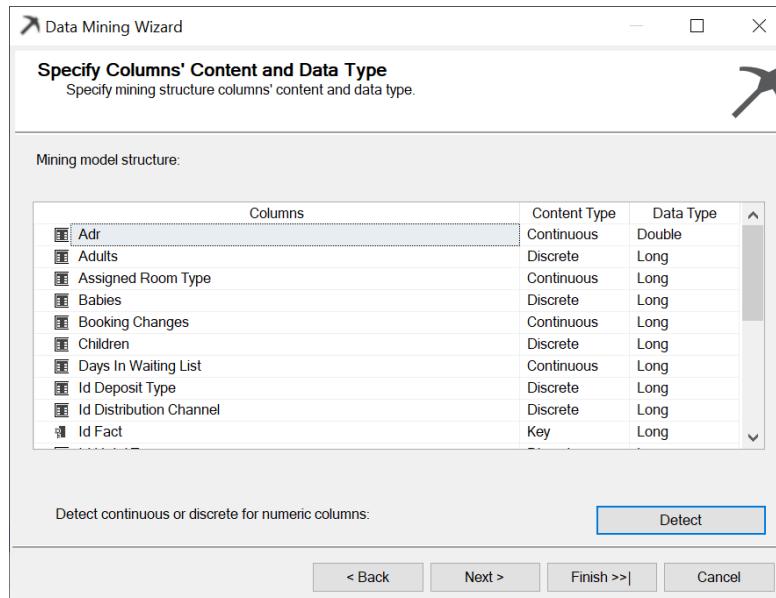
Hình 5.22: Màn hình xem mối tương quan giữa giá trị Predict và các giá trị Input.

- Bước 9 : Chọn các cột input lần lượt là adr, adults, babies, children, days in waiting list, id deposit type, id distribution channel, id hotel type, id market segment, id reservation status, is repeated guest, lead time, meal, required car parking spaces, lead time, meal, required car parking spaces, stays in week nights, status in weekend nights, total of special requests.



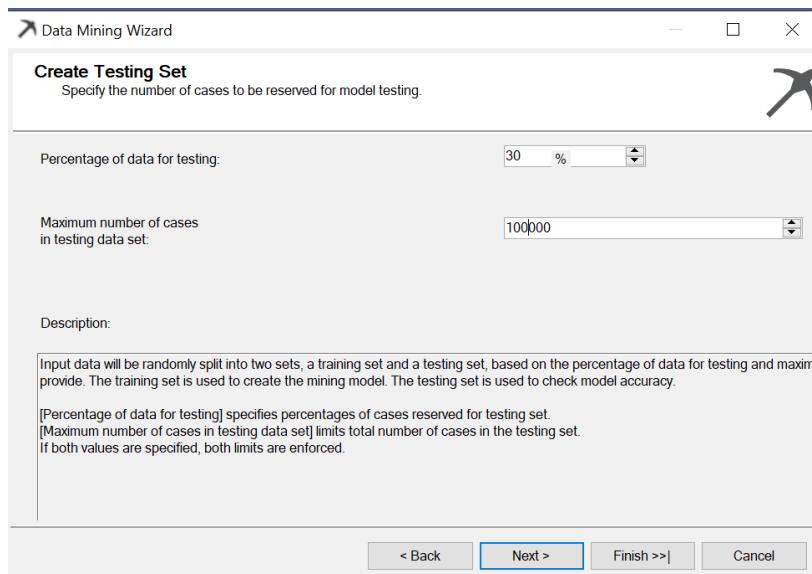
*Hình 5.23: Chọn các Input.*

- Bước 10: Trên trang Specify Columns' Content and Data Type, điều chỉnh lại kiểu dữ liệu và kiểu nội dung, click Detect để xác định hai trường trên. Sau đó chọn Next.



*Hình 5.24: Chọn content Type và data Type cho từng column.*

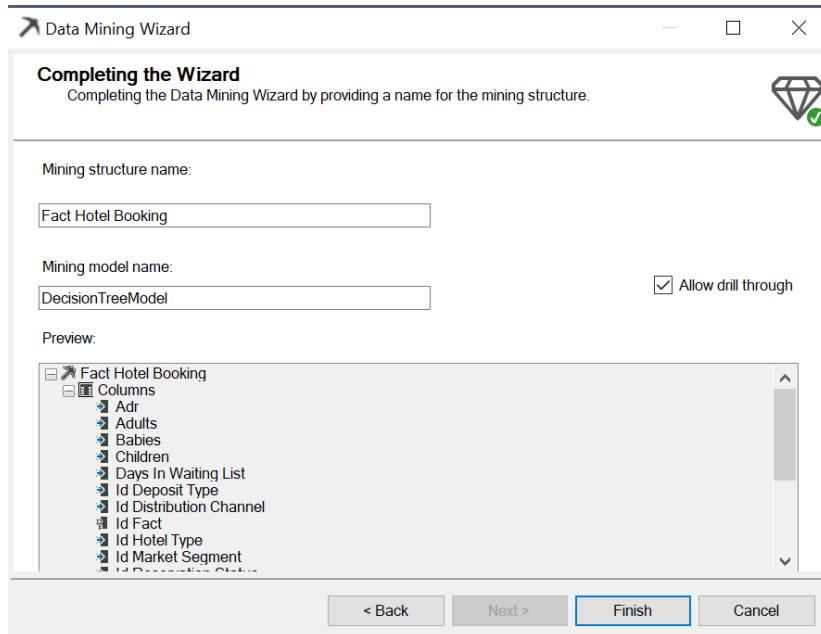
- Bước 11: Trên trang Create Testing Set, Percentage of data for testing mặc định là 30%, Maximum number of cases in testing data set để 1000. Tiếp tục chọn Next.



*Hình 5.25: Thiết lập 70% training và 30% testing và 100 000 trường hợp test trên kho dữ liệu.*

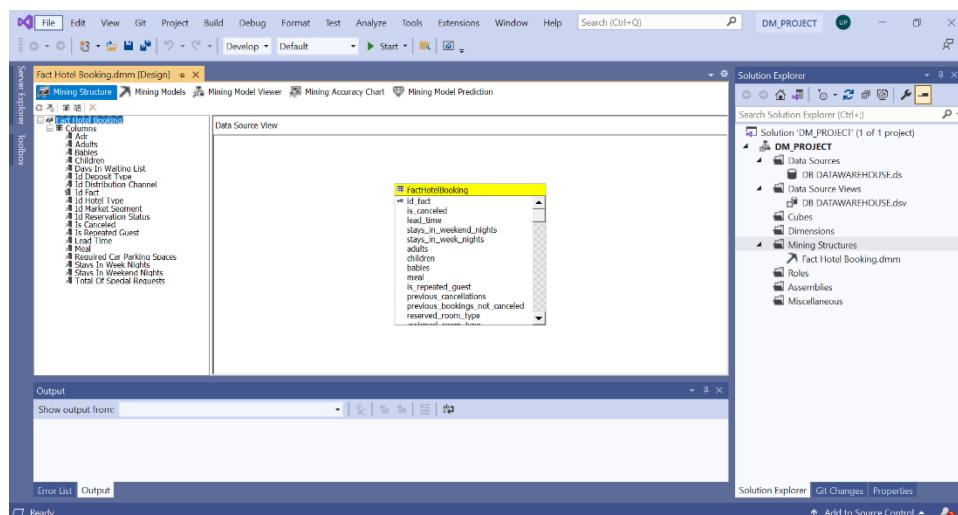
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Bước 12: Trên trang Completing the Wizard, có thể đổi tên model, tick chọn vào ô Allow drill through. Chọn Finish để kết thúc quá trình thiết lập.



Hình 5.26: Đặt tên cho cấu trúc Data Mining.

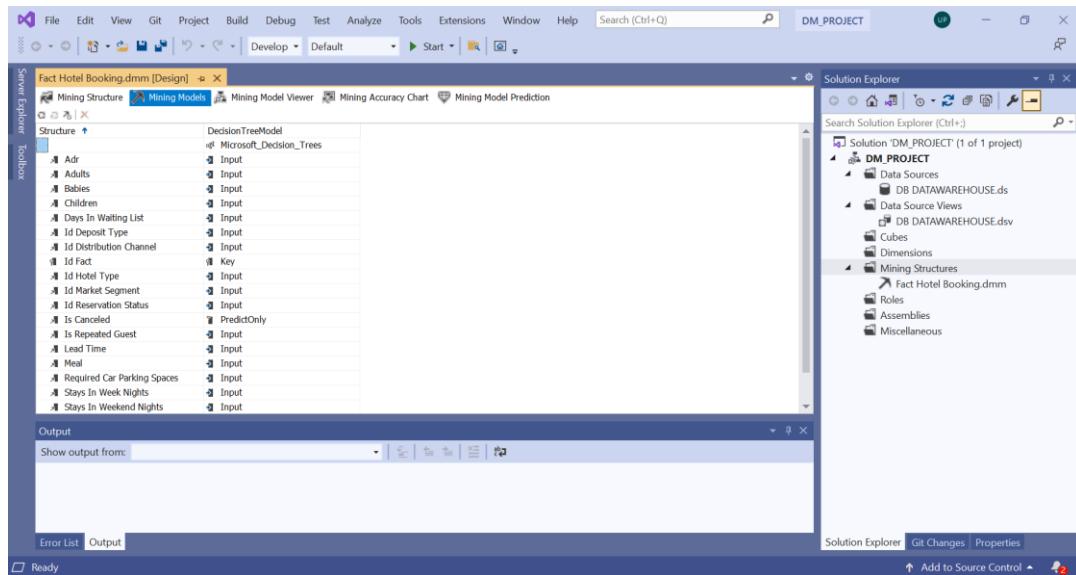
- Bước 13: Kết quả của việc thiết lập cấu trúc cho Data Mining.



Hình 5.27: Màn hình kết thúc quá trình tạo cấu trúc Data Mining.

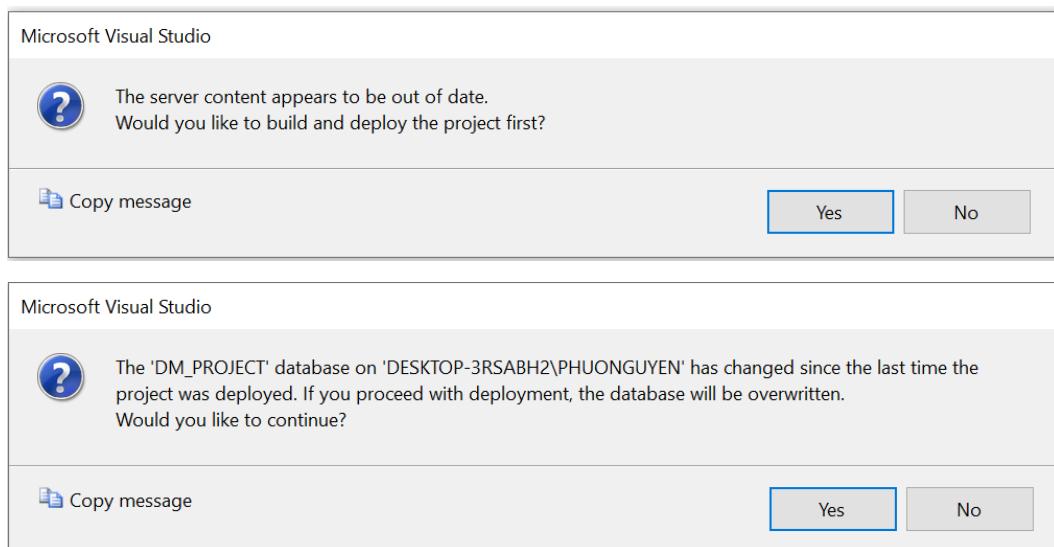
## 5. Deploy project Data Mining

- Bước 1: Chuyển sang tab Mining Models để xem lại thiết lập cấu trúc data mining.



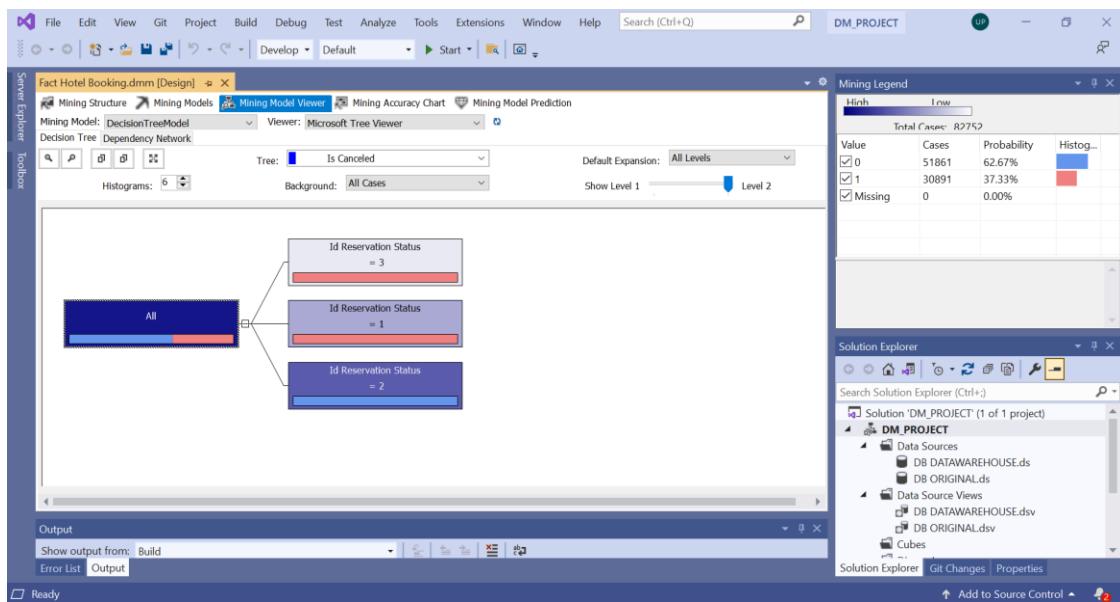
Hình 5.28: Thiết lập cấu trúc Data Mining tại Mining Model

- Bước 2: Deploy project → Yes để tiếp tục.



