



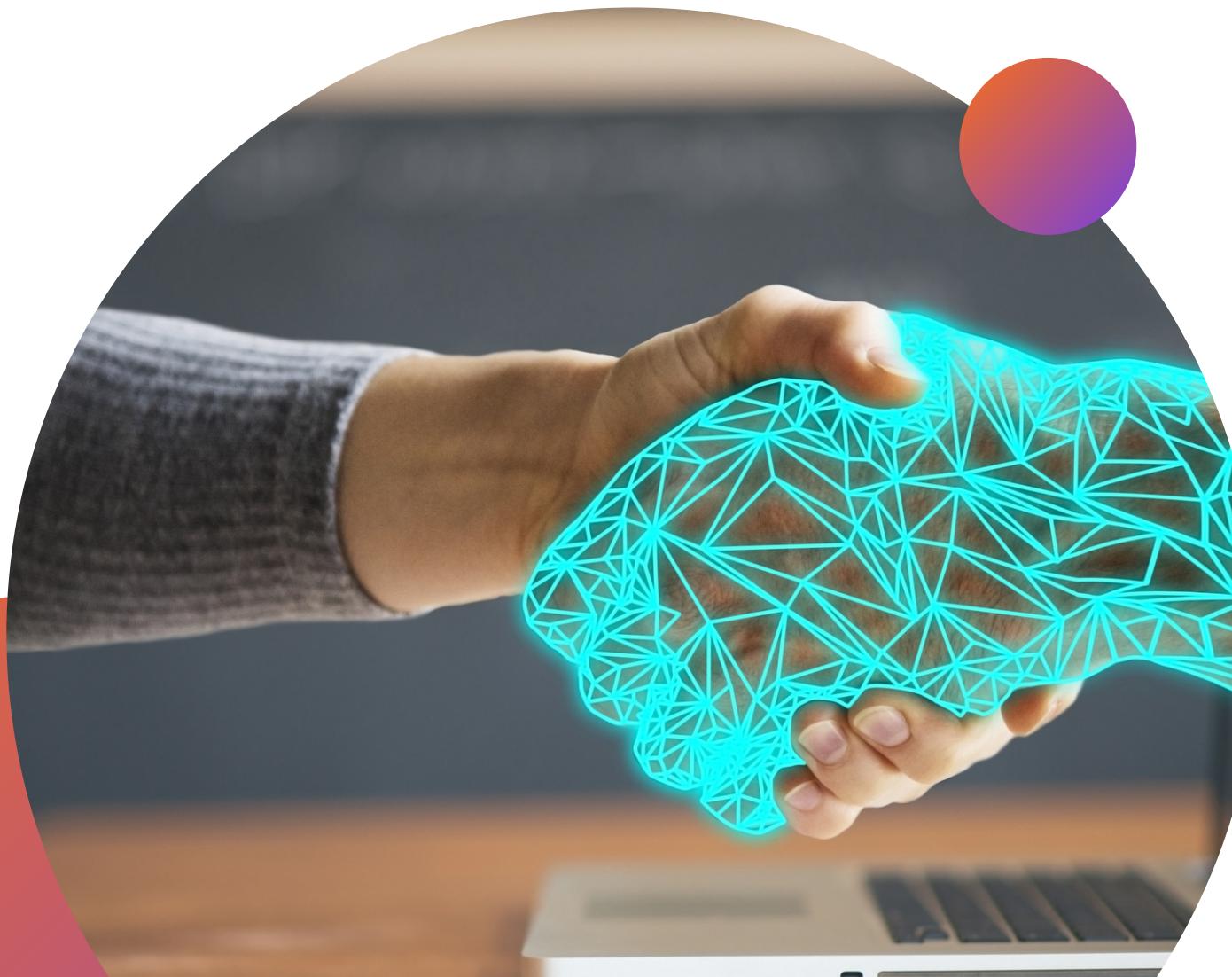
STUDIO SHODWE

BÁO CÁO CUỐI KÌ KHO DỮ LIỆU VÀ OLAP

UKRAINE AIRSPACE



STUDIO SHODWE



NHÓM 18

19522208 NGUYỄN HỮU THẮNG

19522215 PHẠM MINH THẮNG



01

TỔNG QUAN ĐỀ TÀI

02

TÍCH HỢP DỮ LIỆU VÀO
KHO (SSIS)

03

PHÂN TÍCH DỮ LIỆU TRONG
KHO (SSAS)

04

QUÁ TRÌNH LẬP BÁO BIỂU
(SSRS)

05

DATA MINING

1. Tổng quan đề tài

2. Giới thiệu về dataset

3. Xây dựng kho dữ liệu

4. Các câu truy vấn

1. Lý do chọn đề tài

- Ukraine là một trong những điểm đến du lịch nổi tiếng nhất ở châu Âu.
- Trong những ngày gần đây, sự ảnh hưởng từ Nga và Ukraine là rất lớn, có rất nhiều chuyến bay đi qua không phận của Ukraine. Trong đó có cả những chuyến bay đưa các công dân Việt Nam về nước.

2. Giới thiệu về dataset

Tên kho dữ liệu: Ukraine Airspace

Ngày cập nhật gần nhất: 05/03/2022

Kho dữ liệu Ukraine Airspace là kho dữ liệu chuyến bay trong không phận Ukraine từ ngày 22/02/2022 đến ngày 28/03/2022.

Từ kho dữ liệu, người dùng có thể biết được các thông tin như:

ICAO 24-bit.

Ký hiệu Quốc gia xuất phát.

Thời gian ở vị trí quốc gia xuất phát.

Thời gian liên lạc cuối cùng.

Kinh độ

Vĩ độ

Độ cao khí áp

Máy bay có đăng ở mặt đất hay không

Vận tốc

Hướng đi

Tốc độ theo hướng thẳng đứng

Cảm biến

Độ cao địa lý

Squawk

Spi

Nguồn vị trí

Kho dữ liệu gồm 1048576 dòng dữ liệu và 17 cột thuộc tính

4. Xây dựng kho dữ liệu

4.1 Sơ đồ hình bông tuyết

4.2 dim_location

4.3 dim_plane

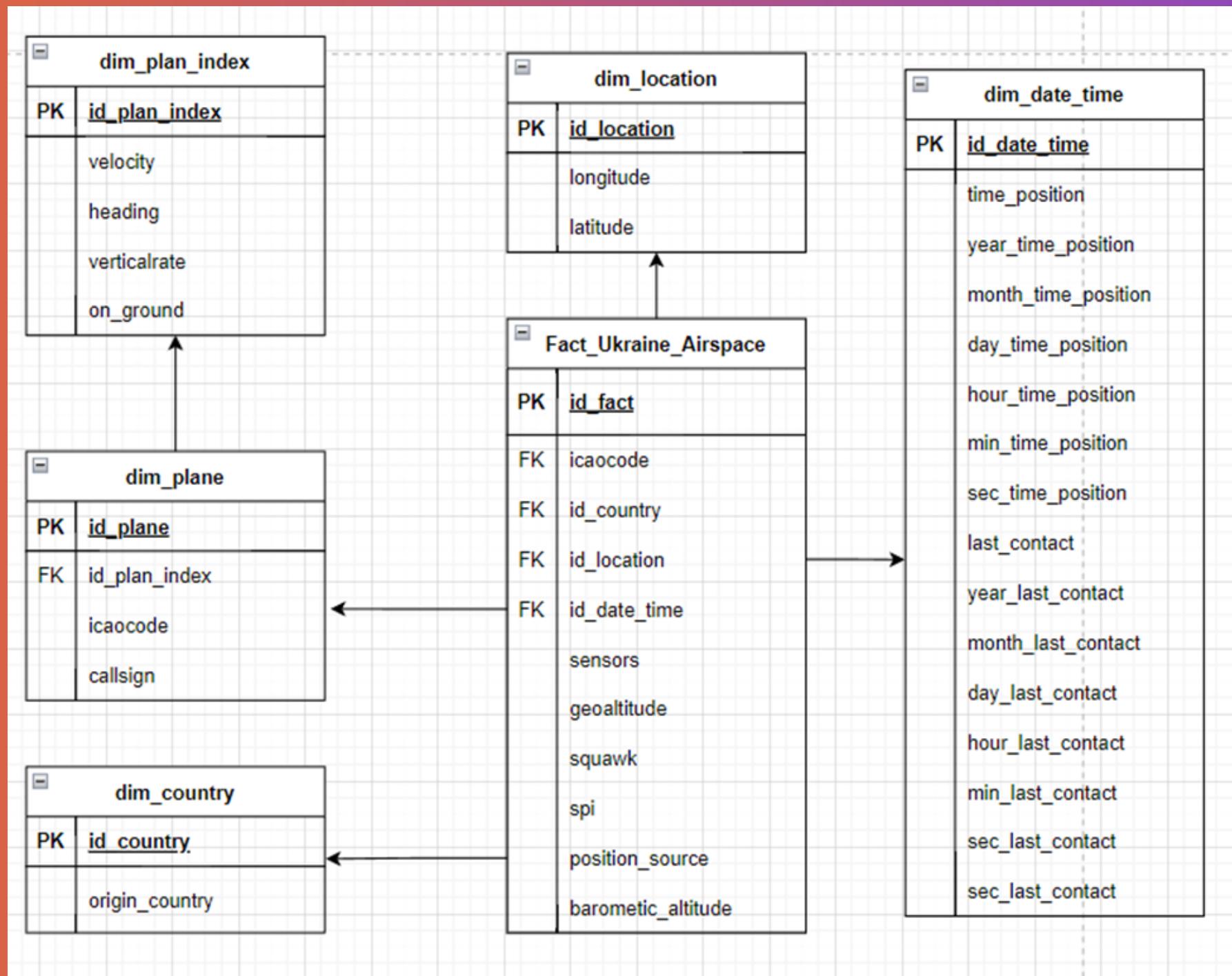
4.4 dim_plane_index

4.5 dim_date_time

4.6 dim_country

4.7 Fact_Ukraine_Airspace

Xây dựng kho dữ liệu



CHƯƠNG 2: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

1. Tạo project và thiết lập kết nối

1.1 Tạo project

1.2 Tạo cơ sở dữ liệu và thiết lập kết nối

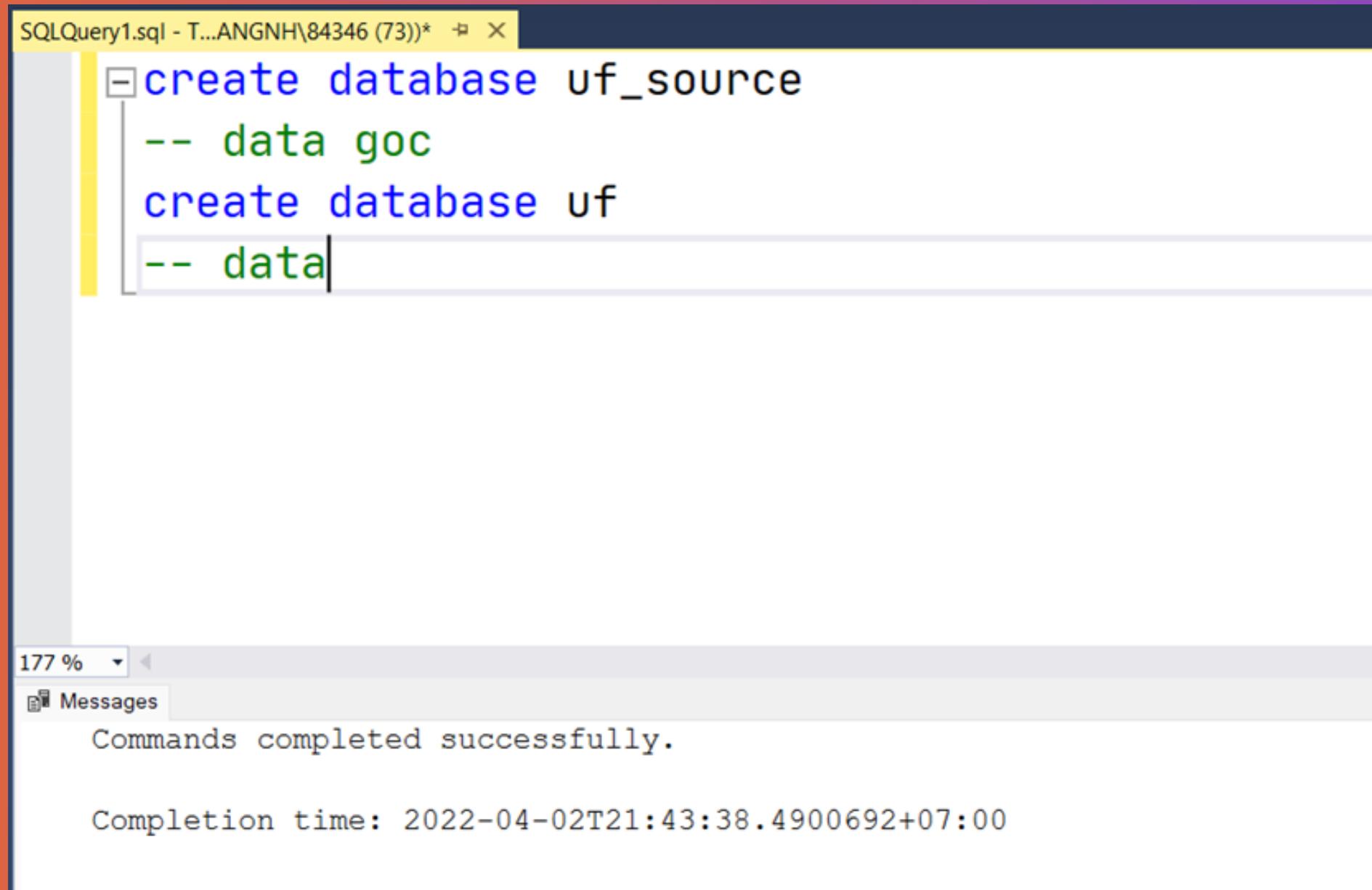
1.2.1 Tạo cơ sở dữ liệu

1.2.2 Thiết lập kết nối

2. Chuẩn bị dữ liệu gốc, import dữ liệu gốc

3. Quá trình làm sạch dữ liệu

Tạo cơ sở dữ liệu và thiết lập kết nối

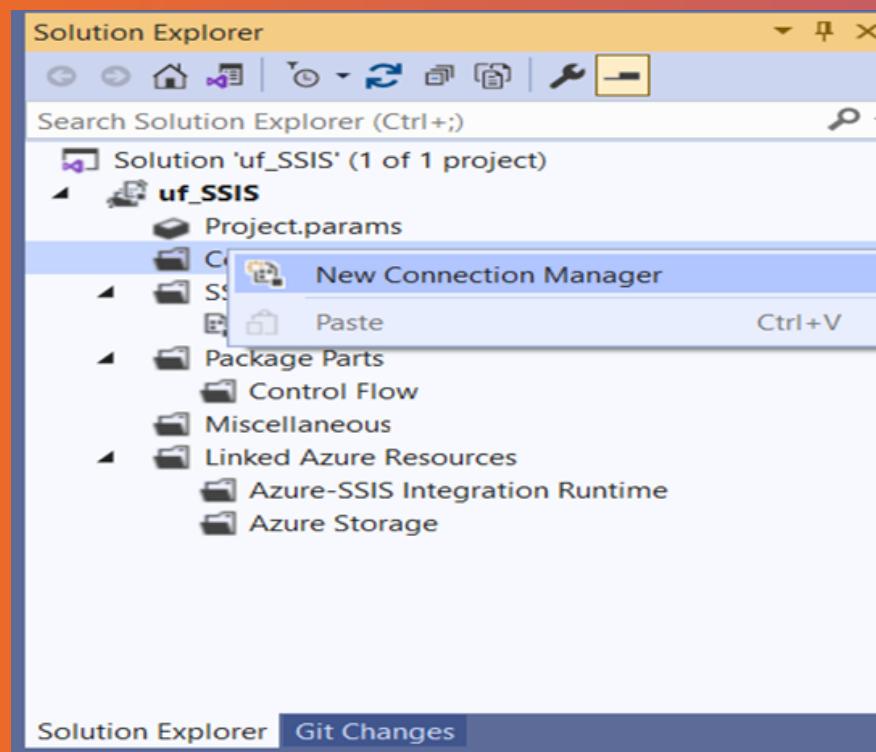


The screenshot shows a SQL query window in SSMS with the following content:

```
SQLQuery1.sql - T...ANGNH\84346 (73)*
create database uf_source
-- data goc
create database uf
-- data|
```

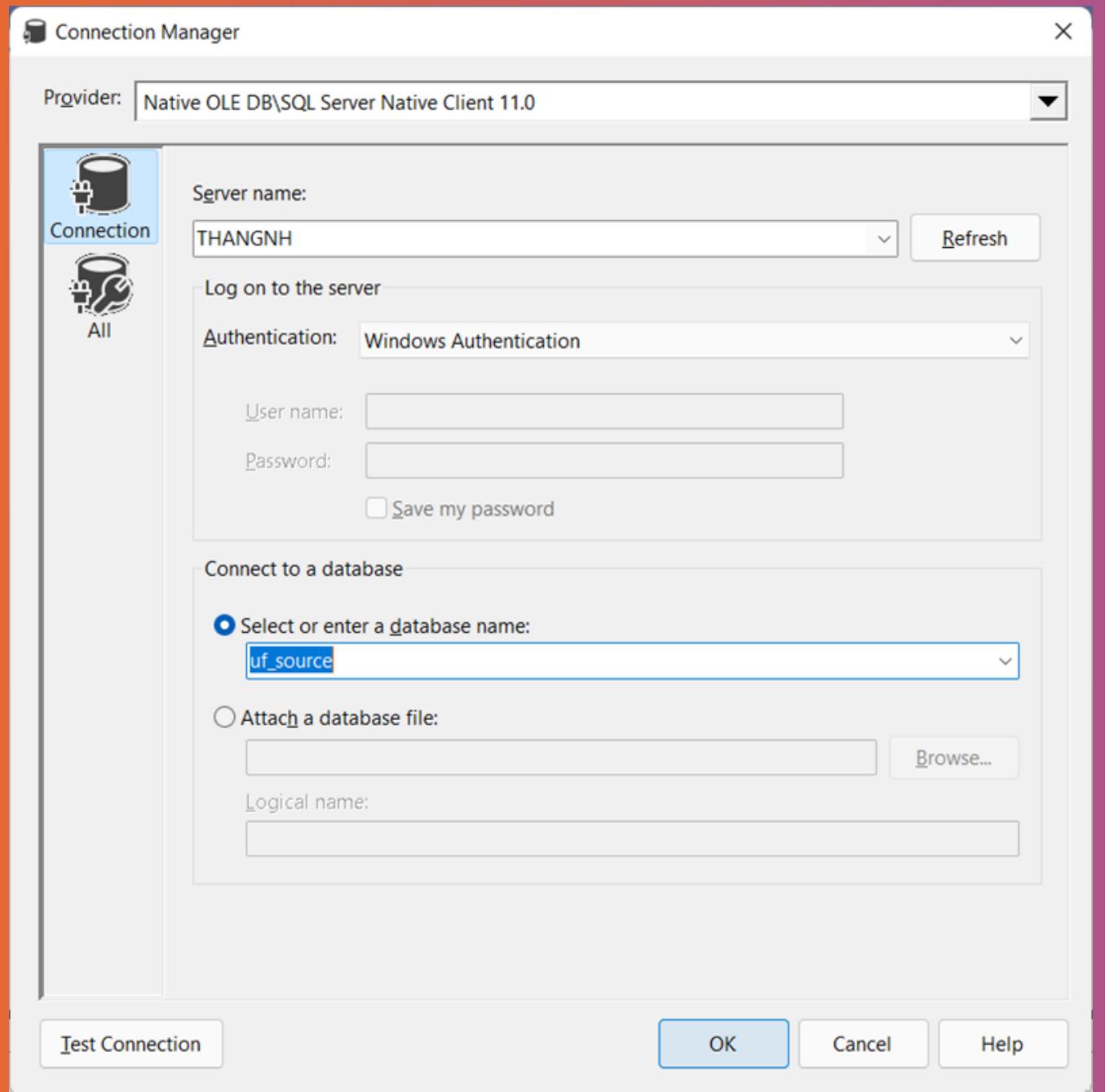
The window includes a status bar at the bottom displaying "177 %", a "Messages" tab, and the message "Commands completed successfully." followed by the completion time.

Thiết lập kết nối

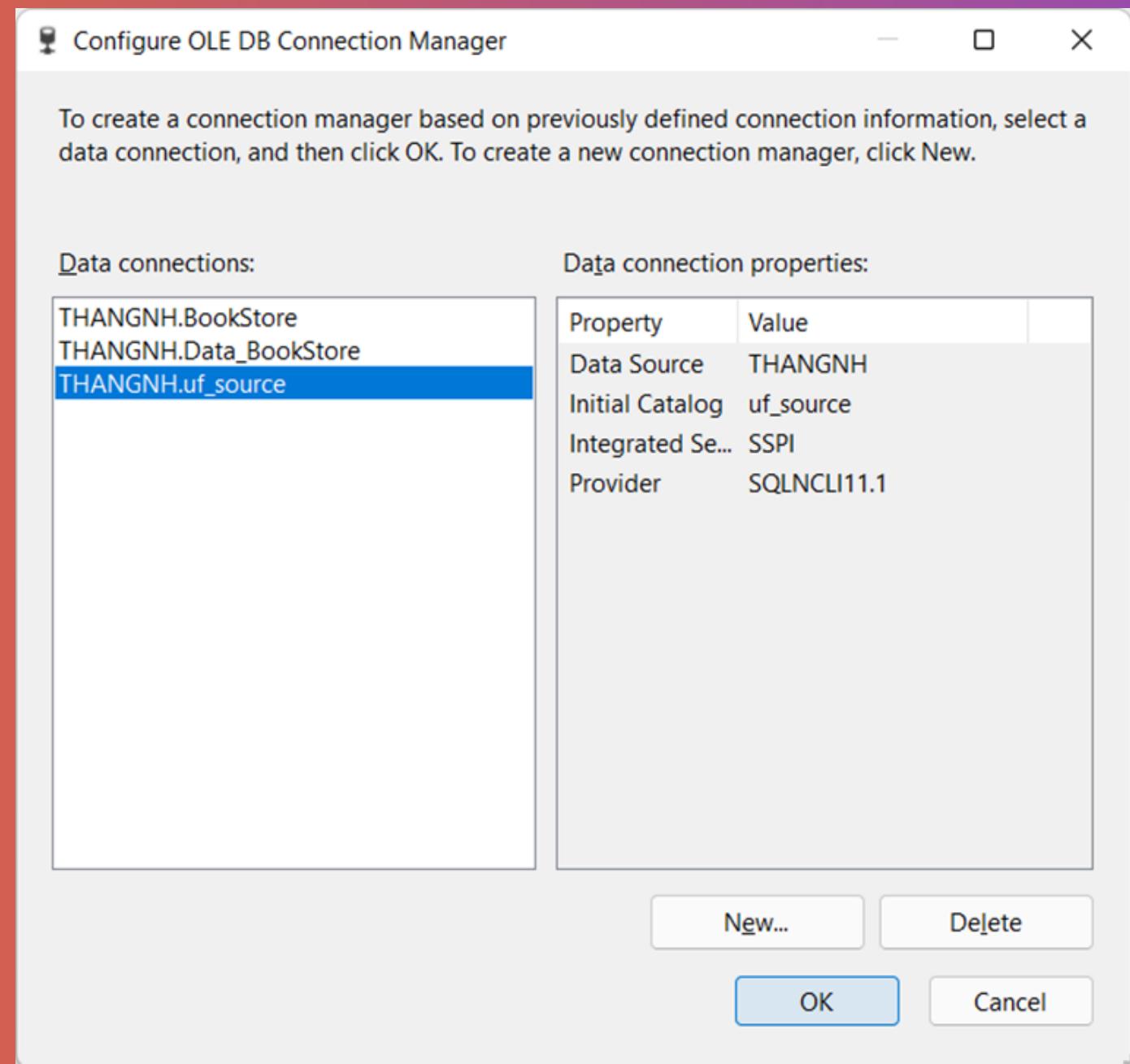


Chọn Configure OLE DB Connection Manager

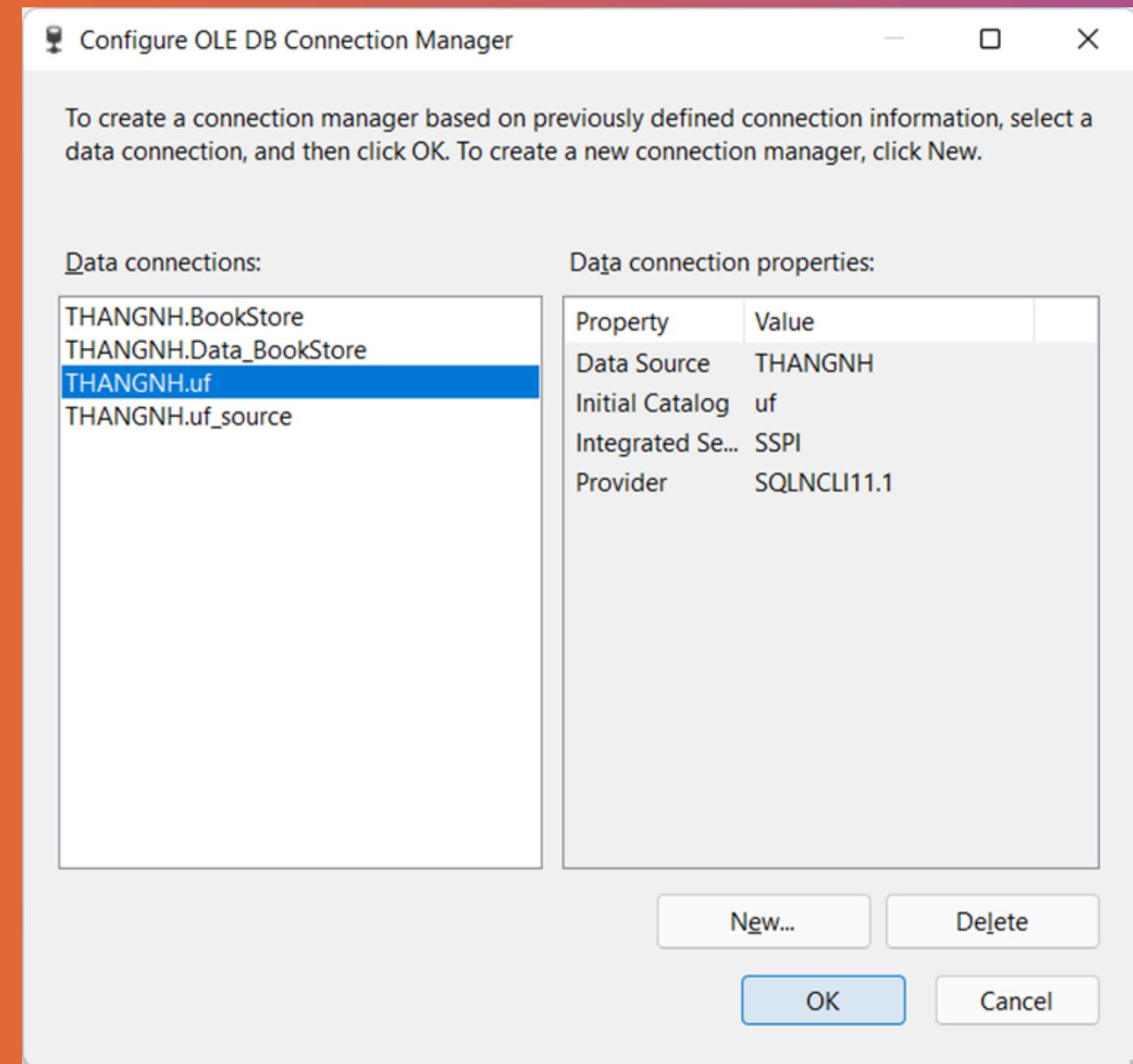
ODATA	Connection Manager for OData
ODBC	Connection manager for ODBC
OLEDB	Connection manager for OLE DB
SMOServer	Connection manager for SQL Server



Chọn database uf_source ở phần
Connect to a database

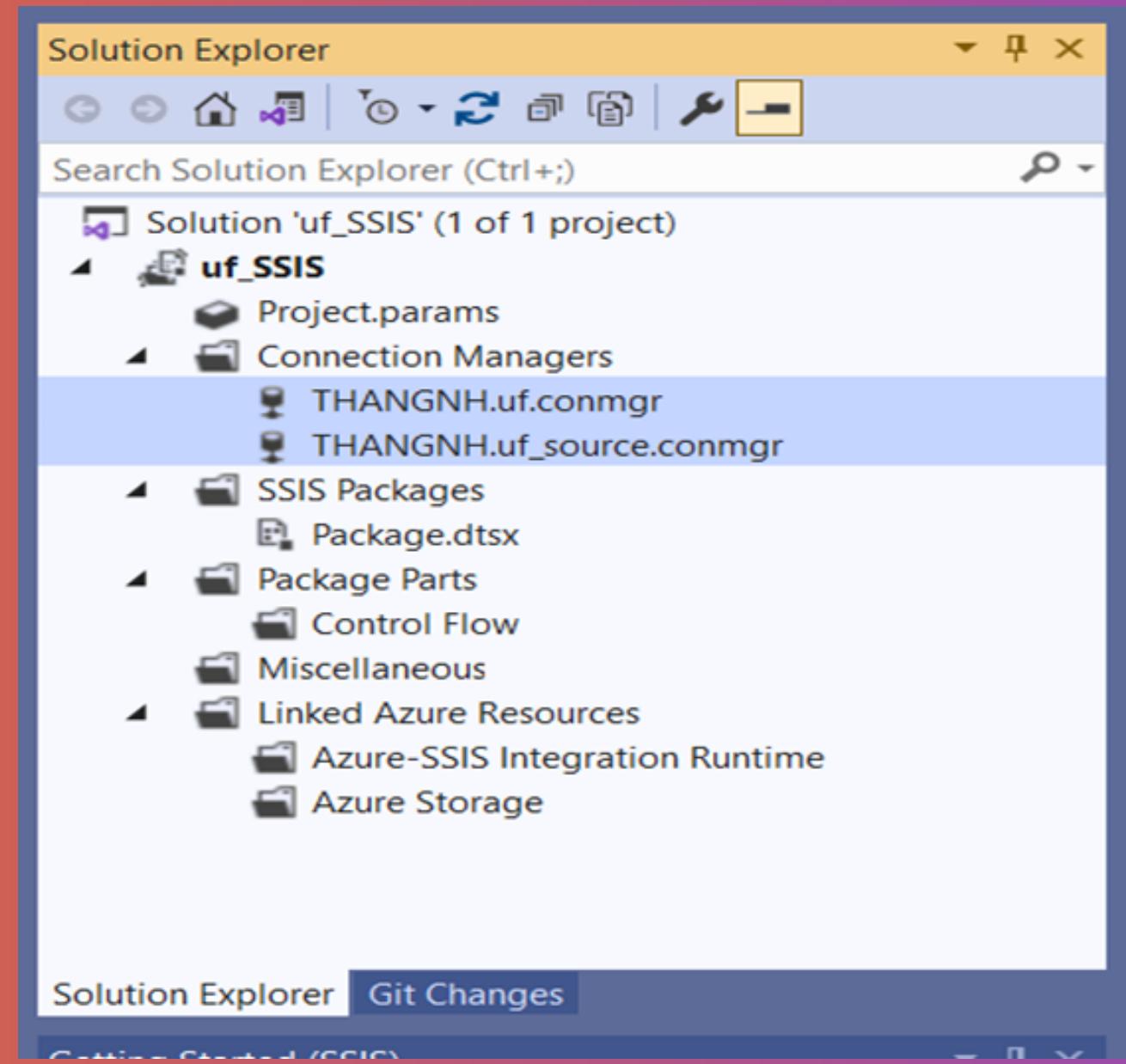


Chọn Data connections vừa thêm xong



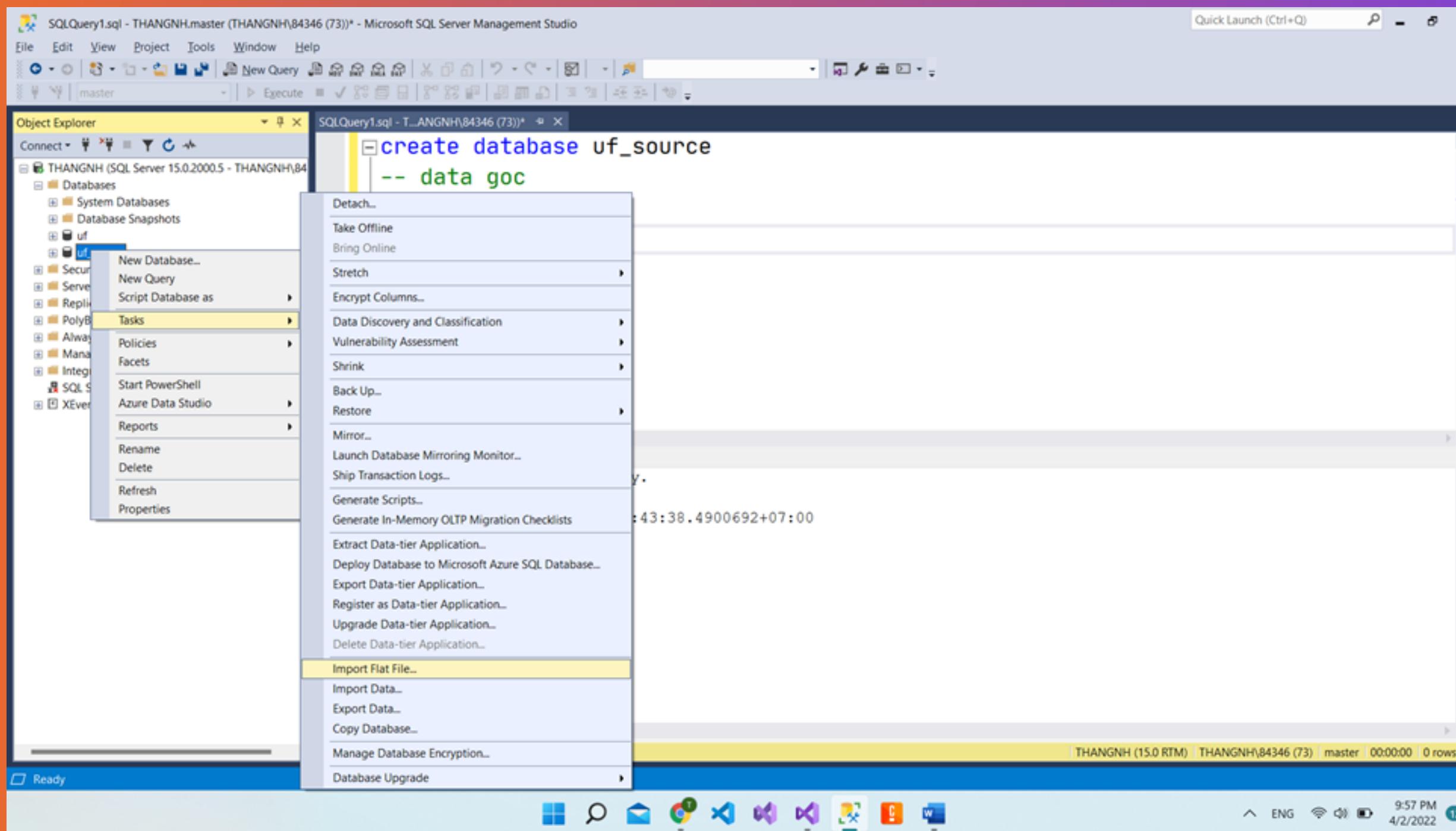
Tương tự ở trên

- Thiết lập kết nối đến database uf



Kết quả

2. Chuẩn bị dữ liệu gốc, import dữ liệu gốc



Preview Data

This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.

icao24	callsign	origin_country	time_position	last_contact	longitude	latit
4248ef	SVR522	United King...	2/22/2022 1...	2/22/2022 1...	37.7853	55.49
471f8c	WZZ1550	Hungary	2/22/2022 1...	2/22/2022 1...	21.2426	45.91
471f89	WZZ2276	Hungary	2/22/2022 1...	2/22/2022 1...	22.2773	46.08
4.25E+09	SDM6413	United King...	2/22/2022 1...	2/22/2022 1...	45.004	56.10
471f92	WZZ1166	Hungary	2/22/2022 1...	2/22/2022 1...	21.5623	46.65
471f57	WZZ4603	Hungary	2/22/2022 1...	2/22/2022 1...	29.9471	44.63
4520b6	GBG672	Bulgaria	2/22/2022 1...	2/22/2022 1...	23.5427	42.68
471f51	WZZ1202	Hungary	2/22/2022 1...	2/22/2022 1...	21.2312	44.52
471f3c	WZZ1575	Hungary	2/22/2022 1...	2/22/2022 1...	24.483	48.36
471f3b	WZZ3095	Hungary	2/22/2022 1...	2/22/2022 1...	25.6659	44.40
407b55	WUK4486	United King...	2/22/2022 1...	2/22/2022 1...	22.6999	50.57
a60693	BRI068	United States	2/22/2022 1...	2/22/2022 1...	26.1559	45.49
505cd8	IAW234	Slovakia	2/22/2022 1...	2/22/2022 1...	21.7485	46.05
463afe	JEF22	Finland	2/22/2022 1...	2/22/2022 1...	30.0547	56.18
3c4aab	CFG2321	Germany	2/22/2022 1...	2/22/2022 1...	29.0851	42.88
4249ce	SVR2841	United King...	2/22/2022 1...	2/22/2022 1...	38.5087	54.13
4249b2	AFL1843	United King...	2/22/2022 1...	2/22/2022 1...	37.7065	56.07
4401dc	EJU48YG	Austria	2/22/2022 1...	2/22/2022 1...	21.7903	43.86
489324	LOT7RM	Poland	2/22/2022 1...	2/22/2022 1...	23.5576	54.68
48b9e6	SPOPG	Poland	2/22/2022 1...	2/22/2022 1...	22.0975	50.24

Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous

Modify Columns

This operation generated the following table schema. Please verify if schema is accurate, and if not, please make any changes.

Column Name	Data Type	Primary Key	<input type="checkbox"/> Allow Nulls
icao24	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
callsign	nvarchar(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
origin_country	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
time_position	datetime2	<input type="checkbox"/>	<input type="checkbox"/>
last_contact	datetime2	<input type="checkbox"/>	<input type="checkbox"/>
longitude	float	<input type="checkbox"/>	<input type="checkbox"/>
latitude	float	<input type="checkbox"/>	<input type="checkbox"/>
barometric_altitude	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
on_ground	bit	<input type="checkbox"/>	<input type="checkbox"/>
velocity	float	<input type="checkbox"/>	<input type="checkbox"/>
heading	float	<input type="checkbox"/>	<input type="checkbox"/>
vertical_rate	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
sensors	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
geo_altitude	float	<input type="checkbox"/>	<input checked="" type="checkbox"/>
squawk	smallint	<input type="checkbox"/>	<input checked="" type="checkbox"/>
spi	bit	<input type="checkbox"/>	<input type="checkbox"/>
position_source	tinyint	<input type="checkbox"/>	<input type="checkbox"/>

Row granularity of error reporting (performance impact with smaller ranges)

Chỉnh sửa Data Type cho các cột thuộc tính

The screenshot shows two windows from Microsoft SQL Server Management Studio.

Import Flat File 'uf_source' (Left Window):

- Results** tab selected.
- Operation Complete**: Summary table:

Name	Result
Insert Data	Warning

SQL Server Management Studio (Right Window):

- Object Explorer** shows the database structure:

 - THANGNH (SQL Server 15.0.2000.5 - THANGNH\84346)
 - Databases
 - uf
 - uf_source
 - Tables
 - dbo.ukraineflights
 - Views
 - External Resources
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
 - Server Objects
 - Replication
 - PolyBase
 - Always On High Availability
 - Management
 - Integration Services Catalogs
 - SQL Server Agent (Agent XPs disabled)
 - XEvent Profiler

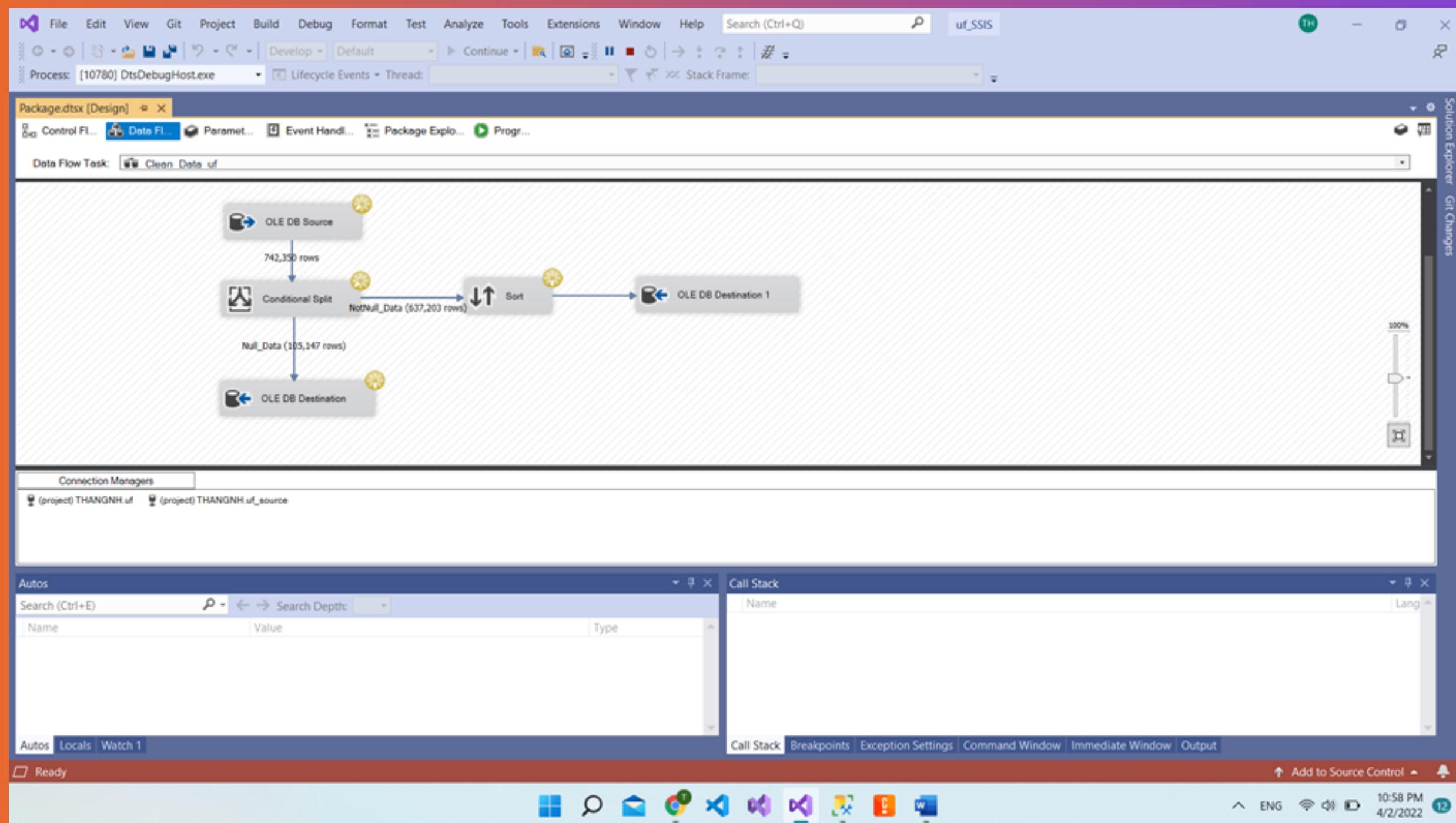
- SQLQuery2.sql** (Top Query):

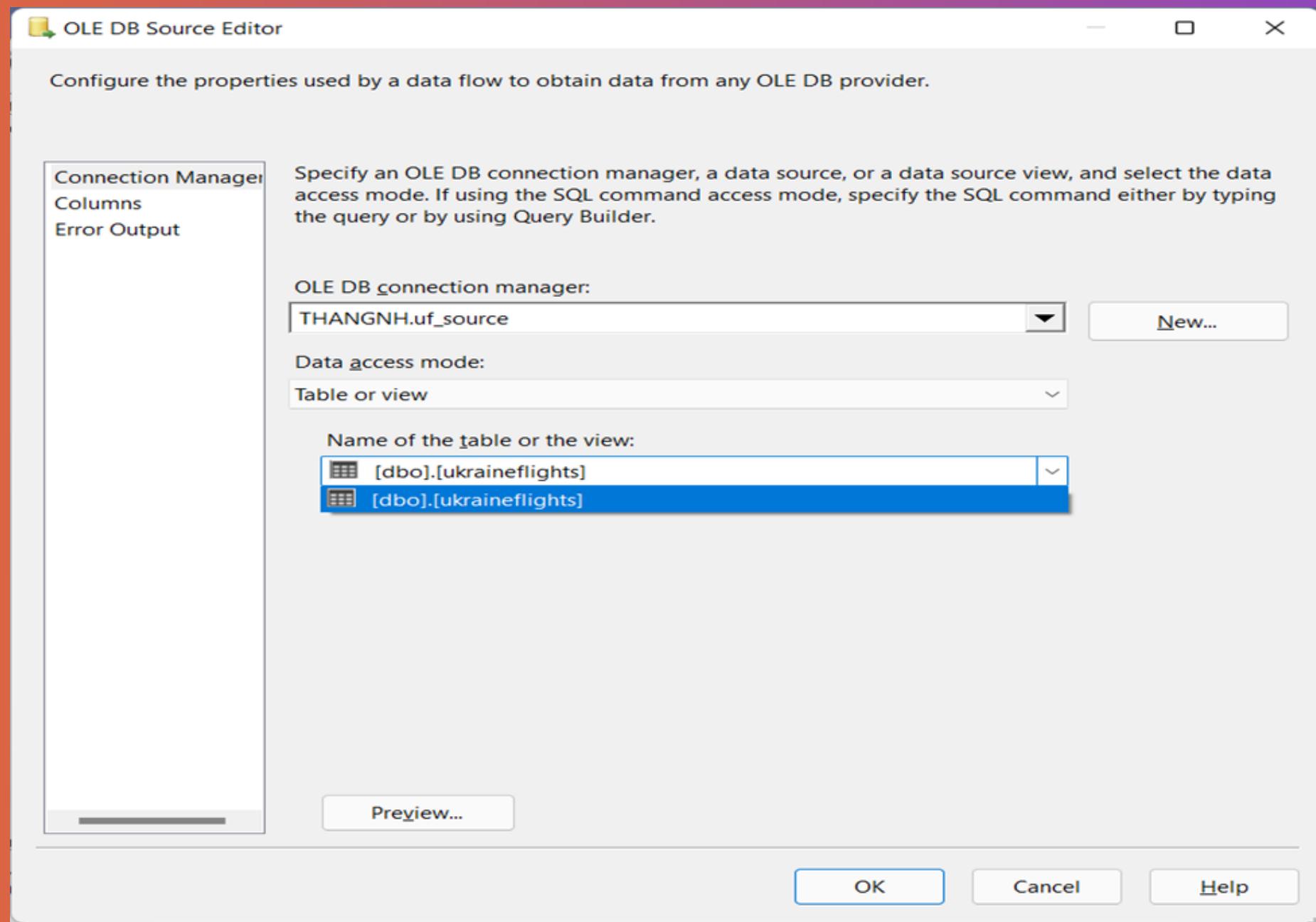
```
***** Script for SelectTopNRows command *****
SELECT TOP (1000) [icao24]
,[callsign]
,[origin_country]
,[time_position]
,[last_contact]
,[longitude]
,[latitude]
,[barometric_altitude]
,[on ground]
```
- Results** tab (Bottom): Shows the output of the query:| icao24 | callsign | origin_country | time_position | last_contact | longitude | latitude | barometric_altitude | on ground |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 4248ef | SVR522 | United Kingdom | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 37.7853012084961 | 5 | 25.6658992767334 | 4 |
| 2 4718c | WIZ21550 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 21.2425994873047 | 4 | 29.9470996856689 | 4 |
| 3 47189 | WZZ22276 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 22.2772998809814 | 4 | 23.5426998138428 | 4 |
| 4 425E+09 | 8DM6413 | United Kingdom | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 45.0040016174316 | 5 | 21.2311992645264 | 4 |
| 5 47192 | WZZ21166 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 21.5622997283936 | 4 | 24.4829998016357 | 4 |
| 6 47157 | WZZ24603 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 29.9470996856689 | 4 | 22.6998996734619 | 5 |
| 7 452086 | GBG672 | Bulgaria | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 26.1558990478516 | 4 | 21.7485008239746 | 4 |
| 8 47151 | WZZ1202 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 21.2311992645264 | 4 | 30.0547008514404 | 5 |
| 9 47193c | WZZ1575 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 24.4829998016357 | 4 | 22.6998996734619 | 5 |
| 10 4719b | WZZ3095 | Hungary | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 25.6658992767334 | 4 | 22.6998996734619 | 5 |
| 11 407b55 | WUK4486 | United Kingdom | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 22.6998996734619 | 5 | 26.1558990478516 | 4 |
| 12 a60693 | BRI068 | United States | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 21.7485008239746 | 4 | 29.9470996856689 | 4 |
| 13 505ed8 | IAW234 | Slovakia | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 21.2311992645264 | 4 | 22.6998996734619 | 5 |
| 14 463afe | JEF22 | Finland | 2022-02-22 15:23:00.0000000 | 2022-02-22 15:23:00.0000000 | 30.0547008514404 | 5 | 22.6998996734619 | 5 |
| 15 3c4aab | CFG2321 | Germany | 2022-02-22 15:28:00.0000000 | 2022-02-22 15:28:00.0000000 | 29.0851001739502 | 4 | 22.6998996734619 | 5 |
- Messages** tab (Bottom):

Query executed successfully.

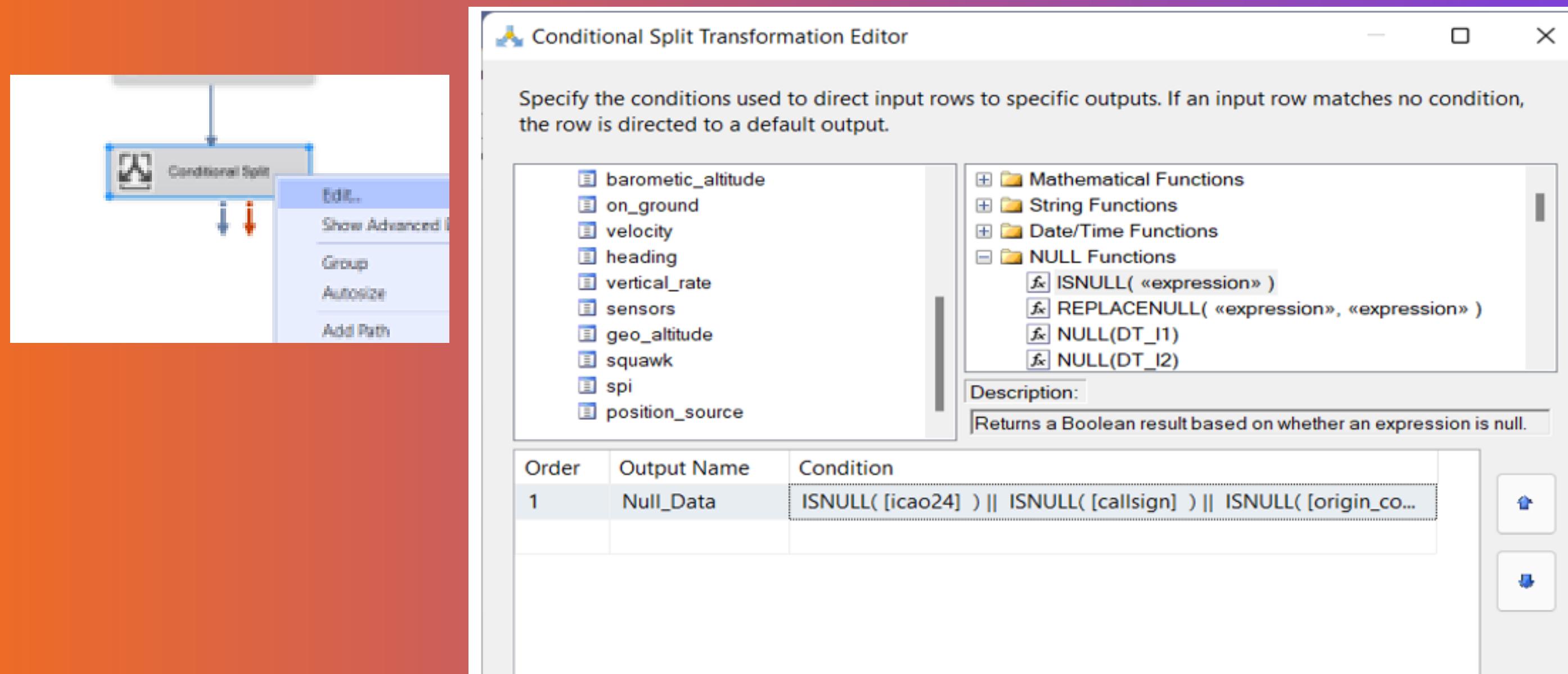
Thành công

3. Quá trình làm sạch dữ liệu





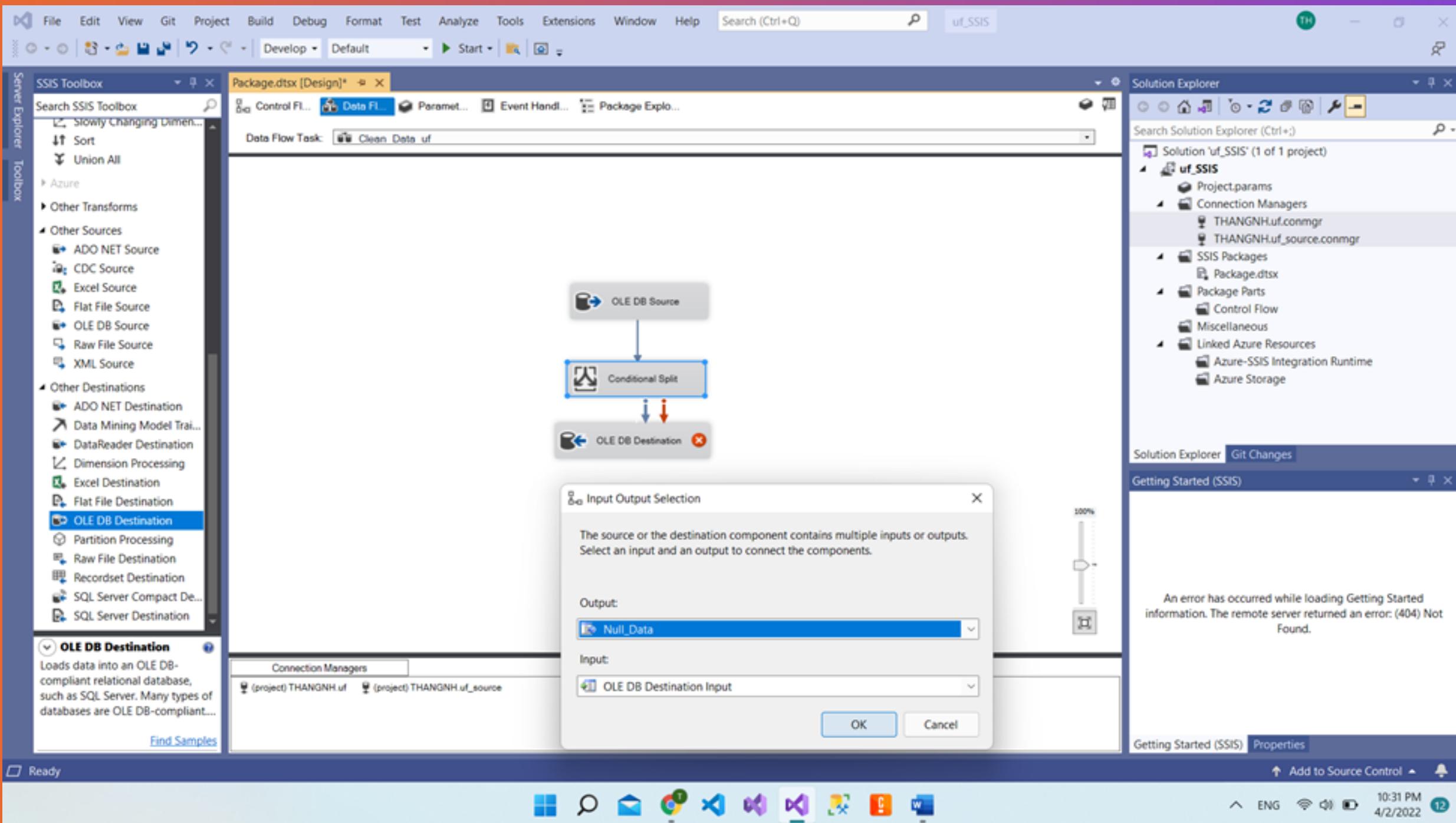
Sử dụng kết nối tới bảng “ukraineflights” trong database uf_source



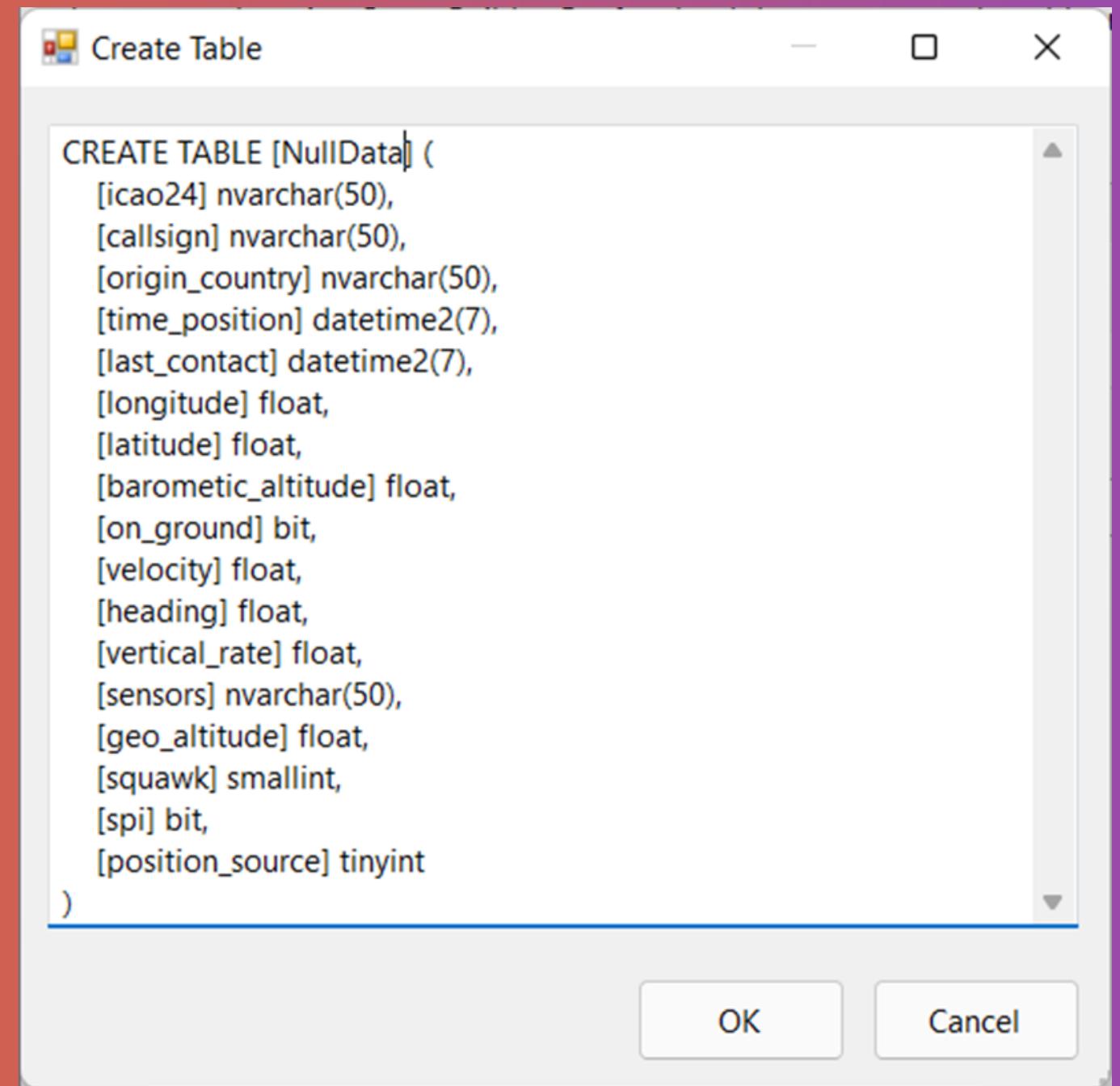
Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.

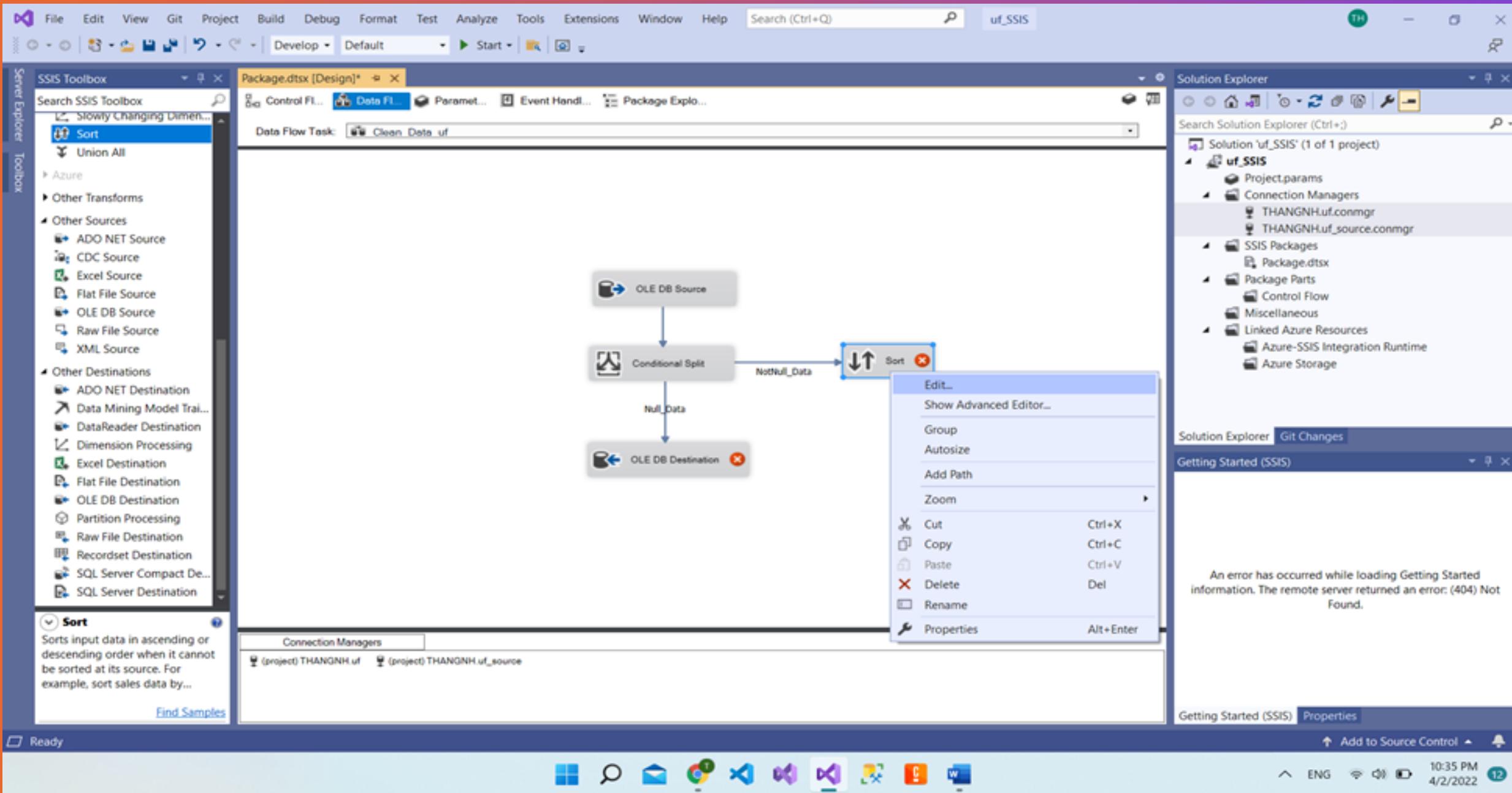
Order	Output Name	Condition
1	Null_Data	ISNULL([icao24]) ISNULL([callsign]) ISNULL([origin_co...])