Hình 5.29: Deploy Project Data Mining

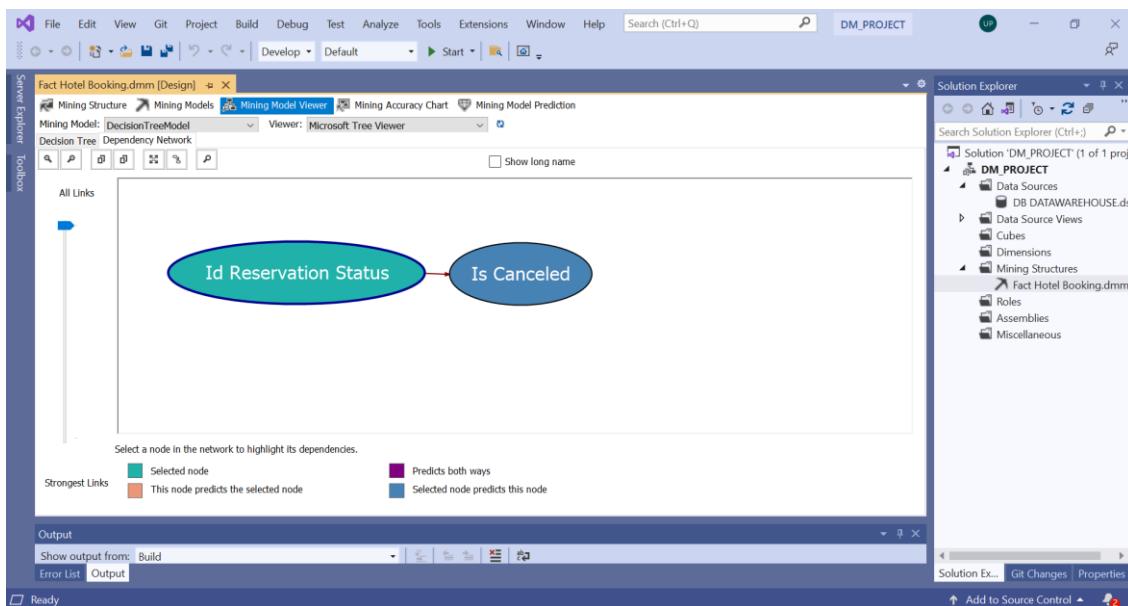
- Bước 3: Kết quả có được mô hình cây quyết định (DecisionTreeModel).



*Hình 5.30 Cấu trúc cây quyết định đã được tạo*

## 6. Phân tích mining bằng cây quyết định và đưa ra luật

- Bước 1: Sau khi cài đặt thành công, chuyển qua tab Dependency Network để xem các cấp phụ thuộc:



*Hình 5.31: Mức phụ thuộc cấp 2.*

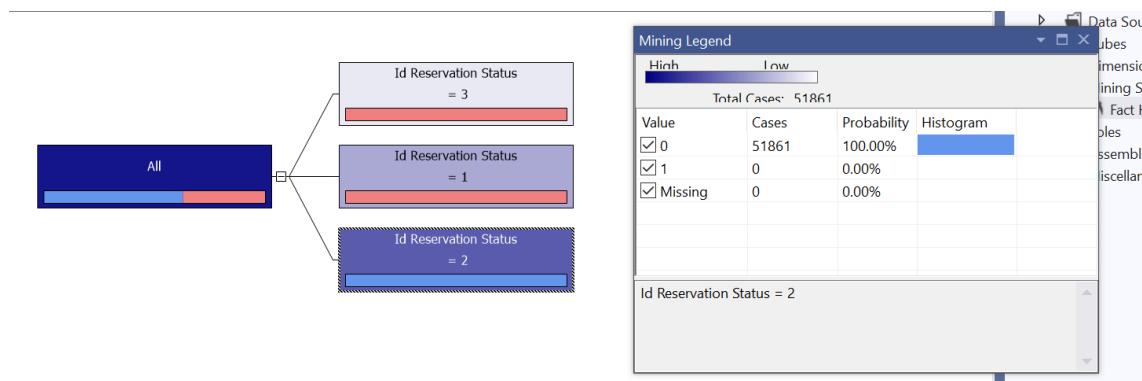
- Bước 2: Ở mức 2, ta dự đoán về trạng thái lần cuối cùng khi thực hiện giao dịch với khách sạn thông qua thuộc tính Id\_Reservation\_Status.

  - Trạng thái là Canceled.



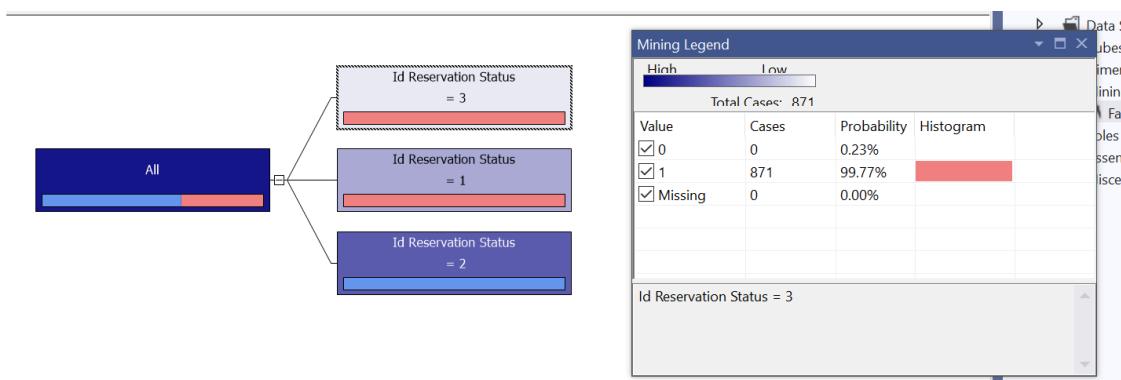
*Hình 5.32: Kết quả với trạng thái cuối cùng là hủy đặt phòng.*

  - Trạng thái là Check-out.



*Hình 5.33: Kết quả với trạng thái cuối cùng là nhận phòng.*

  - Trạng thái là No-show.

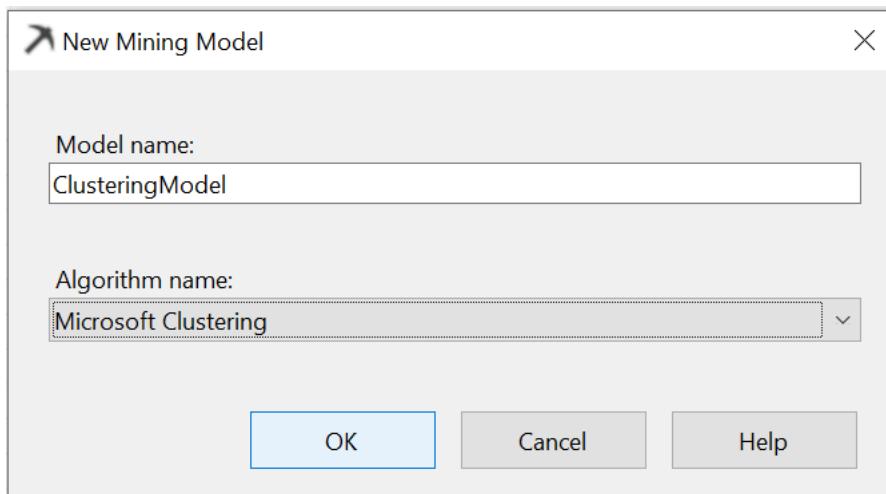


*Hình 5.34: Kết quả với trạng thái cuối cùng là không hiển thị.*

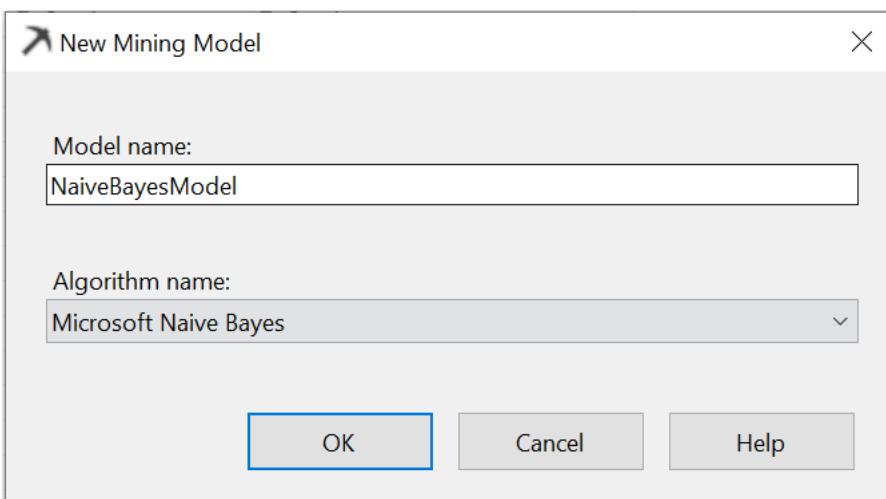
- Rút luật: Các đơn đặt phòng tại khách sạn có trạng thái cuối cùng là hủy đặt phòng hoặc chưa có thông tin thì khả năng hủy đặt phòng cao.  
⇒ Điều này là hiển nhiên do Id\_Reservation\_Status có mối tương quan rất mạnh ( $= 1$ ) với Is\_Canceled nên ta xóa thuộc tính Id\_Reservation\_Status và tiến hành kiểm xây dựng lại model.

## 7. Xây dựng thêm Model

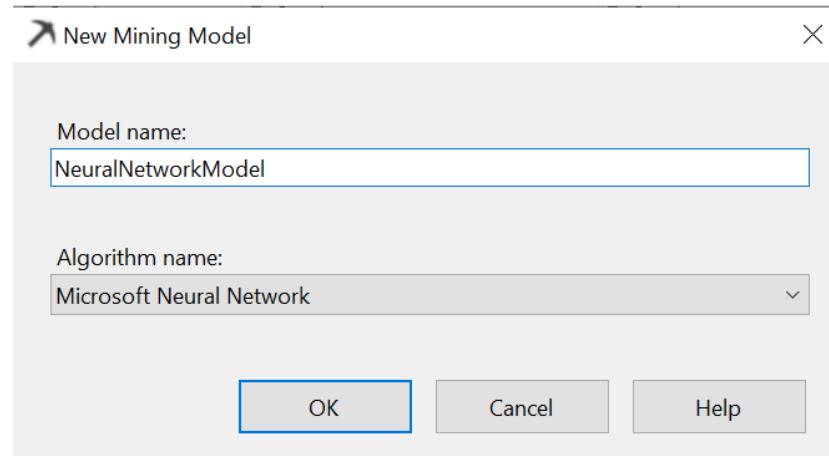
- Bước 1: Click chuột phải vào ô trống chọn New Mining Model