Sử dụng hàm ISNULL (<<expression>>) để lọc giá trị NULL

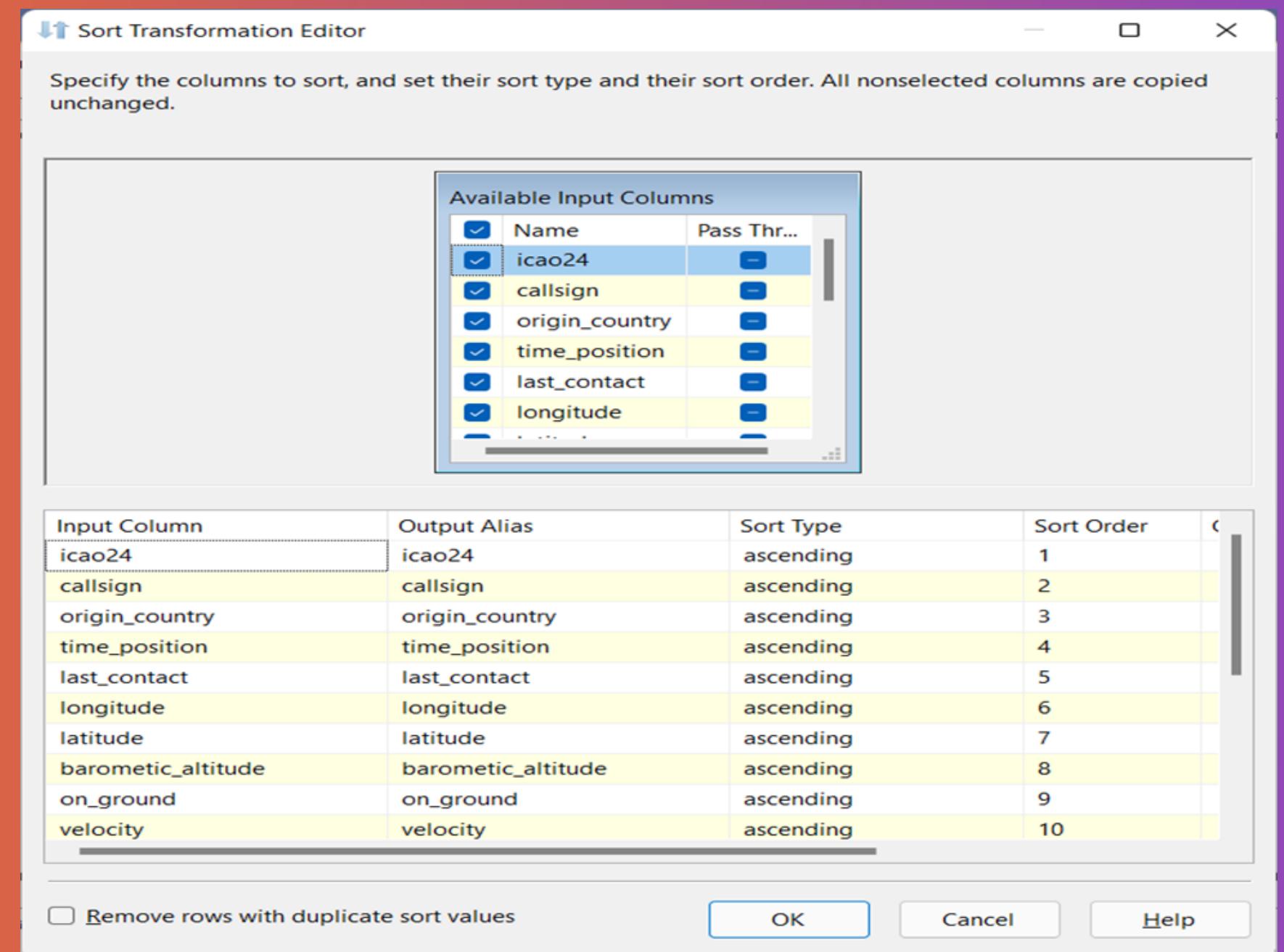


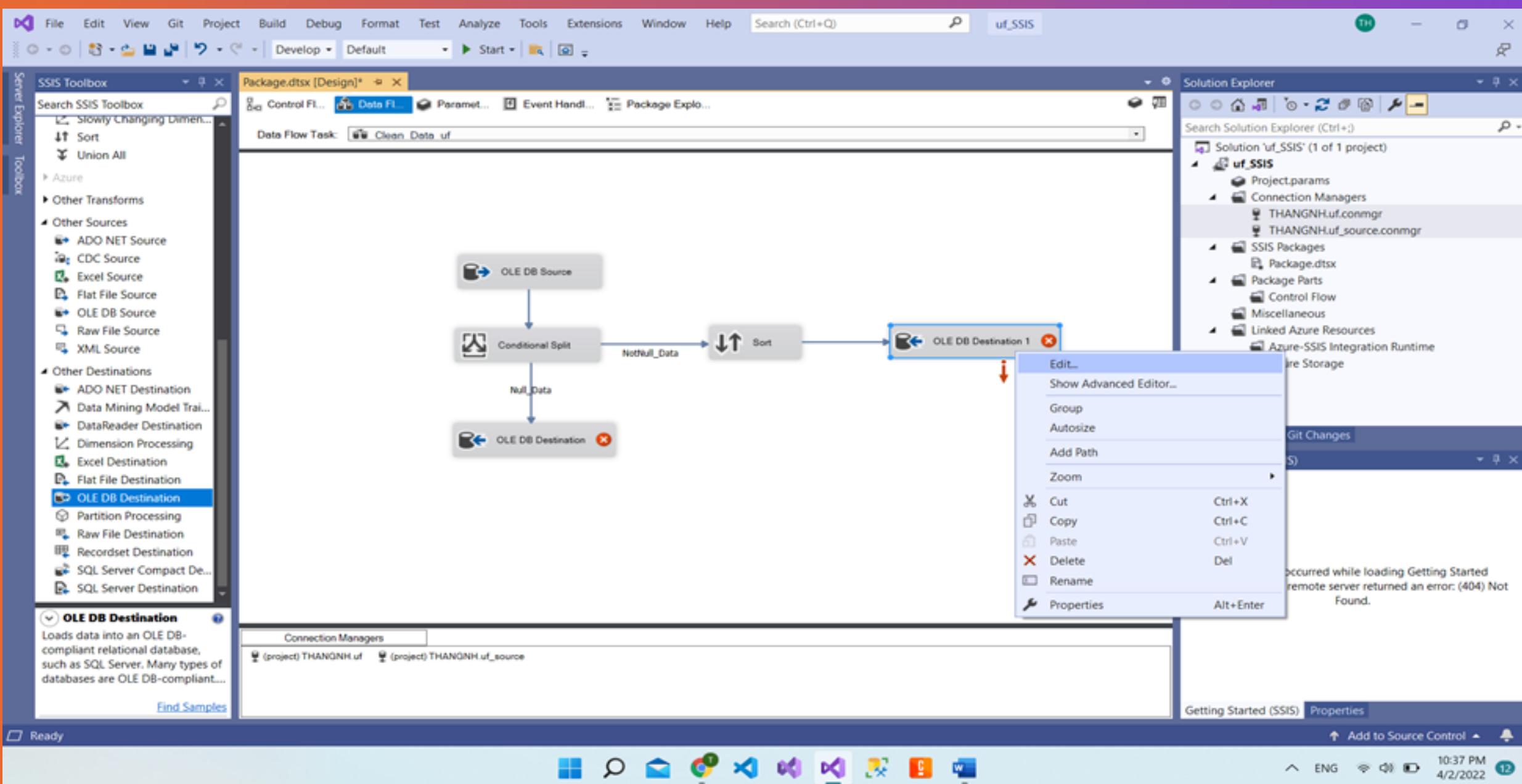
tạo OLE DB Destination chứa Null Data



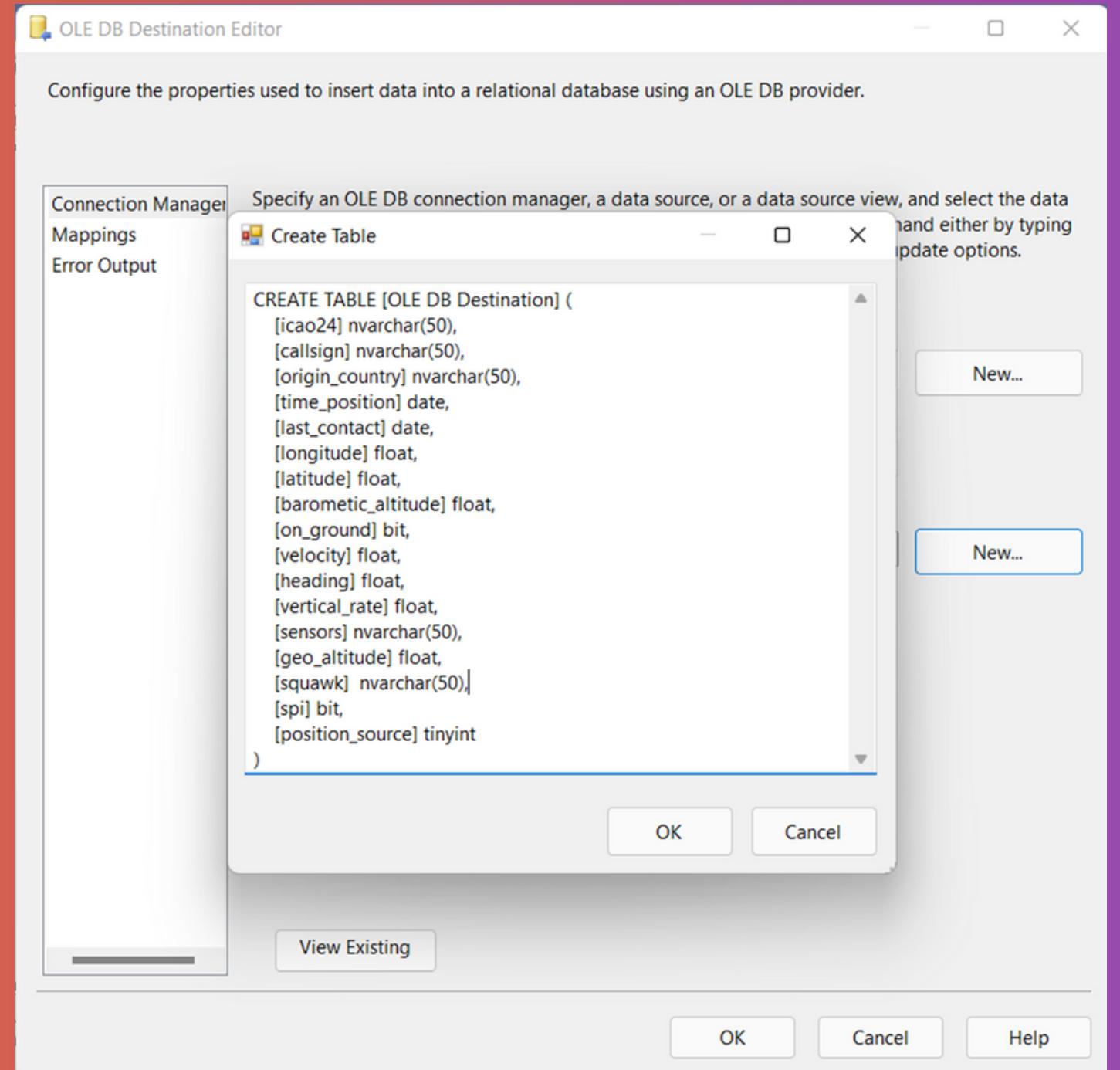


kéo thả và tạo Sort, nhận đầu vào là NotNull_Output từ Conditional Split



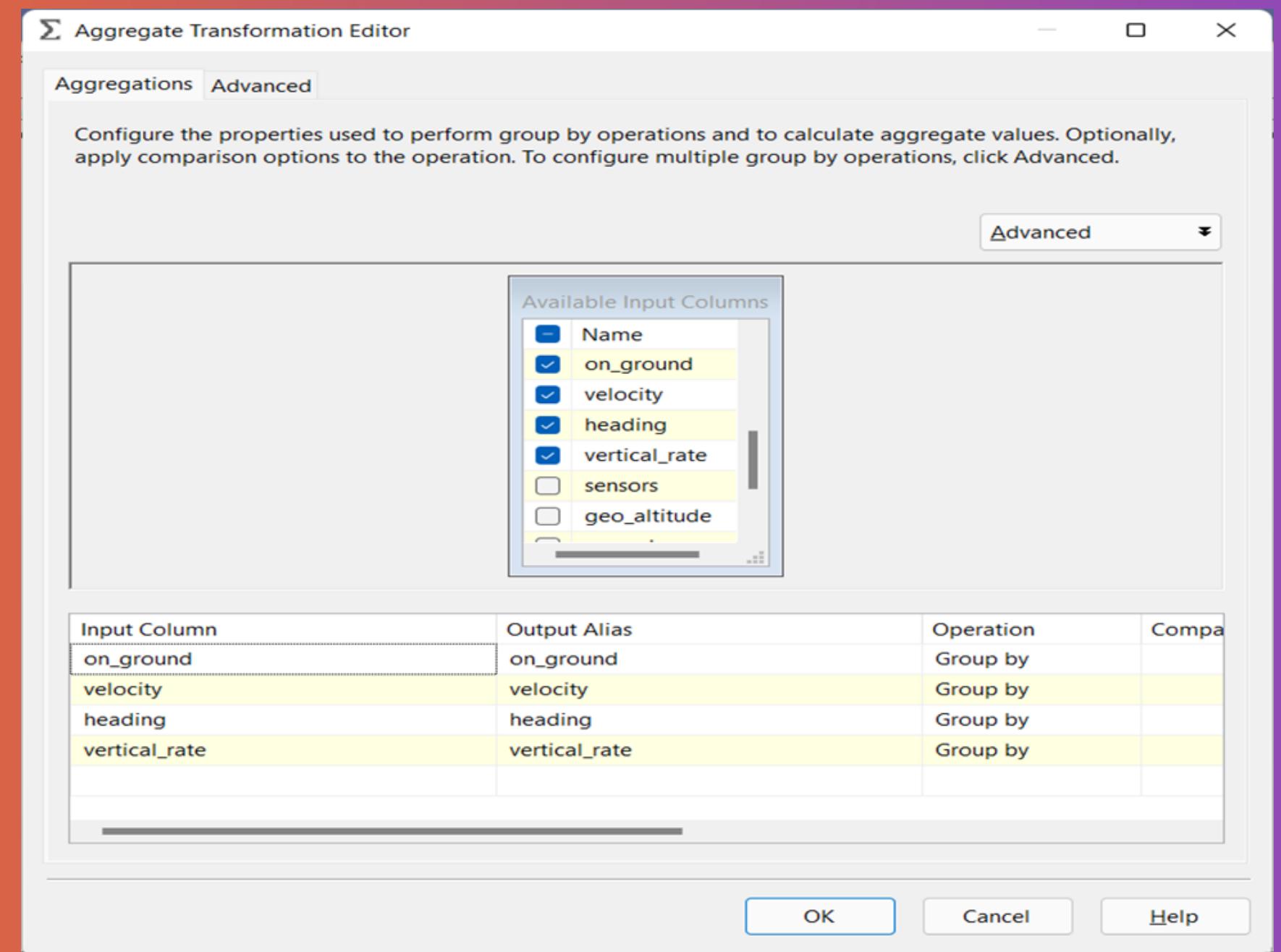


Tạo mới OLE DB Destination

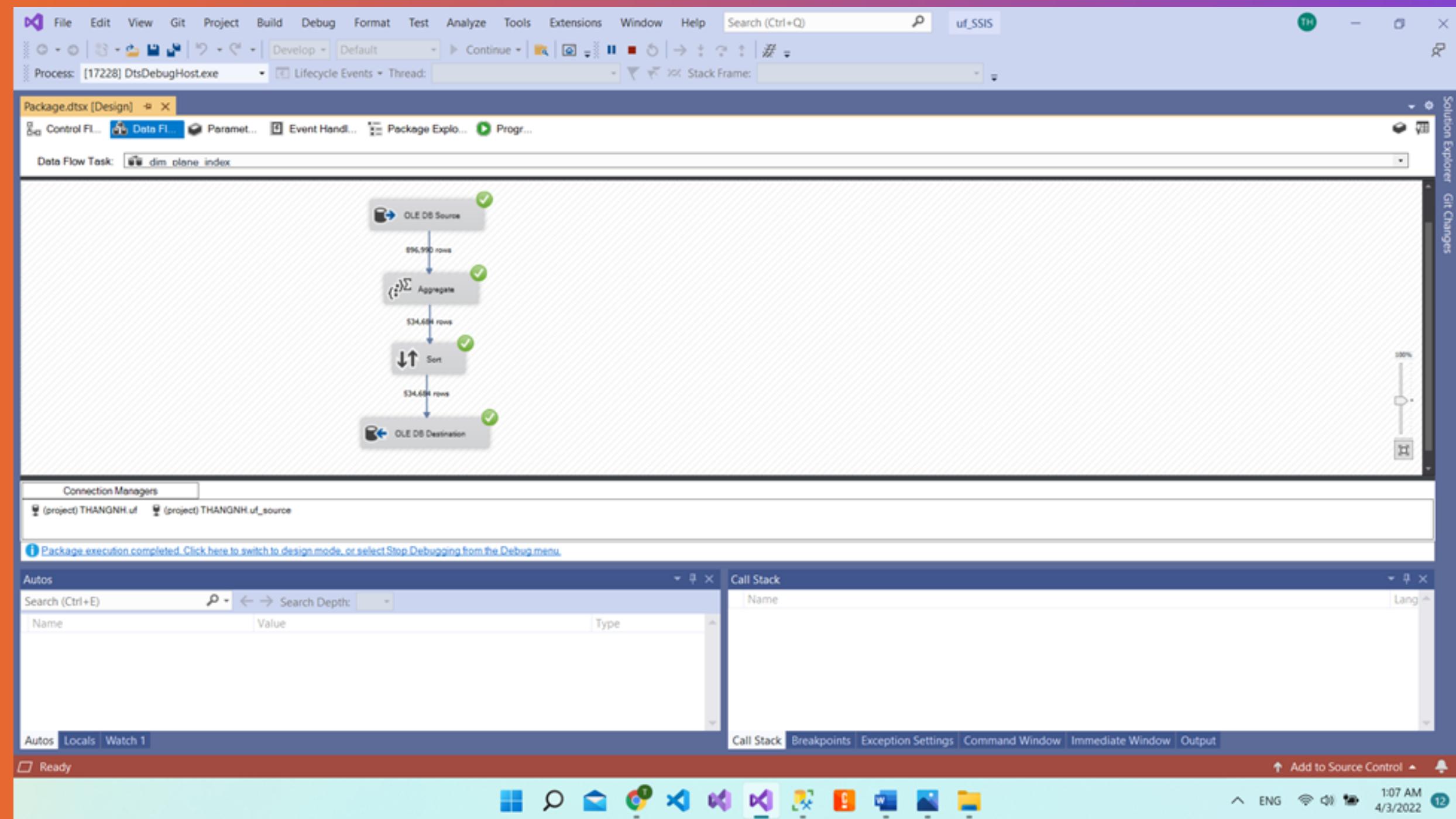


tạo mới bảng CleanData chứa dữ liệu hoàn chỉnh sau khi làm sạch

4. Quá trình tạo các bảng Dimension



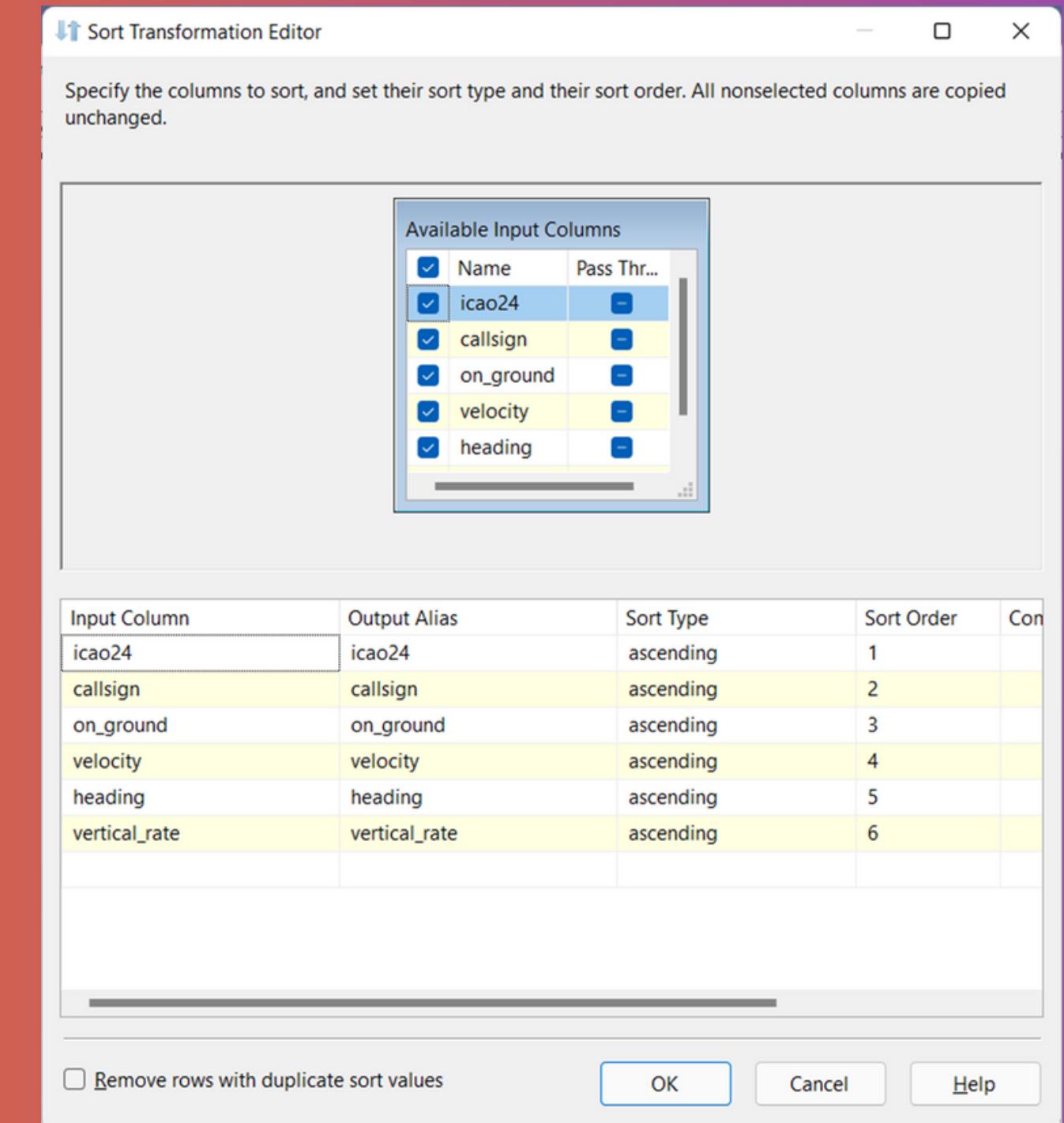
Chọn các thuộc tính tạo bảng



Tạo bảng dim_plane_index

4.2 Tạo bảng dim_plane

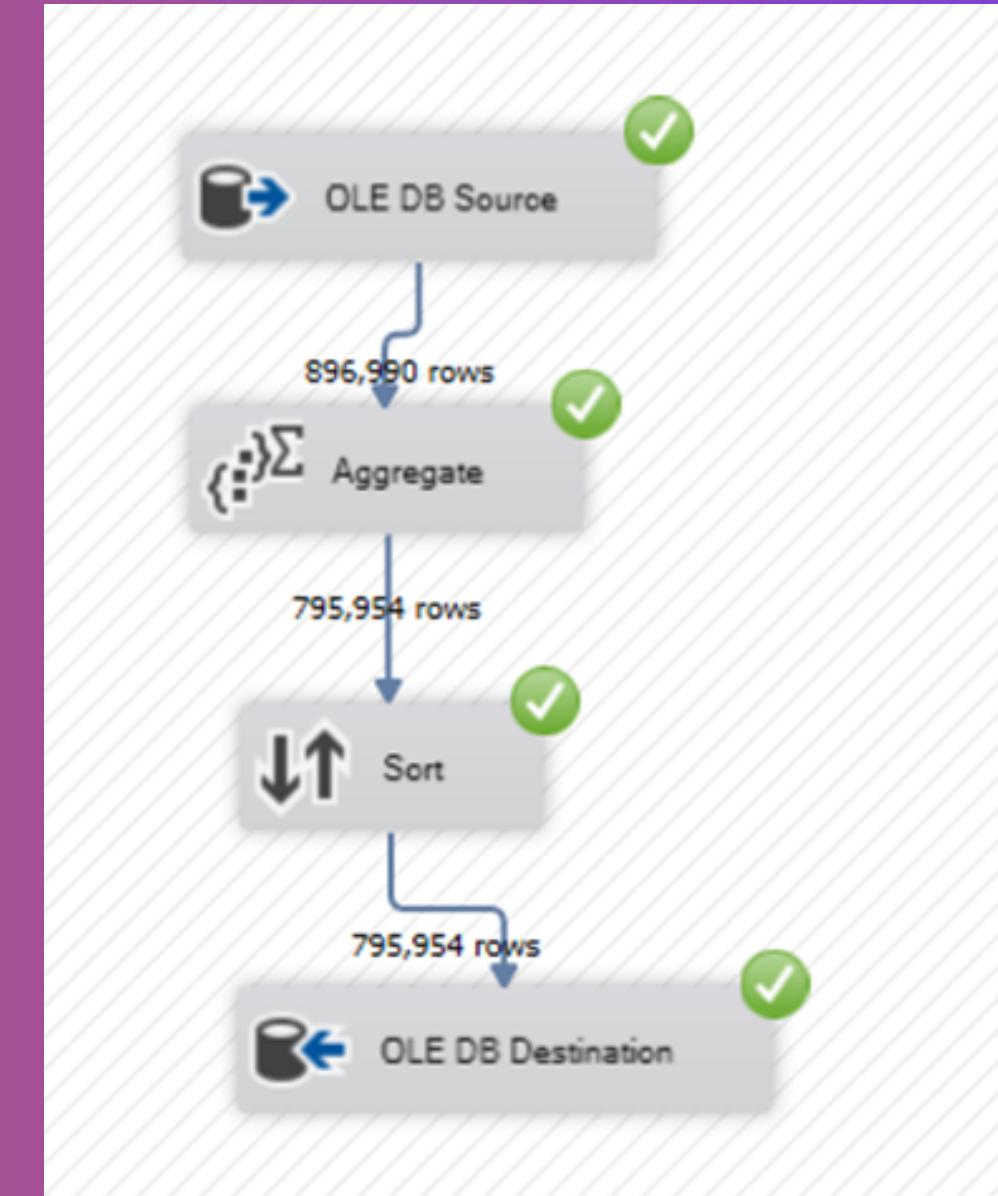
Để tạo bảng dim_plane, trước tiên ta phải tạo 1 bảng có các thuộc tính
ở cả bảng dim_plane_index và bảng dim_plane.



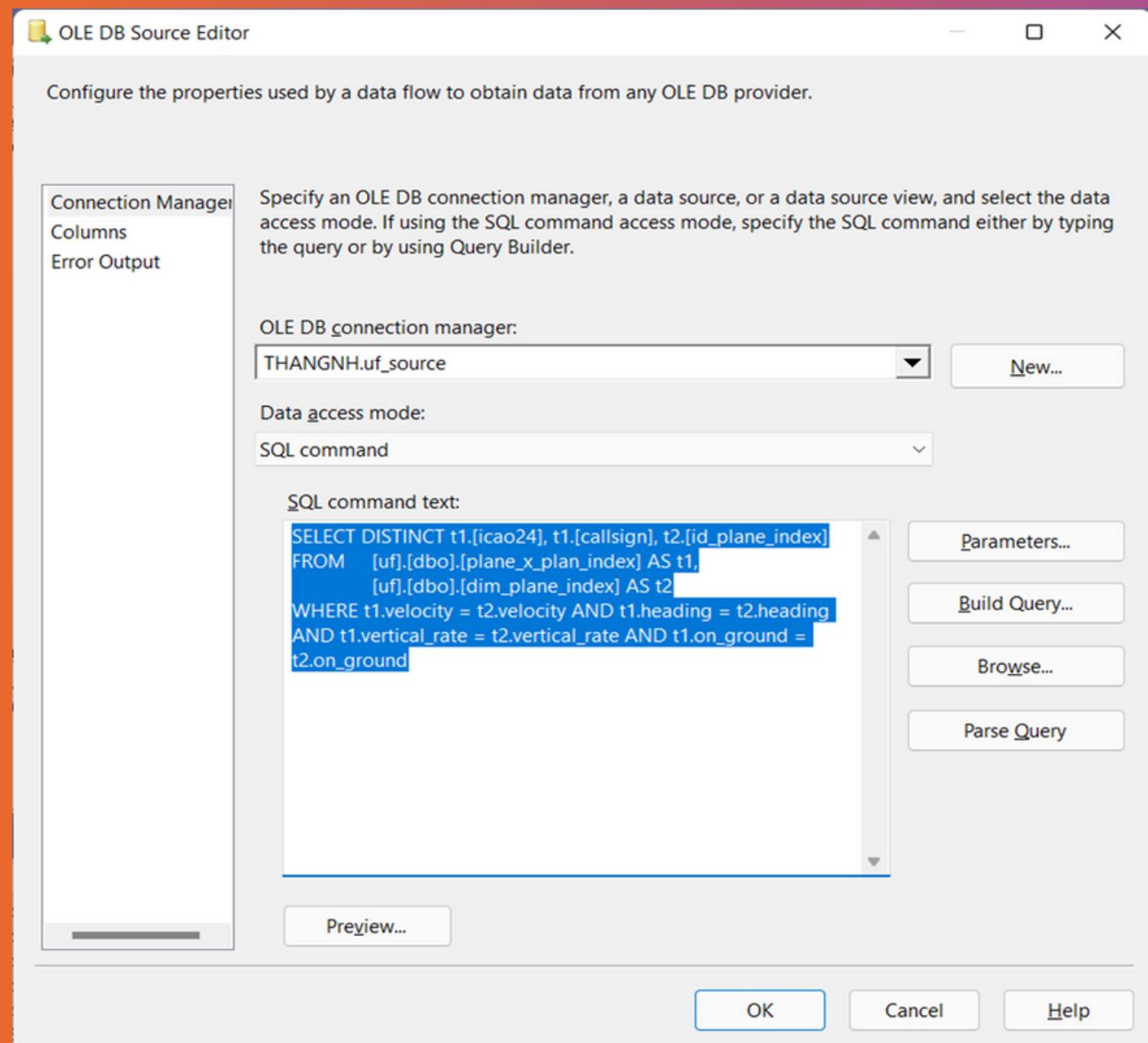
Create Table

```
CREATE TABLE [plan_x_plan_index] (
    [id_plane_index] int primary key identity(1,1),
    [velocity] float,
    [heading] float,
    [vertical_rate] float,
    [on_ground] bit,
    [icao24] nvarchar(50),
    [callsign] nvarchar(50),
)
```

OK Cancel

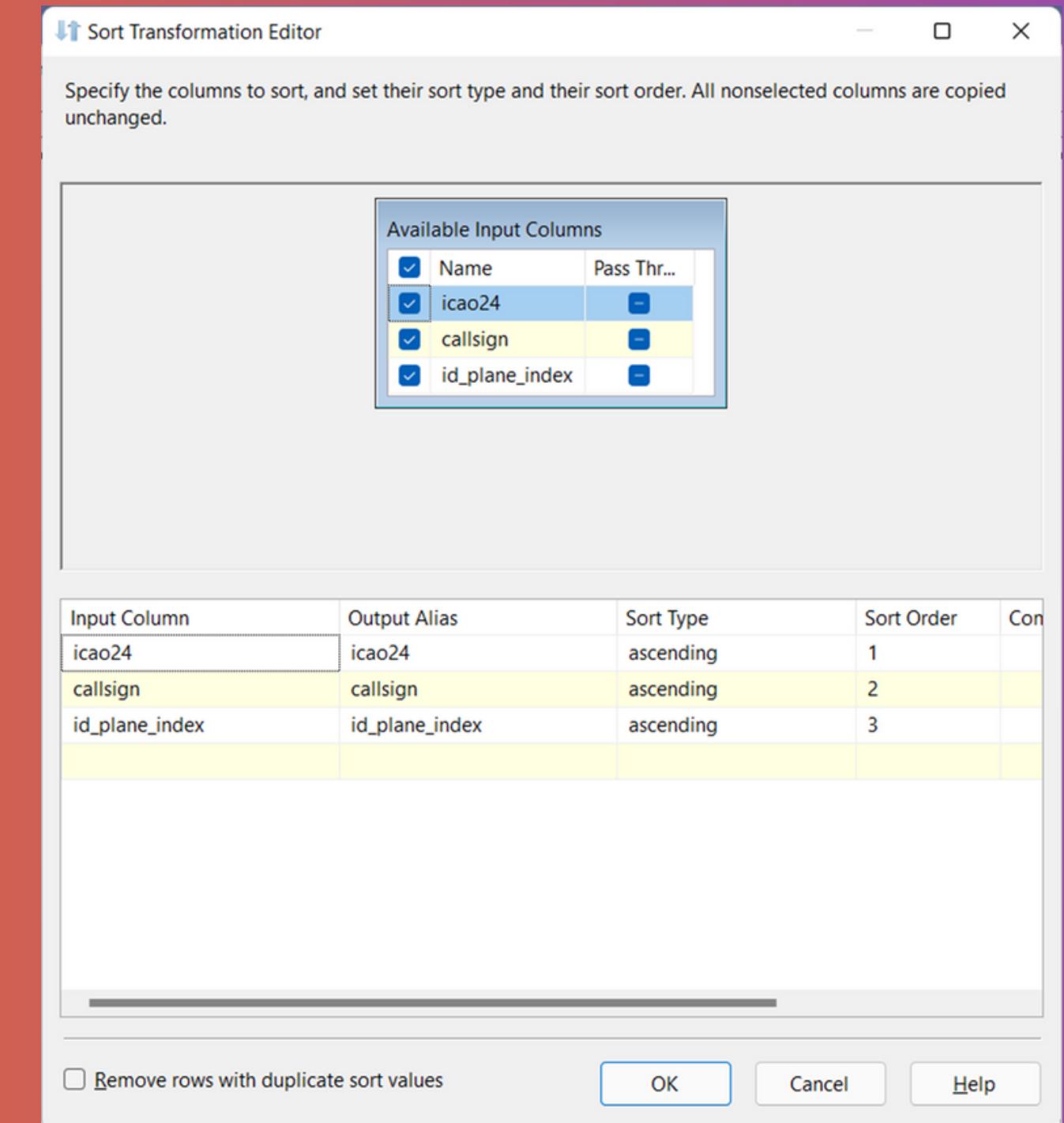


bảng dim_plane_index và bảng dim_plane



chọn SQL Command tại data access mode để dùng SQL lấy dữ liệu

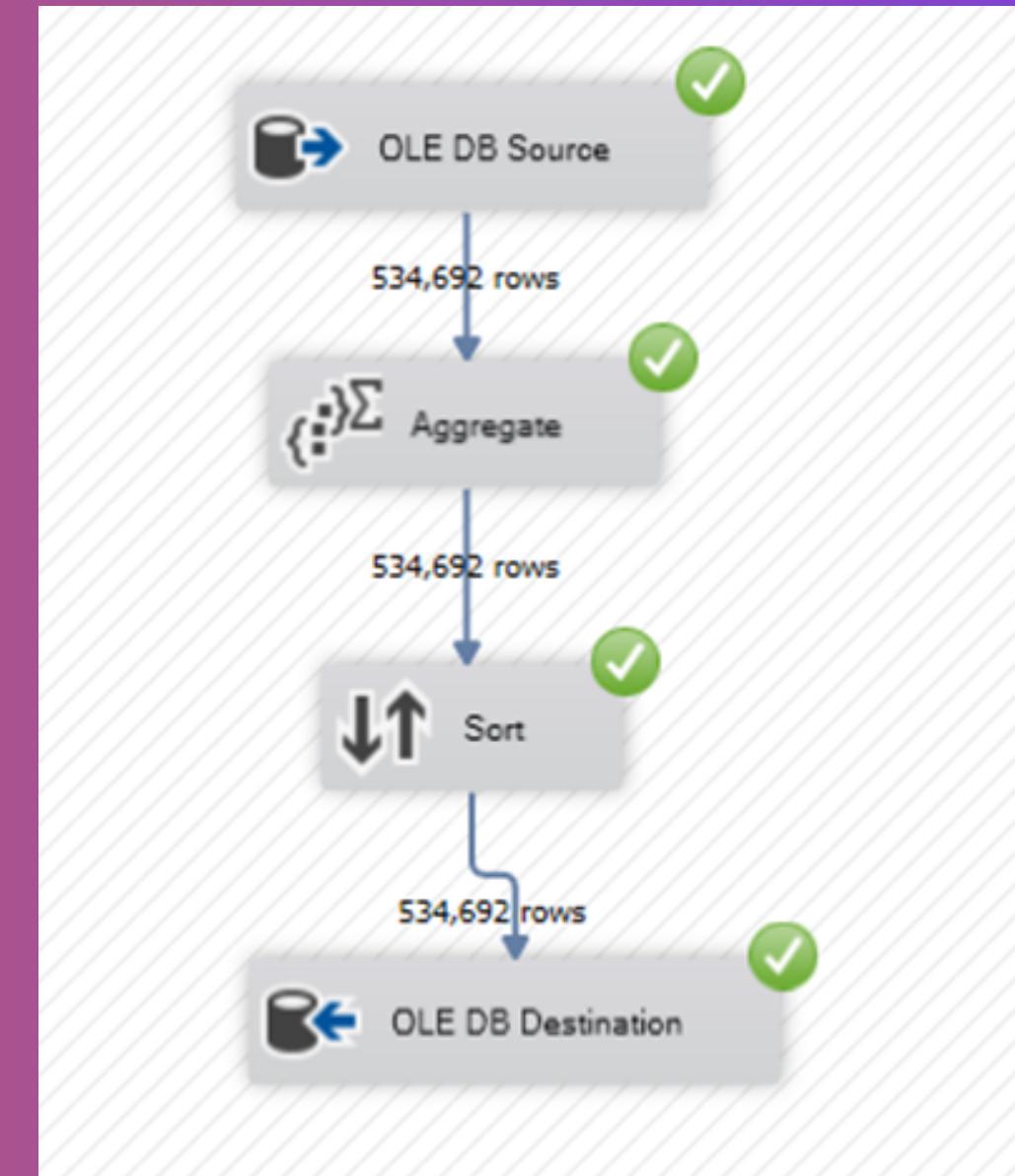
tạo bảng dim_plane



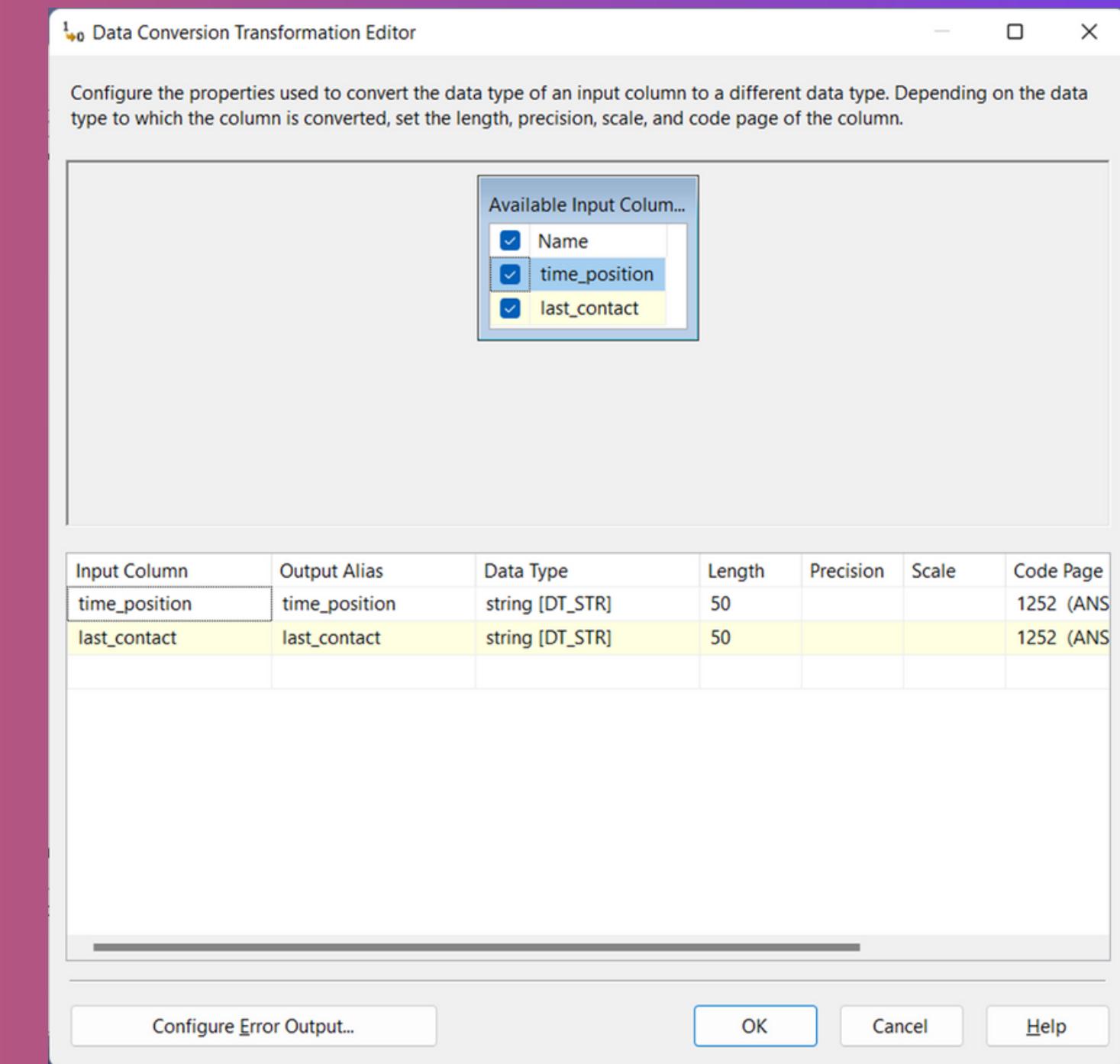
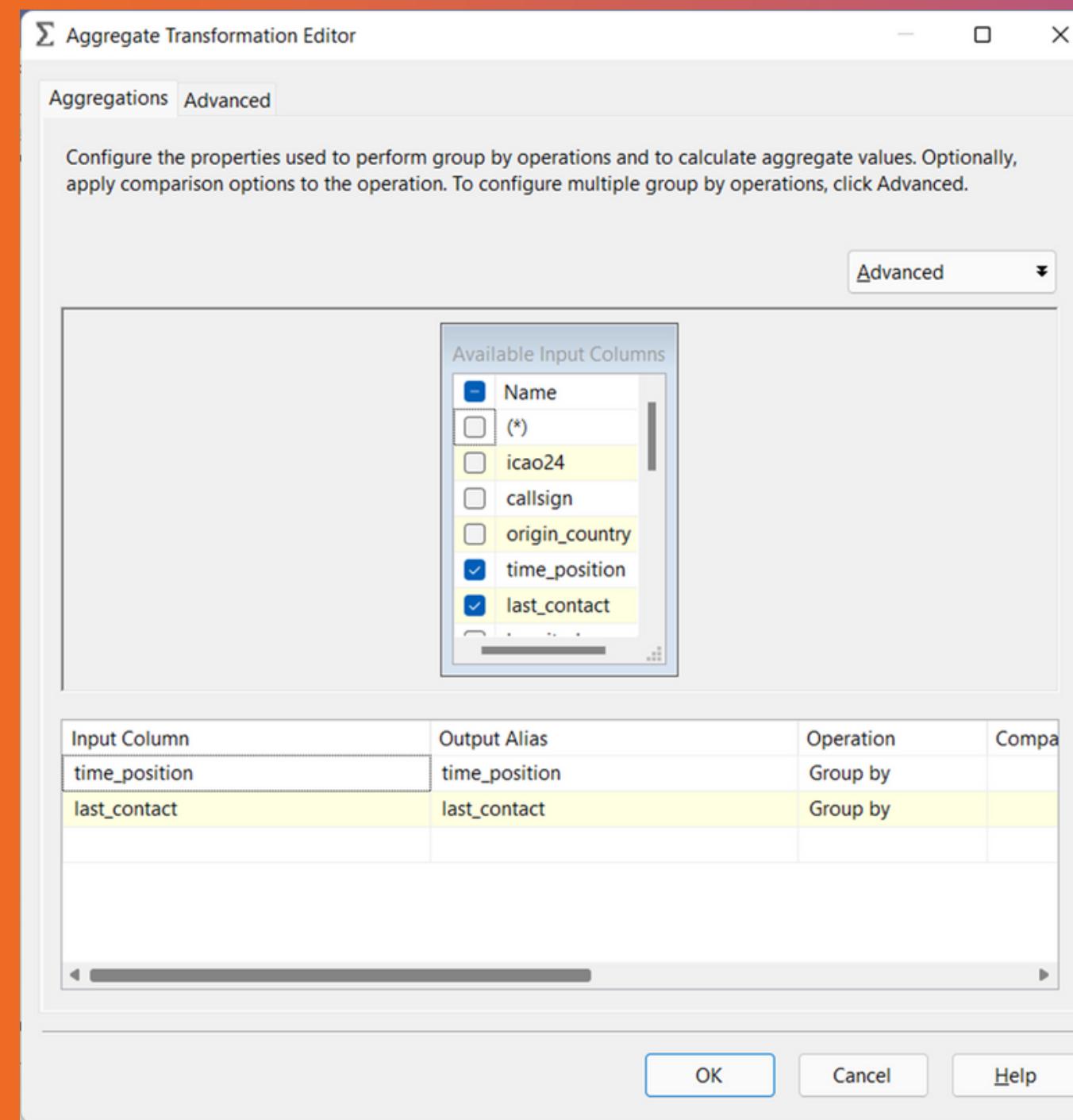
Create Table

```
CREATE TABLE [dim_plane] (
    [id_plane] int primary key identity(1,1),
    [icao24] nvarchar(50),
    [callsign] nvarchar(50),
    [id_plane_index] int
)
```

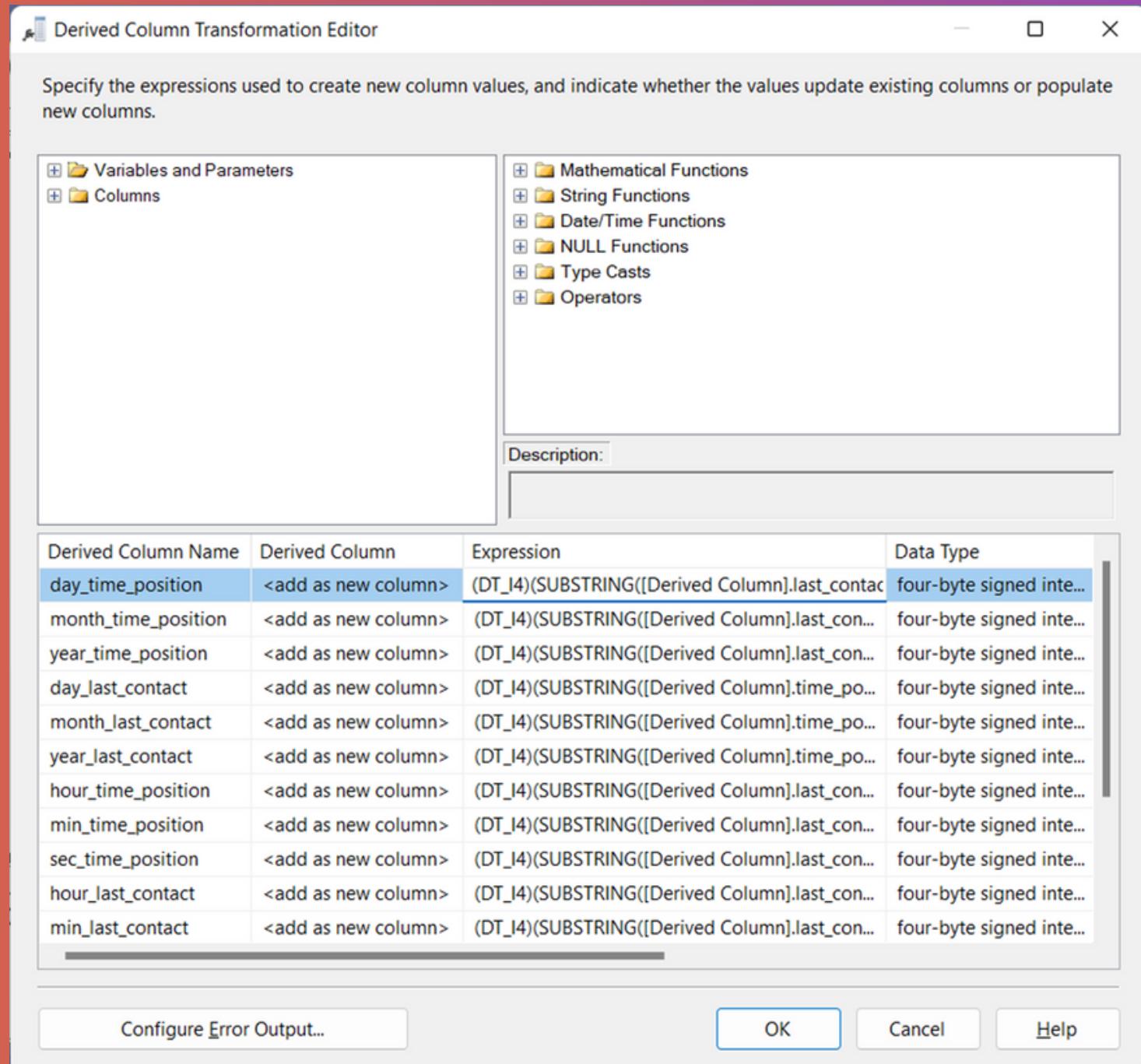
OK Cancel



Tạo bảng dim_date_time



Chuyển sang string

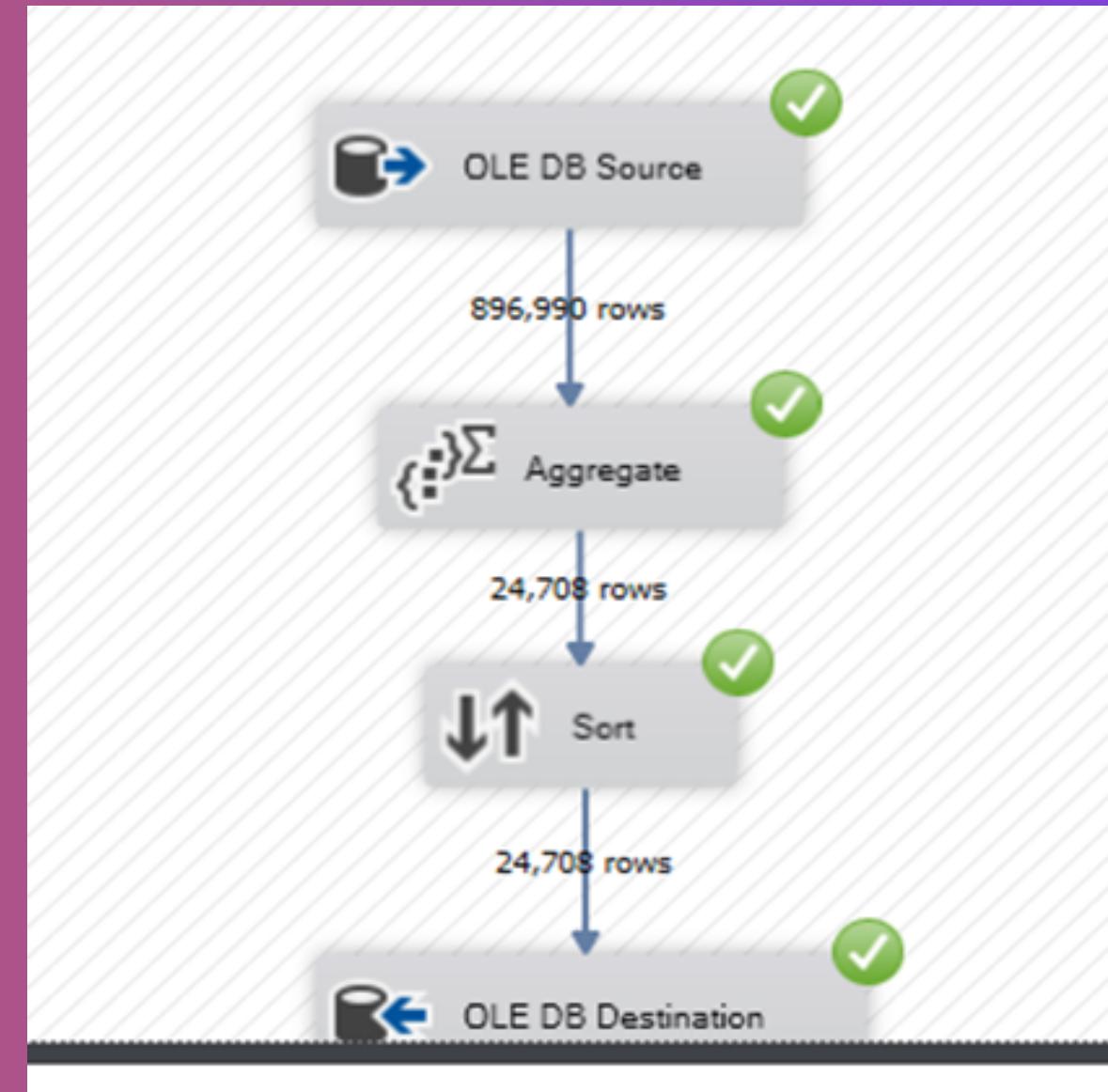


Dùng hàm SubString để cắt và lấy dữ liệu

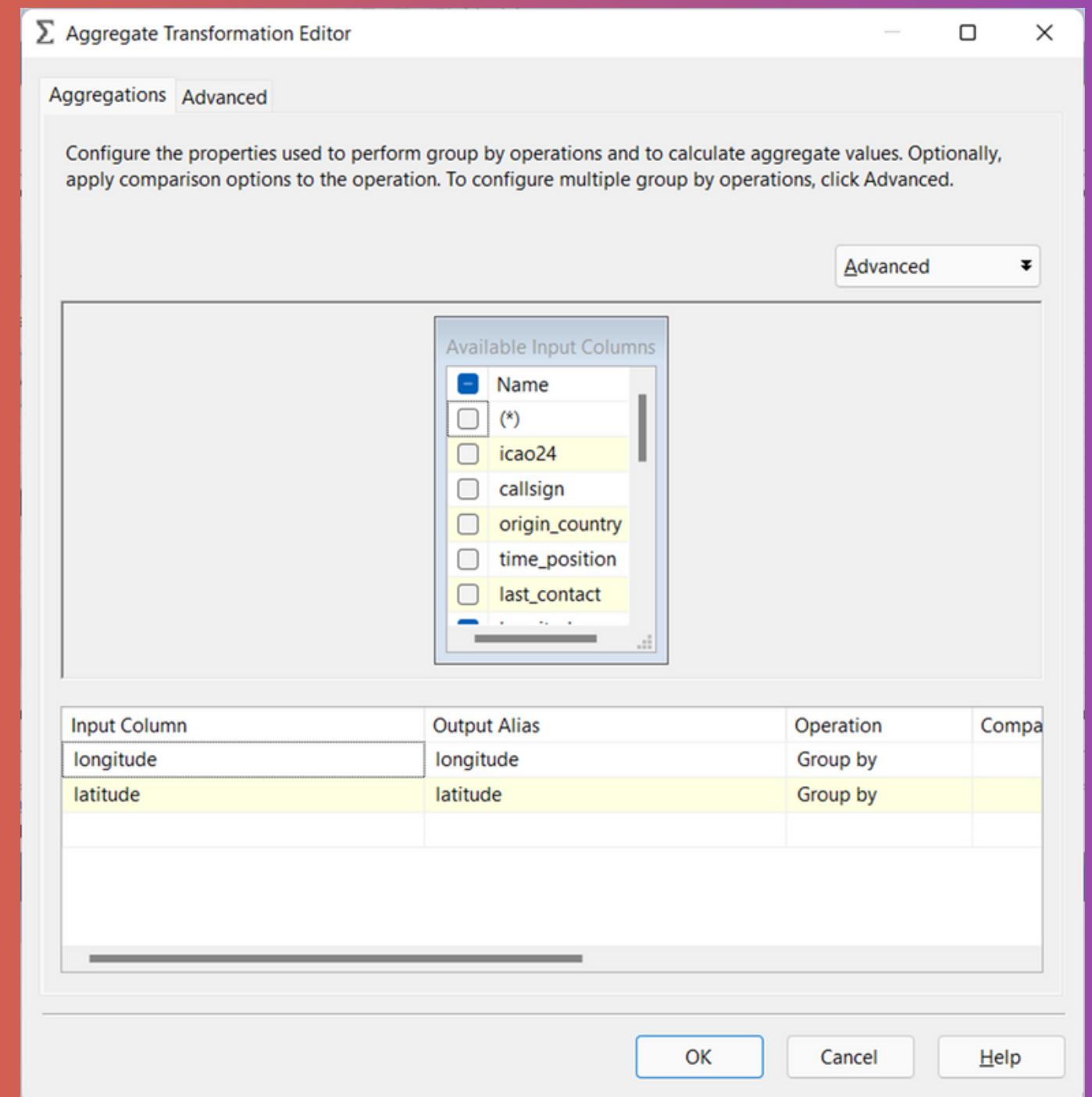
Create Table

```
CREATE TABLE [dim_datetime] (
    [id_date_time] int primary key identity(1,1),
    [day_time_position] int,
    [month_time_position] int,
    [year_time_position] int,
    [day_last_contact] int,
    [month_last_contact] int,
    [year_last_contact] int,
    [hour_time_position] int,
    [min_time_position] int,
    [sec_time_position] int,
    [hour_last_contact] int,
    [min_last_contact] int,
    [sec_last_contact] int,
    [last_contact] datetime2(0),
    [time_position] datetime2(0)
)
```

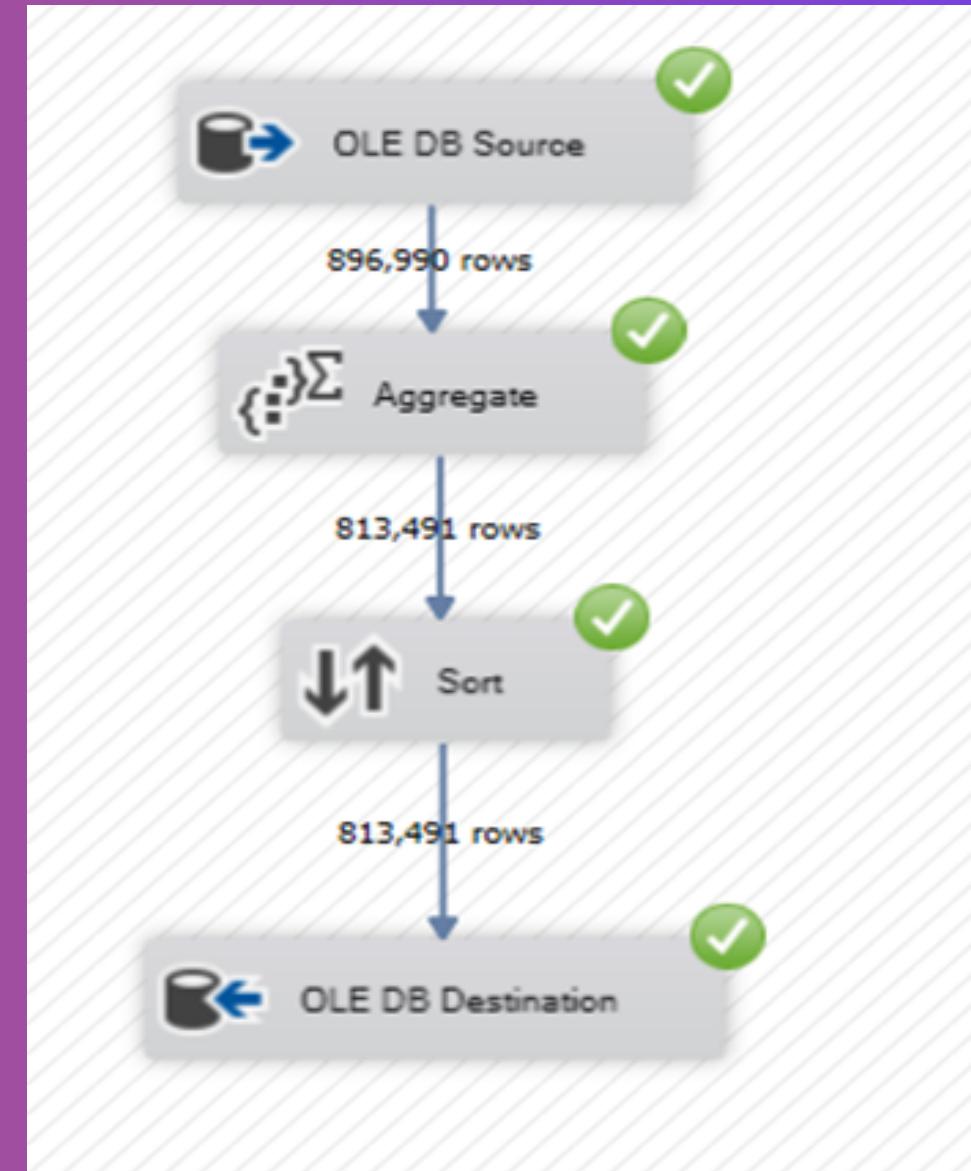
OK Cancel



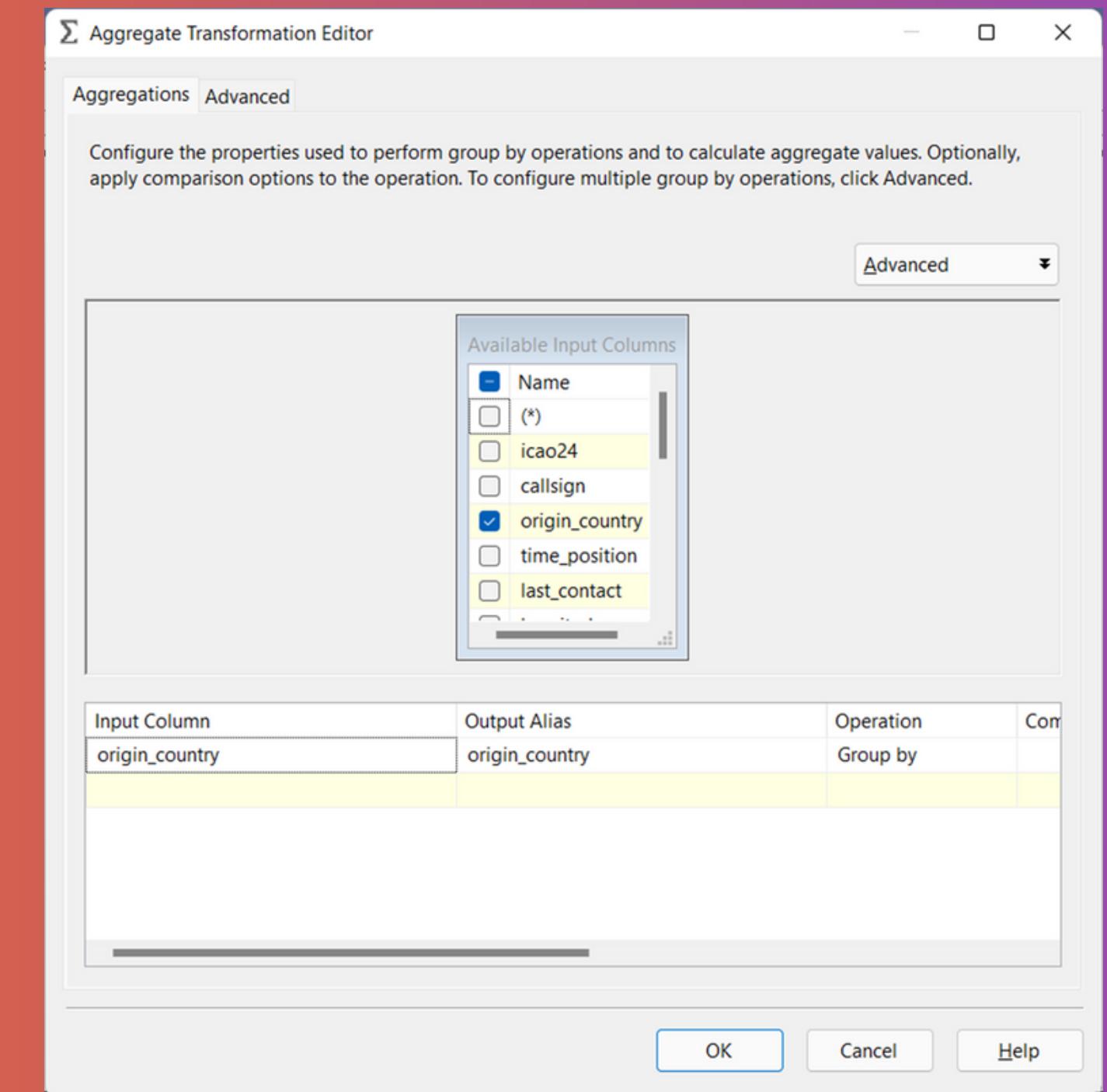
Tạo bảng dim_location

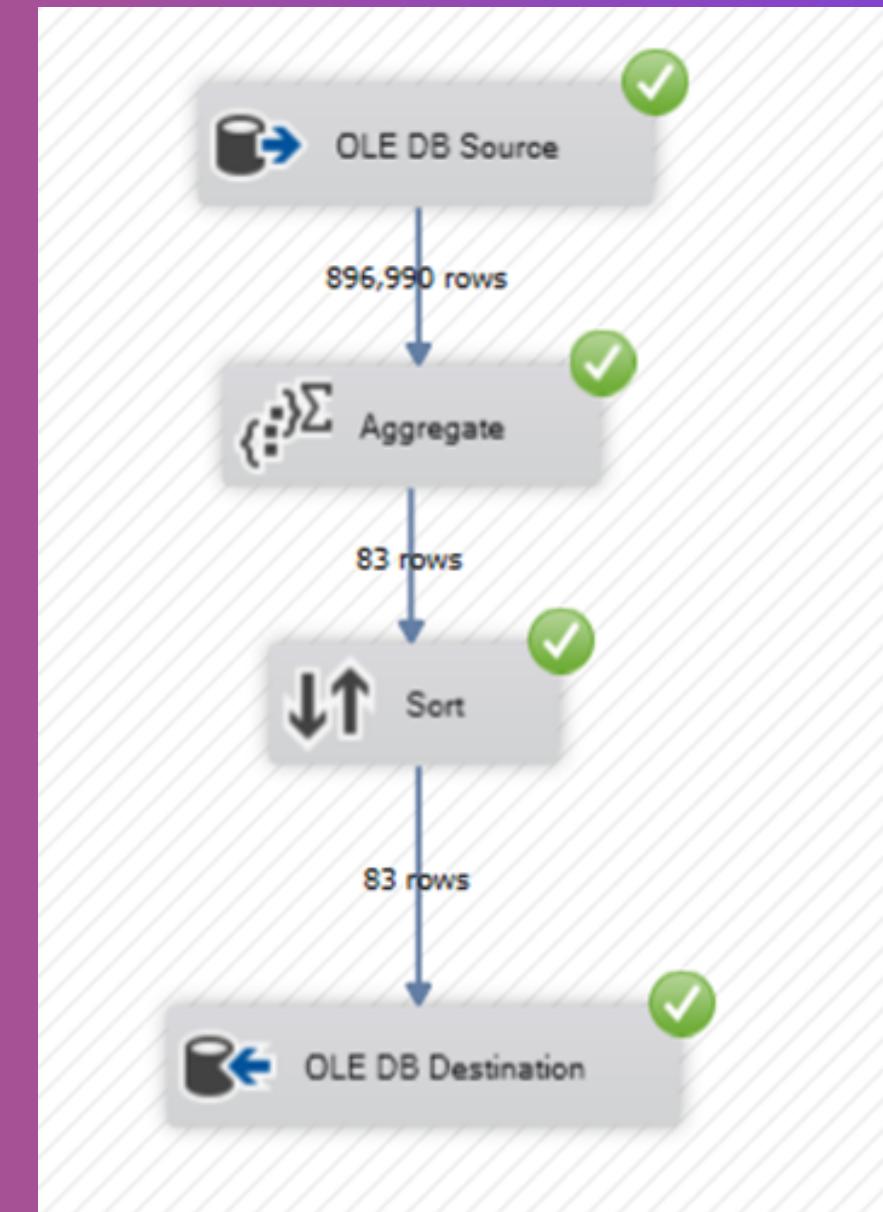
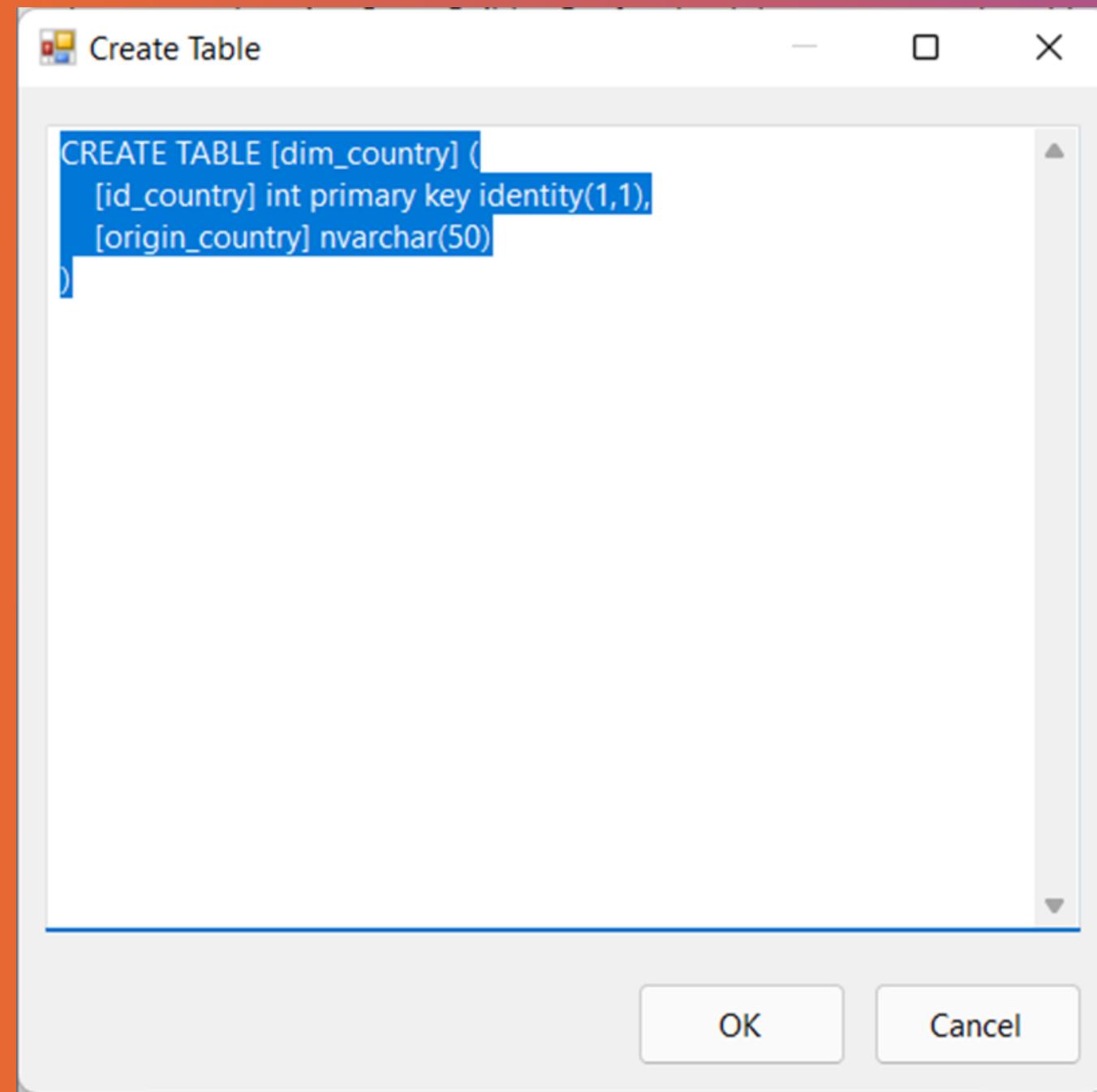


```
CREATE TABLE [dim_location] (
    [id_location] int primary key identity(1,1),
    [longitude] varchar(50),
    [latitude] varchar(50)
)
```

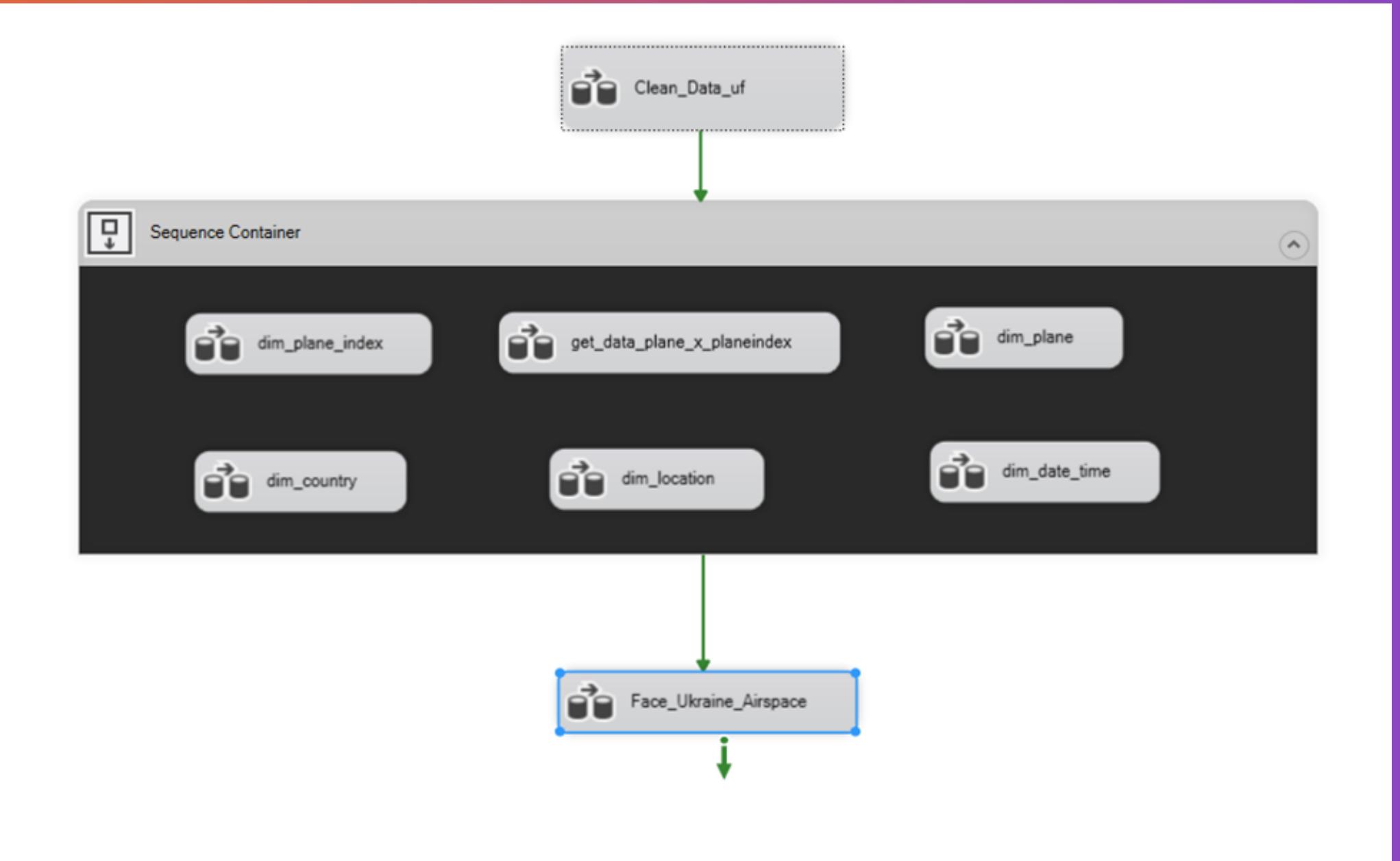


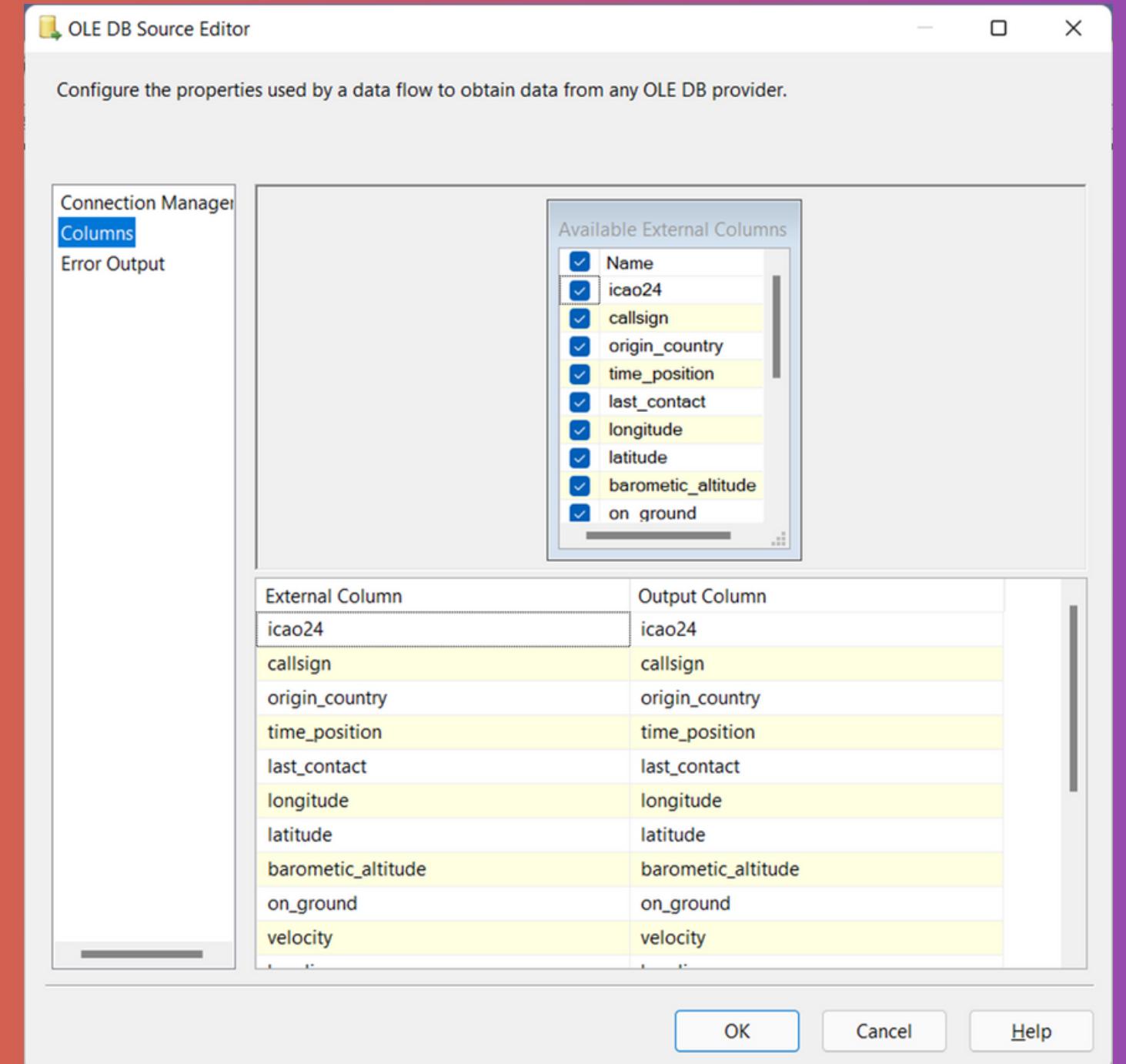
Tạo bảng dim_country

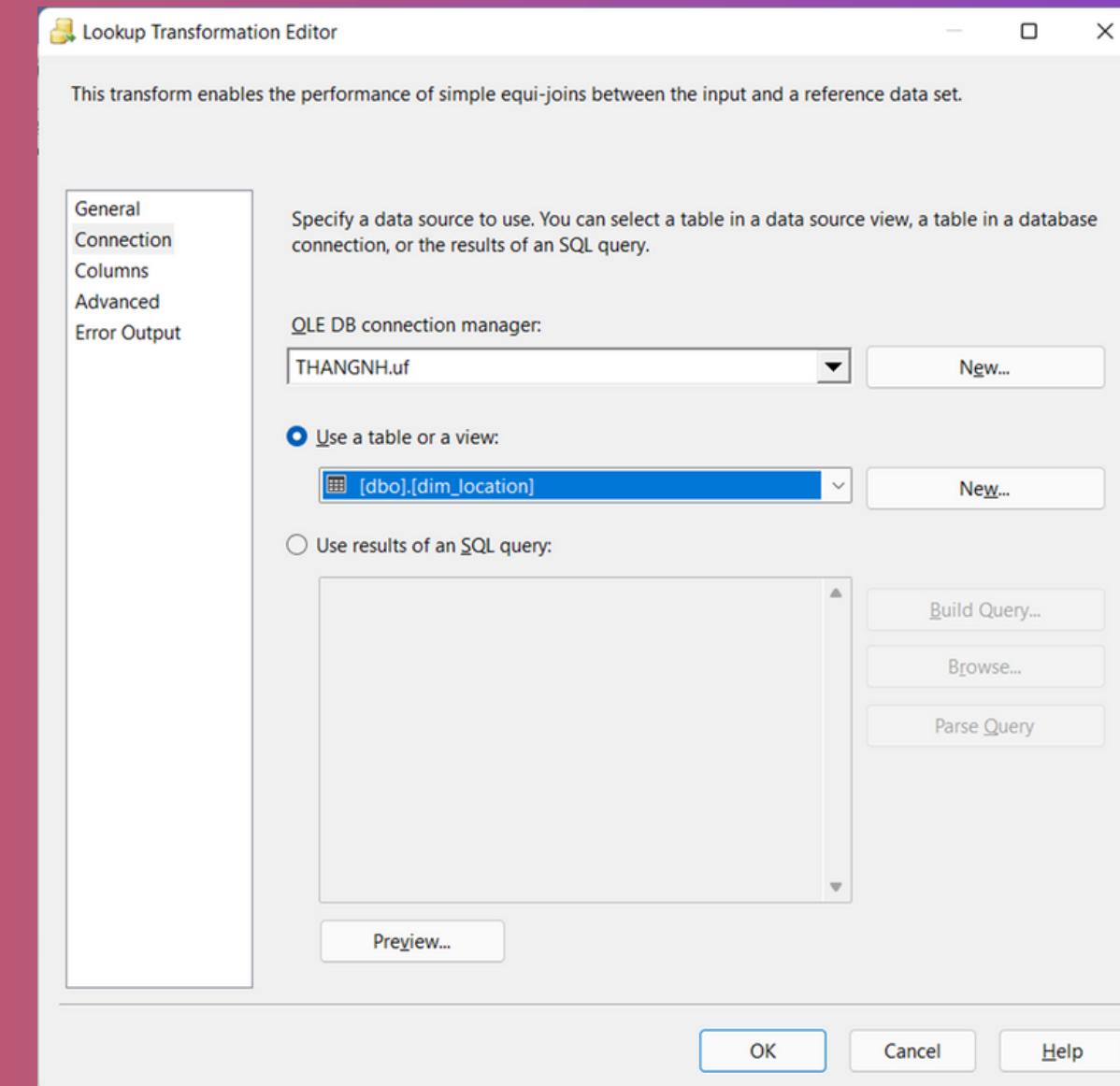
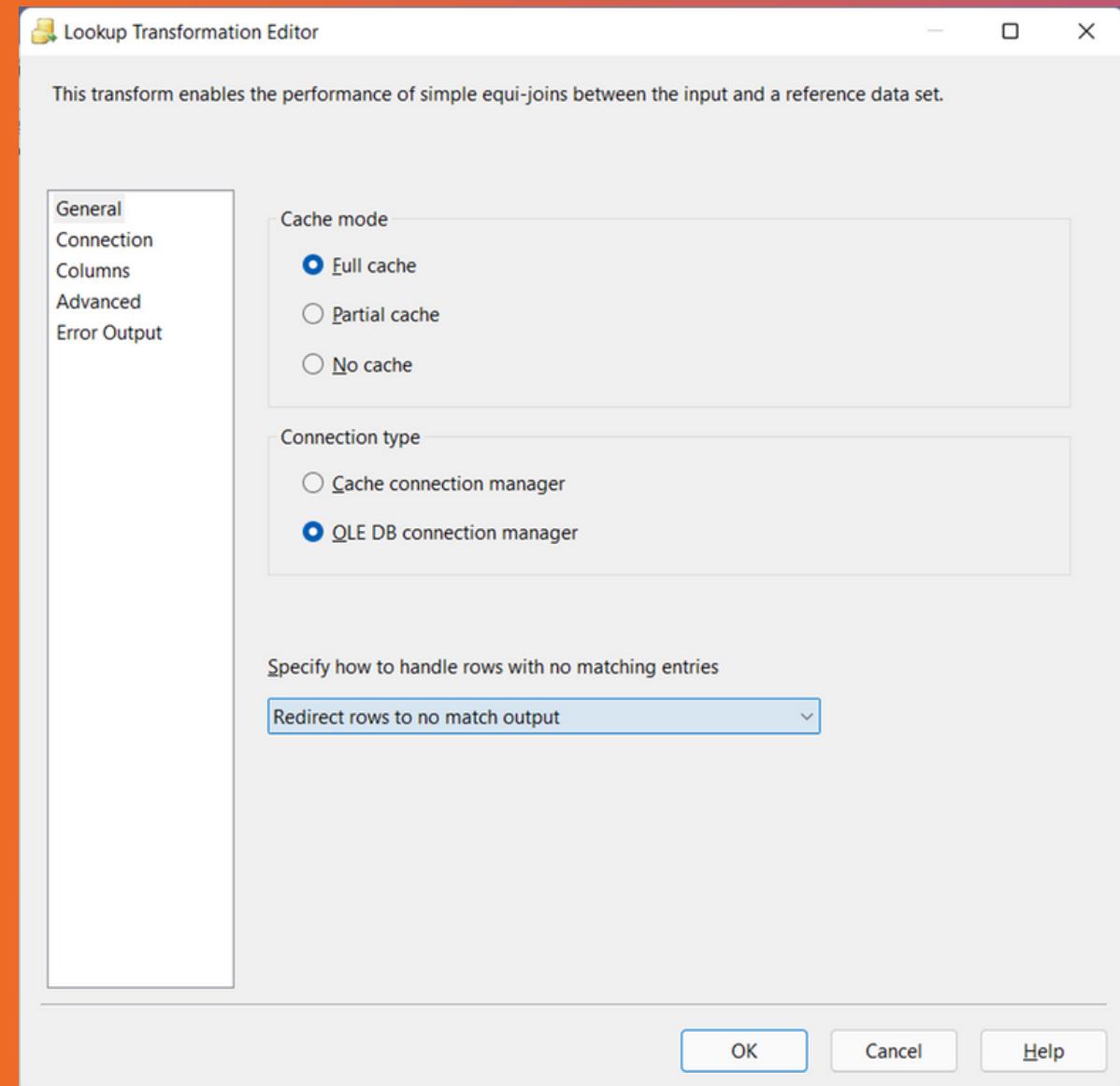


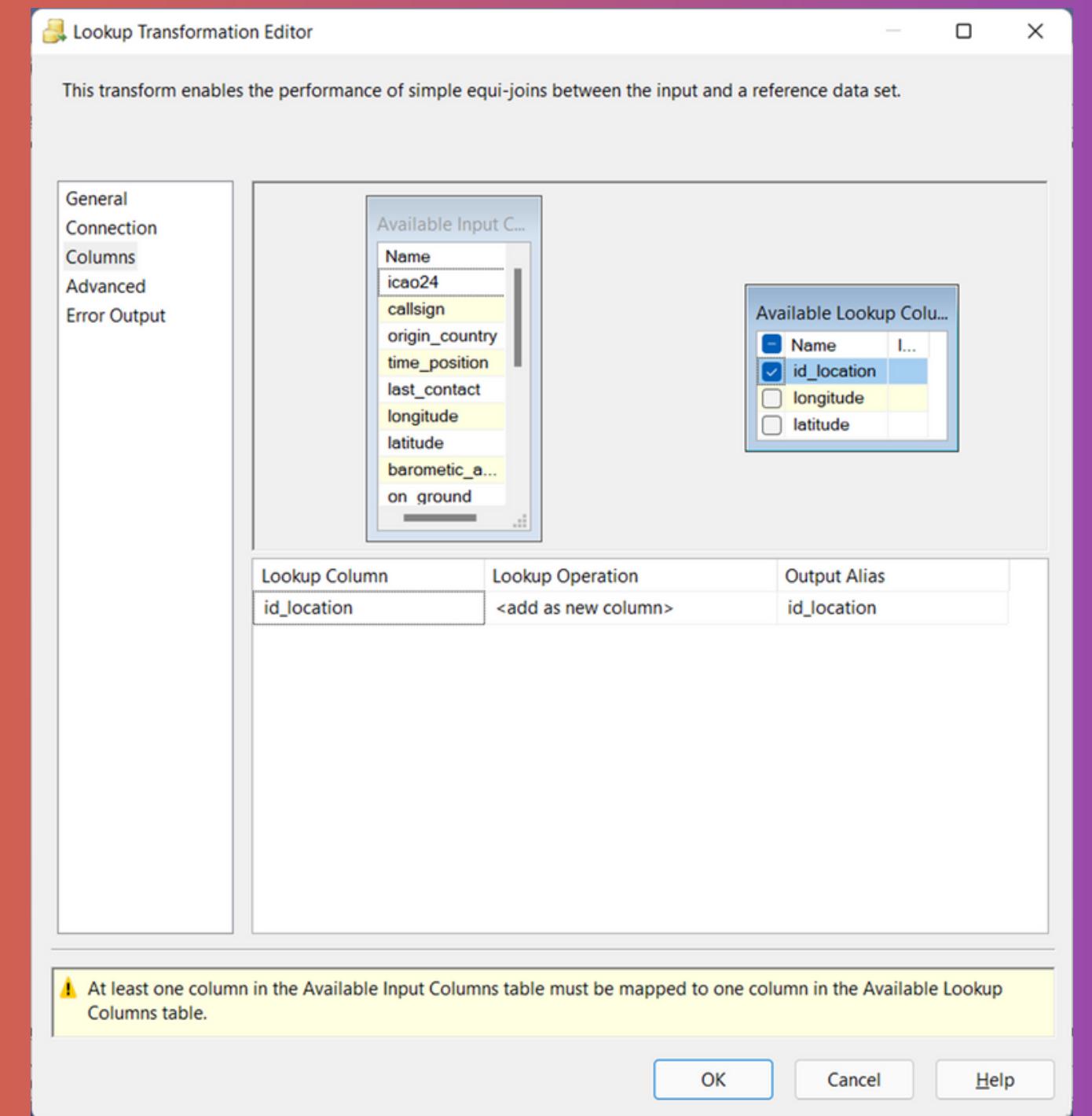


Tạo bảng Fact_Ukraine_Airspace

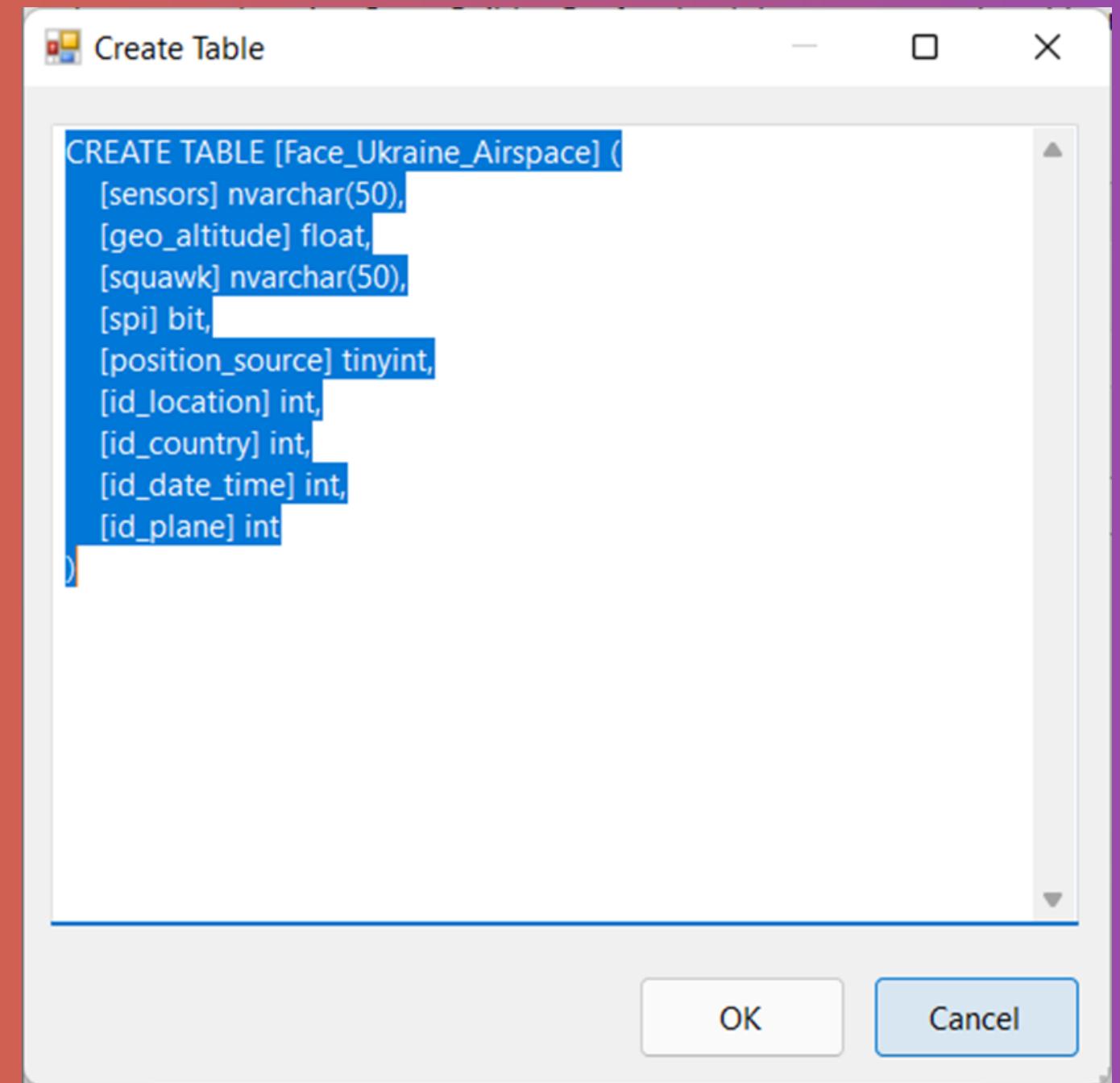


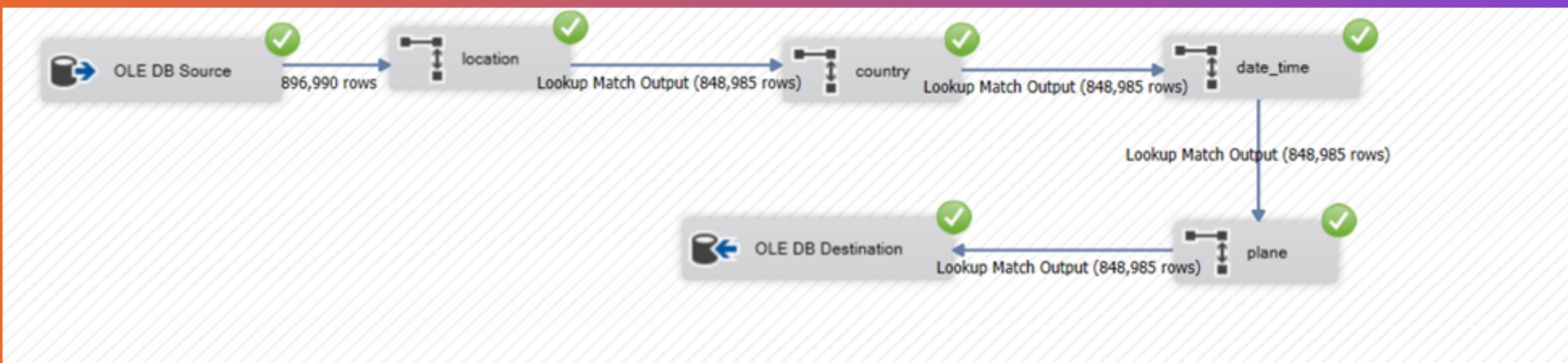






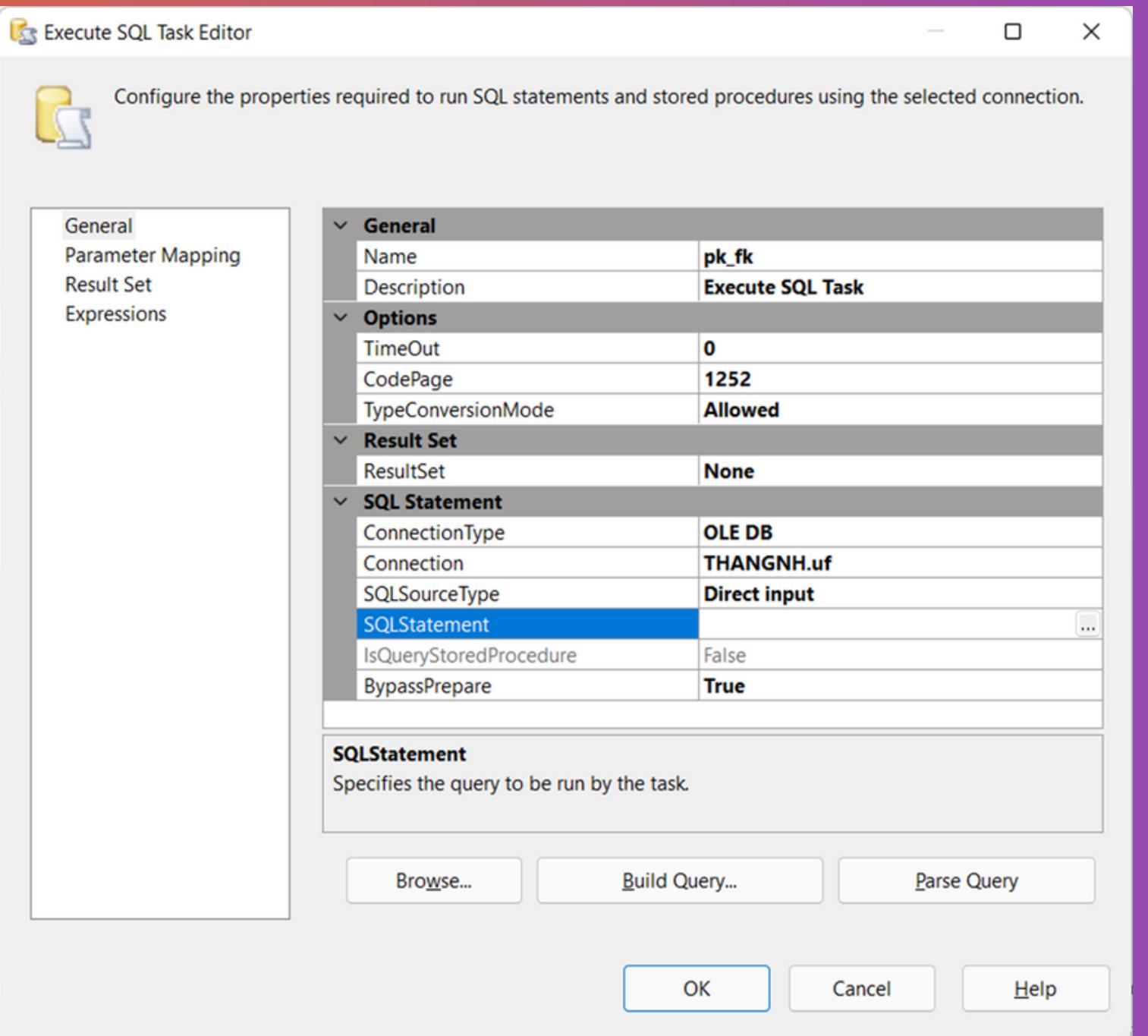
Tương tự với các bảng còn lại





Chạy xong quá trình

Tạo các ràng buộc khóa ngoại giữa bảng Fact và các bảng Dimension

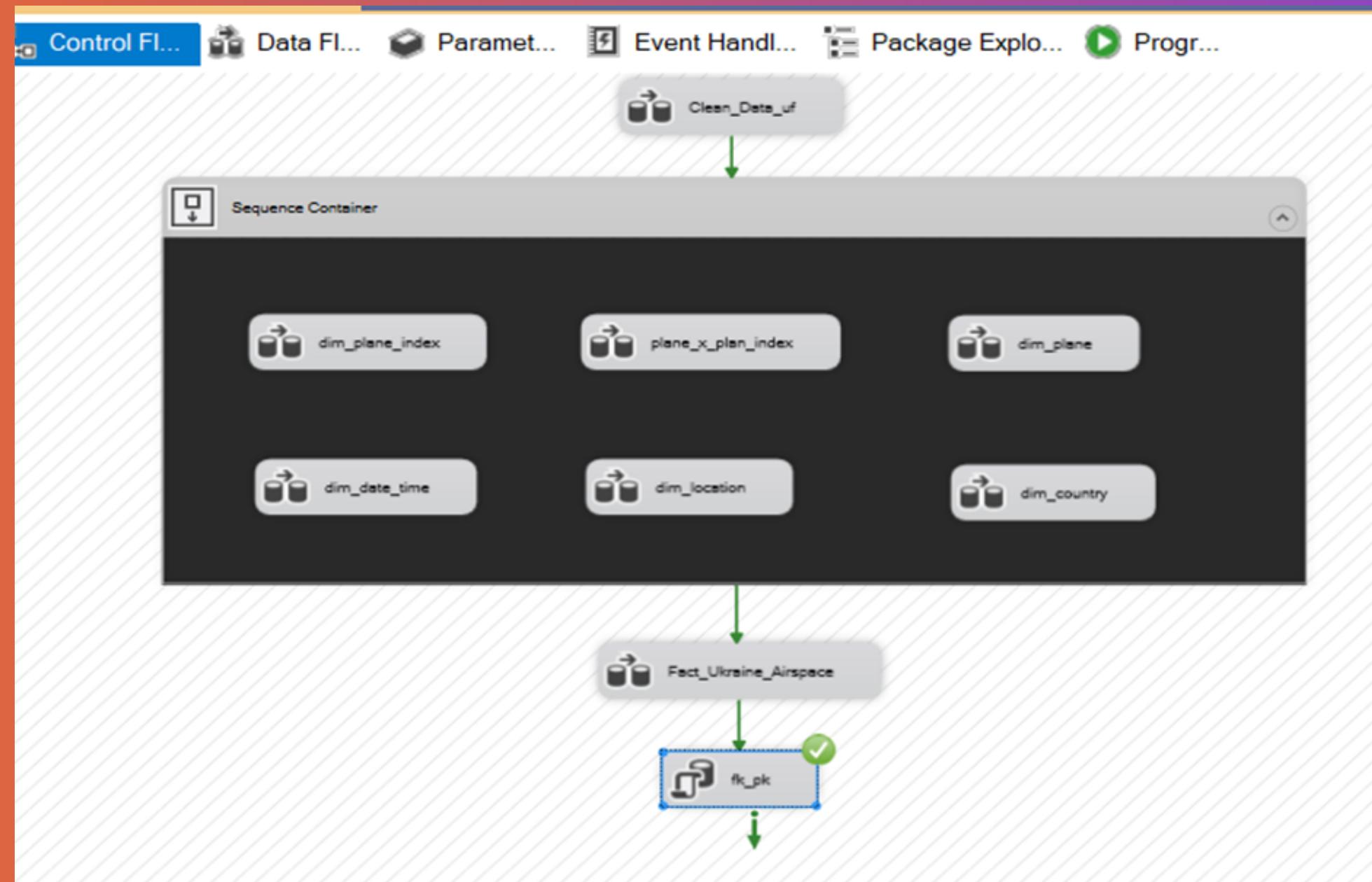


Chọn kết nối tới database uf.