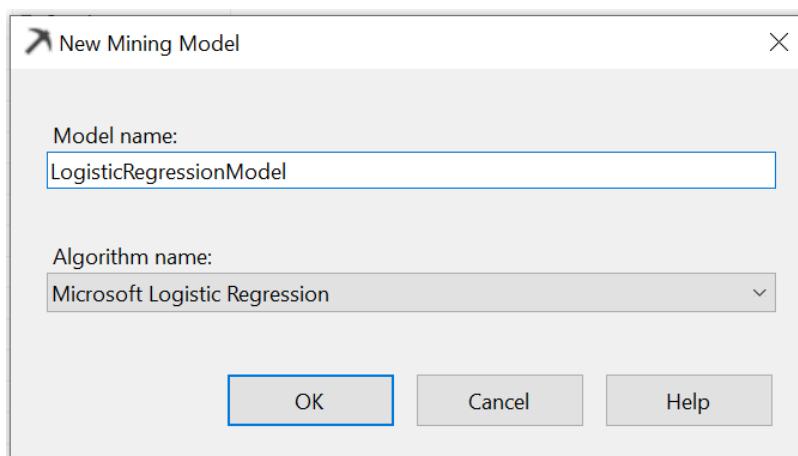
*Hình 5.35: Xây dựng model Clustering*



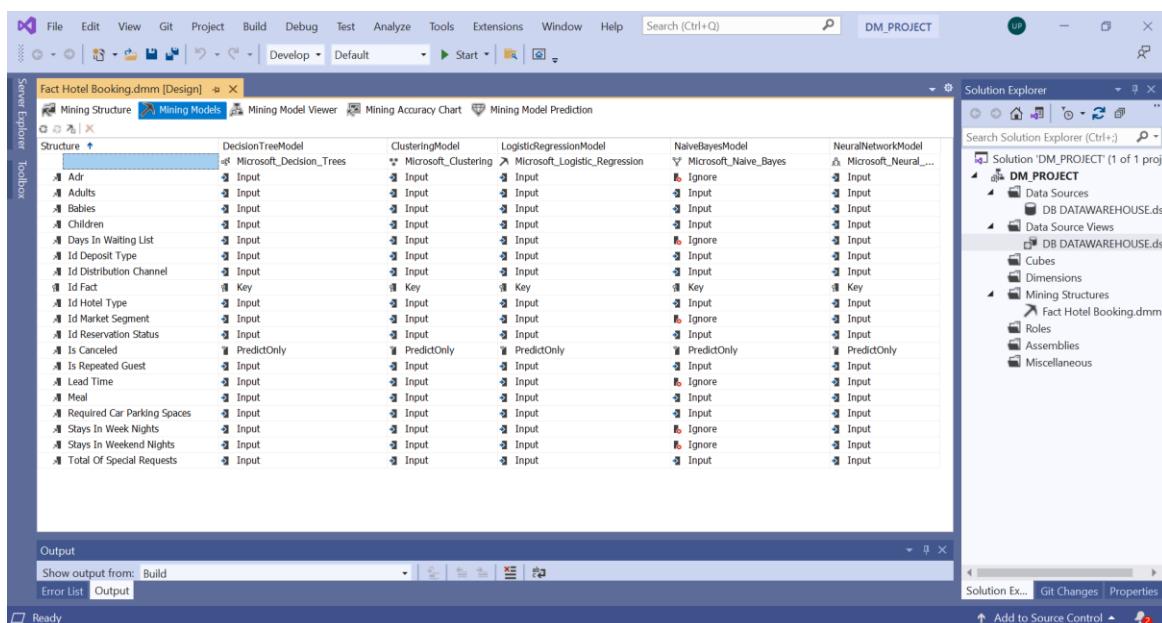
*Hình 5.36: Xây dựng model Naïve Bayes*



Hình 5.37: Xây dựng model Neural Network

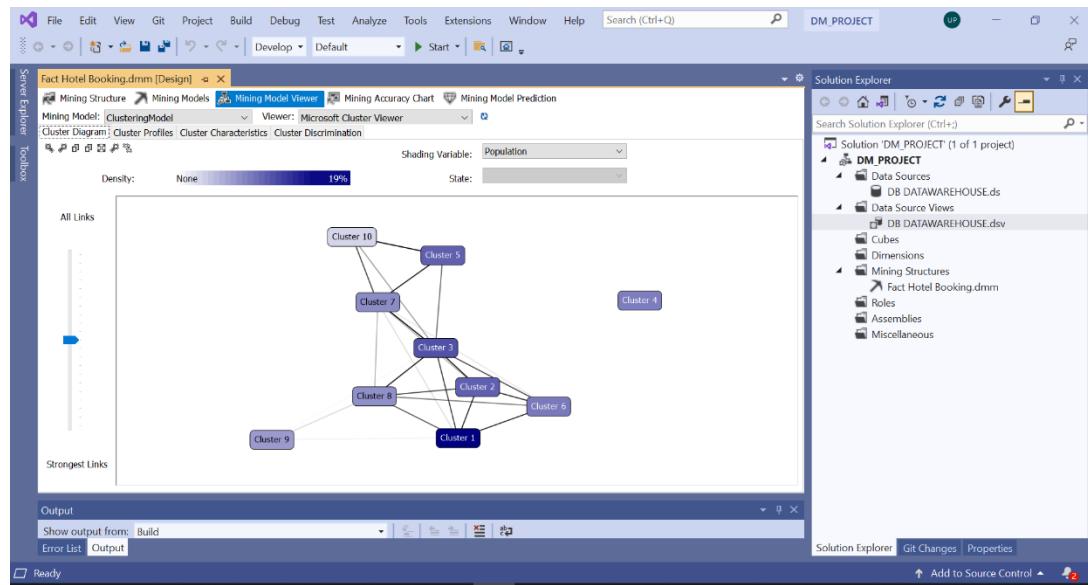


Hình 5.38: Xây dựng model Logistic Regression



Hình 5.39: Thiết lập cấu trúc Data Mining tại Mining Model

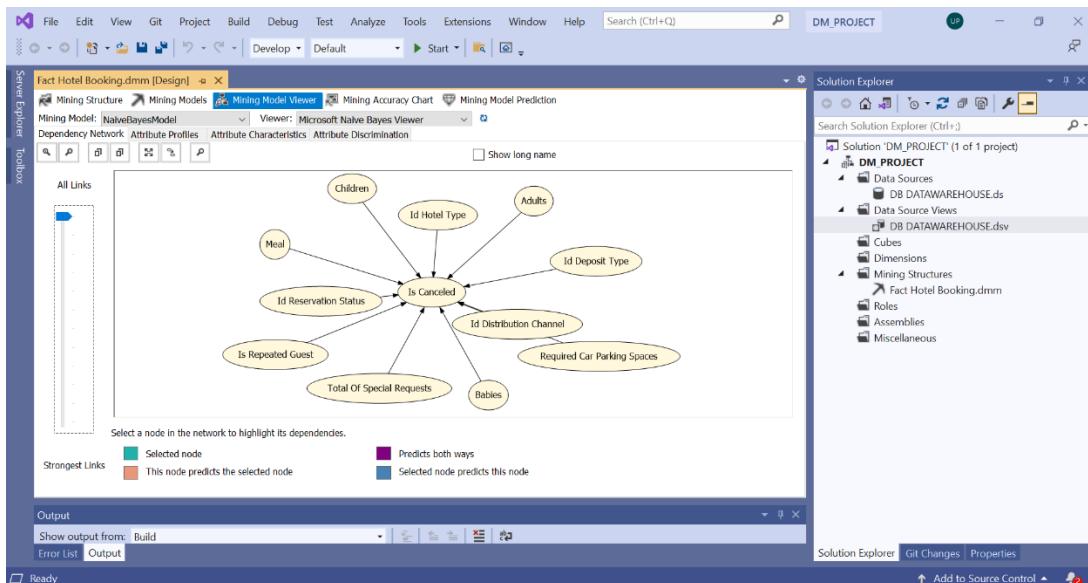
- Bước 2: Chọn thuật toán Microsoft Clustering



Hình 5.40: Chọn thuật toán Microsoft Clustering

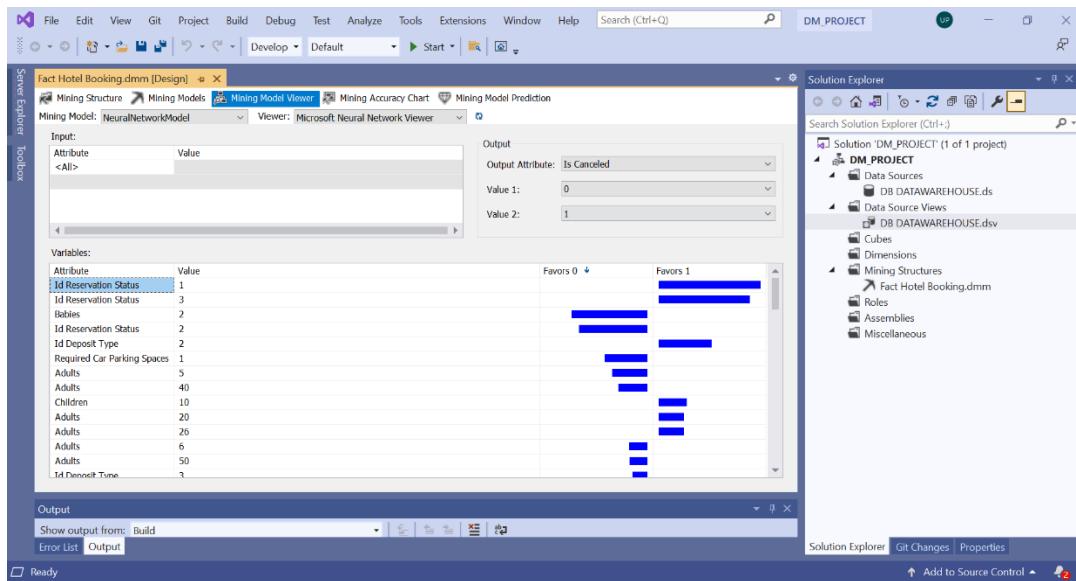
- Bước 3: Tiếp tục tạo thêm thuật toán Mining Microsoft Naive Bayes

Thuật toán này không hỗ trợ mining các cột có kiểu continuous



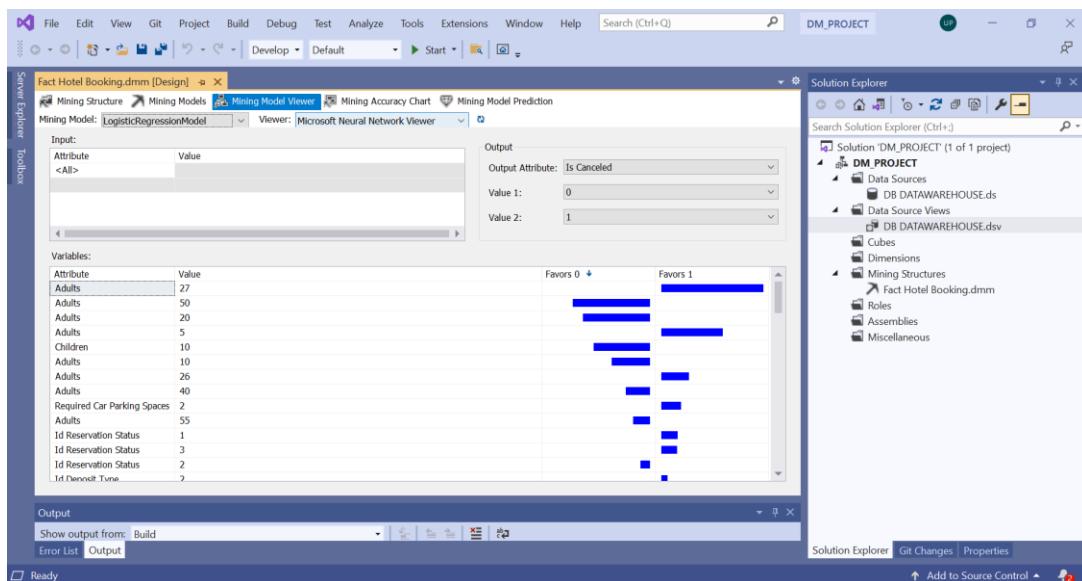
Hình 5.41: Chọn thuật toán Microsoft Naive Bayes

- Bước 4: Chọn thuật toán Microsoft Neural Network



Hình 5.42 : Chọn thuật toán Microsoft Neural Network

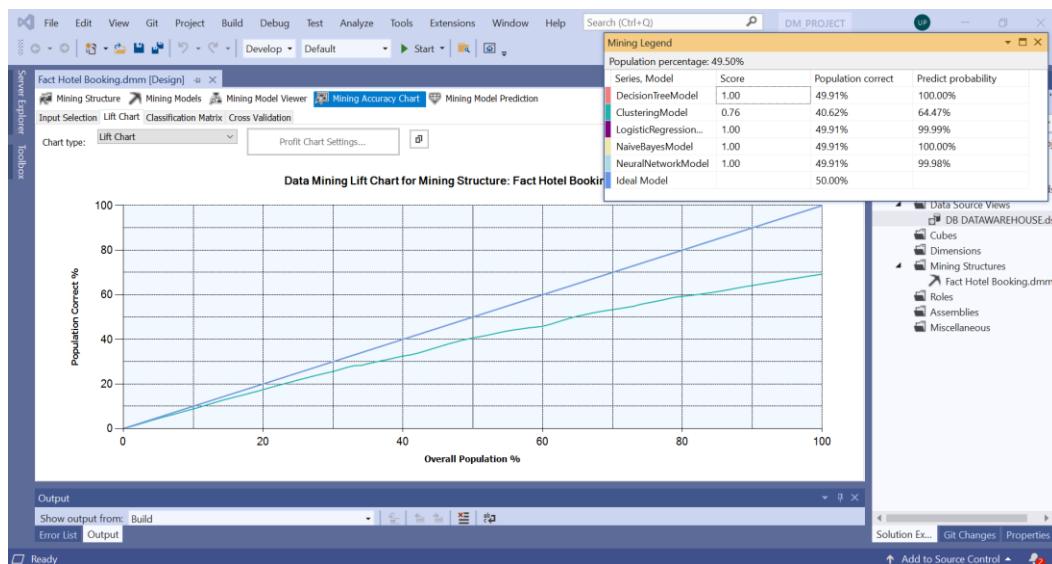
- Bước 5: Chọn thuật toán Microsoft Logistic Regression



Hình 5.43: Chọn thuật toán Microsoft Logistic Regression

## 8. So sánh các thuật toán khác

- Để xem đồ thị Lift chart trong tab Mining Accuracy Chart chọn Lift Chart

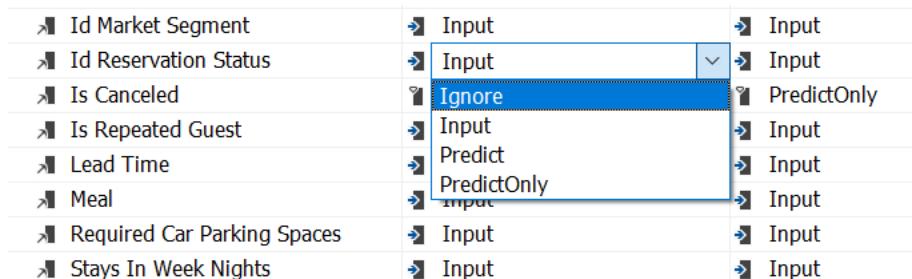


Hình 5.44: Đồ thị Lift chart về so sánh các thuật toán

- Line chart cho ta thấy tỷ lệ chính xác của các thuật toán khi Mining trường hợp dự đoán khả năng hủy đặt phòng của khách hàng
  - Microsoft Decision Tree: 1 Score.
  - Microsoft Clustering: 0.76 Score.
  - Microsoft Naive Bayes: 1 Score.
  - Microsoft Neural Network: 1 Score.
  - Microsoft Logistic Regression: 1 Score.

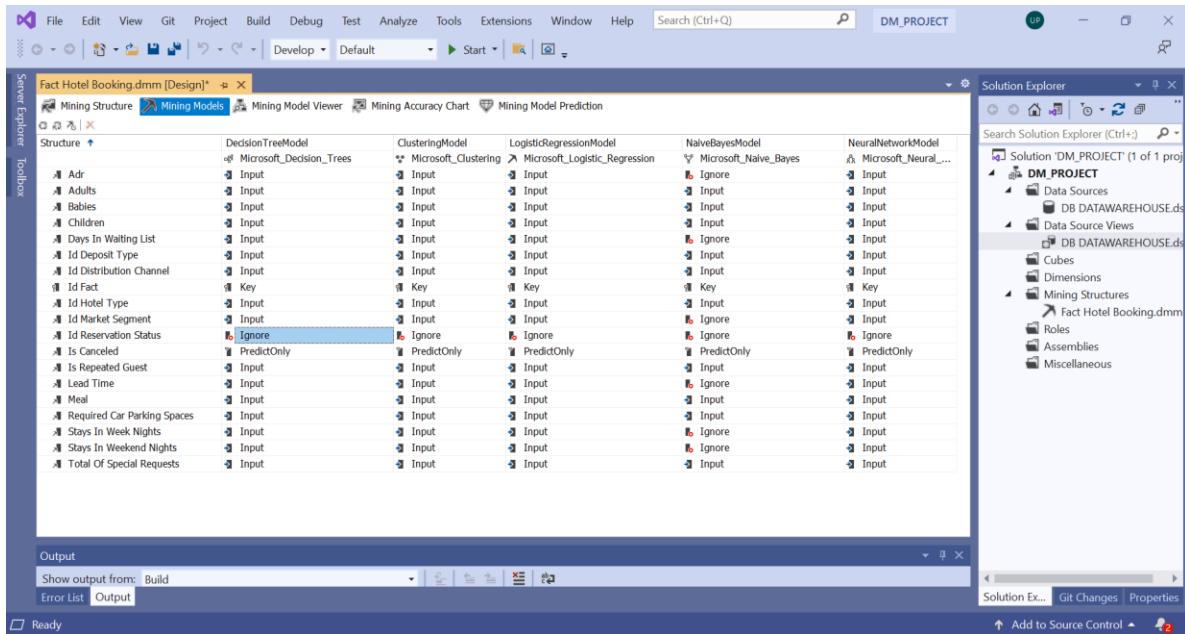
## 9. Thực hiện lại khi xóa thuộc tính Id\_Reservation\_Status

- Bước 1: Loại bỏ thuộc tính Id\_Reservation\_Status tại các Model bằng cách chuyển từ Input sang Ignore.