Enter SQL Query

```
alter table Face_Ukraine_Airspace
add constraint fk_face_country [REDACTED]
foreign key (id_country) references dim_country (id_country)
go [REDACTED]
alter table Face_Ukraine_Airspace
add constraint fk_face_plane [REDACTED]
foreign key (id_plane) references dim_plane (id_plane)
go [REDACTED]
alter table Face_Ukraine_Airspace
add constraint fk_face_date_time [REDACTED]
foreign key (id_date_time) references dim_date_time
(id_date_time)
go [REDACTED]
alter table dim_plane
add constraint fk_plane_plane_index [REDACTED]
foreign key (id_plane_index) references dim_plane_index
(id_plane_index)
```

OK Cancel



CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS)

1. Tạo project SSAS trong Visual Studio 2019

1.1 Tạo project

1.2 Thiết lập kết nối

2. Điều chỉnh các Dimension

2.1 Dim Country

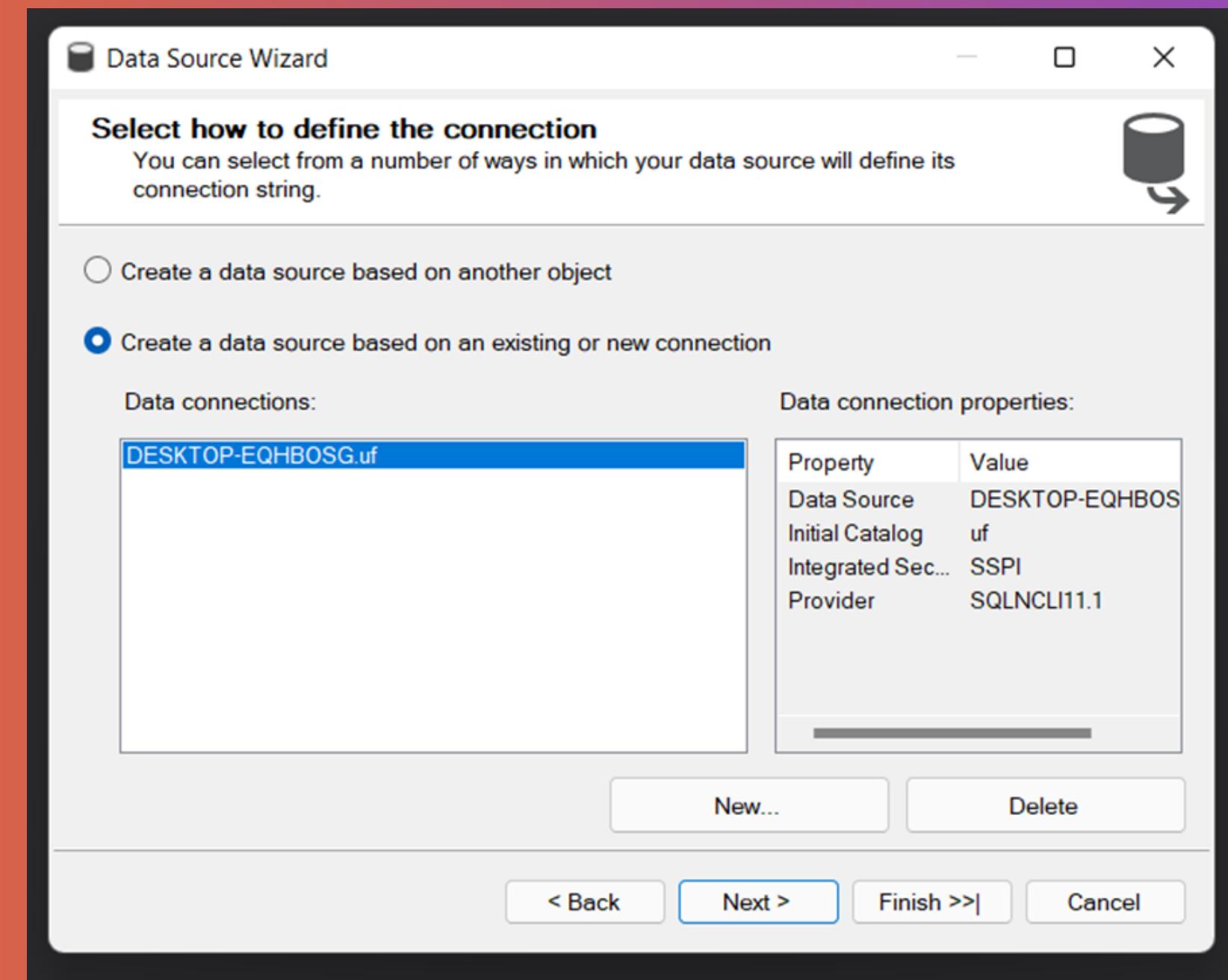
2.2 Dim Location

2.3 Dim Date Time

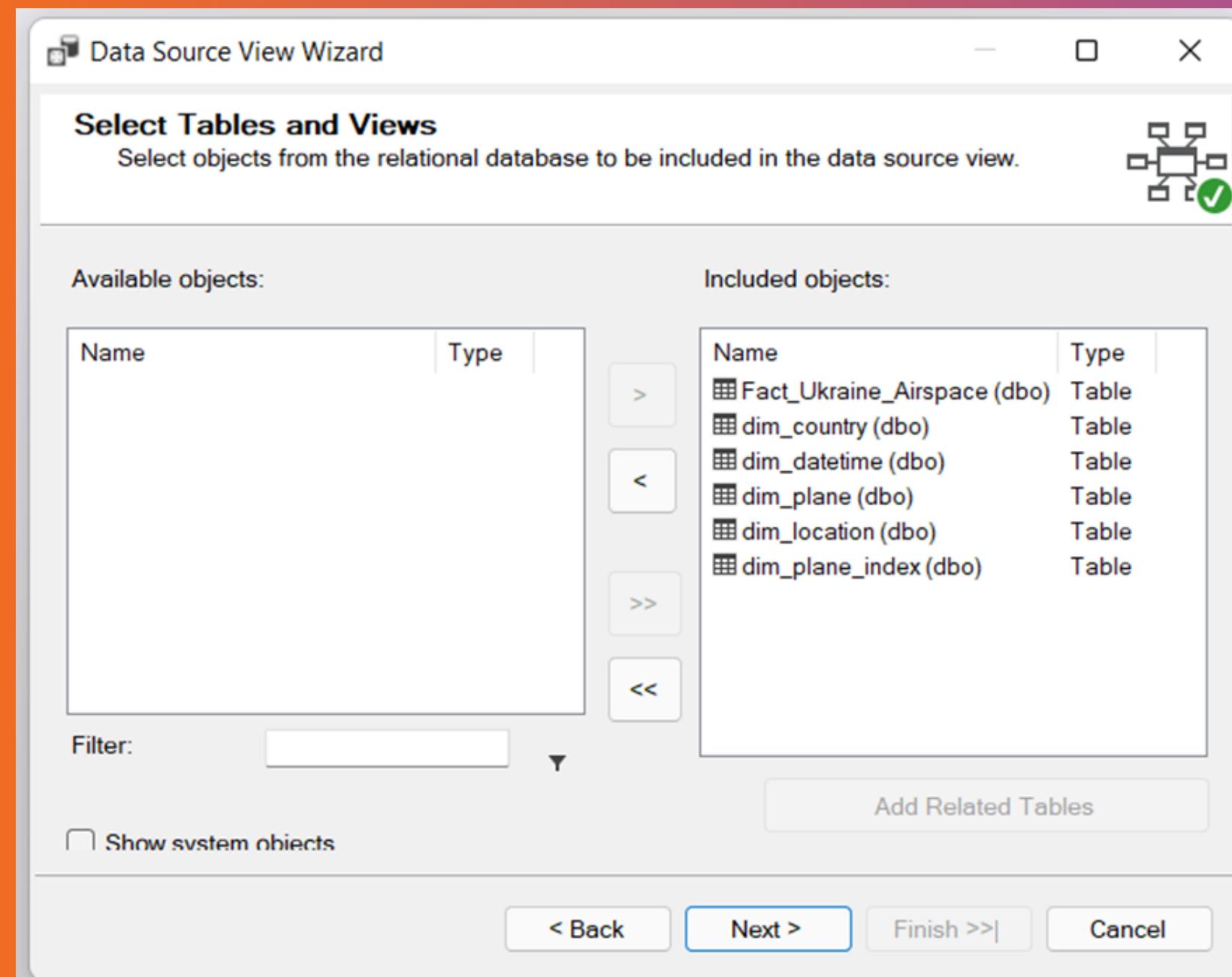
2.4 Dim Plane

2.5 Fact Ukraine Airspace

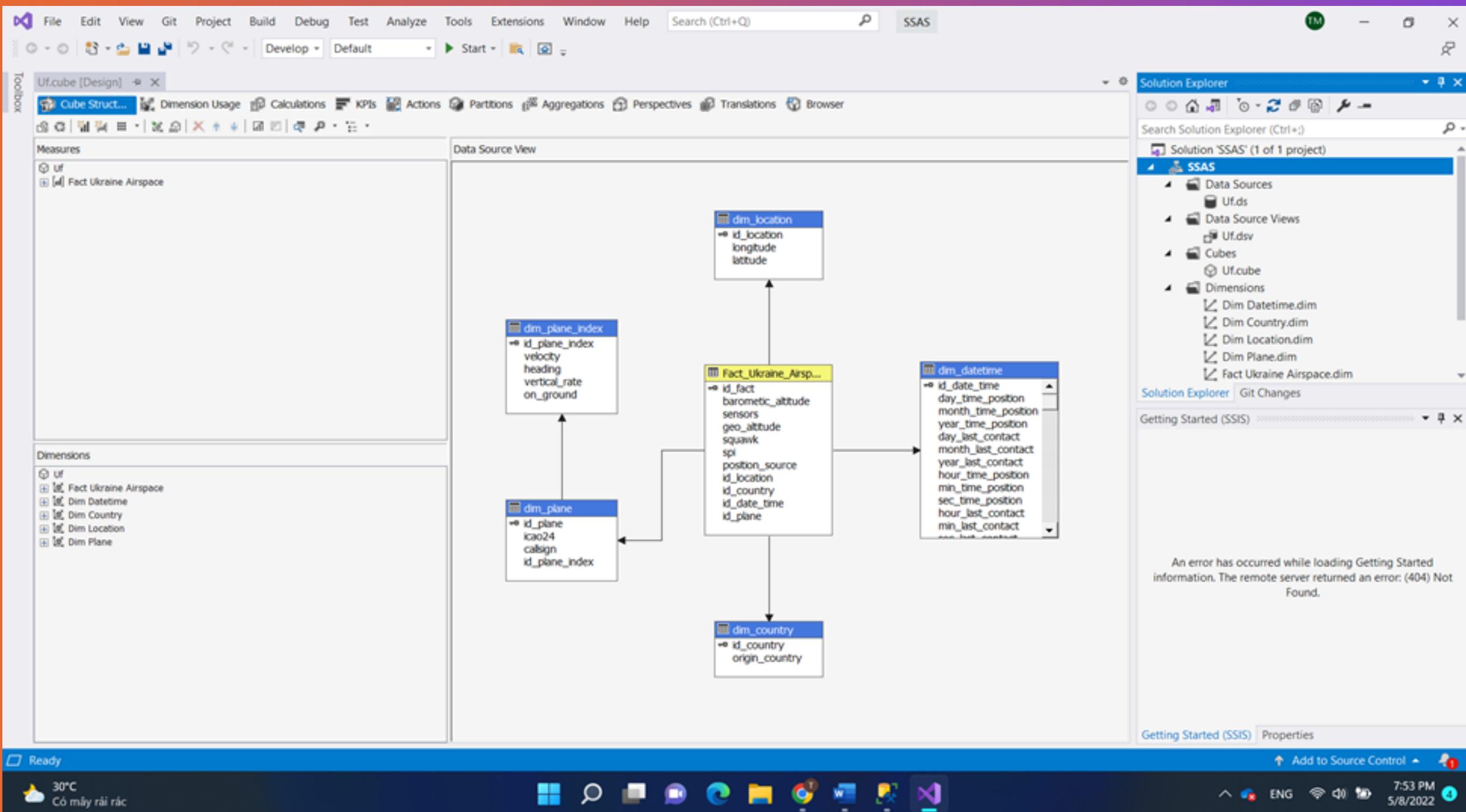
2.6 Dim Plane Index



Create a data source

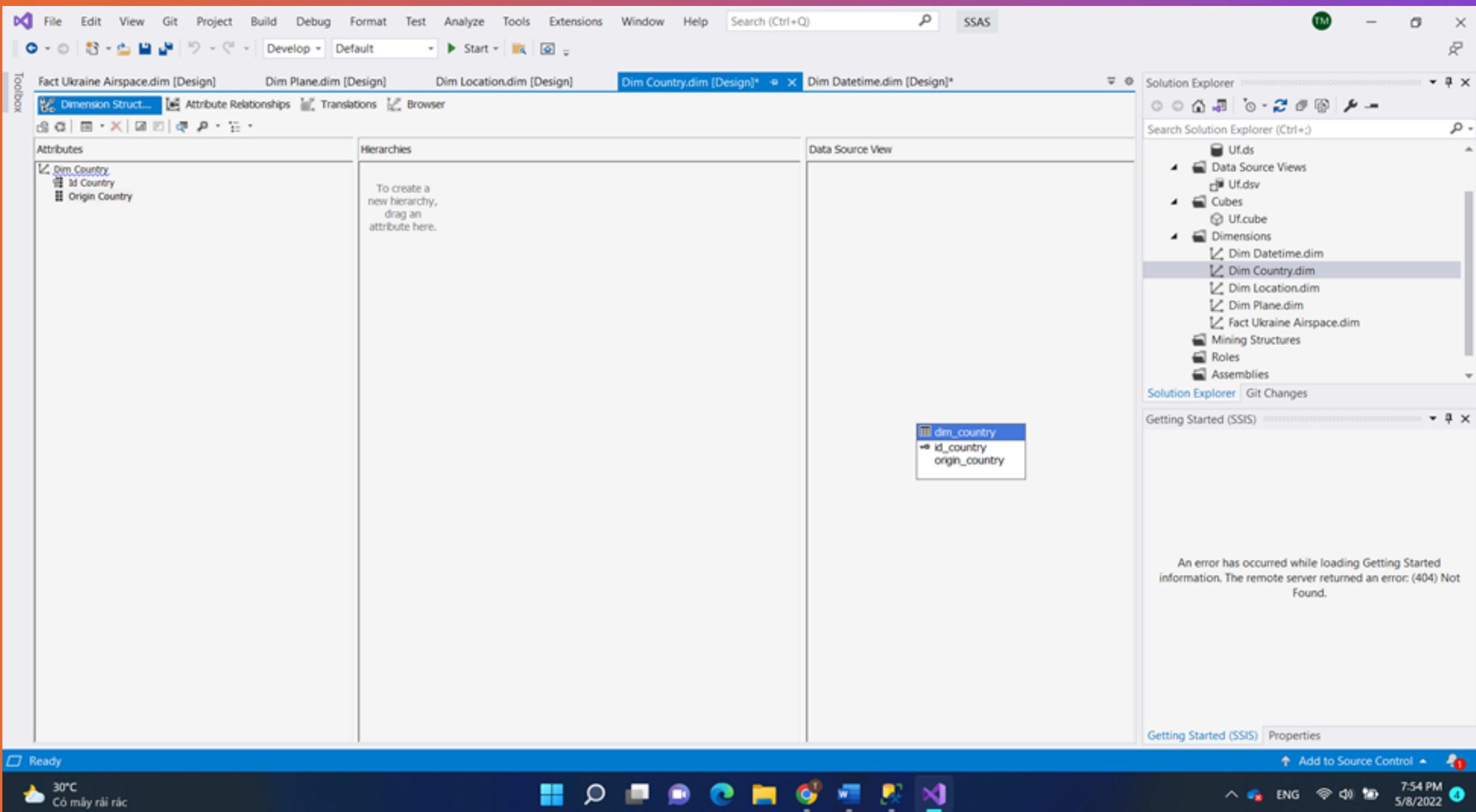


Chọn bảng Fact và click vào Add Related Tables để chọn các bảng có quan hệ với bảng Fact

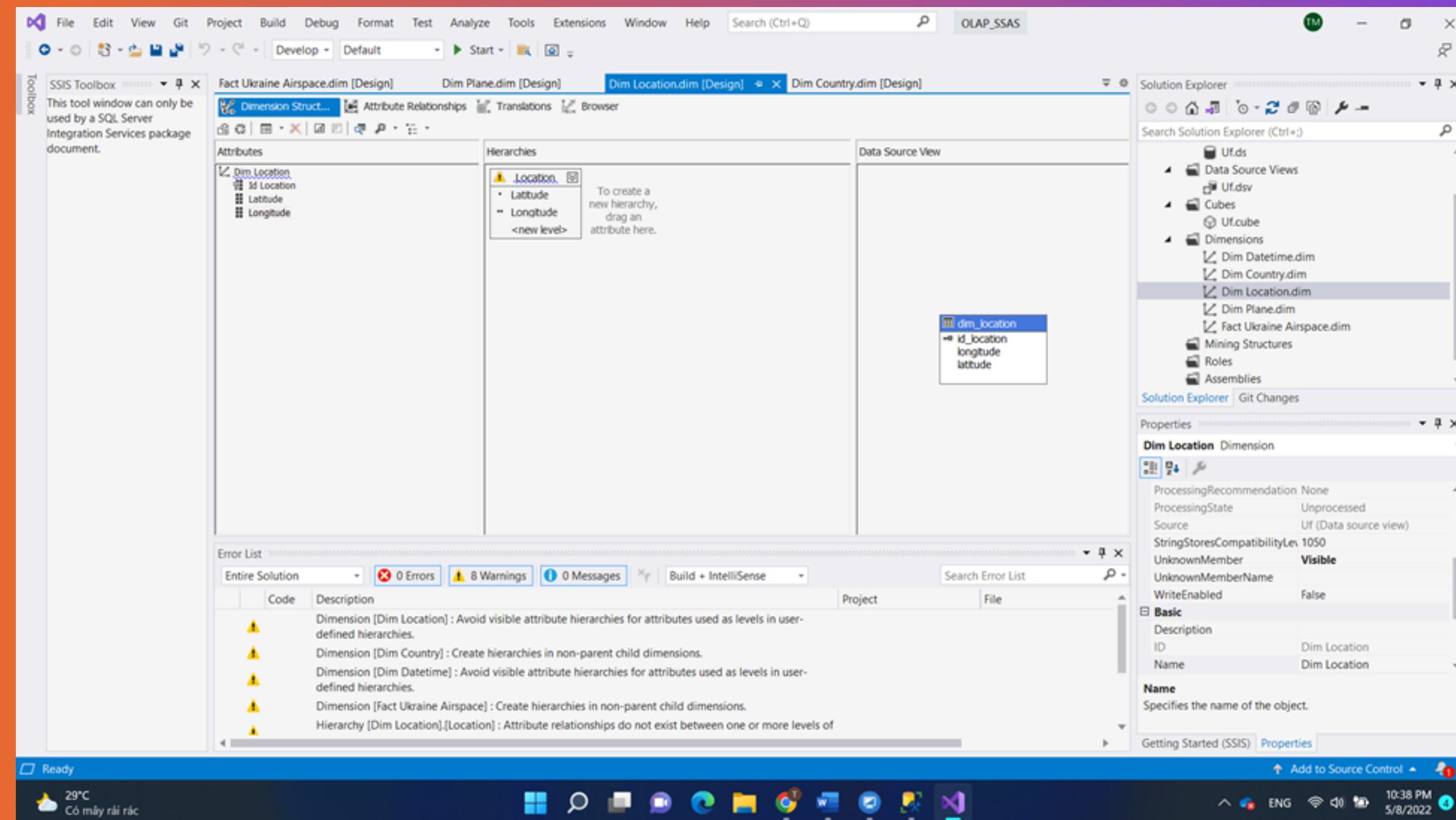


Tại phần Cubes chọn New Cubes

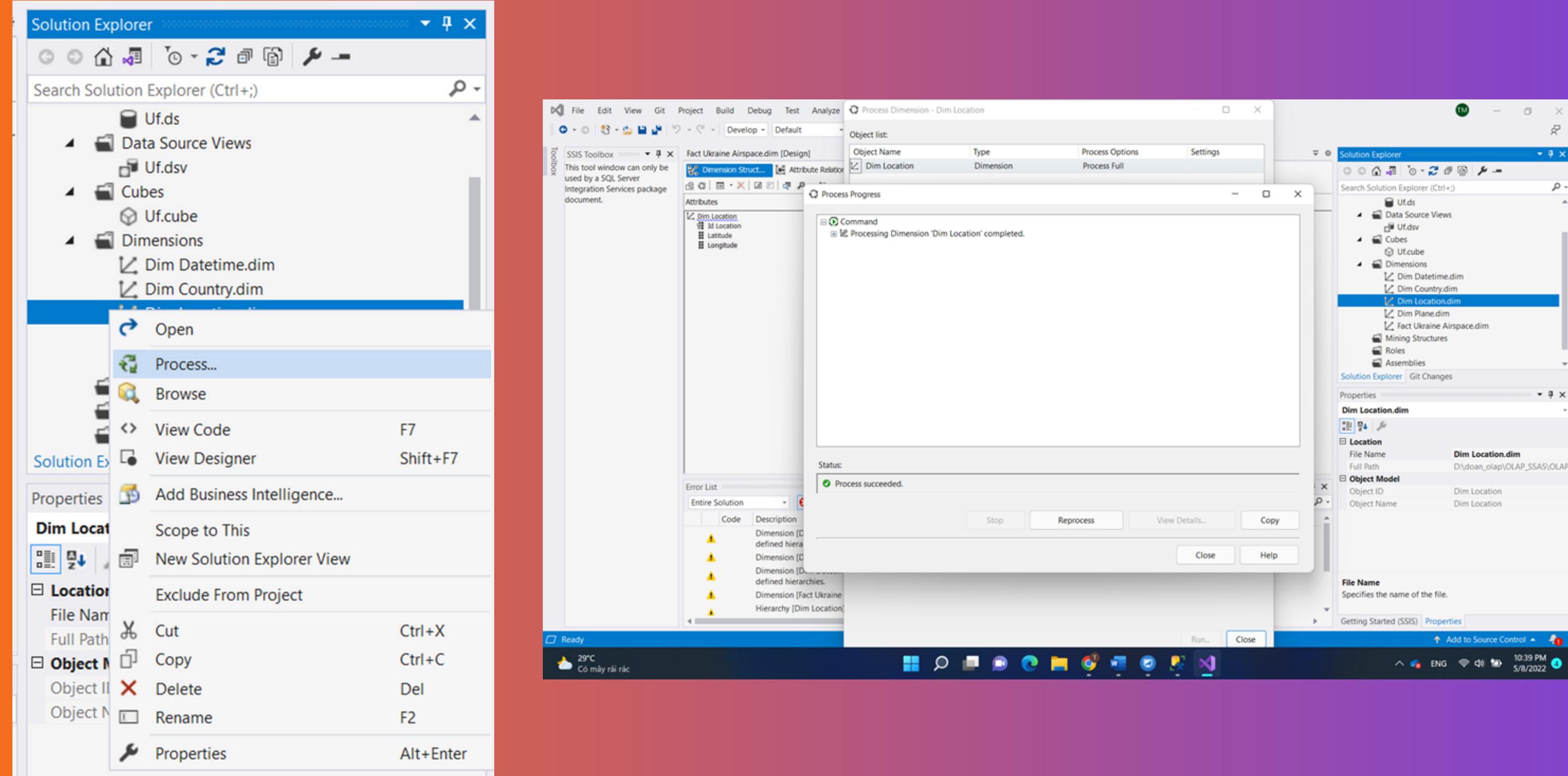
Điều chỉnh các Dimension



Dim Country

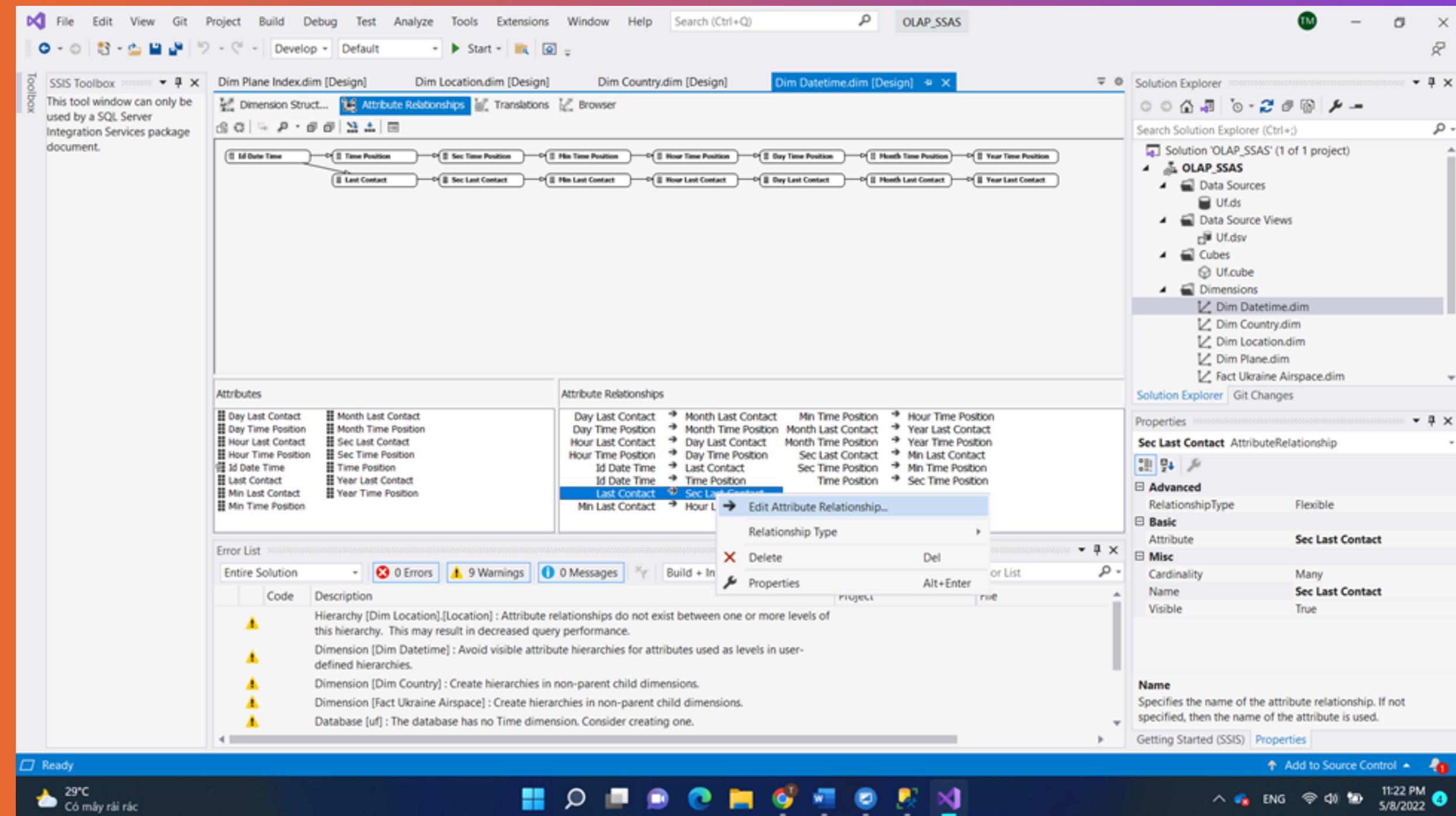


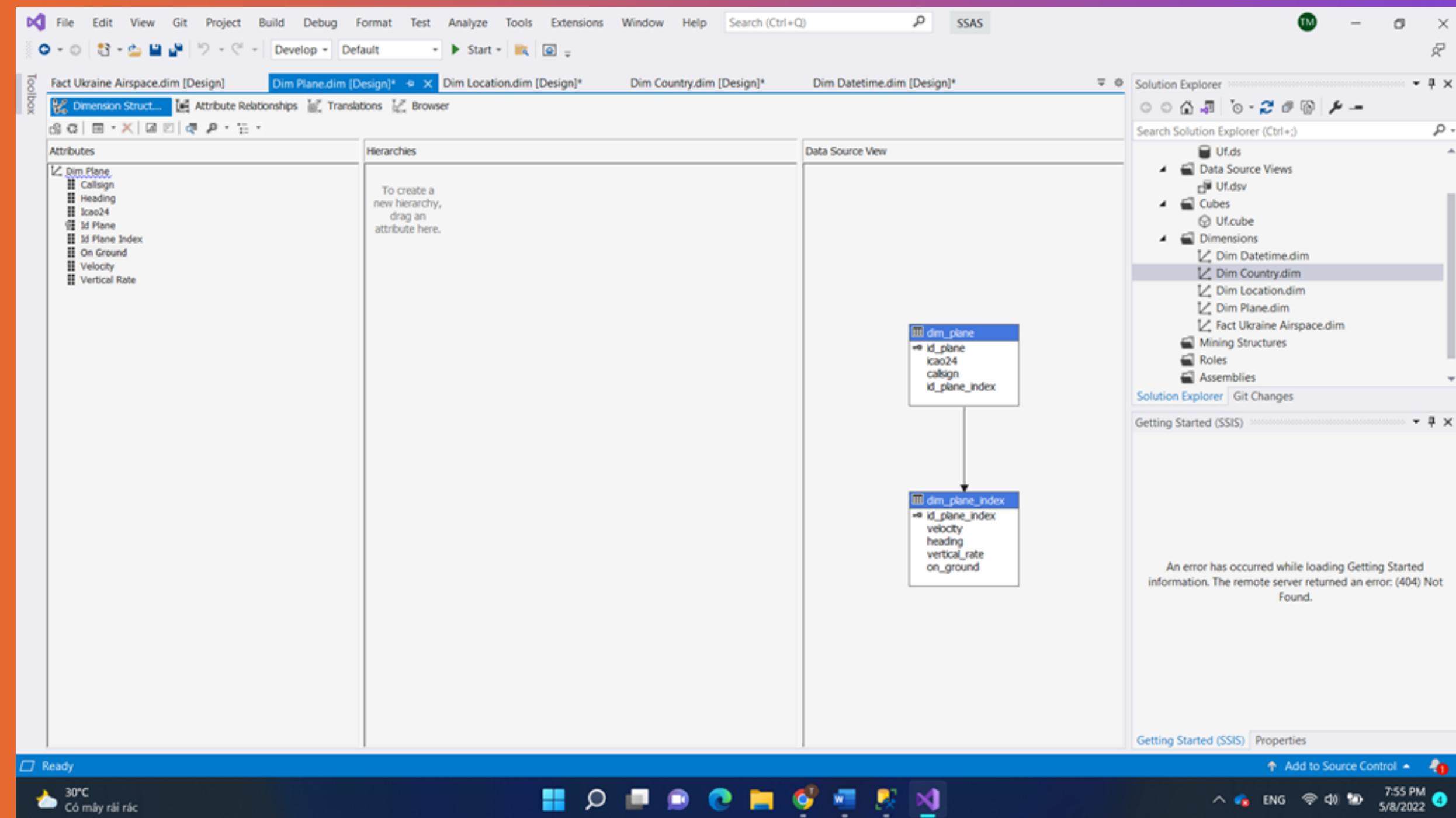
Dim Location



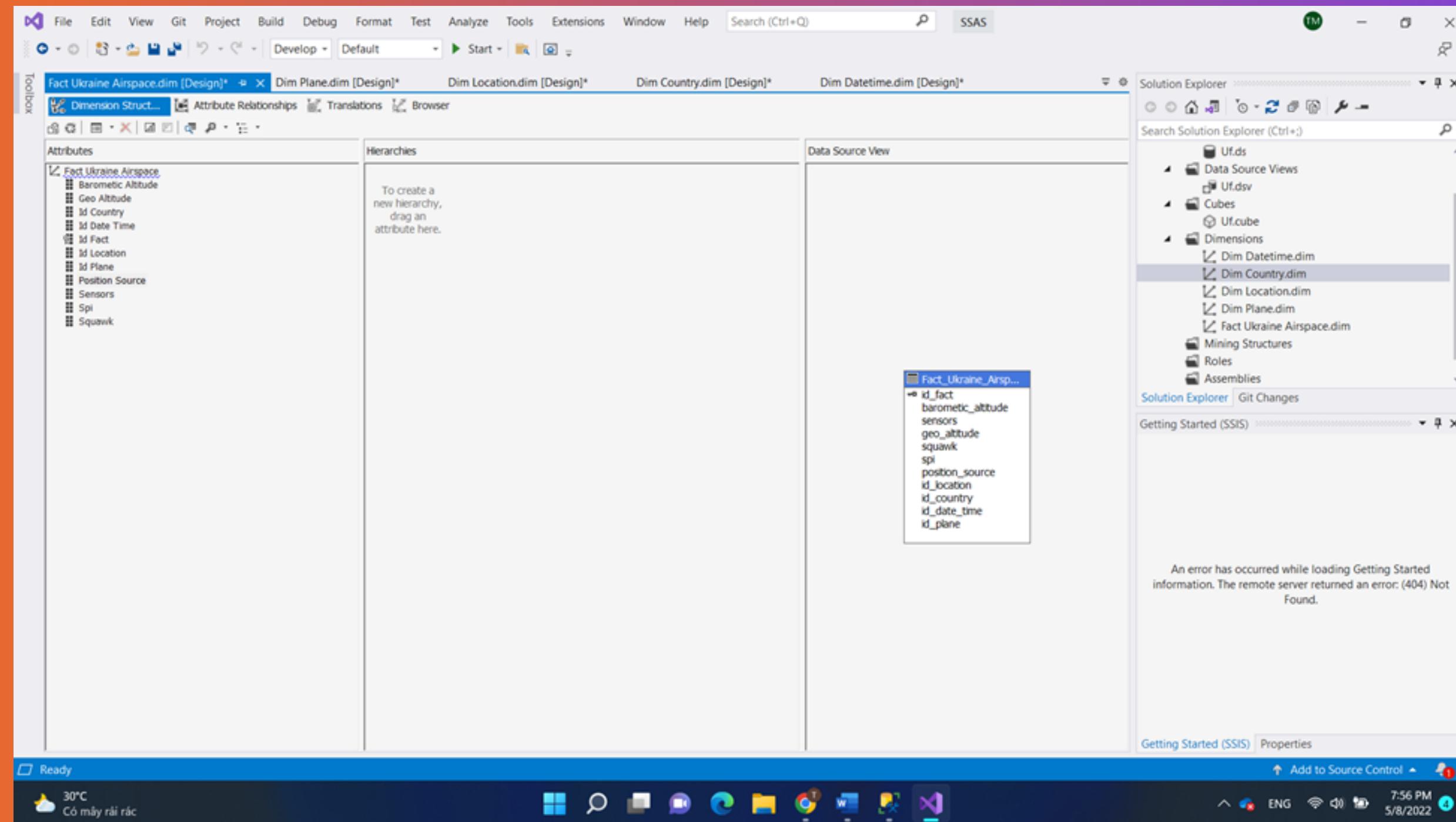
Process và Deploy

Dim Date Time

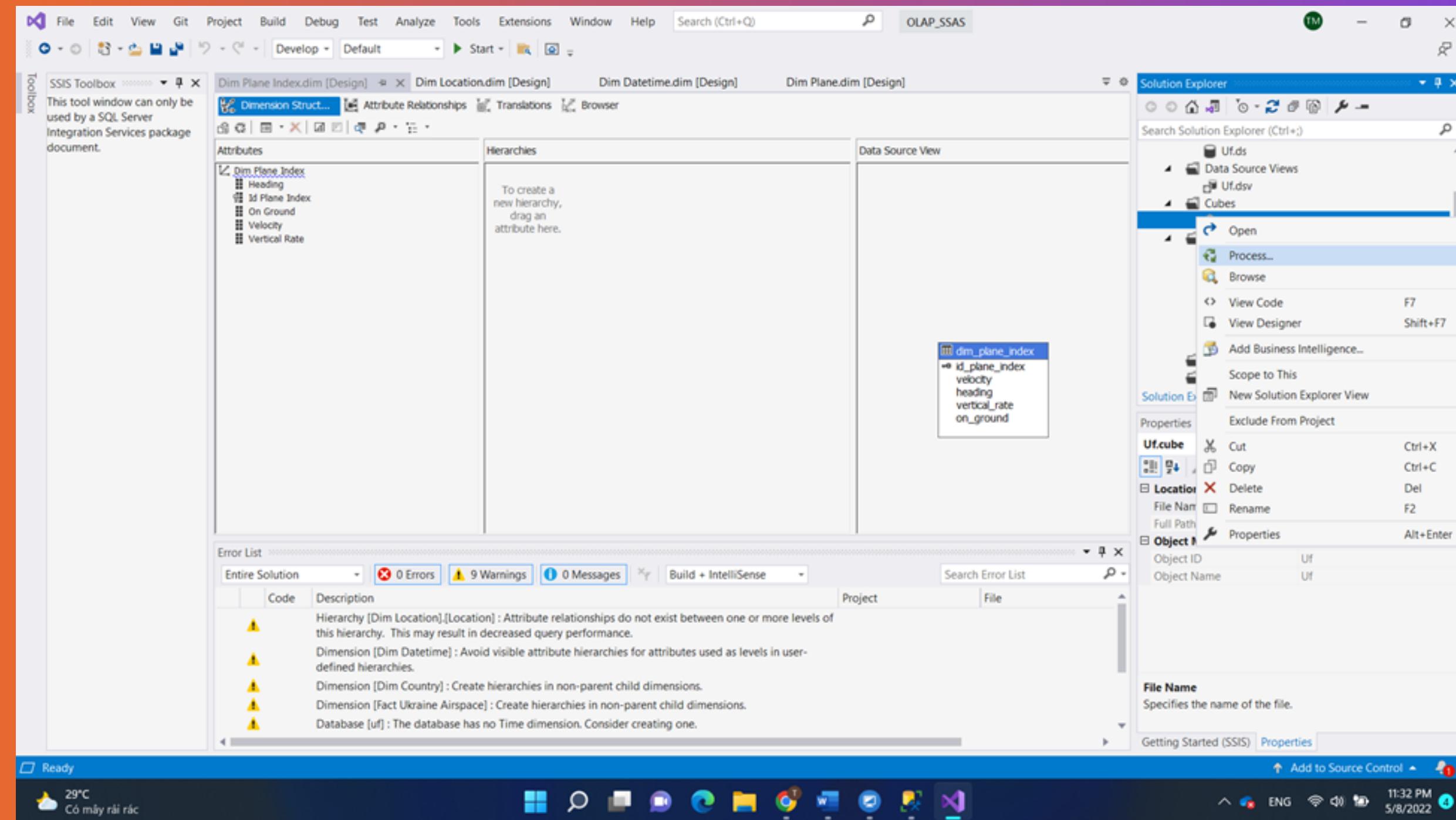




Dim Plane



Fact Ukraine Airspace



Dim Plane Index

3. Các câu truy vấn

CHƯƠNG 4: QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)

1. Cài đặt các công cụ cần thiết

**2. Report Designer
(Reporting Services)**

Report: Thống kê số lượng chuyến bay của Việt Nam

Query Designer

Edit as Text Import... MDX

Dimension Hierarchy Operator Filter Expression Param...