Hình 5.45: Loại bỏ thuộc tính Id\_Reservation\_Status

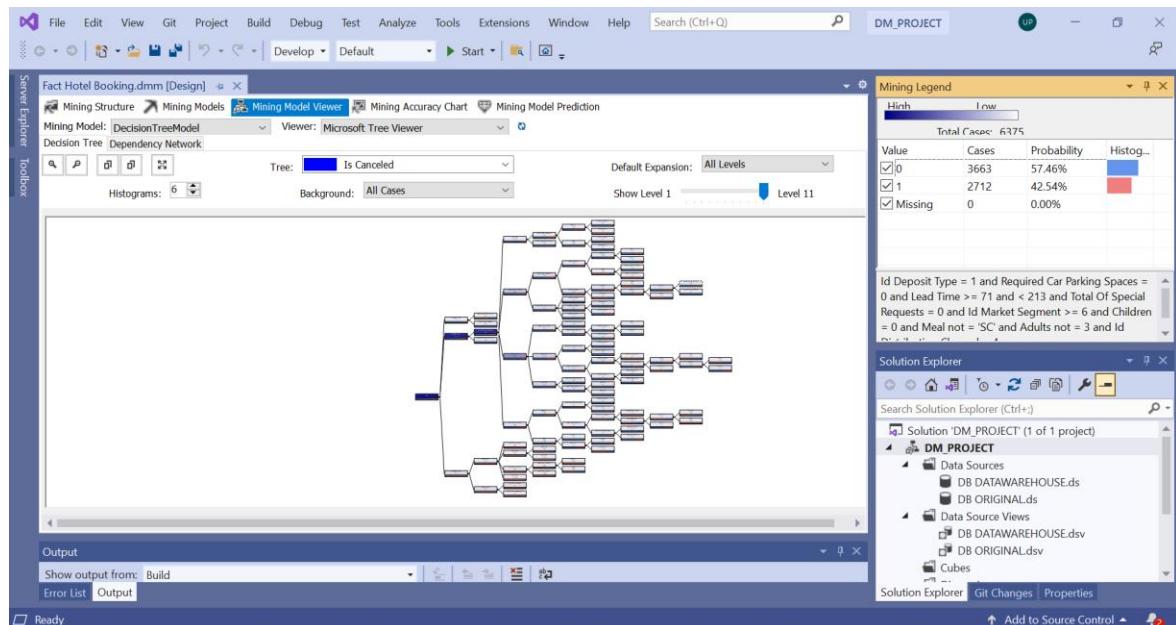
- Bước 2: Lần lượt thực hiện với các Model còn lại ta được:



Hình 5.46: Loại bỏ thuộc tính Id\_Reservation\_Status ở các Model

## 10. Mô hình các thuật toán khi loại bỏ Id\_Reservation\_Status

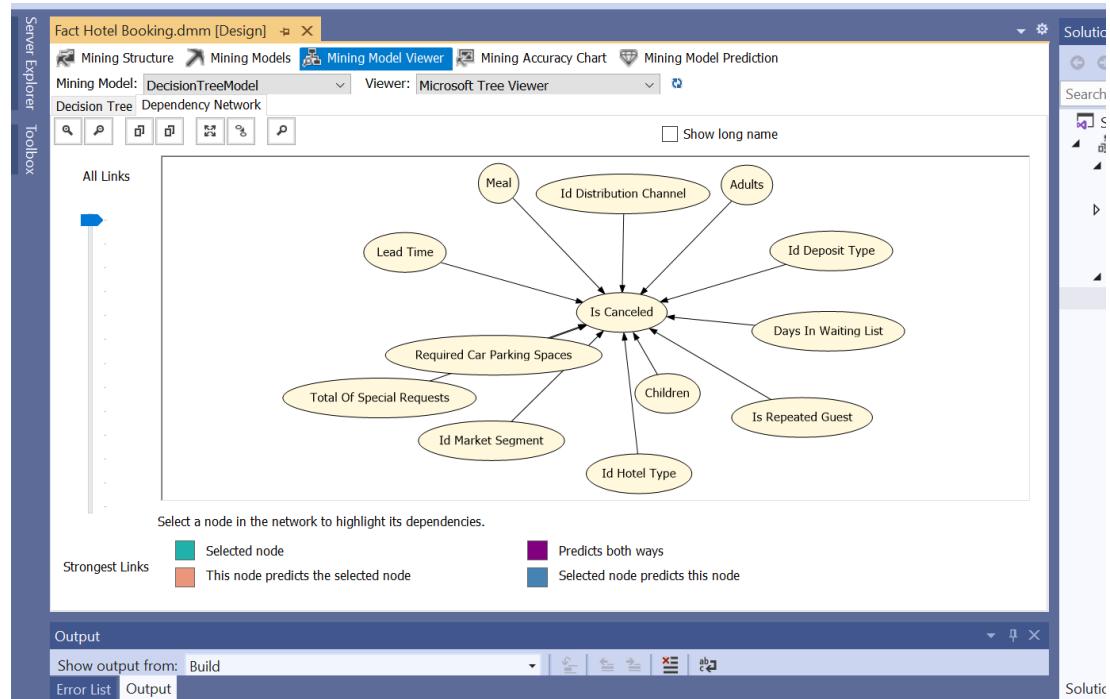
- Bước 1: Chọn thuật toán Microsoft Decision Tree



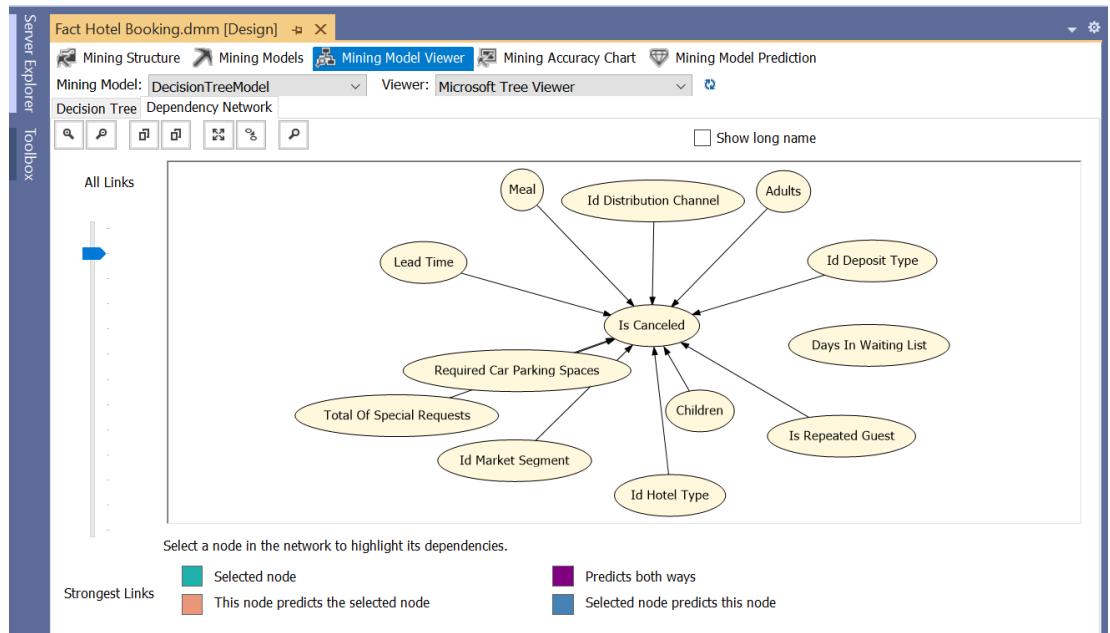
Hình 5.47 : Chọn thuật toán Microsoft Decision Tree

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

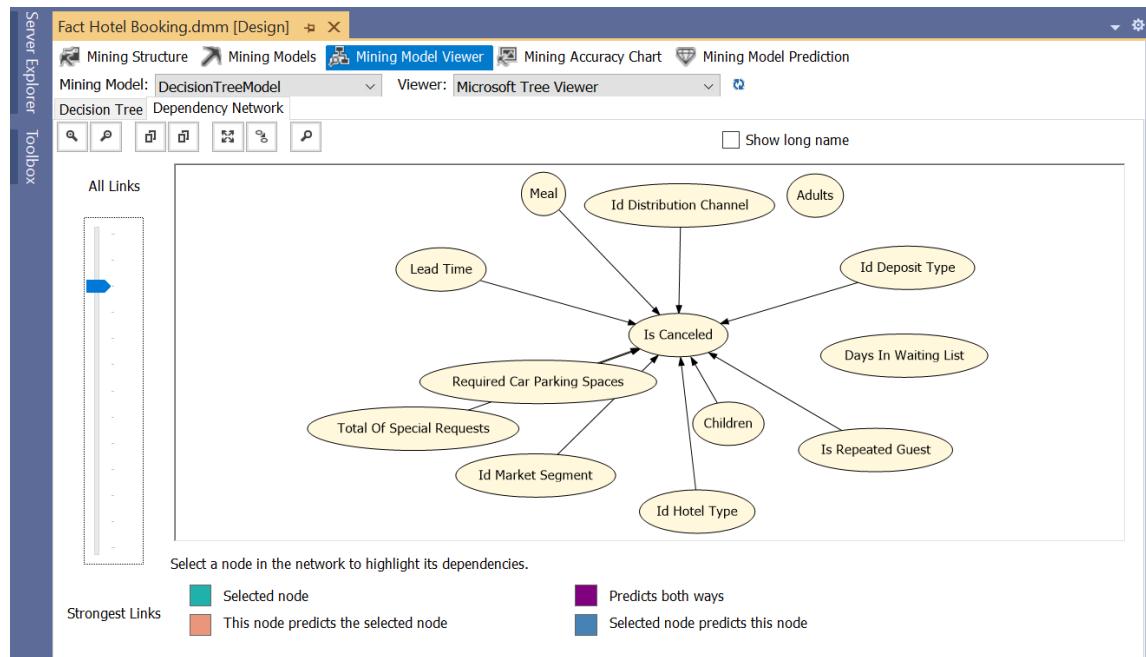
- Sau khi cài đặt thành công, chuyển qua tab Dependency Network để xem các cấp phụ thuộc:



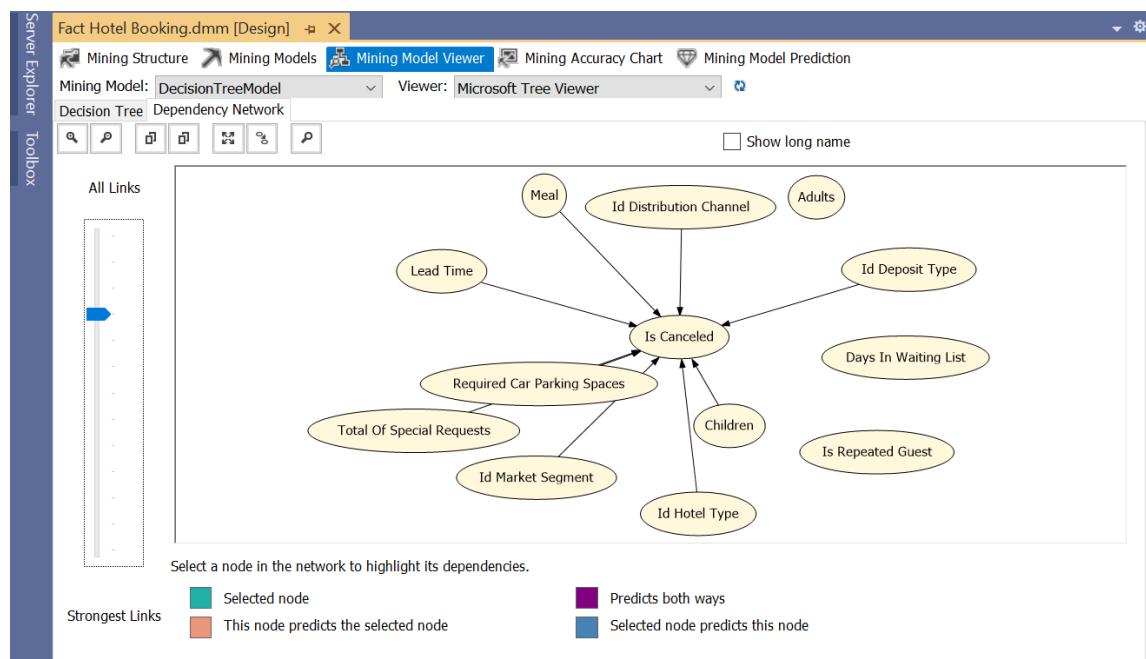
Hình 5.48: Mức phụ thuộc cấp 13.



Hình 5.49: Mức phụ thuộc cấp 12.

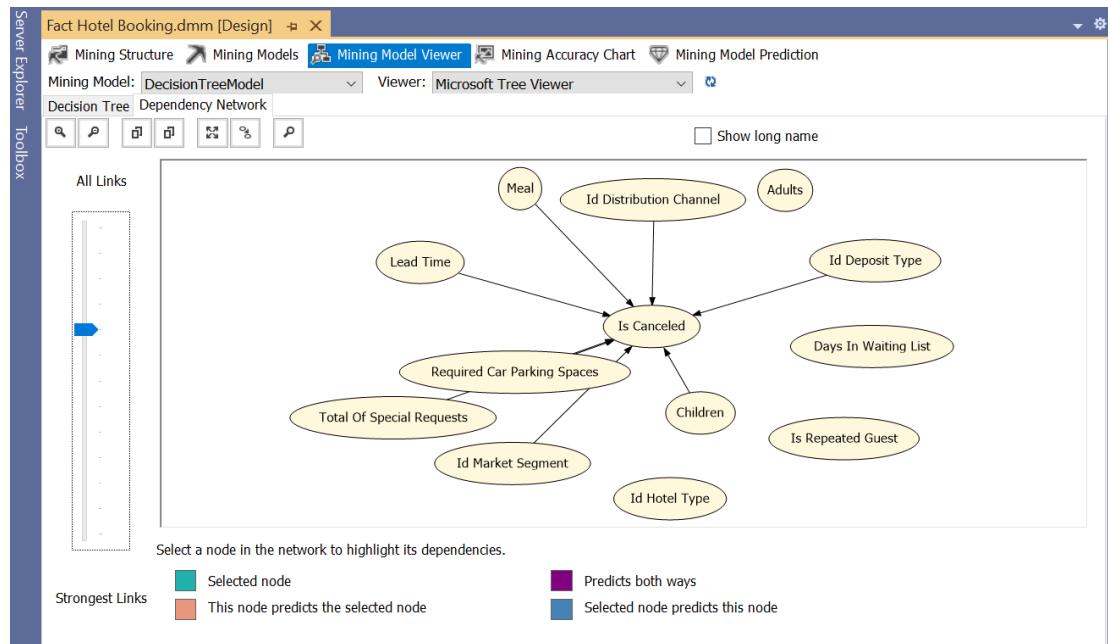


*Hình 5.50: Mírc phụ thuộc cấp 11.*

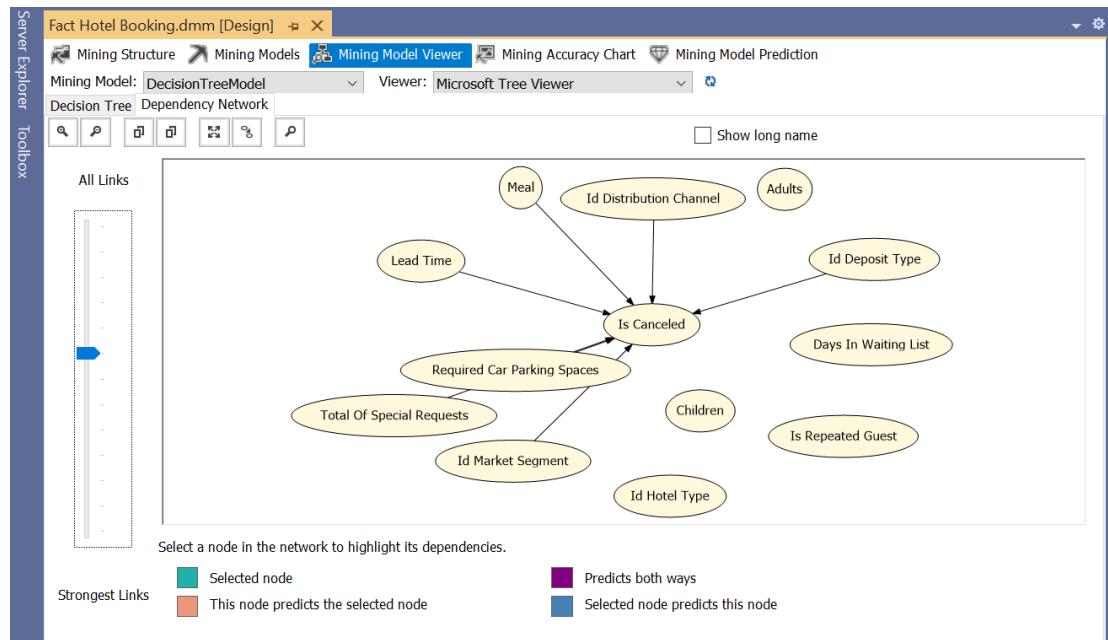


*Hình 5.51: Mírc phụ thuộc cấp 10.*

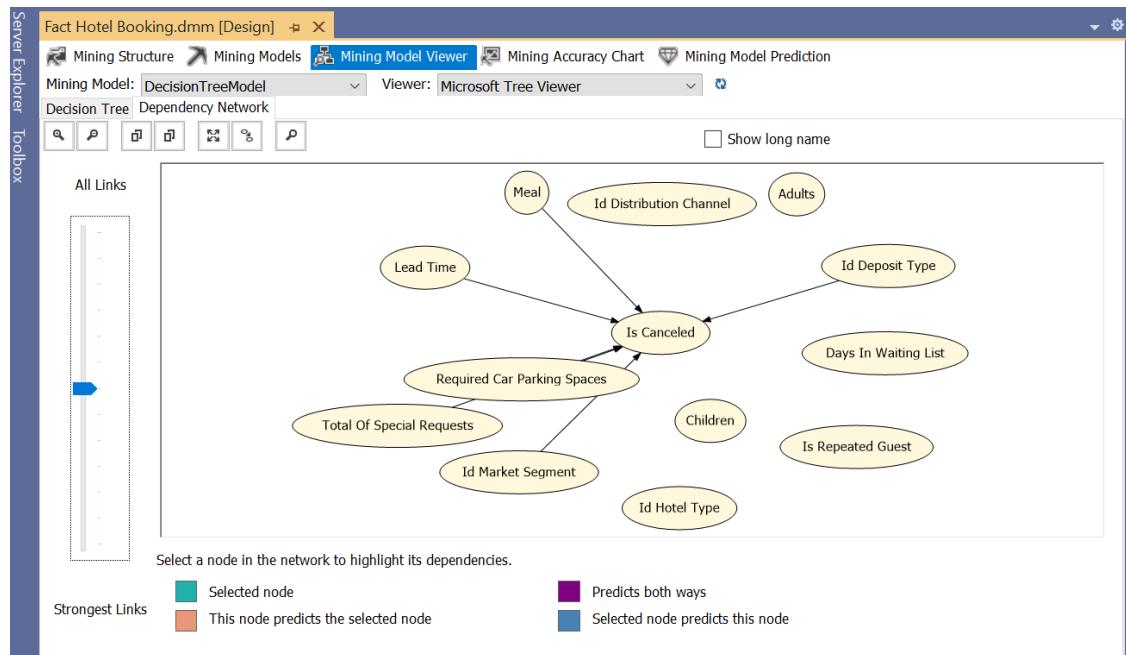
*Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22*