Dim Country Origin Country Equal { Viet Nam }

<Select dimension>

Id	Country	Origin Country	Barometric Altitude	Barometric Altitude	Geo Altitude	Position Source
83	Viet Nam	10096.5	20193	19522.43945...	0	
83	Viet Nam	10119.3603515625	10119.3603515625	9852.660156...	0	
83	Viet Nam	10210.7998046875	10210.7998046875	10066.01953...	0	
83	Viet Nam	10241.2802734375	10241.2802734375	9974.580078...	0	
83	Viet Nam	10317.48046875	10317.48046875	10180.32031...	0	
83	Viet Nam	10408.919921875	10408.919921875	10149.83984...	0	
83	Viet Nam	10447.01953125	10447.01953125	10241.28027...	0	
83	Viet Nam	10530.83984375	10530.83984375	10386.05957...	0	
83	Viet Nam	10576.5595703125	10576.5595703125	10142.21972...	0	
83	Viet Nam	1059.18005371093...	1059.18005371094	1188.719970...	0	
83	Viet Nam	10591.7998046875	10591.7998046875	10325.09960...	0	
83	Viet Nam	10629.900390625	10629.900390625	10485.12011...	0	
83	Viet Nam	10713.7197265625	10713.7197265625	10477.5	0	
83	Viet Nam	10789.919921875	32369.759765625	31630.62011...	0	
83	Viet Nam	10820.400390625	10820.400390625	10668	0	
83	Viet Nam	10965.1796875	10965.1796875	10538.45996...	0	
83	Viet Nam	11102.33984375	11102.33984375	10881.36035...	0	
83	Viet Nam	11254.740234375	11254.740234375	10843.25976...	0	
83	Viet Nam	11269.98046875	349369.39453125	345650.8203...	0	
83	Viet Nam	11277.599609375	575157.580078125	568124.3388...	0	
83	Viet Nam	11369.0400390625	11369.0400390625	11216.63964...	0	

Help OK Cancel

Flight from Vietnam				
	Icao24	Callsign	Origin Country	Location
	[Icao24]	[Callsign]	[Origin_Co]	Latitude: [Latitude] Longitude: [Longitude]
			Total	«Expr» checks
	Total	«Expr» flight		

Flight from Vietnam			
Icao24	Callsign	Origin Country	Location
8.88E+04	HVN5037	Viet Nam	Latitude: 42.832599639892578 Longitude: 26.574399948120117
			Latitude: 43.214199066162109 Longitude: 26.085399627685547
			Latitude: 43.217899322509766 Longitude: 26.079399108886719
			Latitude: 43.421600341796875 Longitude: 25.75160026550293
			Latitude: 43.572498321533203 Longitude: 25.506500244140625
			Latitude: 43.646800994873047 Longitude: 25.389799118041992
			Latitude: 43.664699554443359 Longitude: 25.363899230957031
			Latitude: 44.152999877929688 Longitude: 24.691900253295898

			Latitude: 45.850601196289063 Longitude: 21.778200149536133
			Latitude: 45.978099822998047 Longitude: 21.437099456787109
			Latitude: 46.036899566650391 Longitude: 21.278200149536133
			Latitude: 42.36090087890625 Longitude: 27.23390007019043
			Latitude: 46.150001525878906 Longitude: 20.969999313354492
			Latitude: 42.549999237060547 Longitude: 26.96980094909668
			Latitude: 42.648399353027344 Longitude: 26.831399917602539
Total	1 flight	Total	46 checks

Preview Report

Report: Độ cao trung bình của các chuyến bay theo quốc gia



Average altitude



Origin Country	Icao24	Callsign	Geo Altitude
[Origin_Country]	[Icao24]	[Callsign]	[Max(Geo_Altitude)]
		Total altitude	[Sum(Geo_Altitude)]
		Average altitude	«Expr»
Count Country			[CountDistinct(Origin_Country)]



Average altitude



Origin Country	Icao24	Callsign	Geo Altitude			
Afghanistan	700b20	AZG264	49911.0009765625			
		Total altitude	149656.801757813			
		Average altitude	16628.5335286458			
Albania	501021	ABN1004	58582.55859375			
		Total altitude			Average altitude	16253.8407226563
		Average altitude				
Austria	4.40E+03	ALX279		507c61	QNT7703	21313.1396484375
		Total altitude			QNT7704	22052.279296875
		Average altitude			Total altitude	548792.398864746
	440011	AOJ75L			Average altitude	7127.17401123047
		Total altitude				
		Average altitude				
	4401dc	EJU48YG		Count Country		64
		Total altitude				
		Average altitude	10850.3355887277			

Preview Report

**Report: Thống kê các chuyến bay theo quốc gia
(Nga, Mỹ, Trung Quốc, Việt Nam) trong ngày 25/2/2022**

Ukreina Flights

Biểu đồ số lượng chuyến bay của Mỹ, Trung Quốc, Nga và Việt Nam



- Origin Country 1
- Origin Country 2
- Origin Country 3
- Origin Country 4
- Origin Country 5
- Origin Country 6

	Id Country	Icao24	Origin Country	Callsign
[Id_Country]	[Icao24]	[Origin_Country]	[Callsign]	
	Total Flight			[CountDistinct(Callsign)]
Total Country		[CountDistinct(Origin_Country)]		
Total Flight		[CountDistinct(CallSign)]		



Ukcreina Flights



Biểu đồ số lượng chuyến bay của Mỹ, Trung Quốc, Nga và Việt Nam



China
Russian Federation
United States
Viet Nam

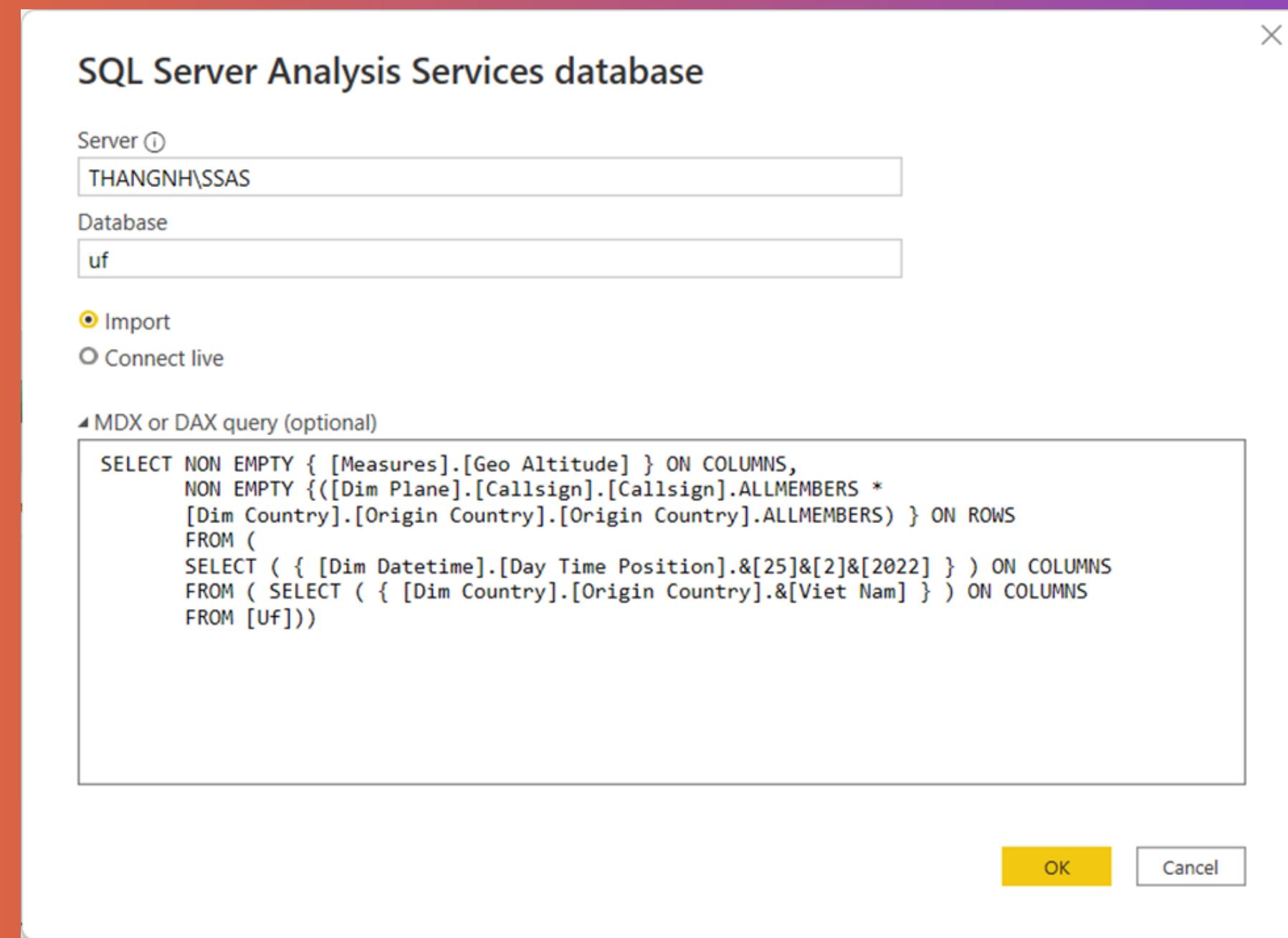
Id	Country	Icao24	Origin Country
14	China	7.81E+08	China
			China
	China	7.81E+09	China
			China
	China	780493	China
			China
	China	7807bb	China
			China
	China	7807d6	China
			China
	China	780850	China
			China
	China	7808a9	China
			China
	China	780b7d	China
			China
	China	780b7e	China
			China
	China	780de6	China
			China

	ae50d0	United States	DUKE48
	ae50e2	United States	DUKE47
	ae50e5	United States	DUKE43
	ae5194	United States	110167
	ae5ad4	United States	DUKE42
	ae5bd3	United States	R20792
	ae5d3e	United States	R20821
	ae6a33	United States	DUKE51
	ae6a35	United States	DUKE51
	Total Flight		53
83	8.88E+04	Viet Nam	HVN5037
	8881b2	Viet Nam	BAV073
	Total Flight		2
	Total Country		4
	Total Flight		240

Preview Report

3 Power BI

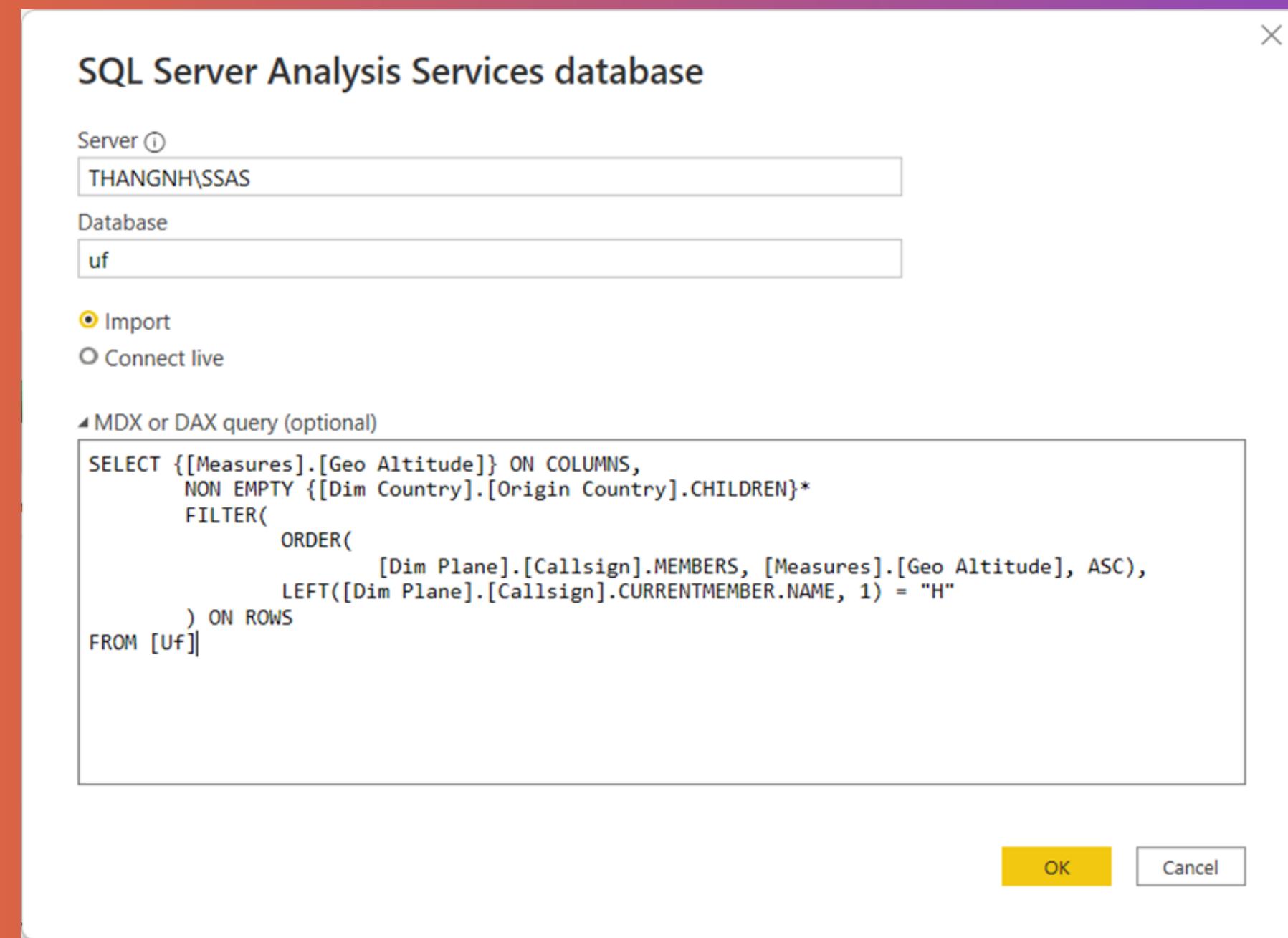
Độ cao lớn nhất của các chuyến bay đến từ Việt Nam vào ngày 25/2



A screenshot of a data visualization interface showing a table with two columns: "Origin Country" and "Count of Callsign". The table lists the following data:

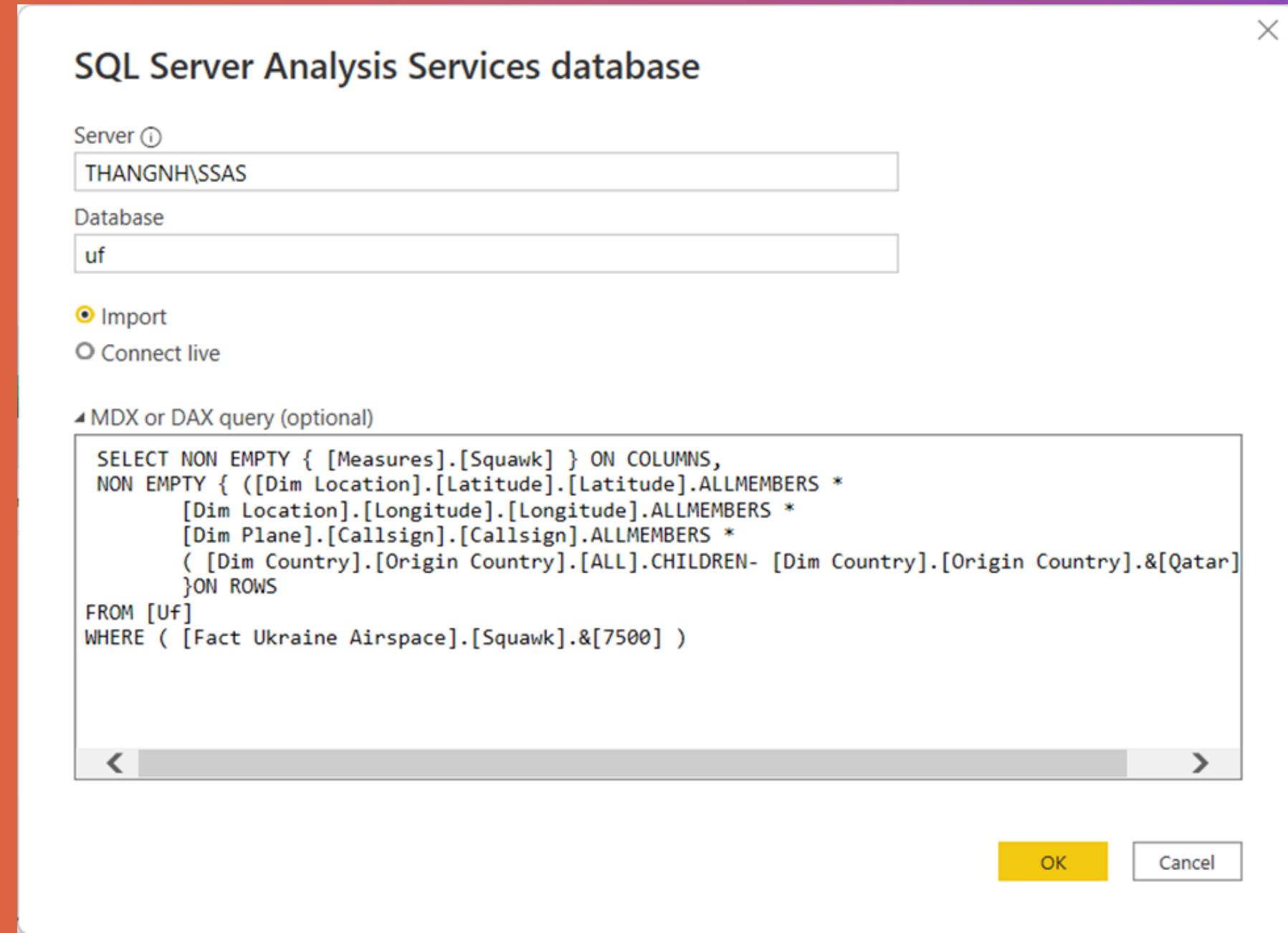
Origin Country	Count of Callsign
Viet Nam	2
BAV073	1
HVN5037	1
Total	2

**Những quốc gia có ký hiệu (Callsign) bắt đầu bằng chữ "H"
và sắp xếp tăng dần theo độ cao vật lý (Geo Altitude)**



Origin Country	Count of Callsign
Bulgaria	1
Germany	1
Hungary	10
Ireland	1
Malta	5
Republic of Korea	1
Romania	14
Saudi Arabia	1
Switzerland	5
United States	7
Viet Nam	7
Total	53

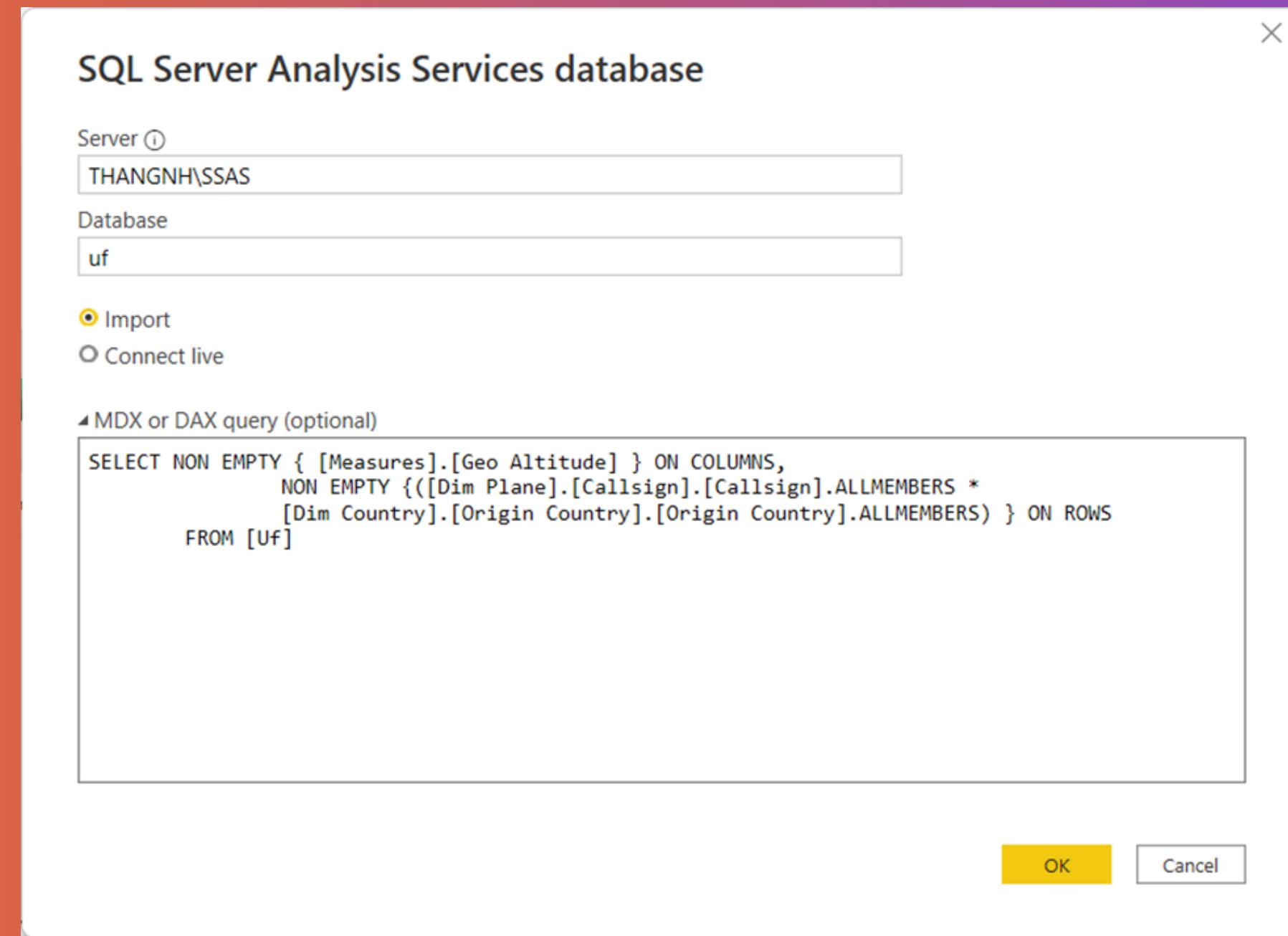
**Vị trí (kinh độ, vĩ độ) của các máy bay đang bị không tặc
tấn công (mã squawk = 7500) nhưng không đến từ Qatar**

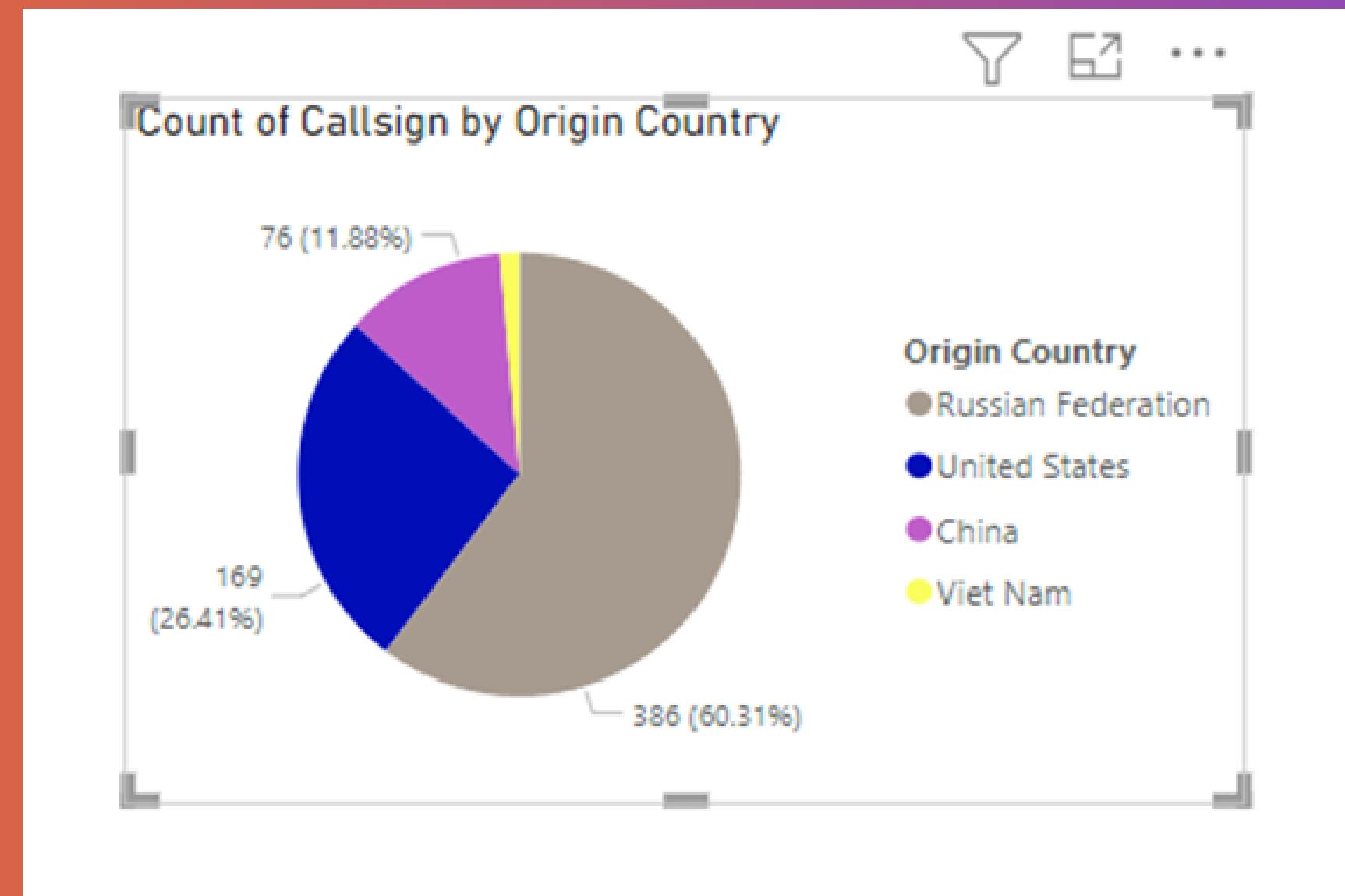


The screenshot shows a Microsoft Power BI report interface with a table visualization. The table has three columns: [Measures].[Squawk], Count of Latitude, and Count of Longitude. The rows show data for different entities, with some being collapsed. The table is set against a background of orange and purple rounded rectangles.

[Measures].[Squawk]	Count of Latitude	Count of Longitude
7500	4	4
Latvia	1	1
BTI792	1	1
Poland	1	1
LOT771	1	1
United Kingdom	2	2
SBI8536	2	2
Total	4	4

Các chuyến bay đến từ Mỹ, Trung Quốc, Nga, Việt Nam





CHƯƠNG 5: QUÁ TRÌNH DATA MINING

1. Chọn lại kho dữ liệu Ukraine_Airspace

2. Data Mining bằng công cụ SSAS trên
Visual Studio

Mô tả về Dataset dùng để mining

- **Tên dataset: Bank Customer Data in VietNam**

Dữ liệu gồm 42640 dòng và 16 cột thuộc tính

Id: Mã khách hàng

Age: Tuổi

Job: Công việc

Marital: Tình trạng hôn nhân

Education: Giáo dục

Default: Vỡ nợ

Balance

Housing

Loan: khoản vay

Day

Month

Duration

Campaign

Pdays

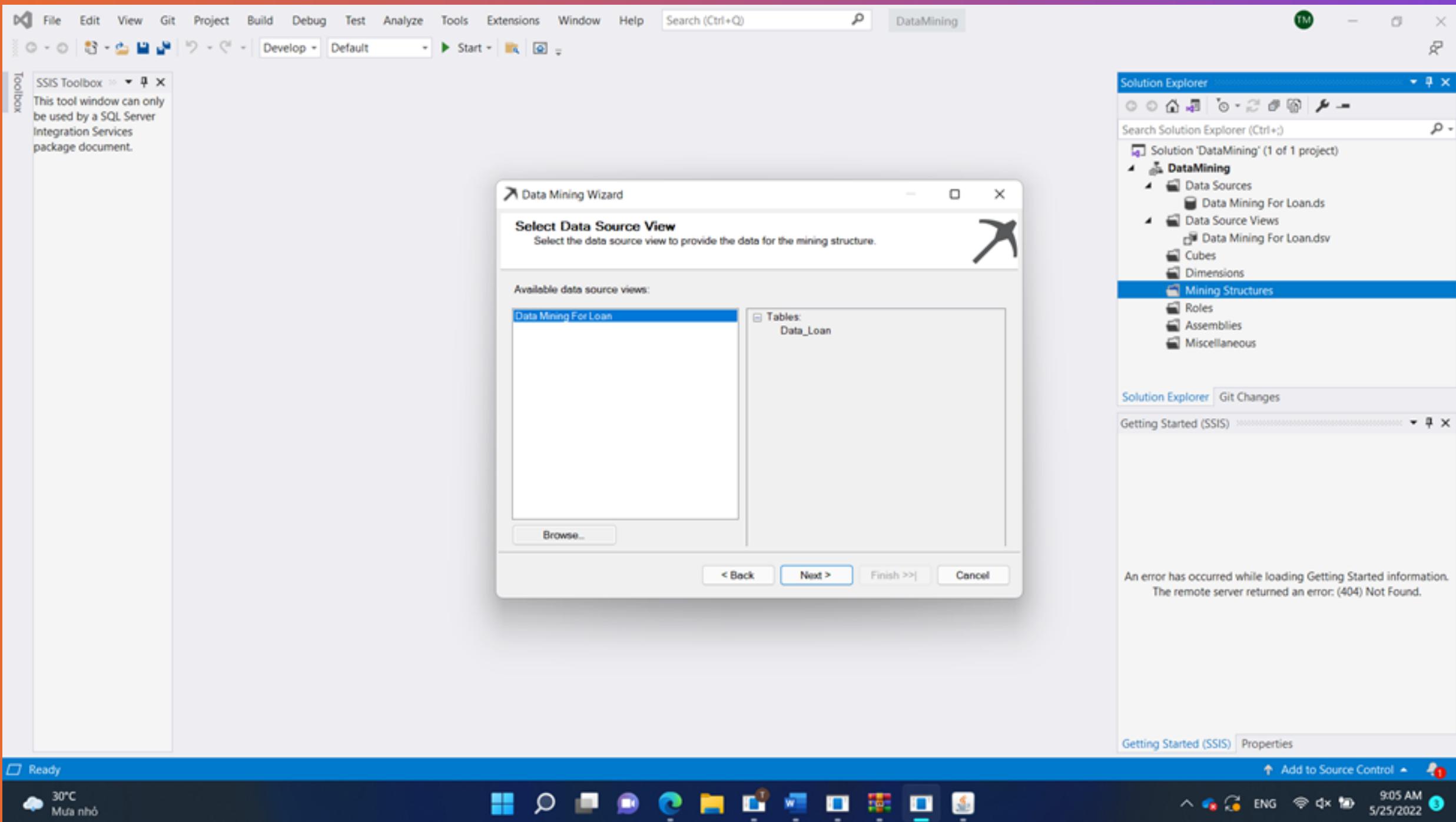
Previous

Term_deposit: khoản tiền gửi có kỳ hạn

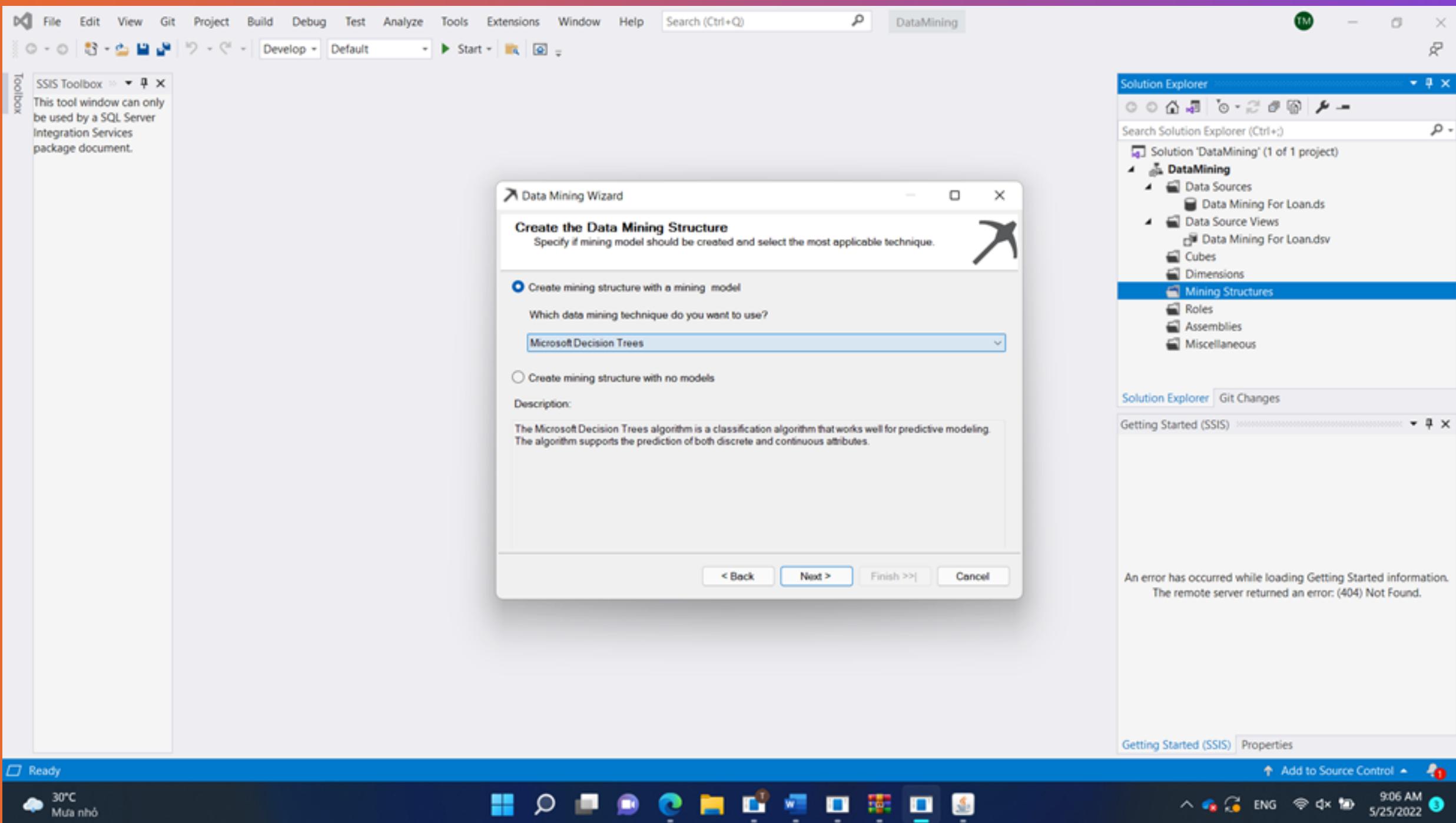
Đổ dữ liệu vào SQL Server bằng công cụ SSIS

Đổ dữ liệu vào bằng công cụ SSIS

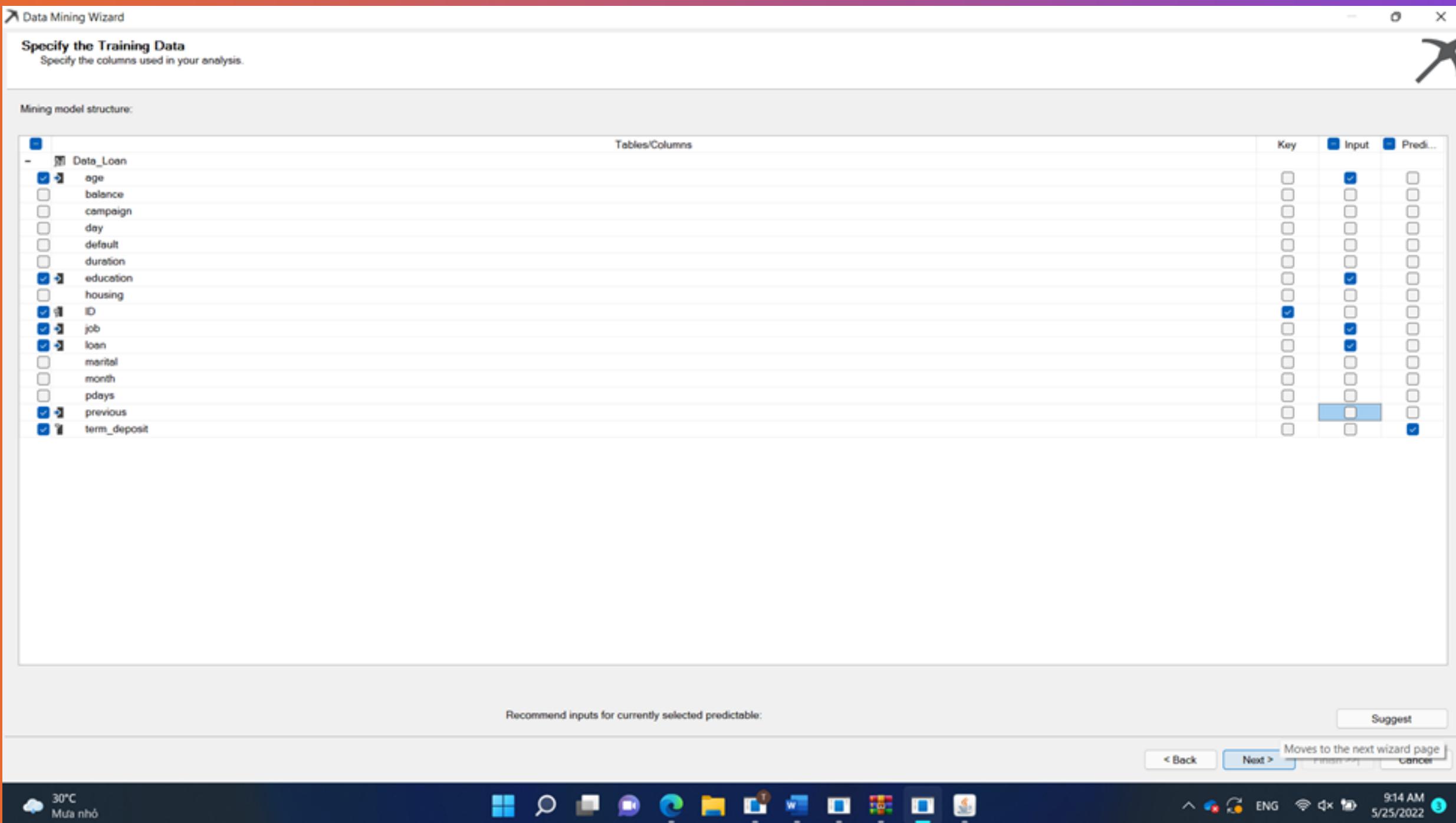
Datamining bằng công cụ SSAS trên Visual Studio



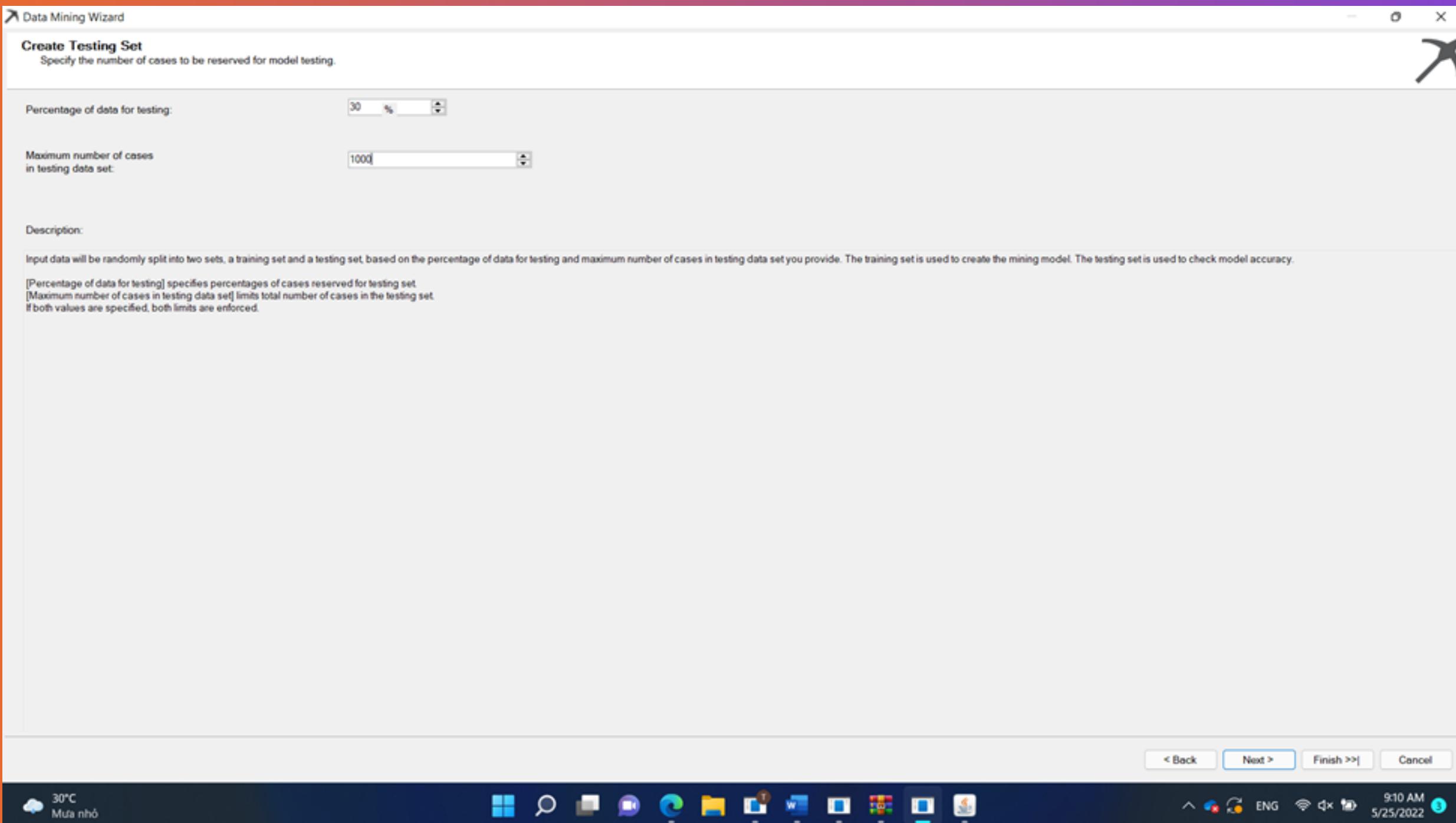
Click phải chuột Mining Structures để tạo mới



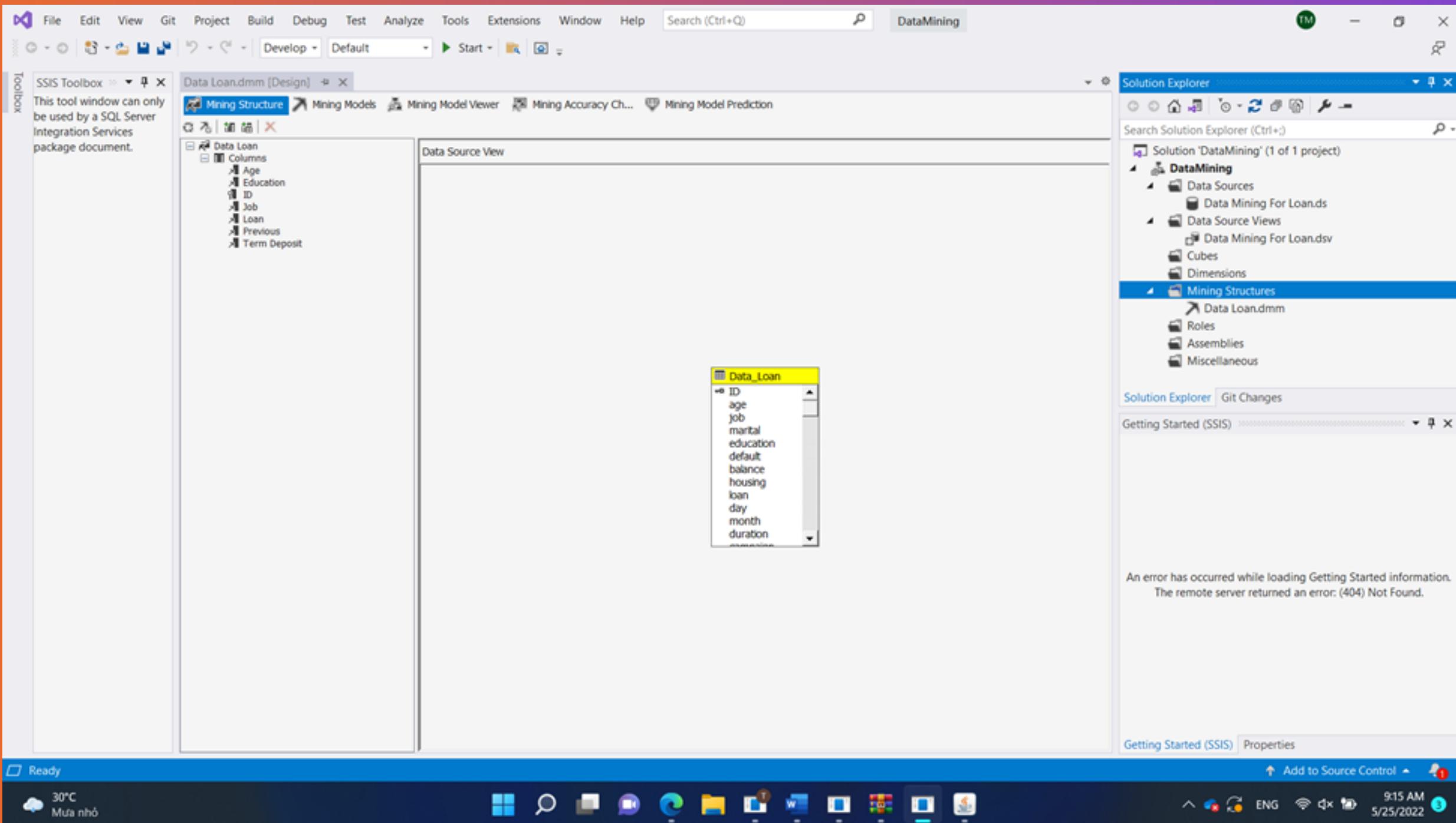
chọn thuật toán cây quyết định Microsoft Decision Tree



Chọn các thuộc tính key, input và predict



Chọn 1000 cases để test



Tiến hành chạy process

Screenshot of the Microsoft Data Mining Model Viewer in Visual Studio, showing a Decision Tree for a "Data Loan" mining model.