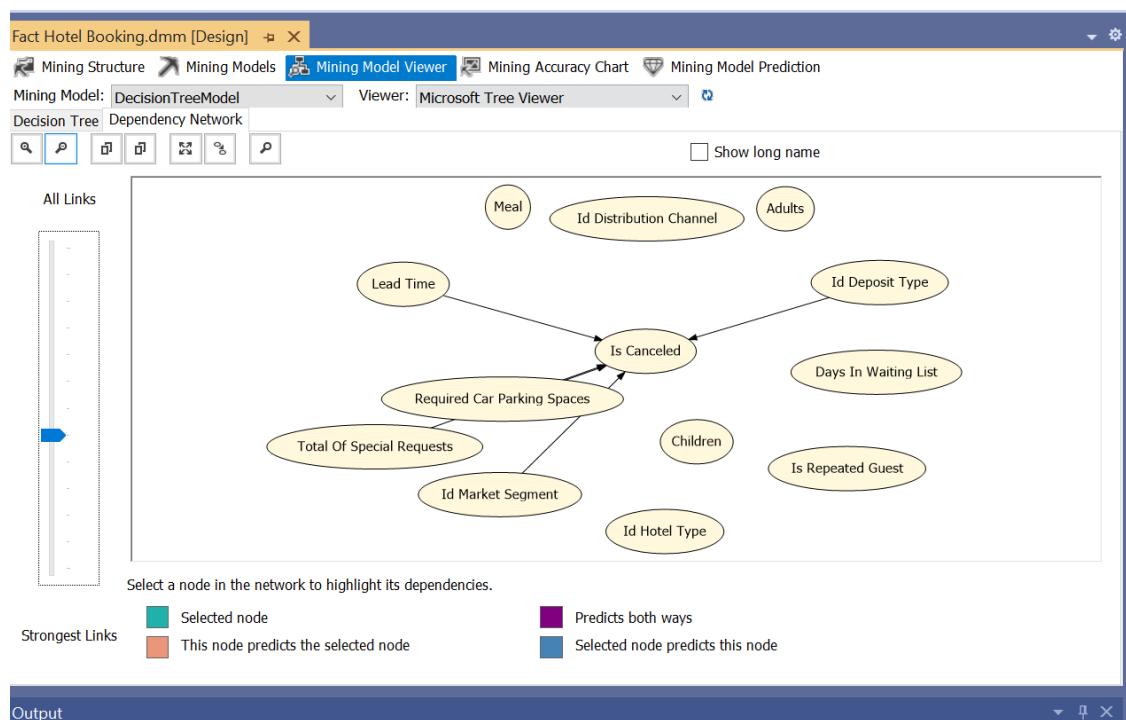
*Hình 5.52: Mírc phụ thuộc cấp 9.*



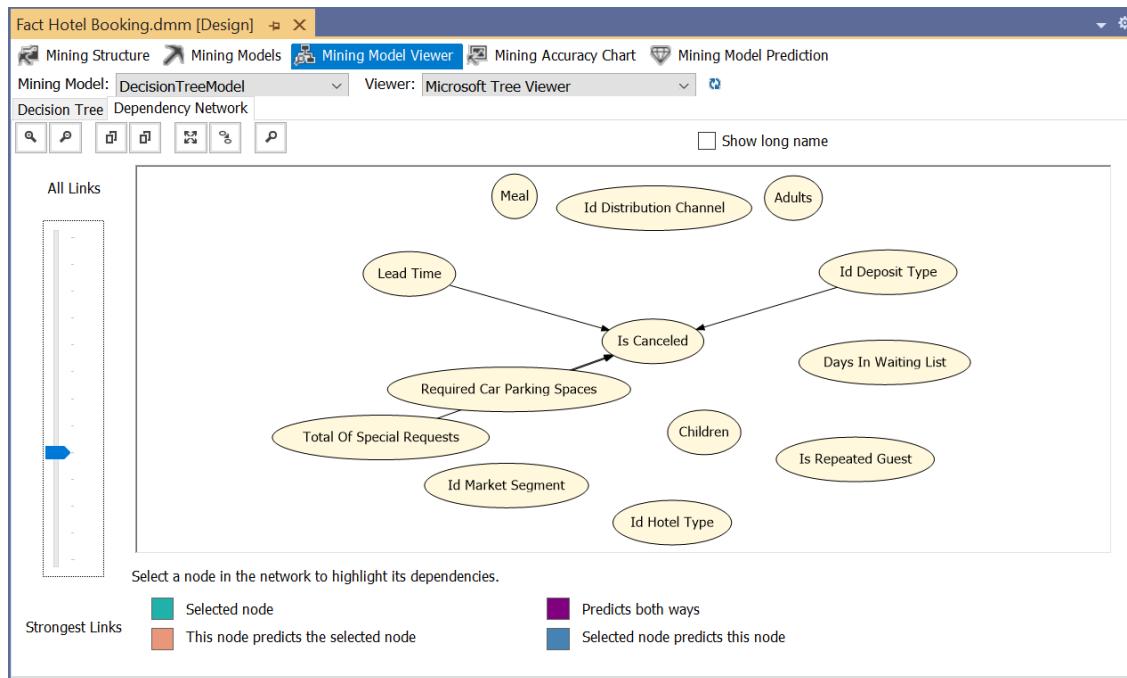
*Hình 5.53: Mírc phụ thuộc cấp 8.*



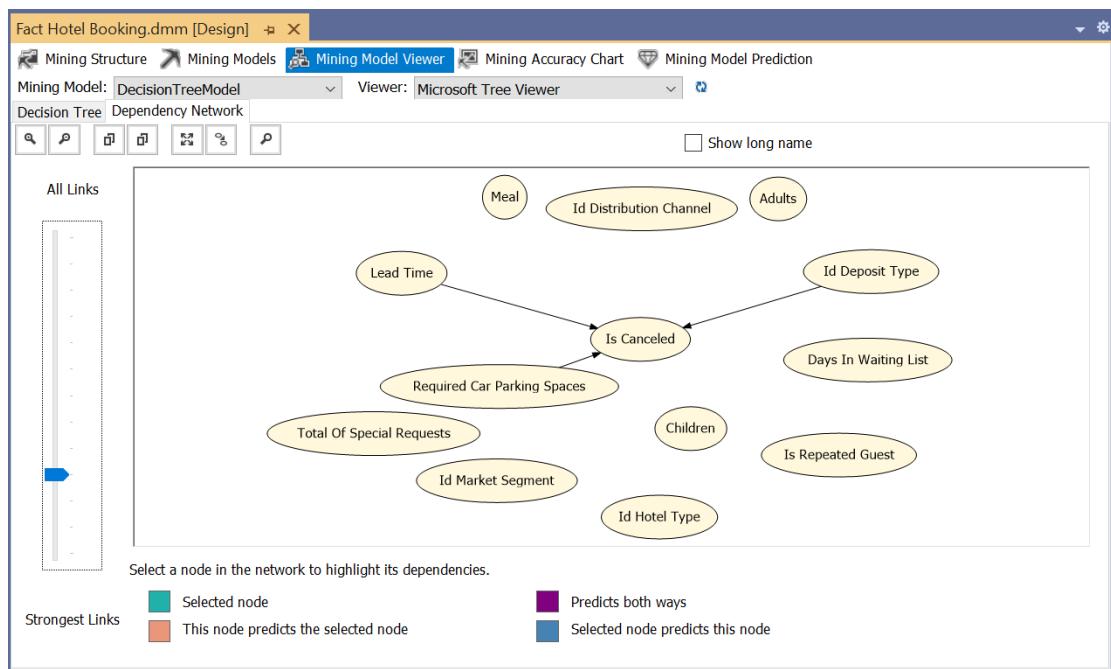
*Hình 5.54: Mírc phụ thuộc cấp 7.*



*Hình 5.55: Mírc phụ thuộc cấp 6.*

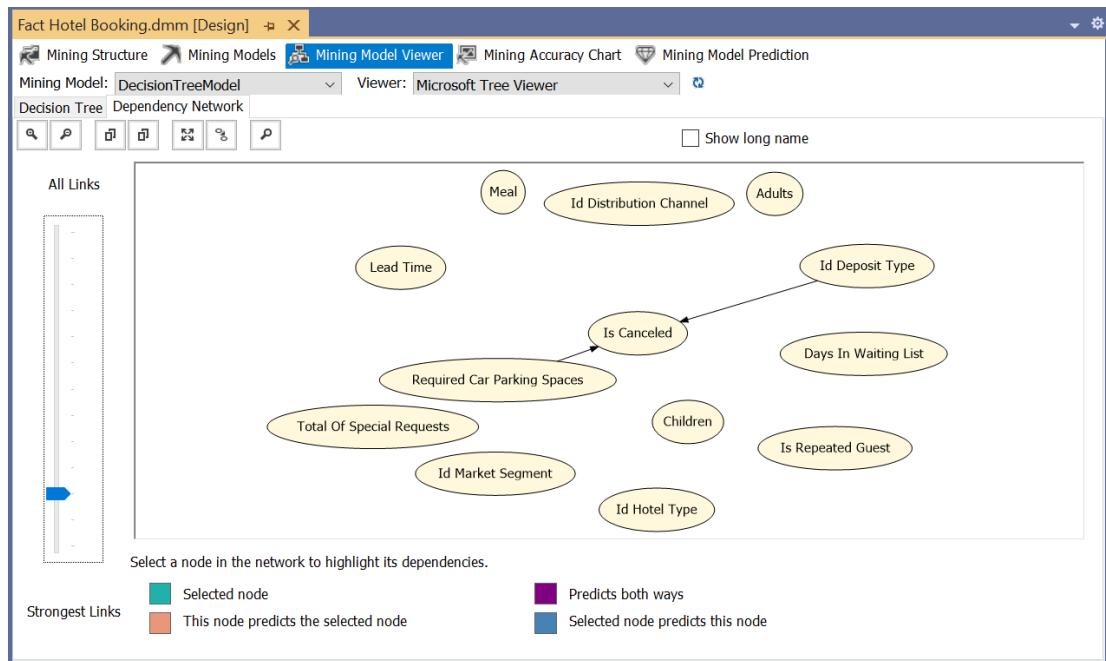


*Hình 5.56: Mírc phu thuộc cấp 5.*

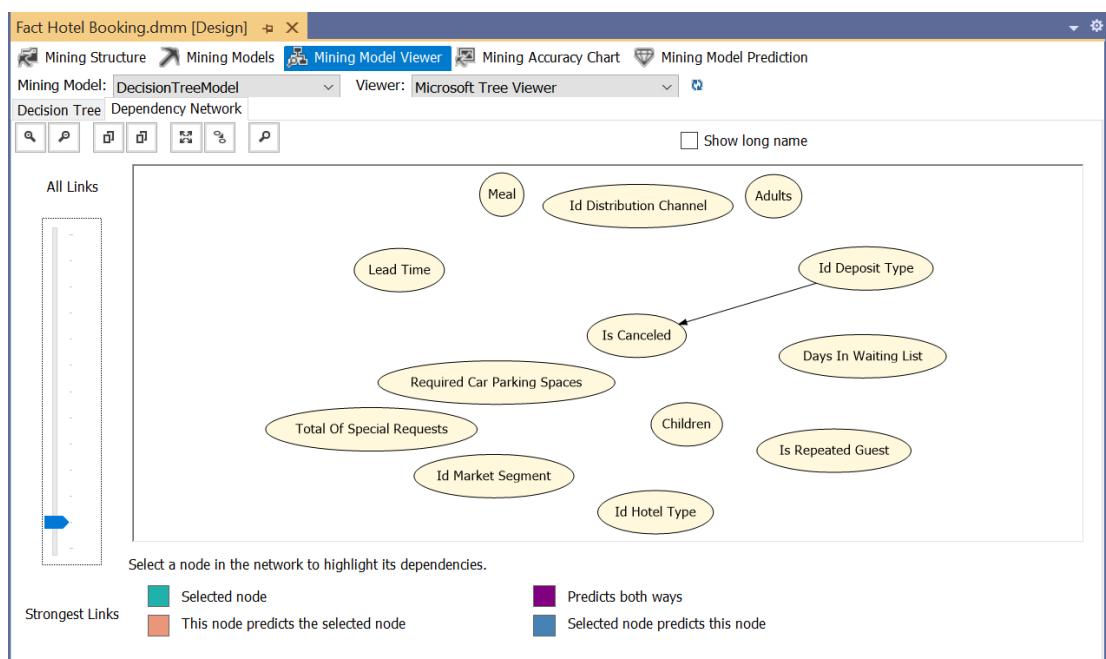


*Hình 5.57: Mírc phu thuộc cấp 4.*

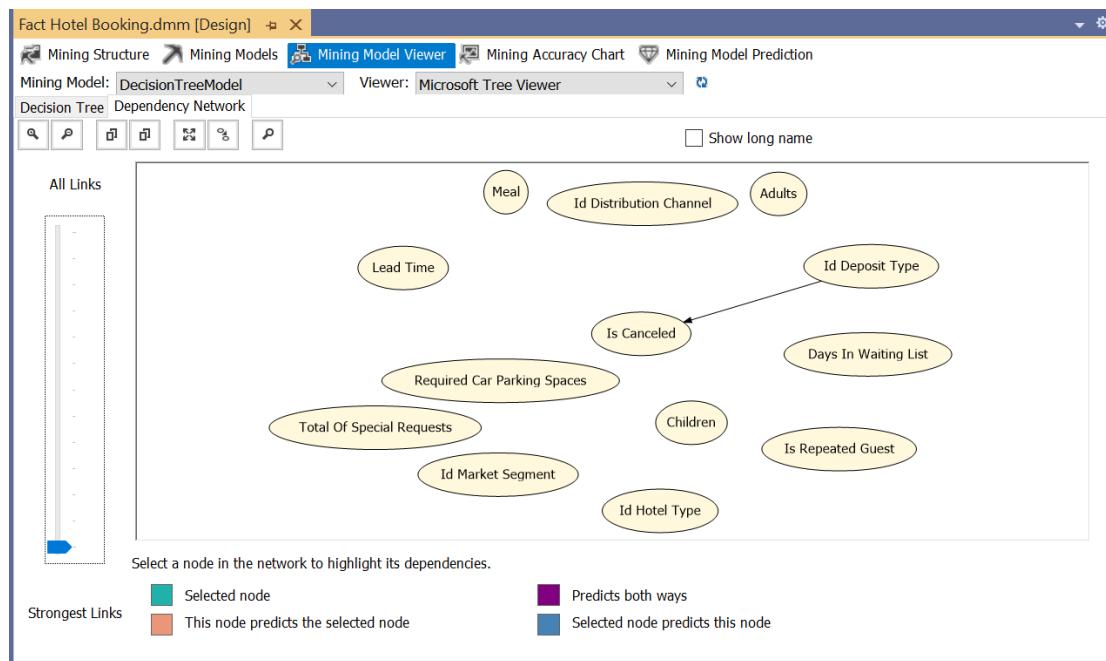
## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 5.58: Mức phụ thuộc cấp 3.

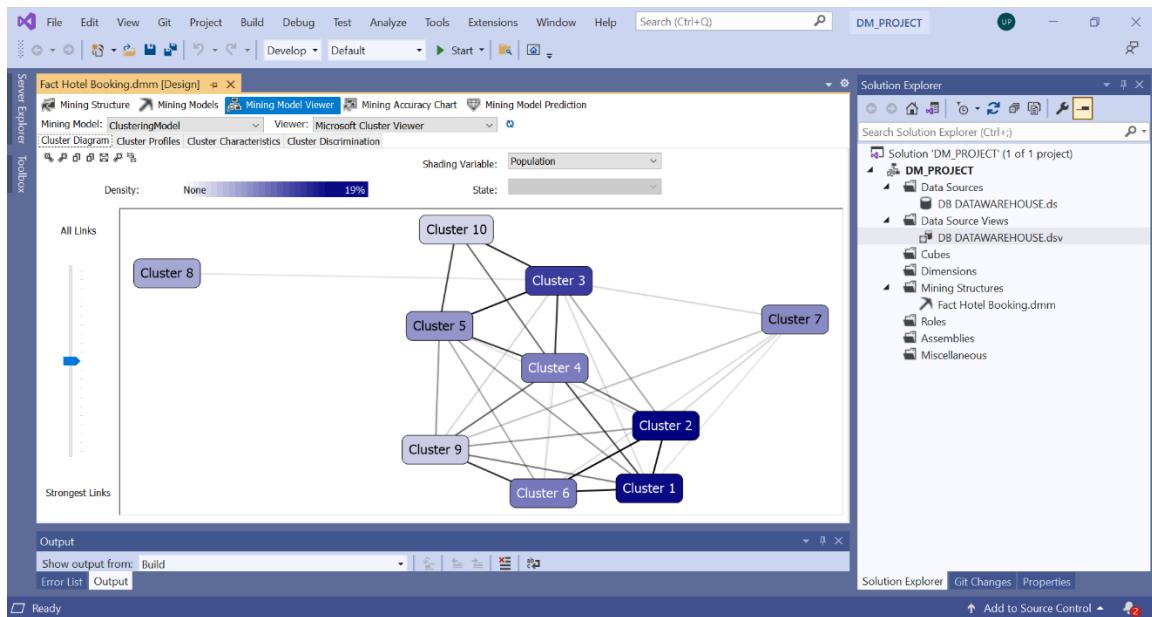


Hình 5.59: Mức phụ thuộc cấp 2.



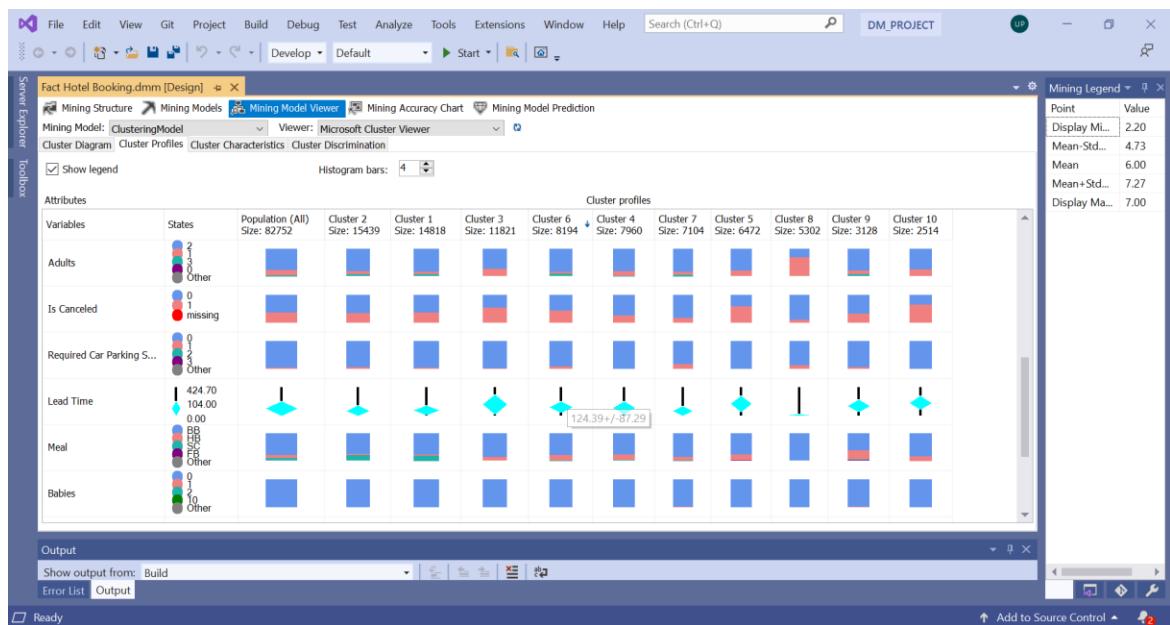
*Hình 5.60: Mức phụ thuộc cấp 1.*

- Bước 2: Chọn thuật toán Microsoft Clustering

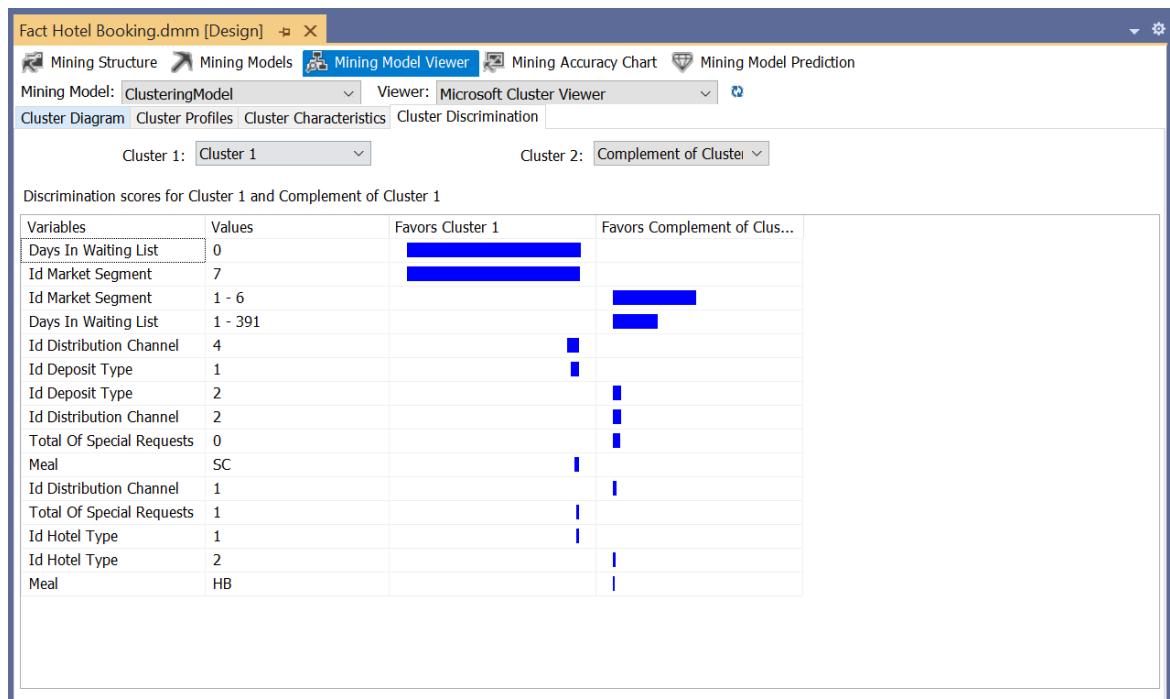


*Hình 5.61 : Chọn thuật toán Microsoft Clustering*

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22



Hình 5.62: Cho thấy khả năng nhận phòng của Cluster 8 cao nhất là 87.9%



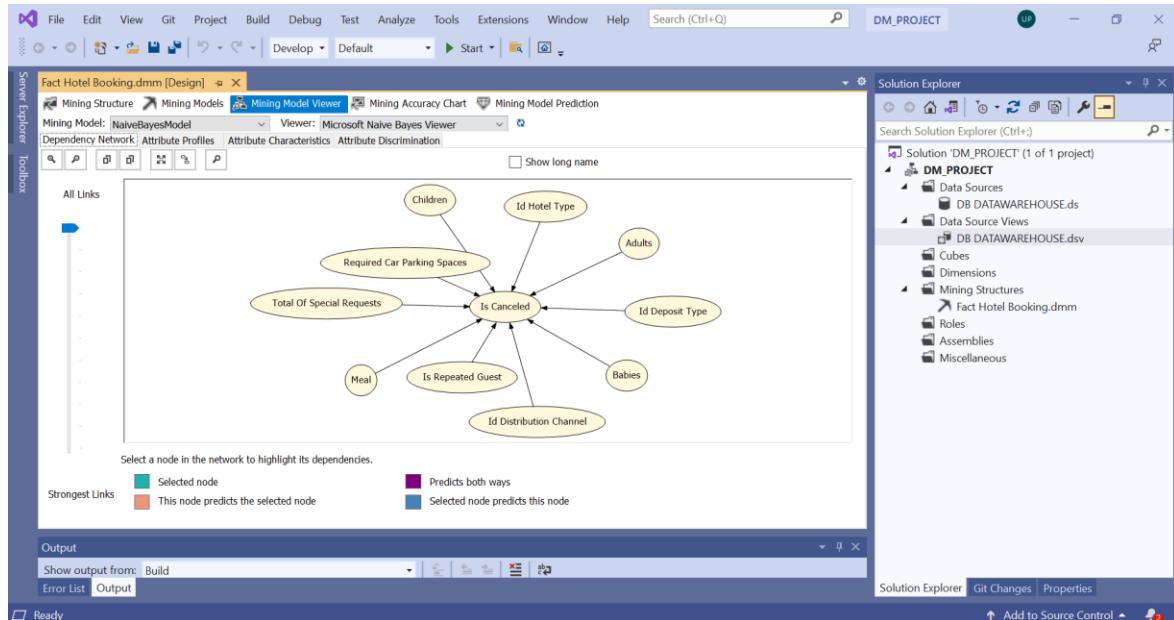
Hình 5.63: Tab Attribution Discrimination của thuật toán Clustering.

### ❖ Nhận xét ở cụm 1:

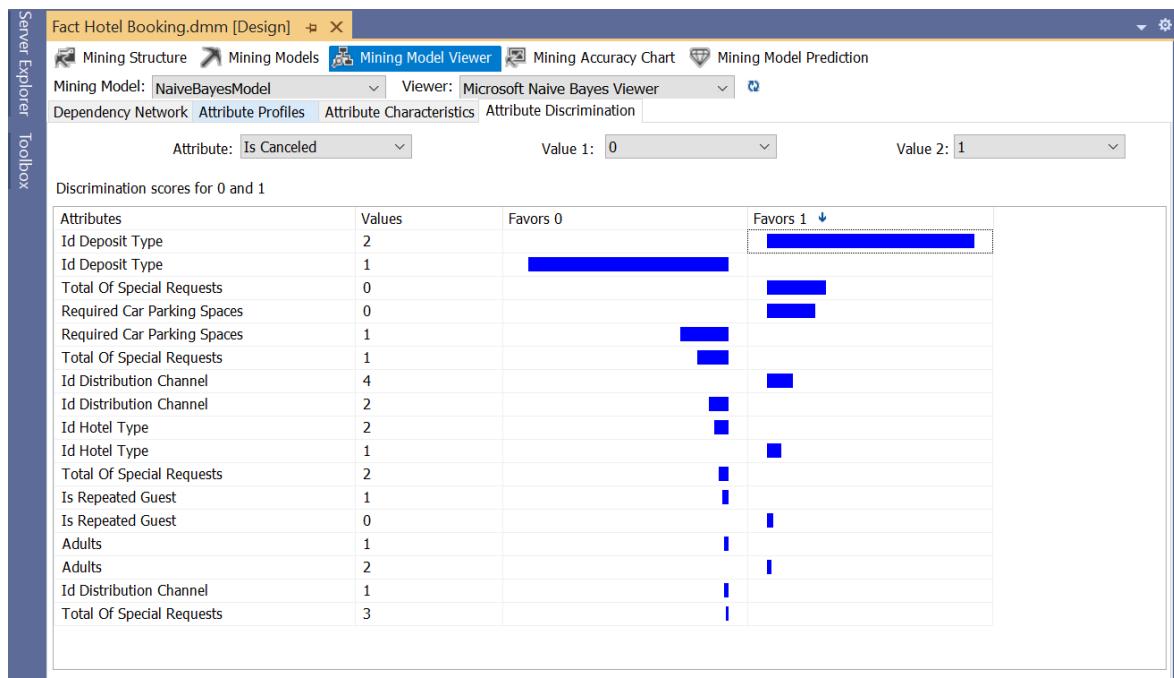
- Những khách hàng đăng ký online có tỷ lệ nhận phòng cao trong khi những phân khúc thị trường còn lại sẽ có khả năng hủy phòng nhiều hơn.

## Dồ án phân tích hoạt động đặt phòng khách sạn – IS217.M22

- Những khách hàng sử dụng bữa ăn “SC” sẽ có khả năng nhận phòng nhưng không cao.
- Bước 3: Tiếp tục tạo thêm thuật toán Mining Microsoft Naive Bayes



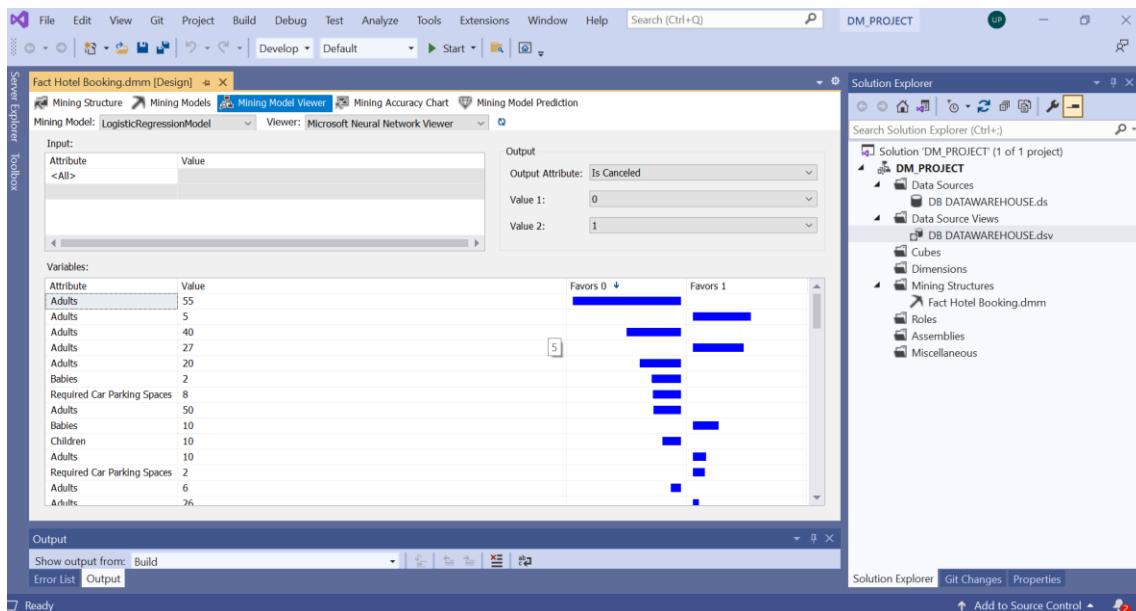
Hình 5.64 : Chọn thuật toán Microsoft Naive Bayes



Hình 5.65: Tab Attribution Discrimination của thuật toán Naïve Bayes.

❖ Nhận xét:

- Những khách hàng có hình thức đặt cọc là không đặt cọc, có yêu cầu đặc biệt, có bãі đỗ xe và đi một người thuê phòng tại khách sạn nghỉ dưỡng thường sẽ không hủy đặt phòng.
- Những khách hàng có hình thức thanh toán là không hoàn lại, không có yêu cầu hêm và ở khách sạn trong thành phố thường hủy đặt phòng.
- Bước 4: Chọn thuật toán Microsoft Logistic Regression

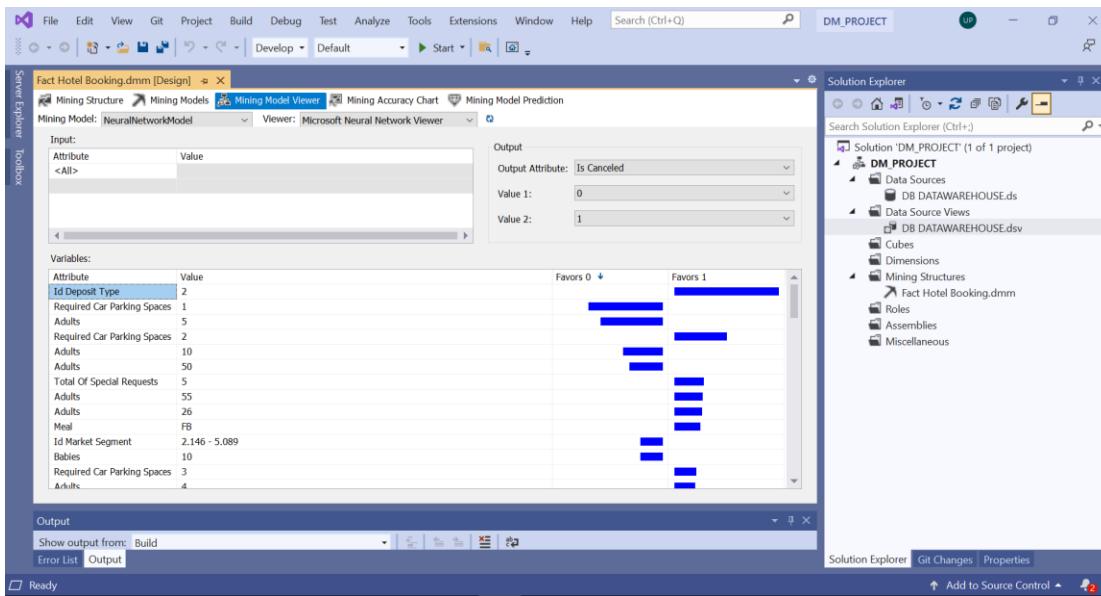


Hình 5.66: Chọn thuật toán Microsoft Logistic Regression

❖ Nhận xét:

- Khi đi với số lượng lớn ( $> 30$  người) thì khách hàng thường nhận phòng với độ chính xác hơn 50%.
- Ngược lại, khi đi với số lượng ít hơn 30 người, tỷ lệ hủy đặt phòng cao hơn rất nhiều so với tỷ lệ nhận phòng.

- Bước 5: Chọn thuật toán Microsoft Neural Network



Hình 5.67: Chọn thuật toán Microsoft Neural Network

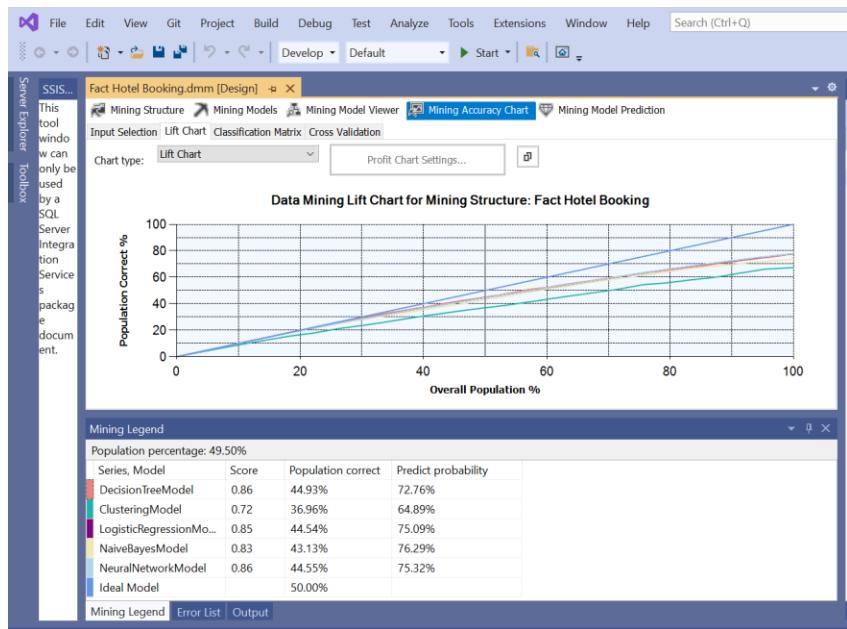
❖ **Nhận xét:**

- Những khách hàng đi đông thường hủy đặt phòng.
- Những khách hàng đi cùng gia đình (có trẻ em hoặc em bé) có 22% khả năng nhận phòng.