Mining Model: Data Loan

Viewer: Microsoft Tree Viewer

Tree: Term Deposit

Default Expansion: 3 Levels

Histograms: 6

Background: All Cases

Show Level 1 to **Level 5**

Mining Legend:

Value	Cas...	Prob...	Histogram
0	377...	90.65%	Blue bar
1	3885	9.35%	Red bar
Missing	0	0.00%	Grey bar

Total Cases: 41639

Decision Tree Structure:

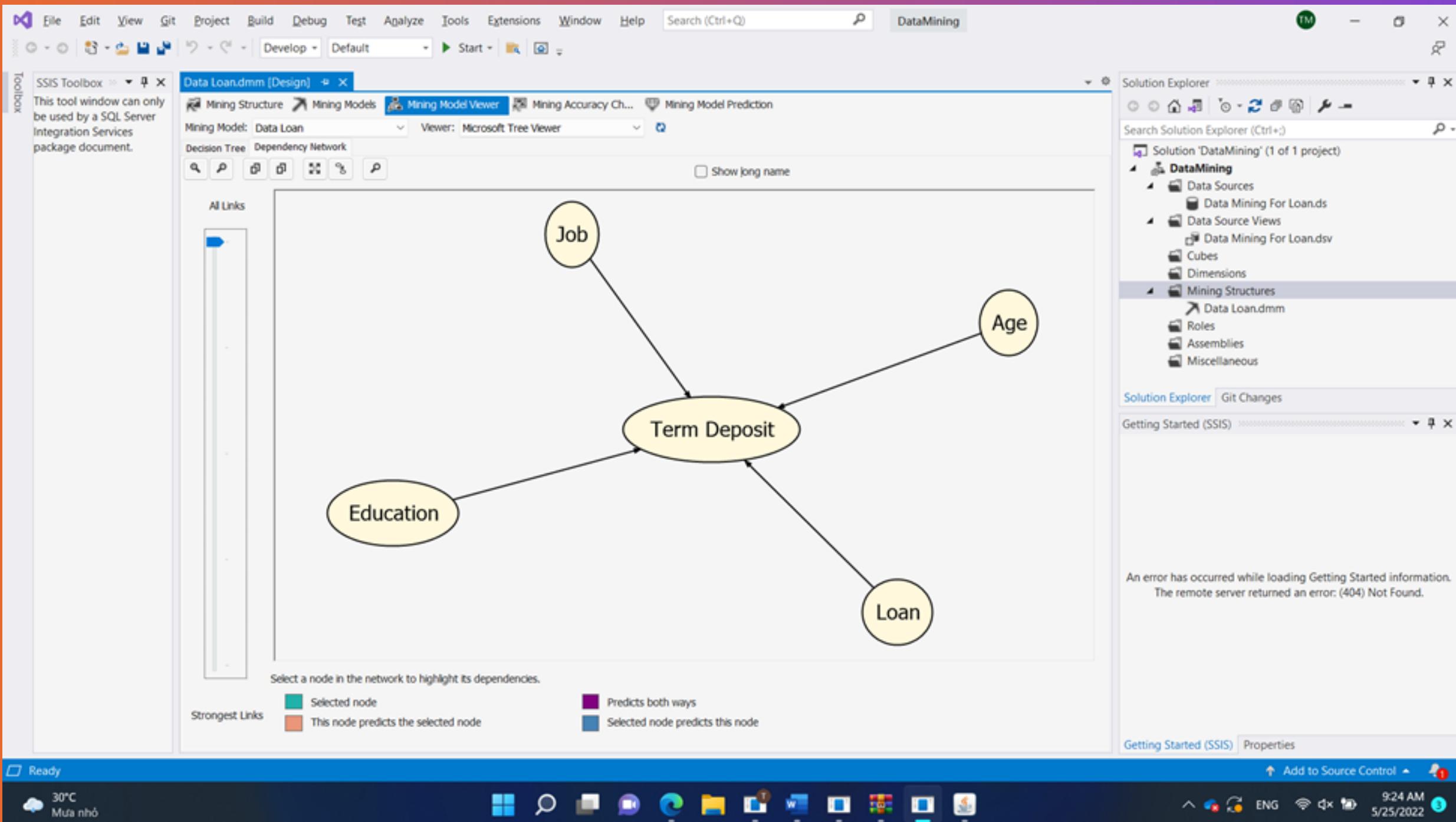
- All
 - Age ≥ 64.200
 - Age ≥ 56.500 and < 64.200
 - Age ≥ 33.400 and < 56.500
 - Education not = 'tertiary'
 - Education = 'tertiary'
 - Age ≥ 25.700 and < 33.400
 - Education not = 'tertiary'
 - Education = 'tertiary'
 - Age < 25.700

Solution Explorer:

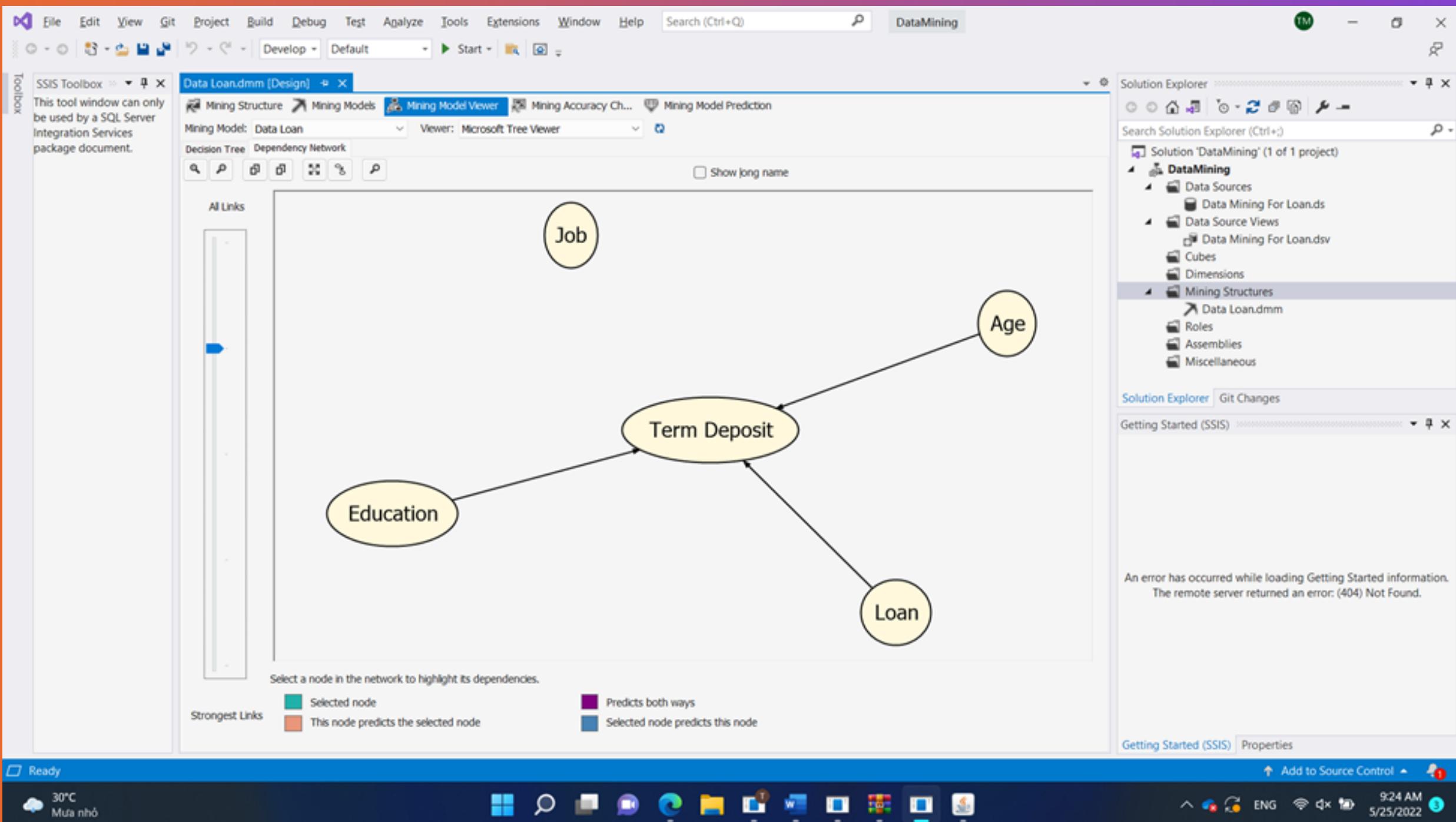
- Solution 'DataMining' (1 of 1 project)
 - DataMining
 - Data Sources
 - Data Mining For Loan.ds
 - Data Source Views
 - Data Mining For Loan.dsv

Properties:

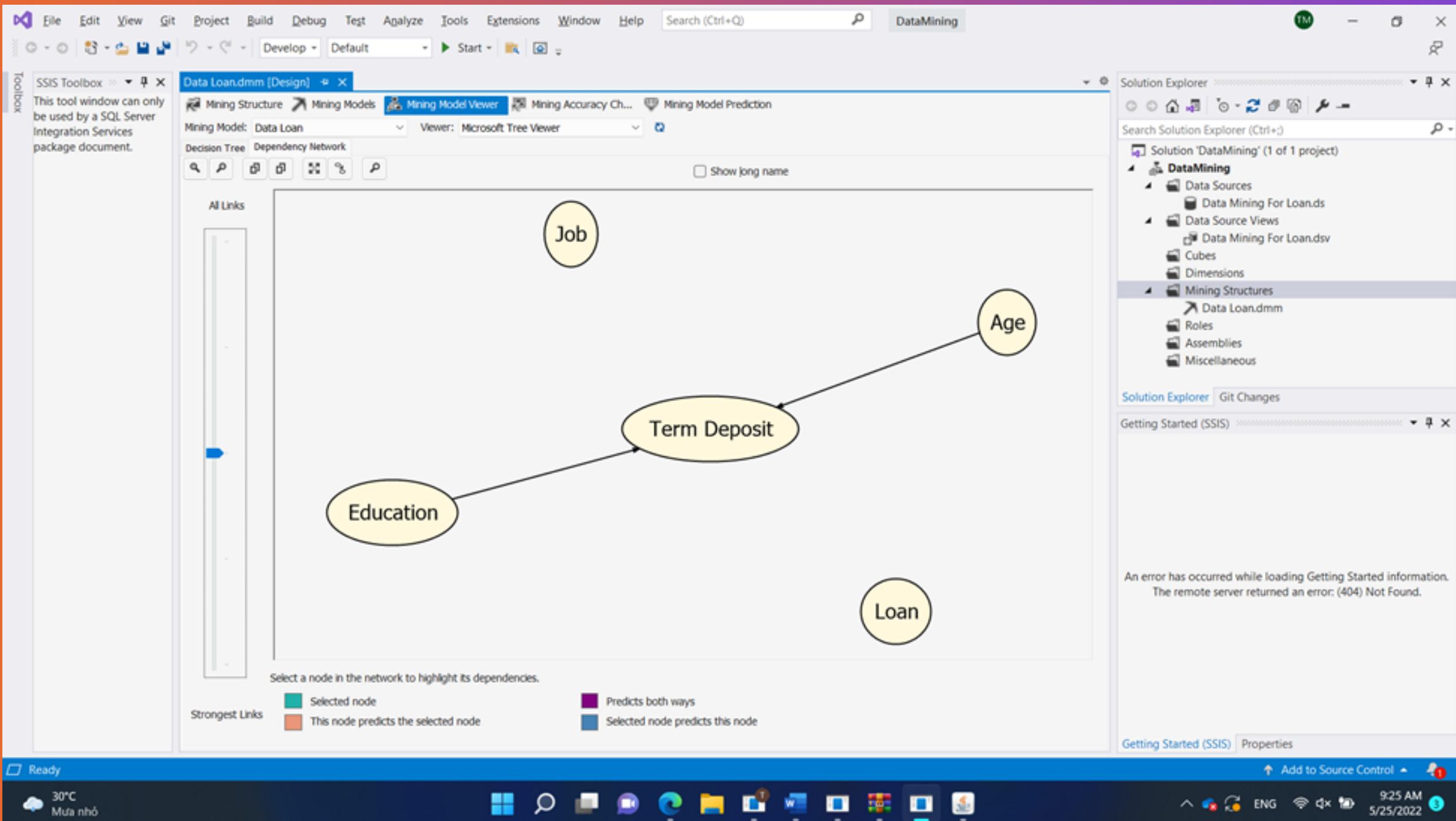
- Ready
- 30°C Mưa nhỏ
- 9:20 AM 5/25/2022



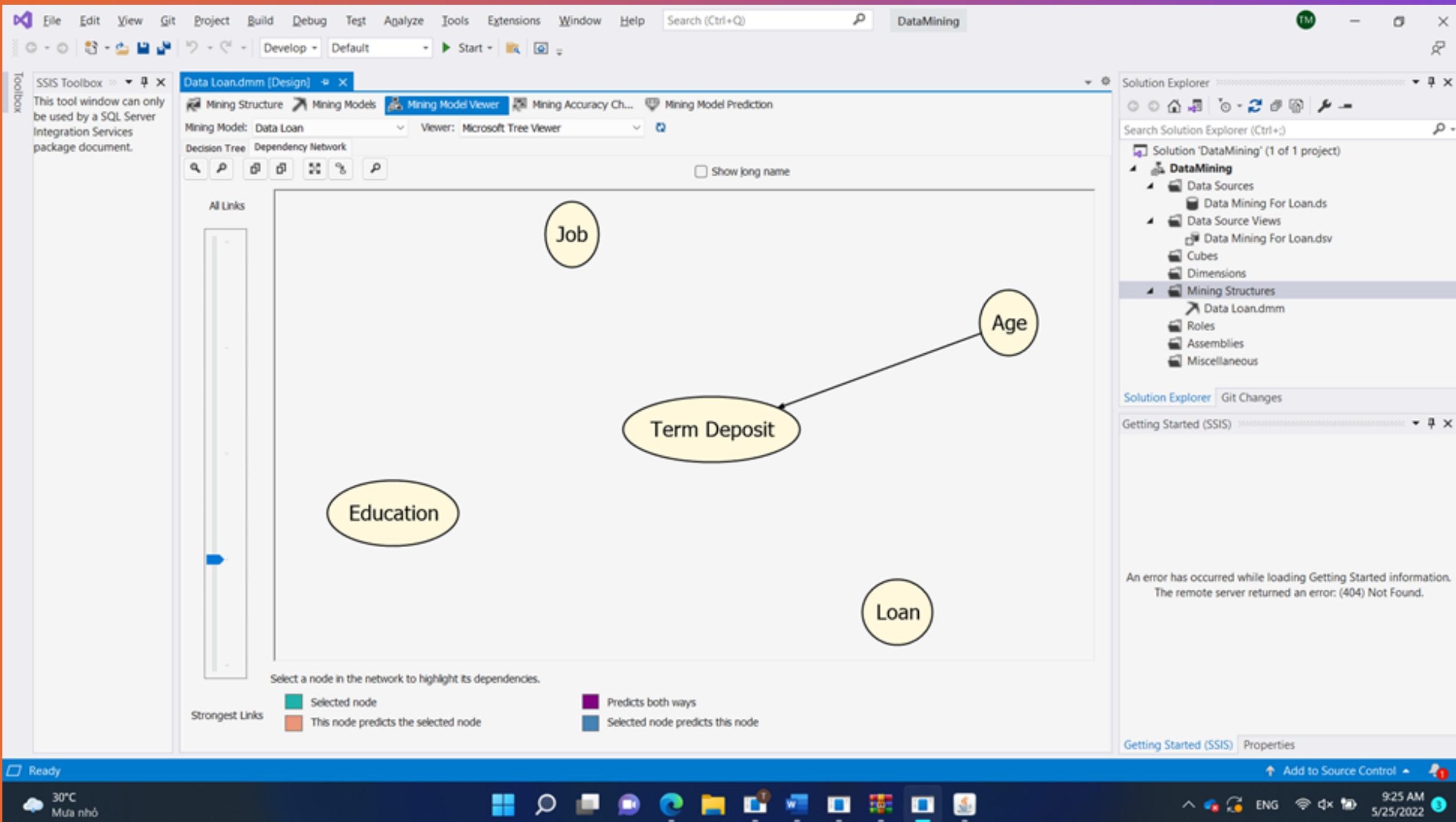
Mức phụ thuộc cấp 5
Chọn Dependency Network để xem các mức phân cấp



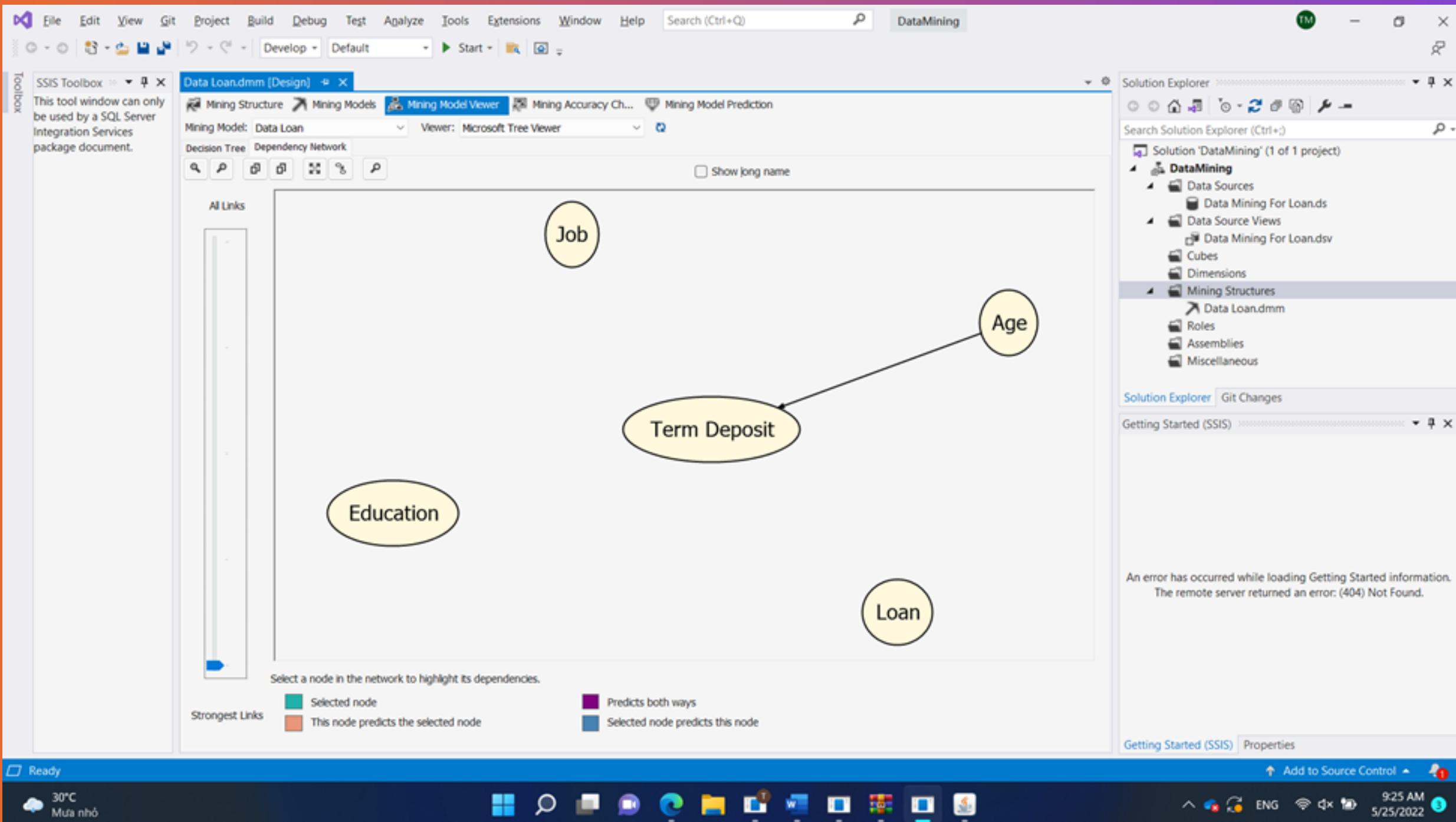
Mức phụ thuộc cấp 4



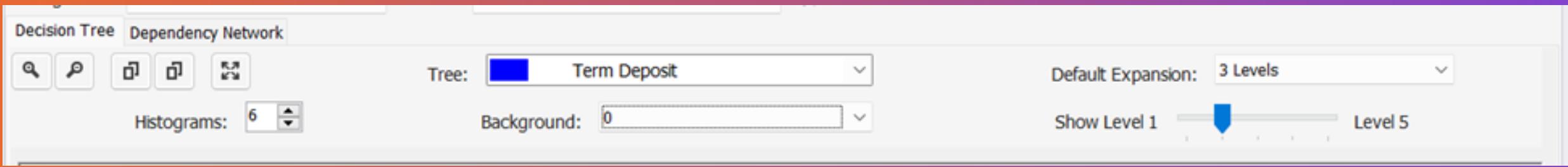
Mức phụ thuộc cấp 3



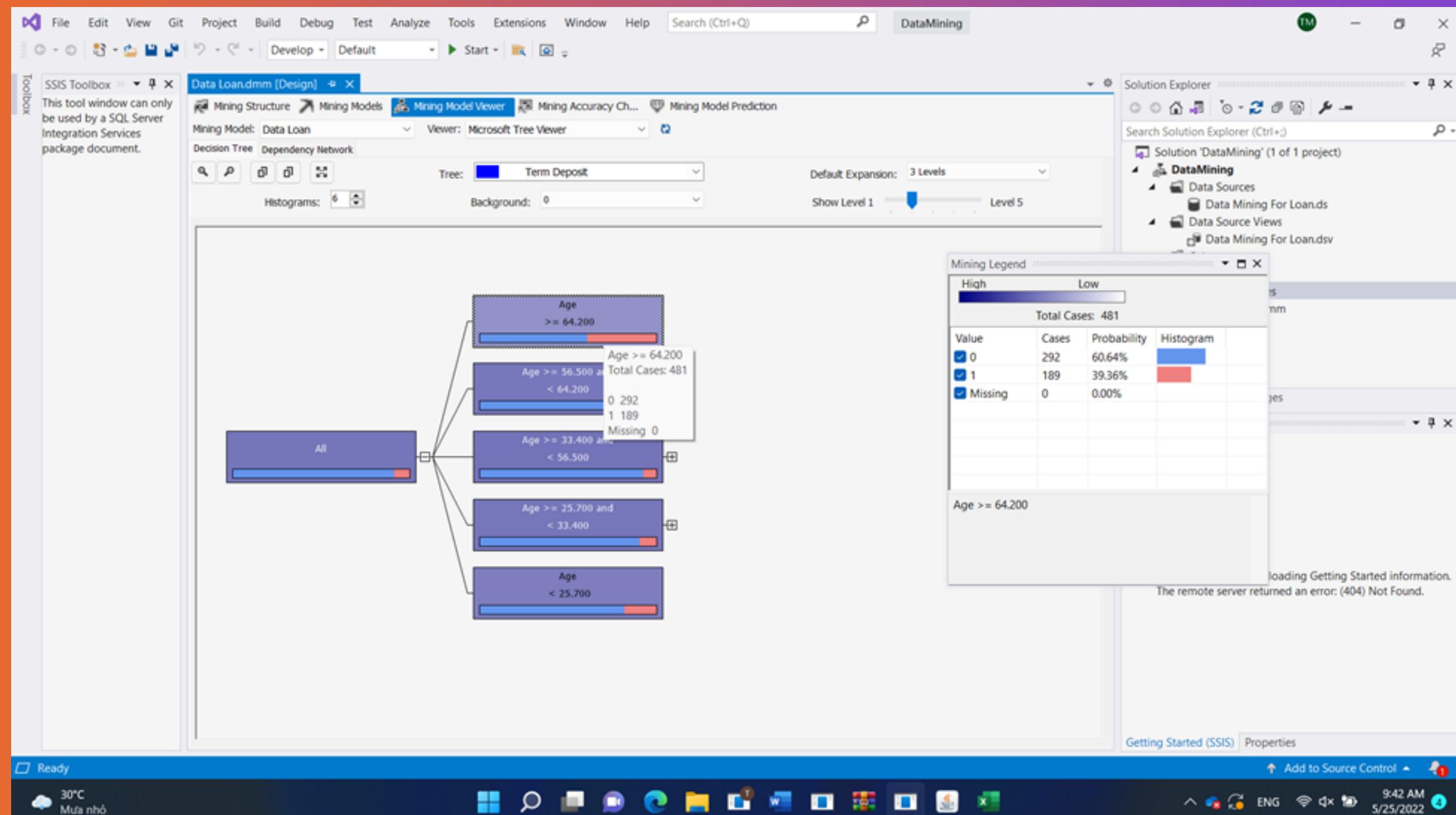
Mức phụ thuộc cấp 2



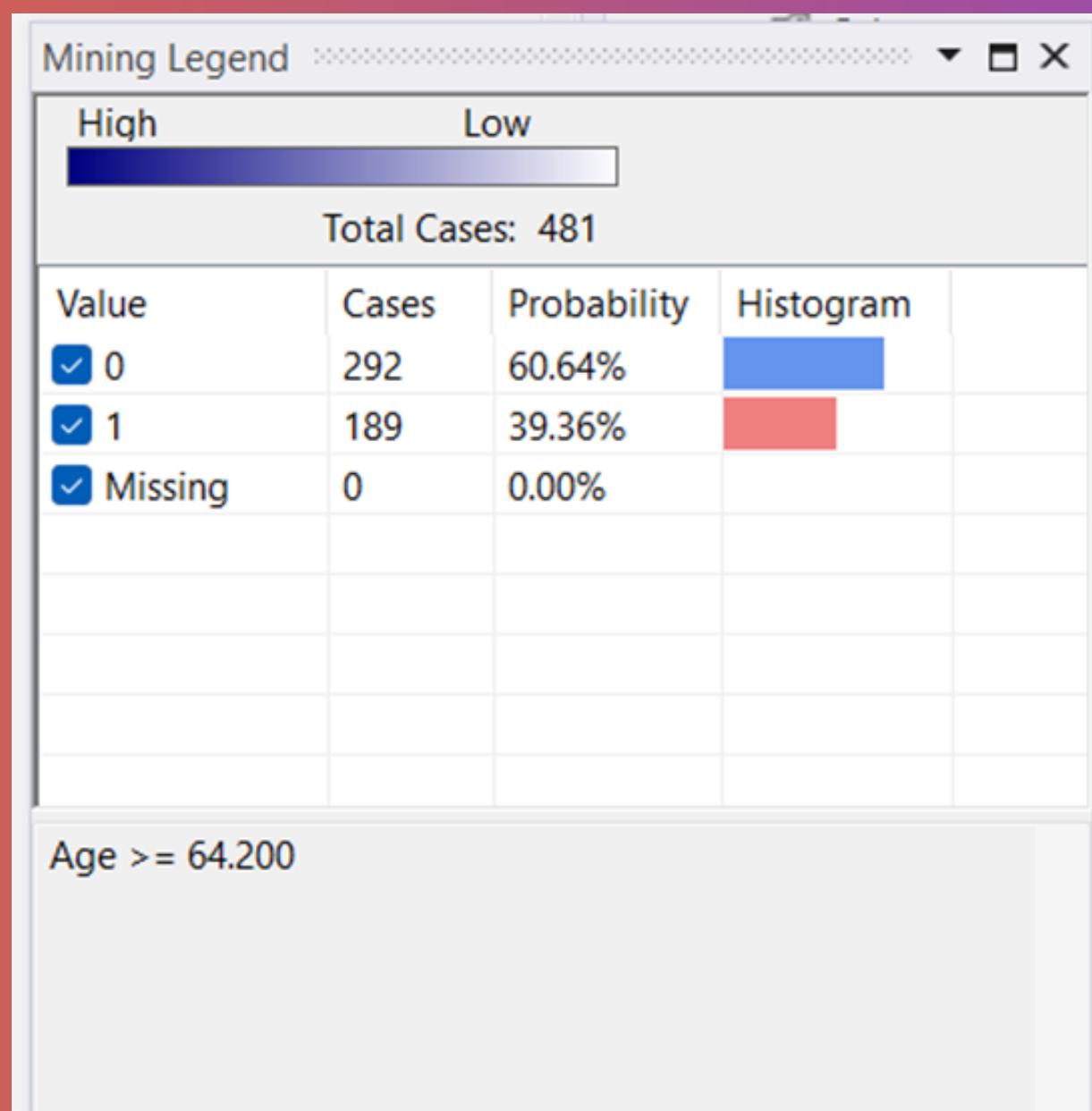
Mức phụ thuộc cấp 1



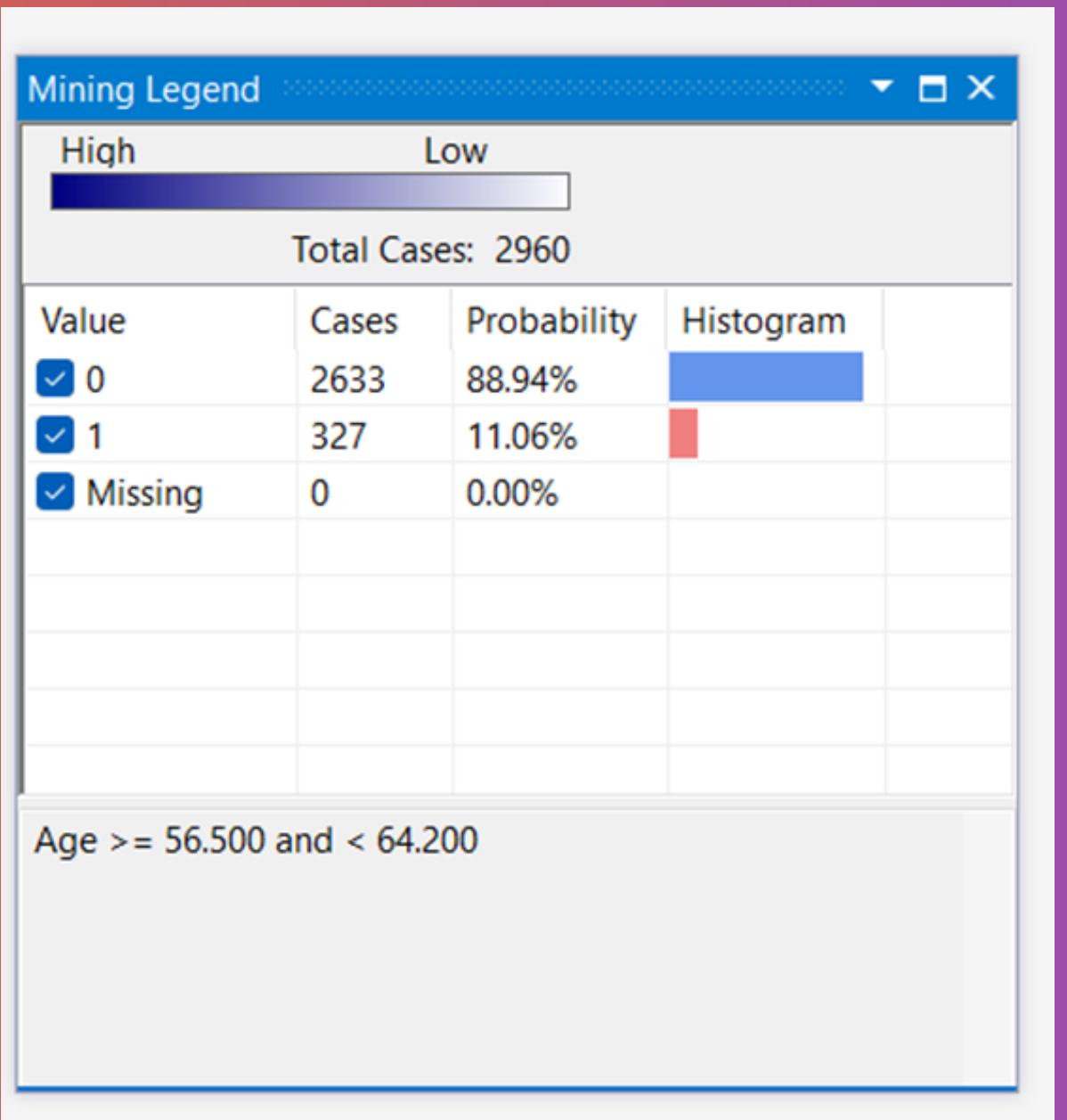
Chọn Background là 0 để dự đoán



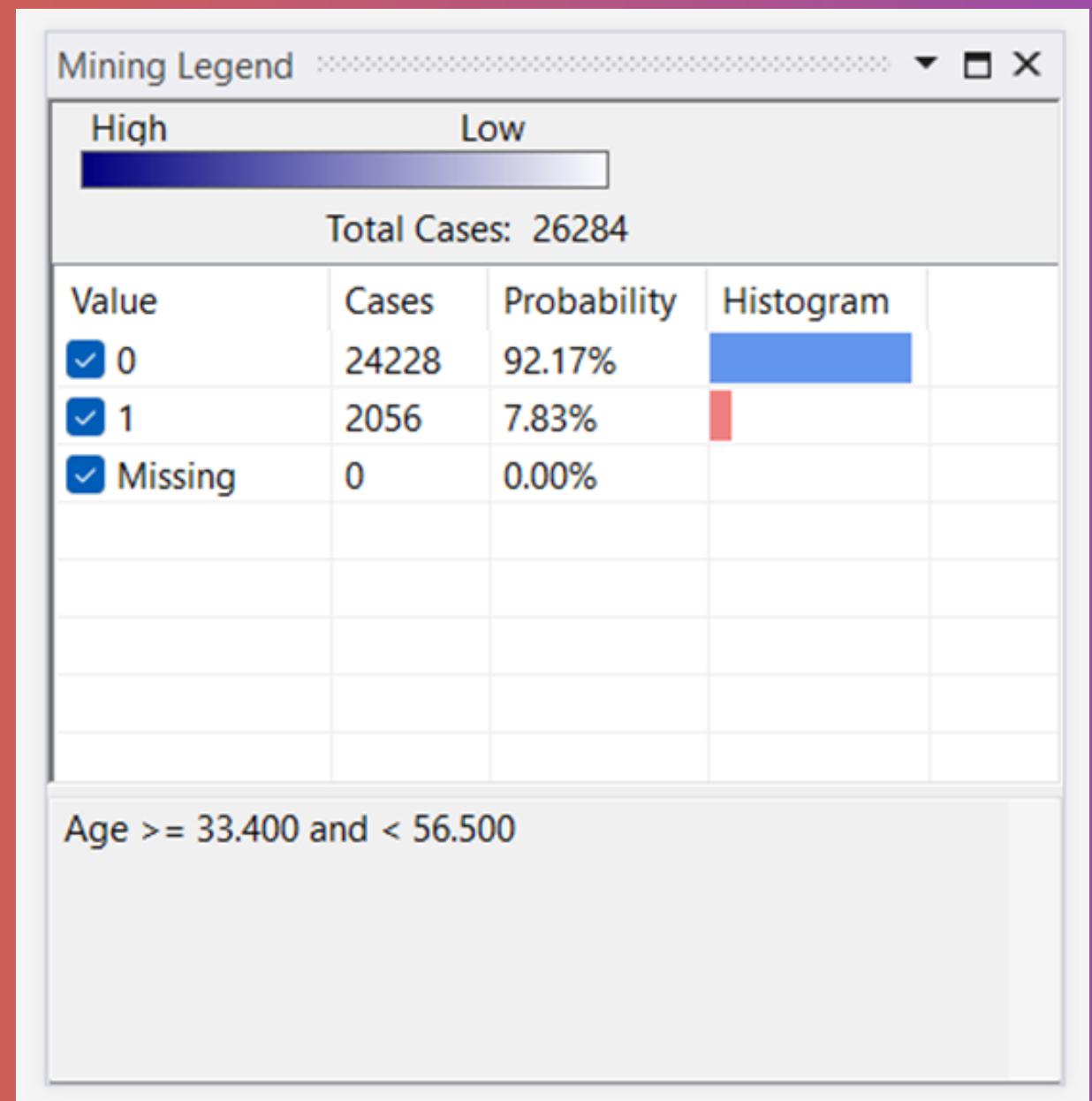
Ở mức 2 dự đoán về độ tuổi của khách hàng



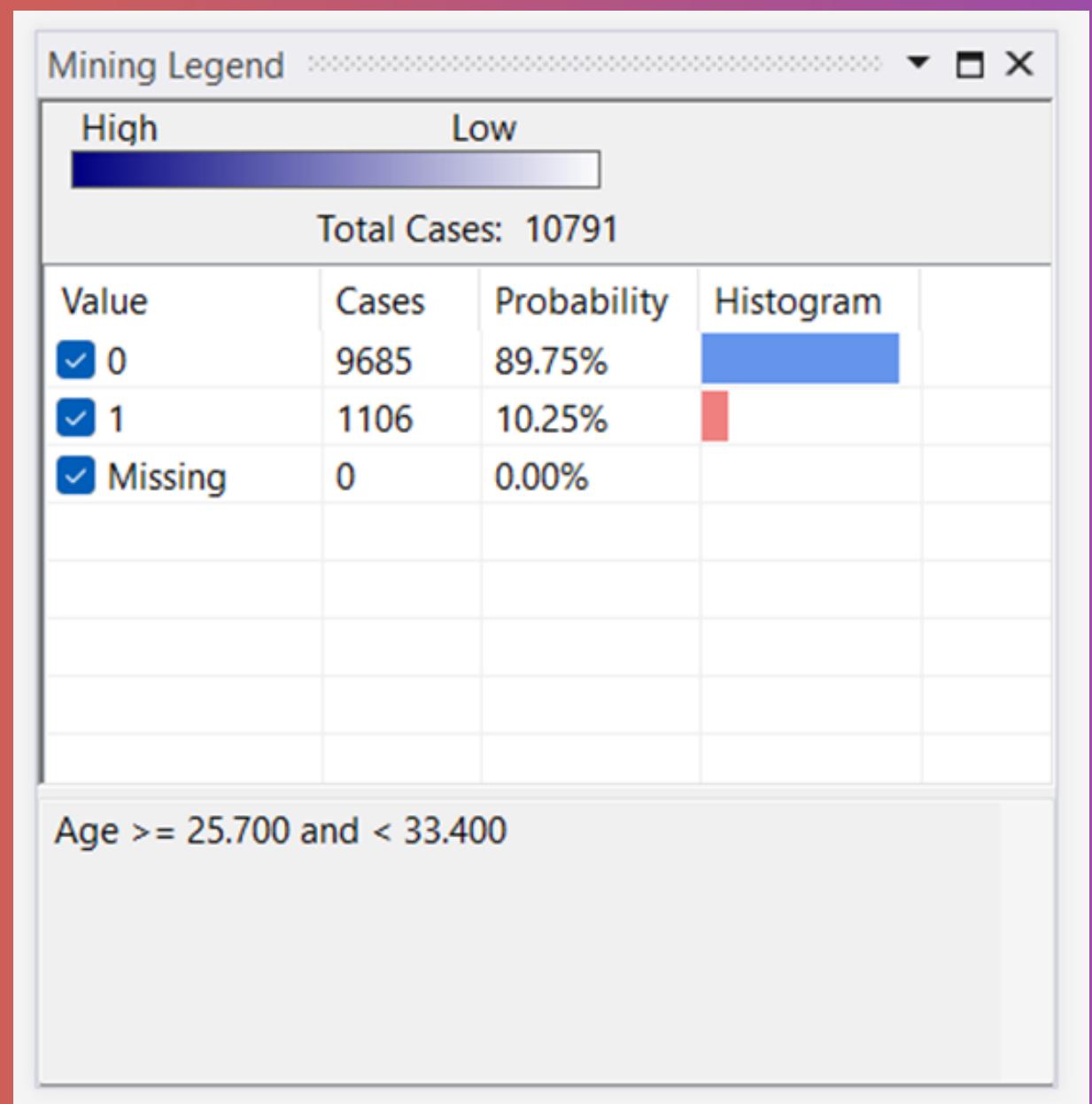
Độ tuổi lớn hơn 64.2 chiếm
60.64%



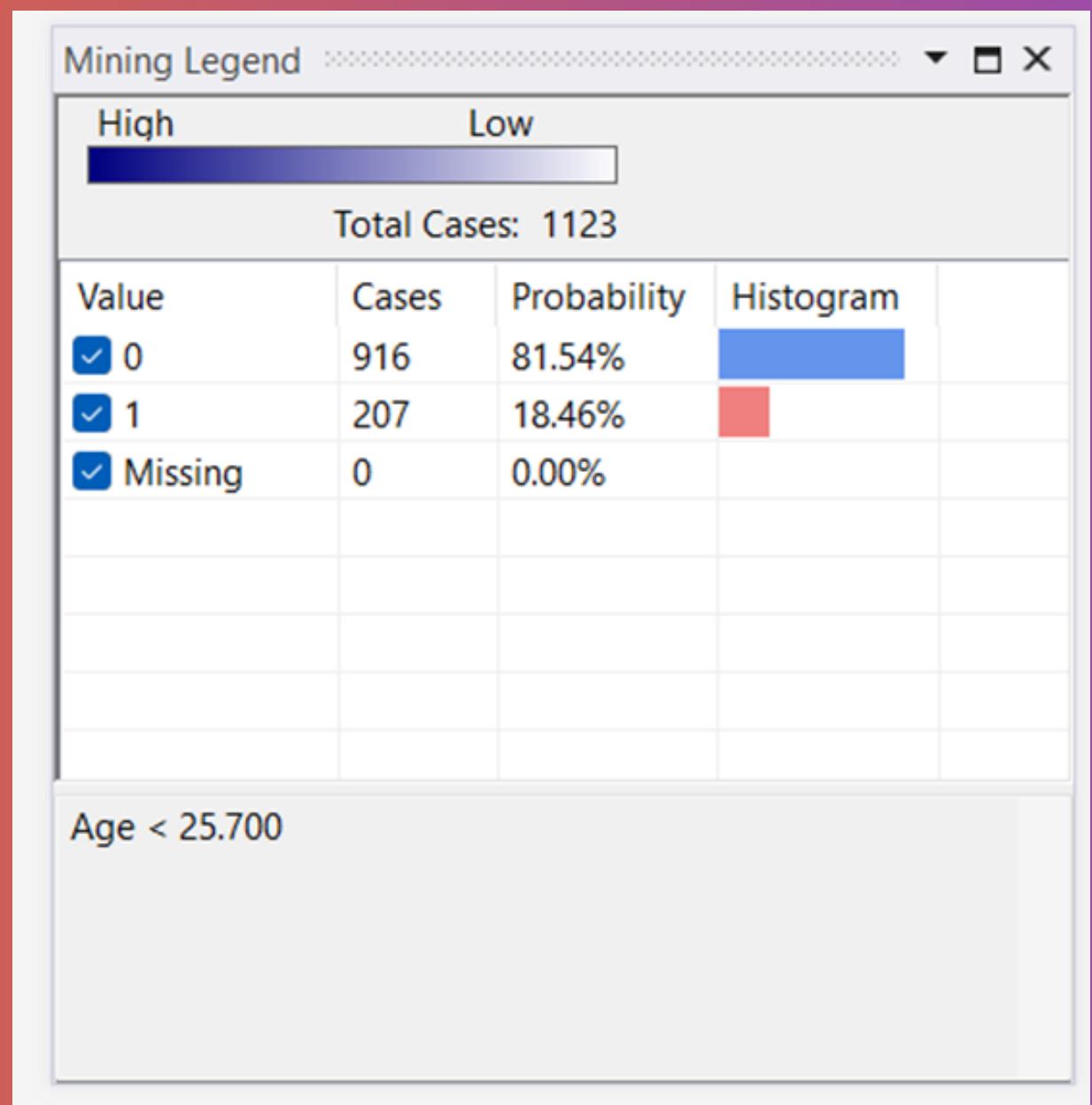
Độ tuổi lớn hơn 56.5 và nhỏ
hơn 64.2 chiếm 88.94%