## 11. So sánh kết quả giữa hai lần chạy

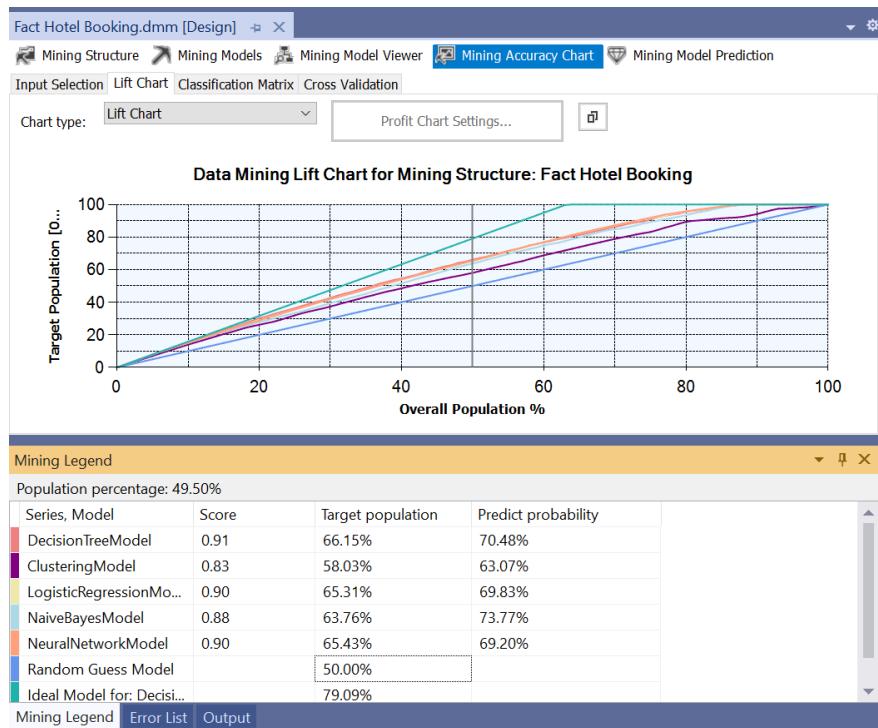
### 11.1. So sánh và đánh giá các thuật toán bằng đồ thị Lift

- Để xem đồ thị Lift chart trong tab Mining Accuracy Chart chọn Lift Chart with No Target Value.
  - Đồ thị Lift chart with no target value xảy ra khi ta không lựa chọn giá trị dự đoán. Biểu đồ sẽ thể hiện tất cả các trạng thái có thể dự đoán (0: nhận phòng và 1: hủy đặt phòng).



Hình 5.68: *Đồ thị Lift chart with no target value* về so sánh các thuật toán

- Line chart cho ta thấy tỷ lệ chính xác của các thuật toán khi Mining trường hợp dự đoán khả năng hủy đặt phòng của khách hàng
  - Microsoft Decision Tree: 0.86 Score.
  - Microsoft Clustering: 0.72 Score.
  - Microsoft Naive Bayes: 0.83 Score.
  - Microsoft Neural Network: 0.86 Score.
  - Microsoft Logistic Regression: 0.85 Score.
- Khi tổng độ chính xác là 50% thì có thể thấy được rằng, thuật toán Decision Tree đạt được độ chính xác cao nhất.
- Bảng dữ liệu cho thấy, ở 50% sự phân bố của dữ liệu, mô hình Decision Tree có độ chính xác 44.93%. Do đó, mô hình có thể dự đoán chính xác rằng khoảng 90% khách hàng nhận phòng thành công.
- Để xem đồ thị Lift chart trong tab Mining Accuracy Chart chọn Lift Chart with Target Value.
  - Đồ thị Lift chart with target value xảy ra khi ta lựa chọn giá trị dự đoán. Biểu đồ sẽ thể hiện trạng thái cụ thể (0: nhận phòng và 1: hủy đặt phòng).

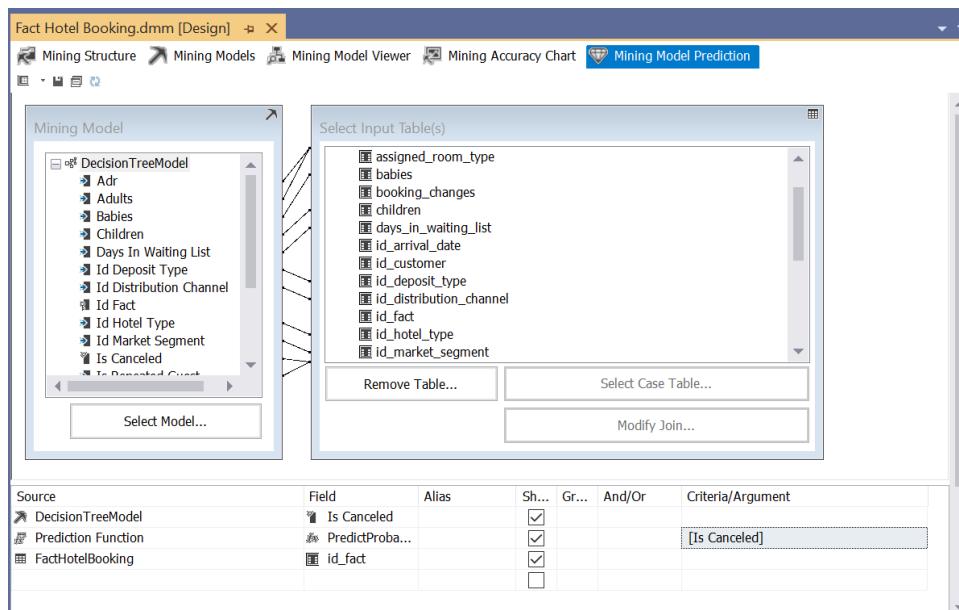


Hình 5.69: Đồ thị Lift chart with target value về so sánh các thuật toán và giá trị mục tiêu là 0 (nhận phòng).

- Line chart cho ta thấy tỷ lệ chính xác của các thuật toán khi Mining trường hợp dự đoán khả năng hủy đặt phòng của khách hàng
  - Microsoft Decision Tree: 0.91 Score.
  - Microsoft Clustering: 0.83 Score.
  - Microsoft Naive Bayes: 0.9 Score.
  - Microsoft Neural Network: 0.88 Score.
  - Microsoft Logistic Regression: 0.9 Score.
- Khi tổng độ chính xác là 79.09% thì có thể thấy được rằng, thuật toán Decision Tree đạt được độ chính xác cao nhất.
- Bảng dữ liệu cho thấy, ở 79.09% sự phân bố của dữ liệu, mô hình Decision Tree có độ chính xác 66.15%. Do đó, mô hình có thể dự đoán chính xác rằng khoảng 90% khách hàng nhận phòng thành công.

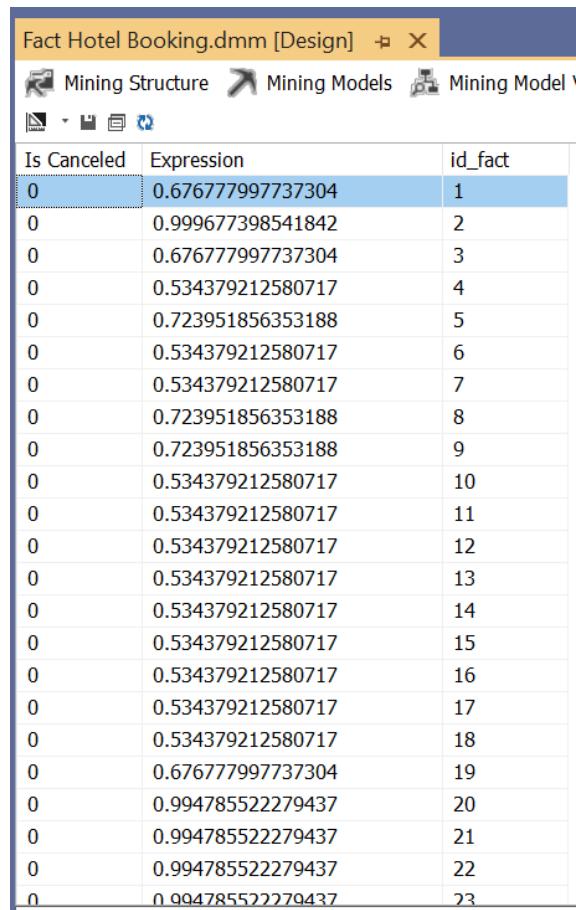
### 11.2. Dự đoán kết quả

- Sau khi chọn được mô hình phù hợp (tốt nhất), bước tiếp theo là sử dụng mô hình đó cho việc dự đoán. Bạn có thể sử dụng ngôn ngữ DMX để thiết lập các query cho dự đoán.
- Các bước sử dụng model để dự đoán như sau:
  - Giả sử ta sử dụng mô hình **Decision Tree** (Mô hình có dự đoán chính xác cao nhất trong các Mining models đã xây dựng) để dự đoán khả năng khách hàng không hủy đặt phòng như sau:
    - + Trong Tab **Mining Model Prediction**, chọn **Select case table**, chọn **FactHotelBooking**
    - + Chọn **Decision Tree Model** trong mục **Select Model** Và thiết kế một prediction Query như sau:



Hình 5.70: Thay đổi các thông số trong Mining Model Prediction.

- Để xem kết quả của dự đoán, click chuột phải chọn Result, ta thu được kết quả như sau:



The screenshot shows a Microsoft SQL Server Data Mining interface titled "Fact Hotel Booking.dmm [Design]". The window has tabs for "Mining Structure", "Mining Models", and "Mining Model". The "Mining Model" tab is selected. Below the tabs is a toolbar with icons for back, forward, search, and refresh. The main area displays a table with three columns: "Is Canceled", "Expression", and "id\_fact". The "Is Canceled" column contains binary values (0 or 1). The "Expression" column contains numerical values representing predicted probabilities. The "id\_fact" column contains integer values from 1 to 23. The table has 23 rows.

Is Canceled	Expression	id_fact
0	0.676777997737304	1
0	0.999677398541842	2
0	0.676777997737304	3
0	0.534379212580717	4
0	0.723951856353188	5
0	0.534379212580717	6
0	0.534379212580717	7
0	0.723951856353188	8
0	0.723951856353188	9
0	0.534379212580717	10
0	0.534379212580717	11
0	0.534379212580717	12
0	0.534379212580717	13
0	0.534379212580717	14
0	0.534379212580717	15
0	0.534379212580717	16
0	0.534379212580717	17
0	0.534379212580717	18
0	0.676777997737304	19
0	0.994785522279437	20
0	0.994785522279437	21
0	0.994785522279437	22
0	0.994785522279437	23

*Hình 5.71: Kết quả của quá trình dự đoán.*

- Các cột id\_fact, Is Canceled và Expression được dùng để xác định khách hàng hủy hoặc nhân phòng và xác suất của các dự đoán chính xác. Để từ đó, khách sạn có thể đưa ra các quyết định hoặc phương pháp để có thể đạt được lợi nhuận cao nhất.

## BẢNG PHÂN CÔNG CÔNG VIỆC

Công việc	Thành viên	Đặng Vũ Phương Uyên 19520345	Nguyễn Thị Thu Phương 19522066
<b>Chọn dữ liệu Hotel Booking Demand.</b>		x	x
<b>Đưa dữ liệu vào kho (SSIS).</b>		x	Hỗ trợ
<b>Quá trình phân tích dữ liệu</b> <ul style="list-style-type: none"><li>• Tạo Project SSAS.</li><li>• Tạo Nameset, Calculate.</li><li>• Phân tích 15 câu truy vấn bằng Manual.</li><li>• Phân tích 15 câu truy vấn bằng công cụ Pivot Excel.</li></ul>		x	x
<b>Quá trình phân tích dữ liệu</b> <ul style="list-style-type: none"><li>• Ngôn ngữ MDX.</li><li>• So sánh các kết quả.</li></ul>		x	x
<b>Quá trình SSRS</b> <ul style="list-style-type: none"><li>• Tạo Project báo biểu.</li><li>• Đưa ra 3 report bằng công cụ Google Data Studio.</li><li>• Đưa Project SSRS lên Server.</li></ul>		x	x
<b>Quá trình Data Mining</b> <ul style="list-style-type: none"><li>• Tạo Projet Datamining.</li><li>• Tạo cấu trúc mining.</li><li>• Đưa ra input, predict.</li><li>• Đưa ra tập luật.</li></ul>		Hỗ trợ	x
<b>Tổng kết báo cáo</b>		Hỗ trợ	x

## **TÀI LIỆU THAM KHẢO**

1. Slide bài giảng môn Kho dữ liệu và OLAP . Tác giả : ThS. Đỗ Thị Minh Phụng.

2. Đồ án tham khảo môn Kho dữ liệu và OLAP.

3. Cài đặt công cụ SQL server, SSDT

<https://youtube.com/watch?v=cVjJswbptMQ&feature=share>

4. Quá trình SSAS

<https://youtube.com/watch?v=CgtCRK9rTGs&feature=share>

5. Quá trình SSIS

[https://youtube.com/playlist?list=PLNIs-AWhQzcmPg\\_uV2BZi\\_KRG4LKs6cRs](https://youtube.com/playlist?list=PLNIs-AWhQzcmPg_uV2BZi_KRG4LKs6cRs)

6. Tạo report bằng Report Builder

[https://youtube.com/playlist?list=PLNIs-AWhQzcmEFHyxCRwA\\_gb29WOz5SJU](https://youtube.com/playlist?list=PLNIs-AWhQzcmEFHyxCRwA_gb29WOz5SJU)

7. Tạo report bằng Google Data Studio

<https://youtube.com/watch?v=DLbwdLpqZD0&feature=share>