Độ tuổi lớn hơn 33.4 và nhỏ
hơn 56.5 chiếm 92.17%

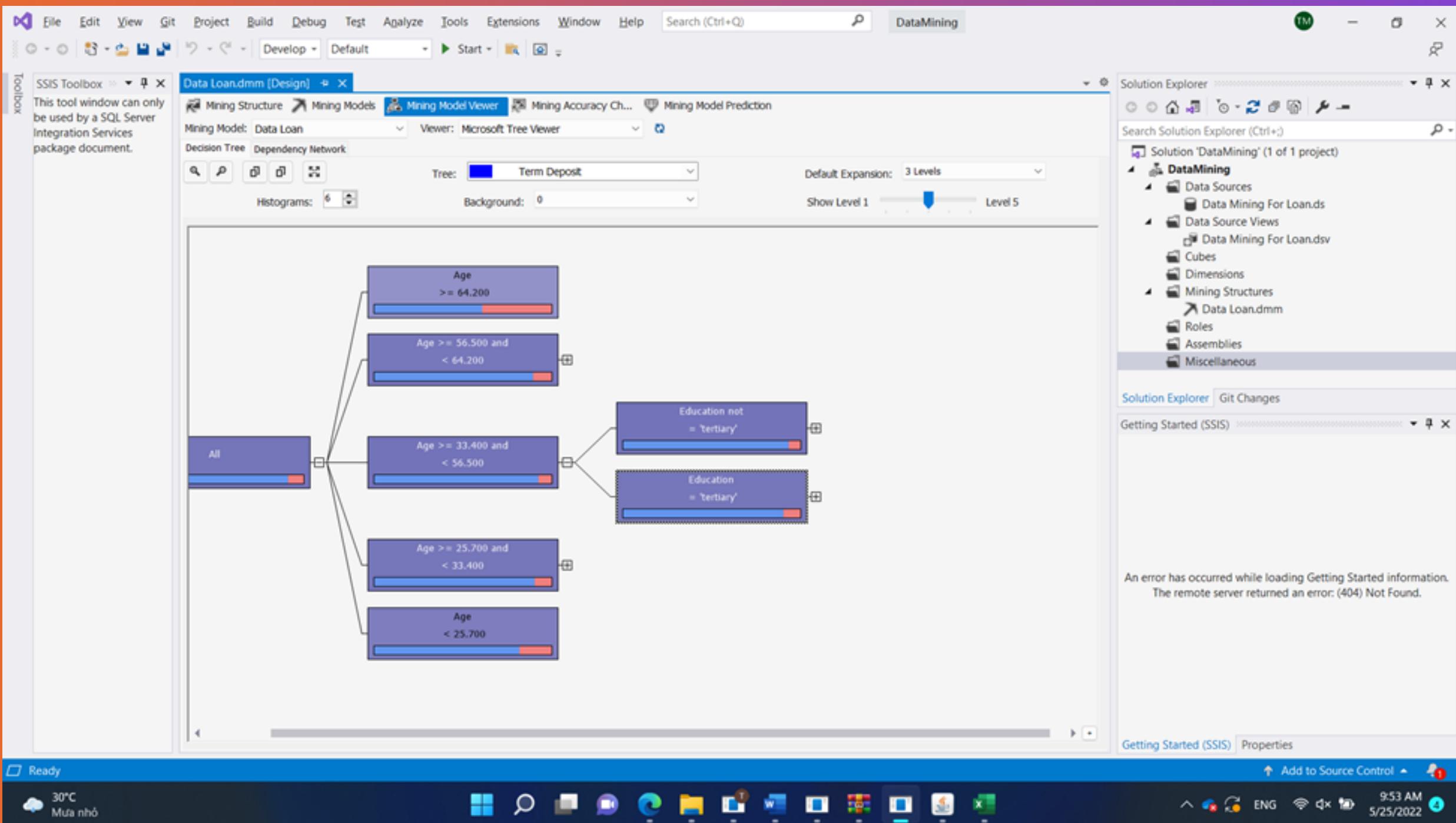


Độ tuổi lớn hơn 25.7 và nhỏ
hơn 33.4 chiếm 89.75%

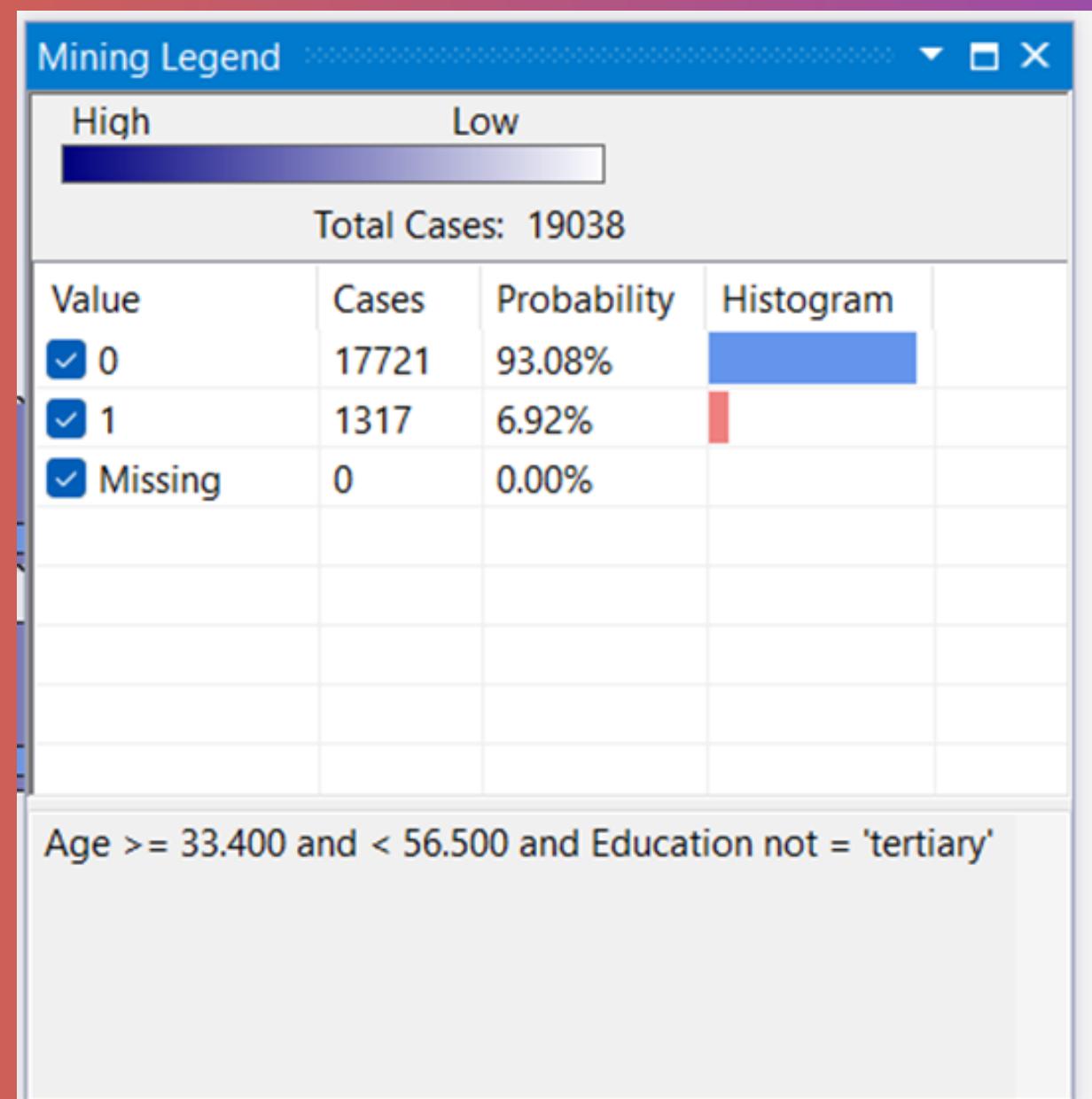


Độ tuổi nhỏ hơn 25.7 chiếm
81.54%

Vì độ tuổi lớn hơn 33.4 và nhỏ hơn 56.5 chiếm 92.17% là cao nhất nên ta quyết định đi theo nhánh này và tiếp tục dự đoán mức tiếp theo.

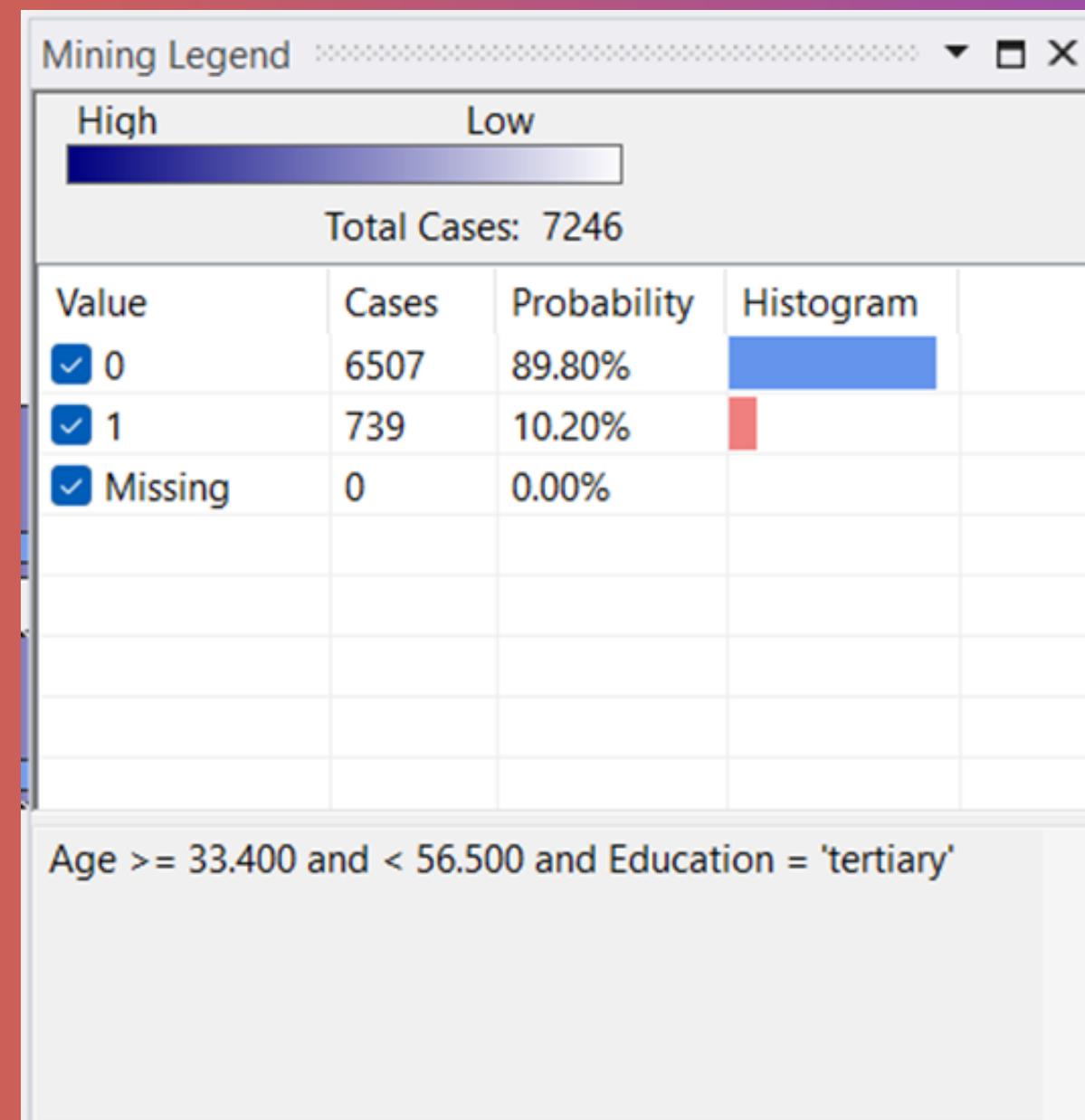


Ở mức 3 dự đoán về giáo dục của khách hàng (education)

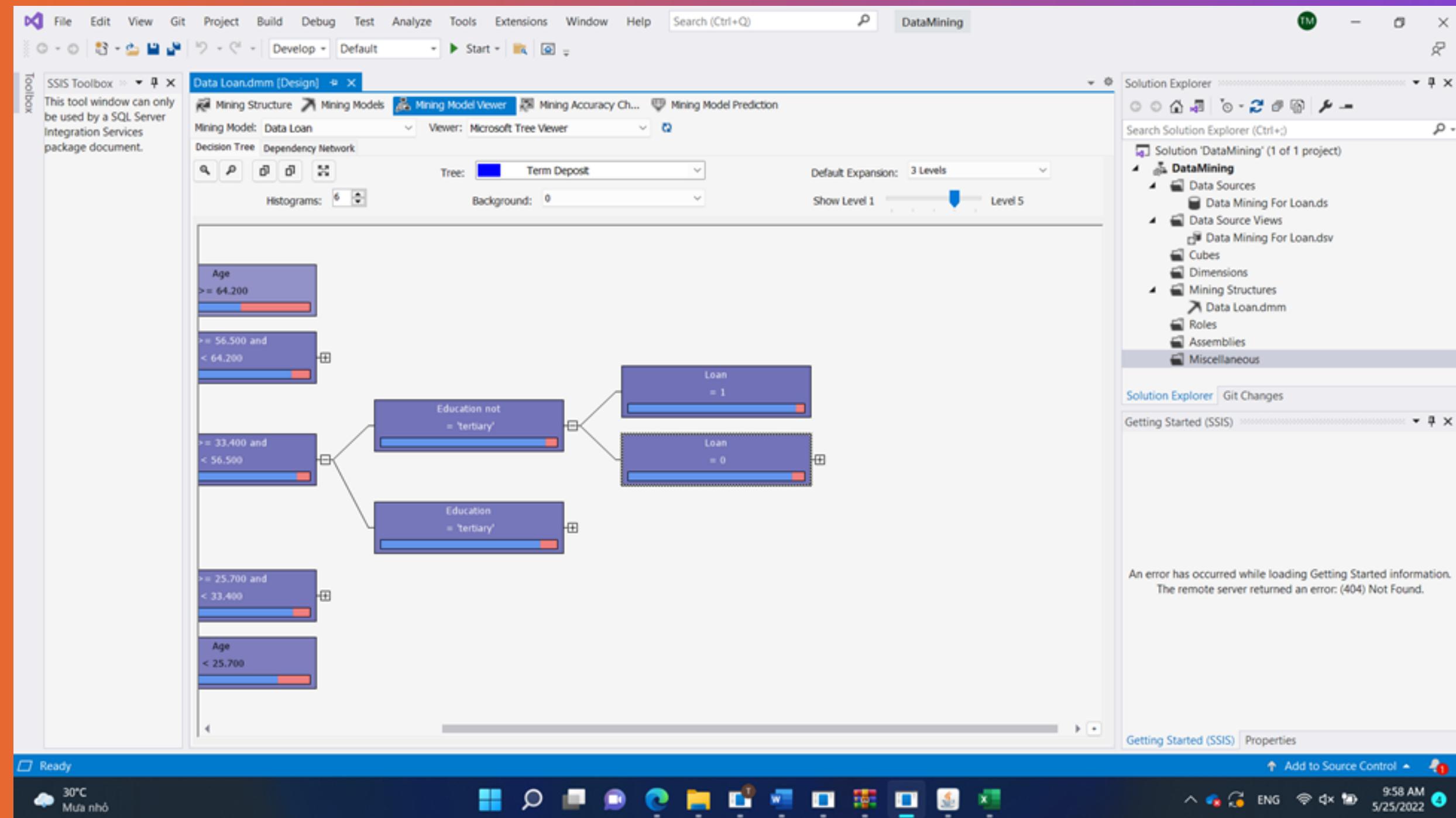


Education not = 'tertiary' chiếm
93.08%

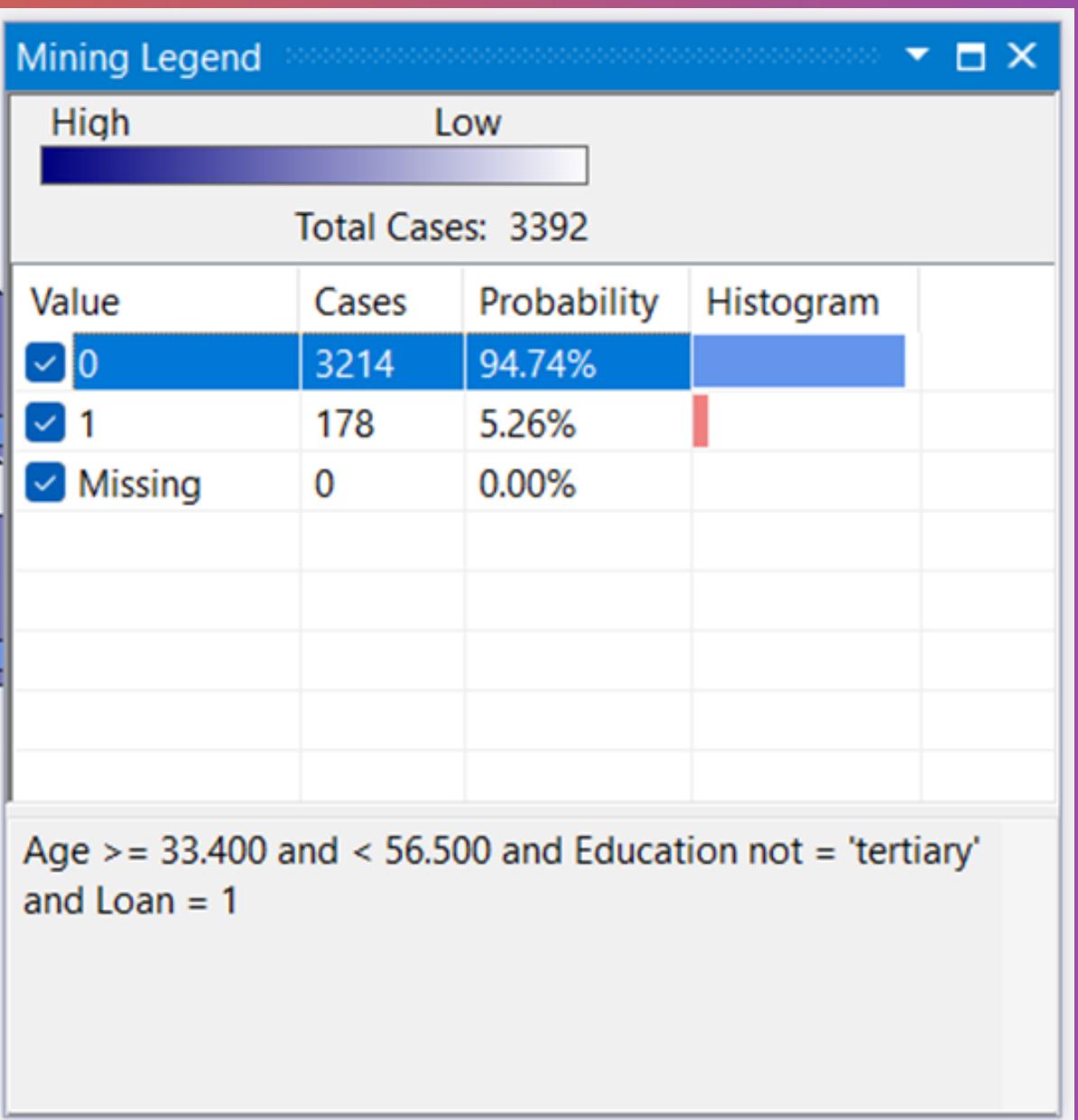
Vì Education not =
'tertiary' chiếm
93.08%
cao hơn nên ta
quyết định đi theo
nhánh này và dự
đoán mức tiếp
theo



Education = 'tertiary' chiếm
89.8%

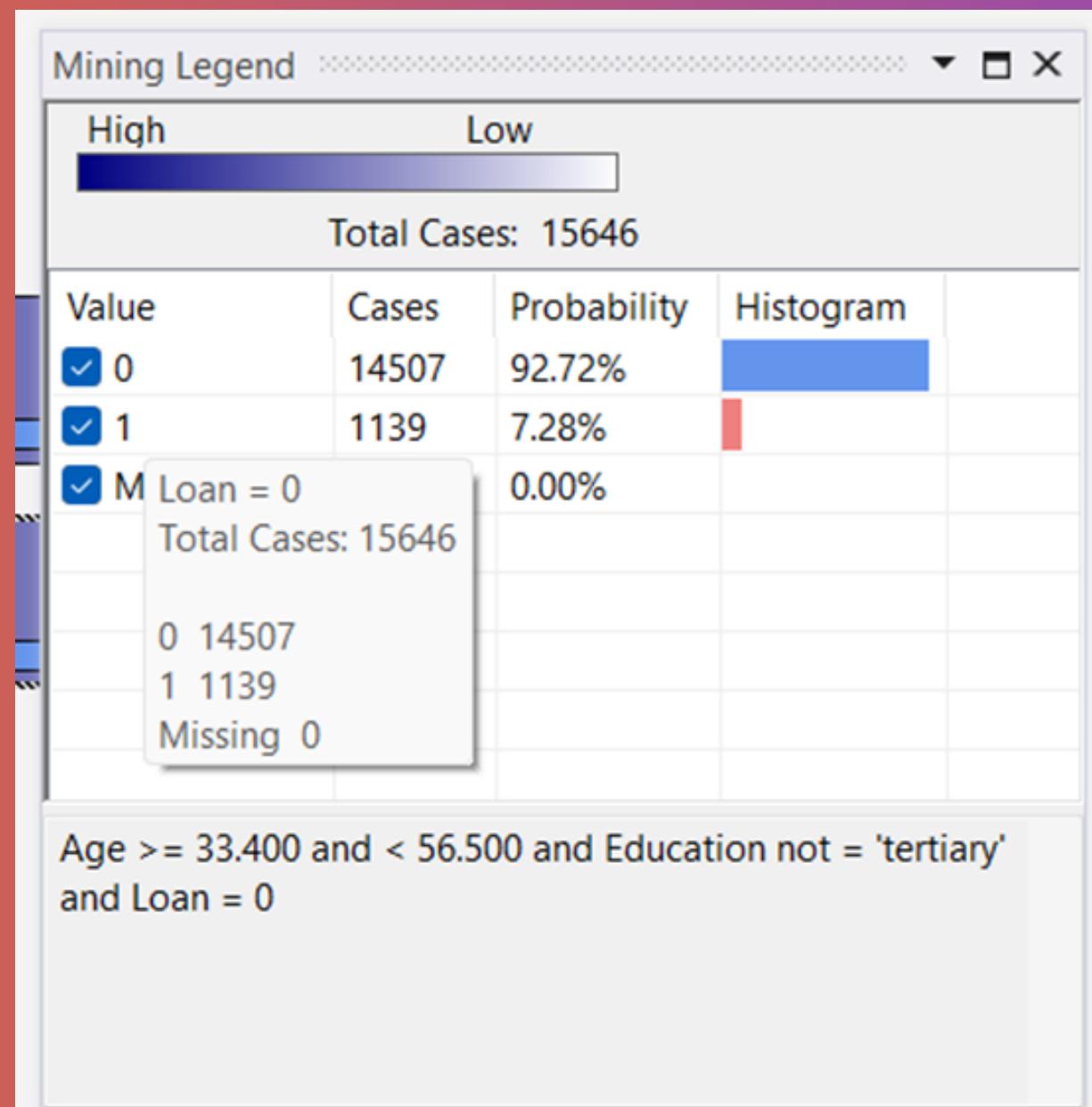


Ở mức 4 ta dự đoán khoản vay cá nhân của khách hàng

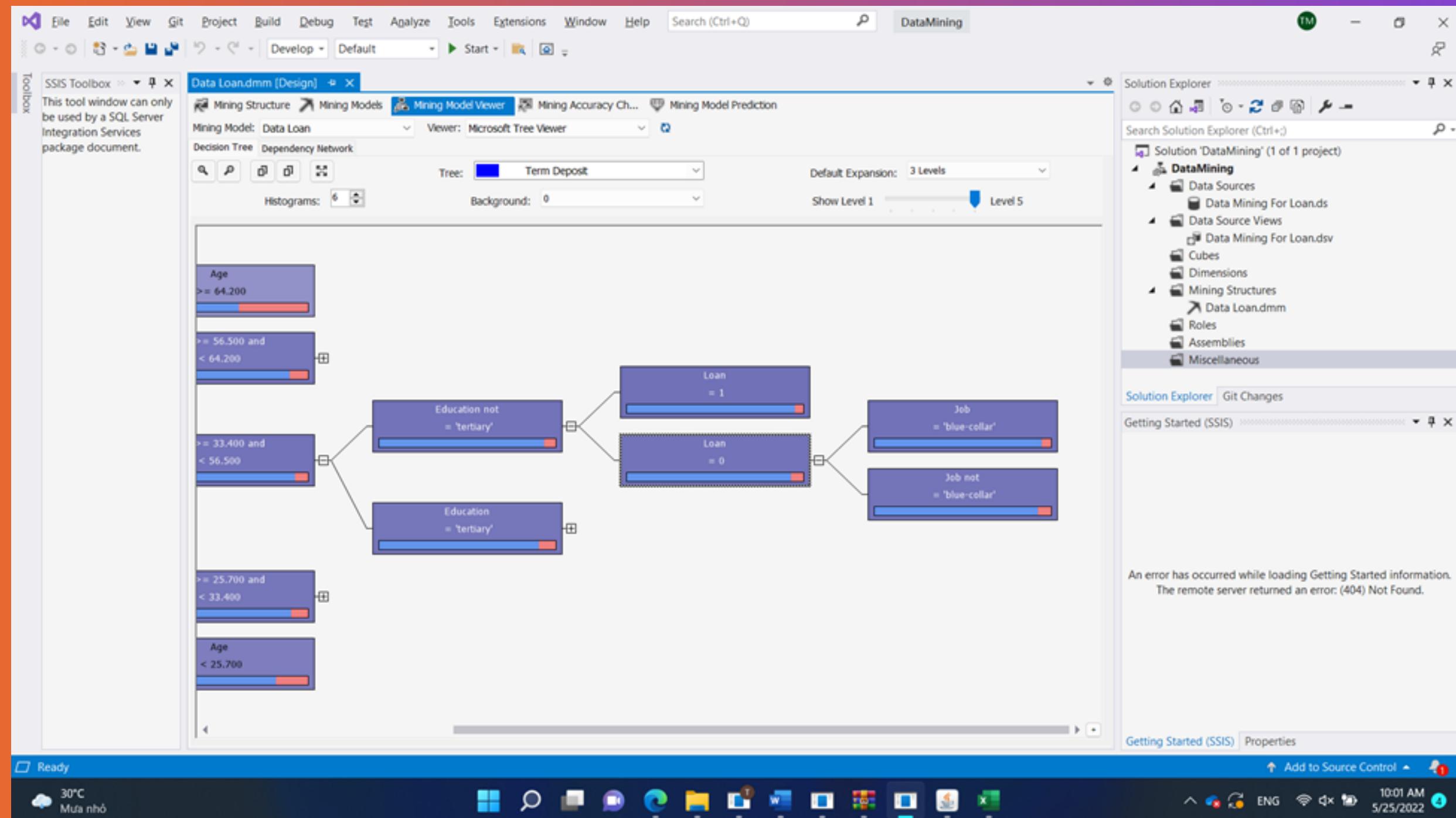


Khoản vay loan = 1 chiếm

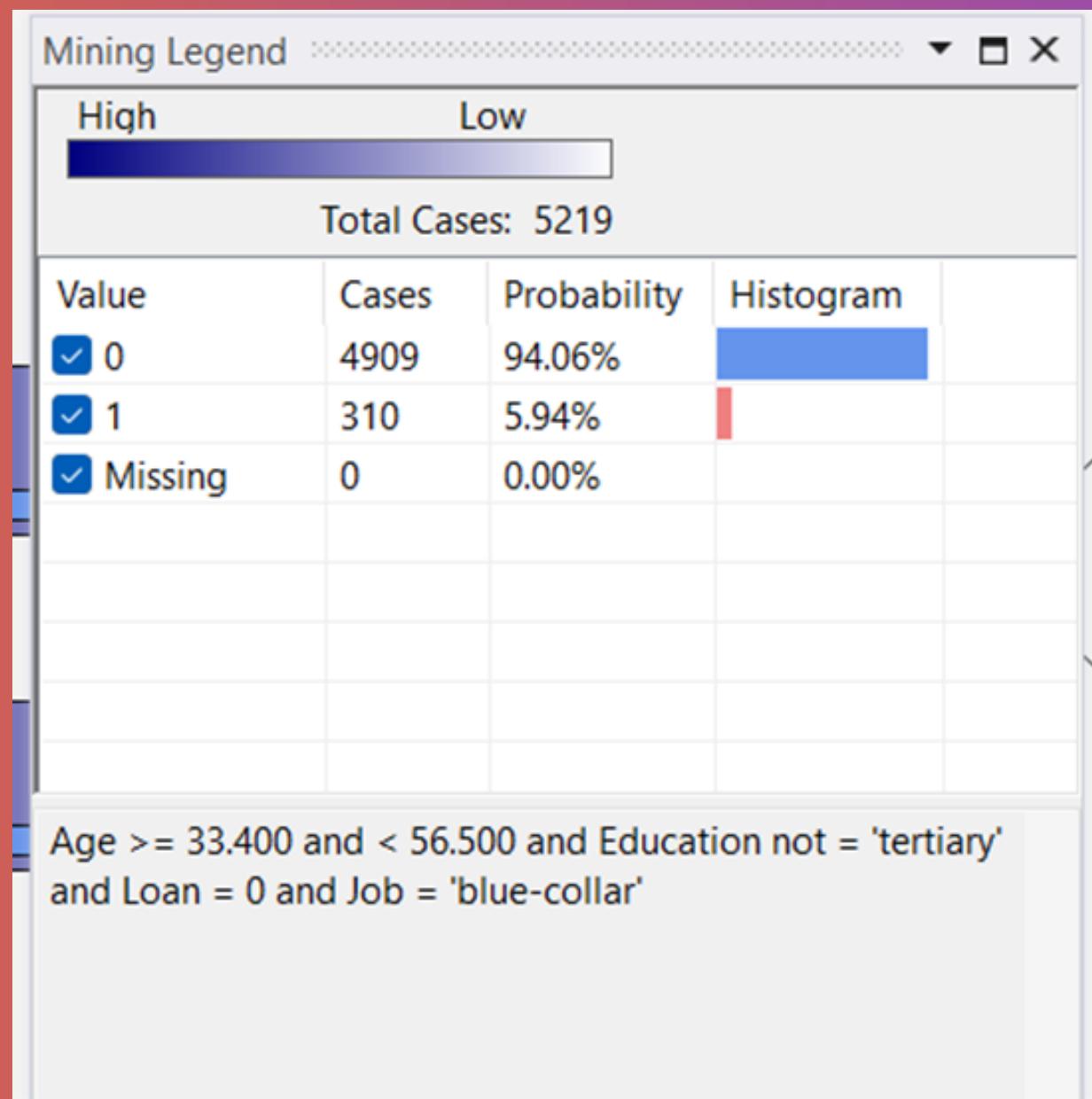
94.74%



Khoản vay loan = 0 chiếm
92.97%

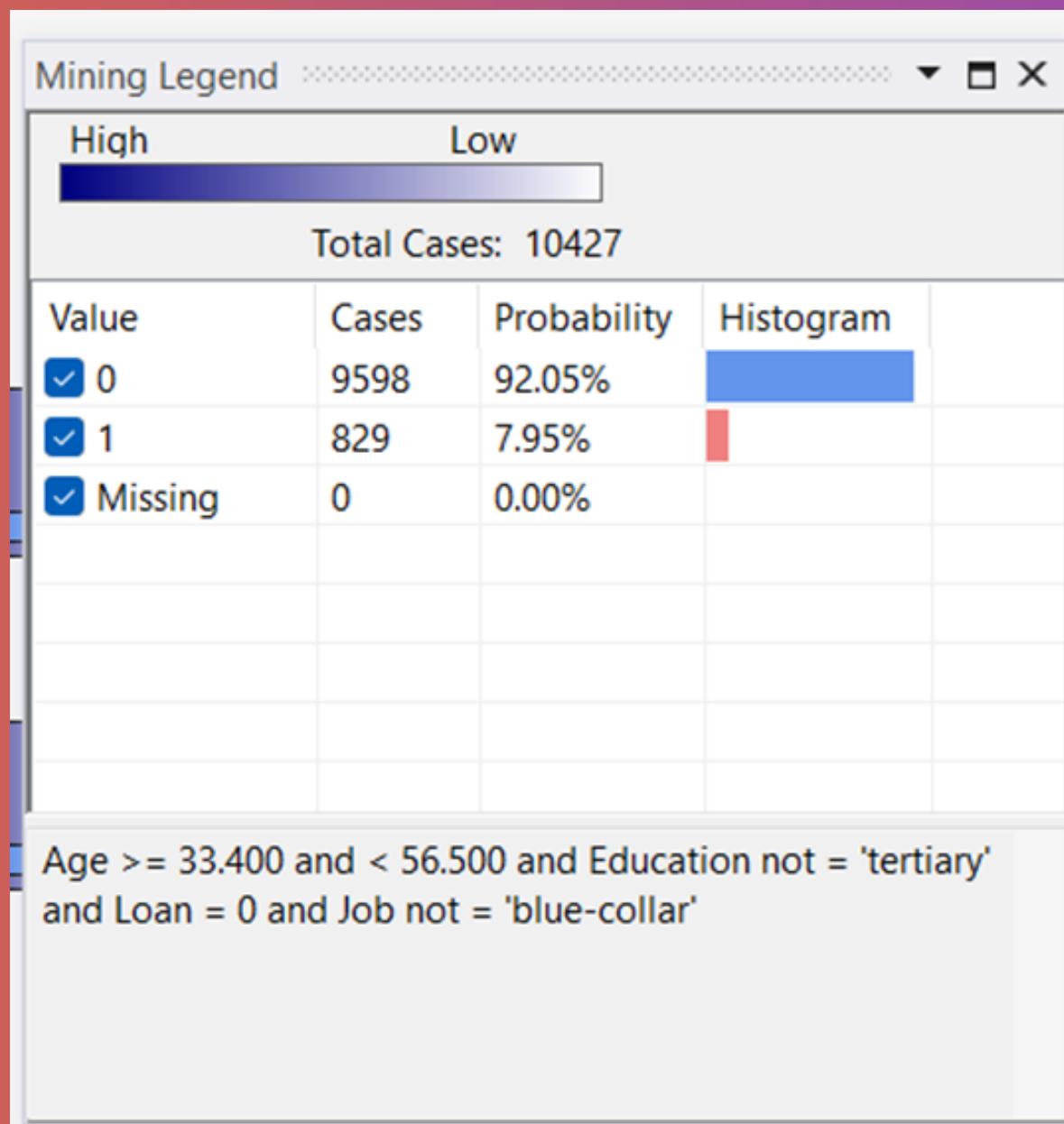


Ở mức 5 ta dự đoán công việc của khách hàng (job)



Công việc job = 'blue-collar'
chiếm 94.06%

Vì phần trăm của job = 'blue-collar' cao nhất nên ta chọn

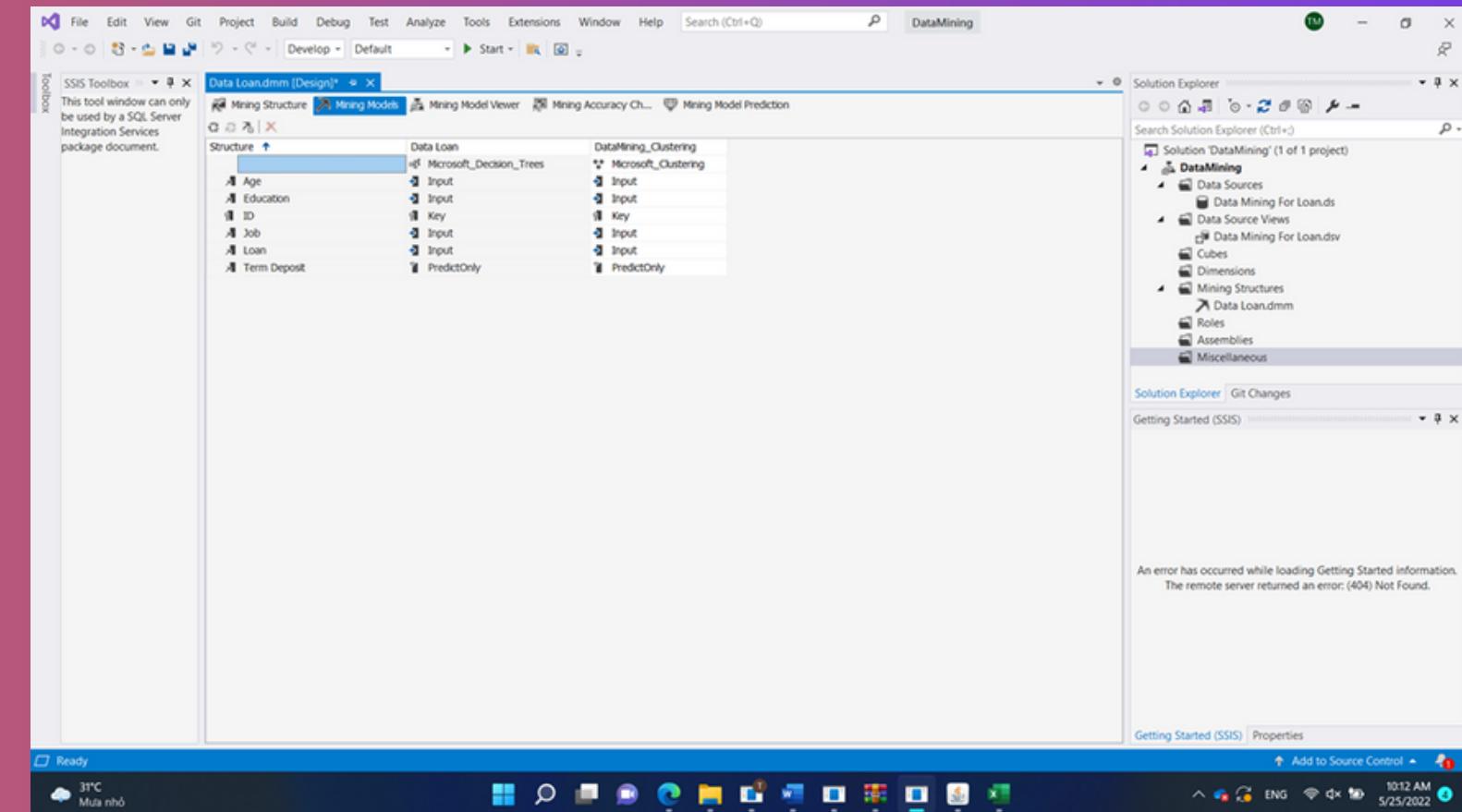
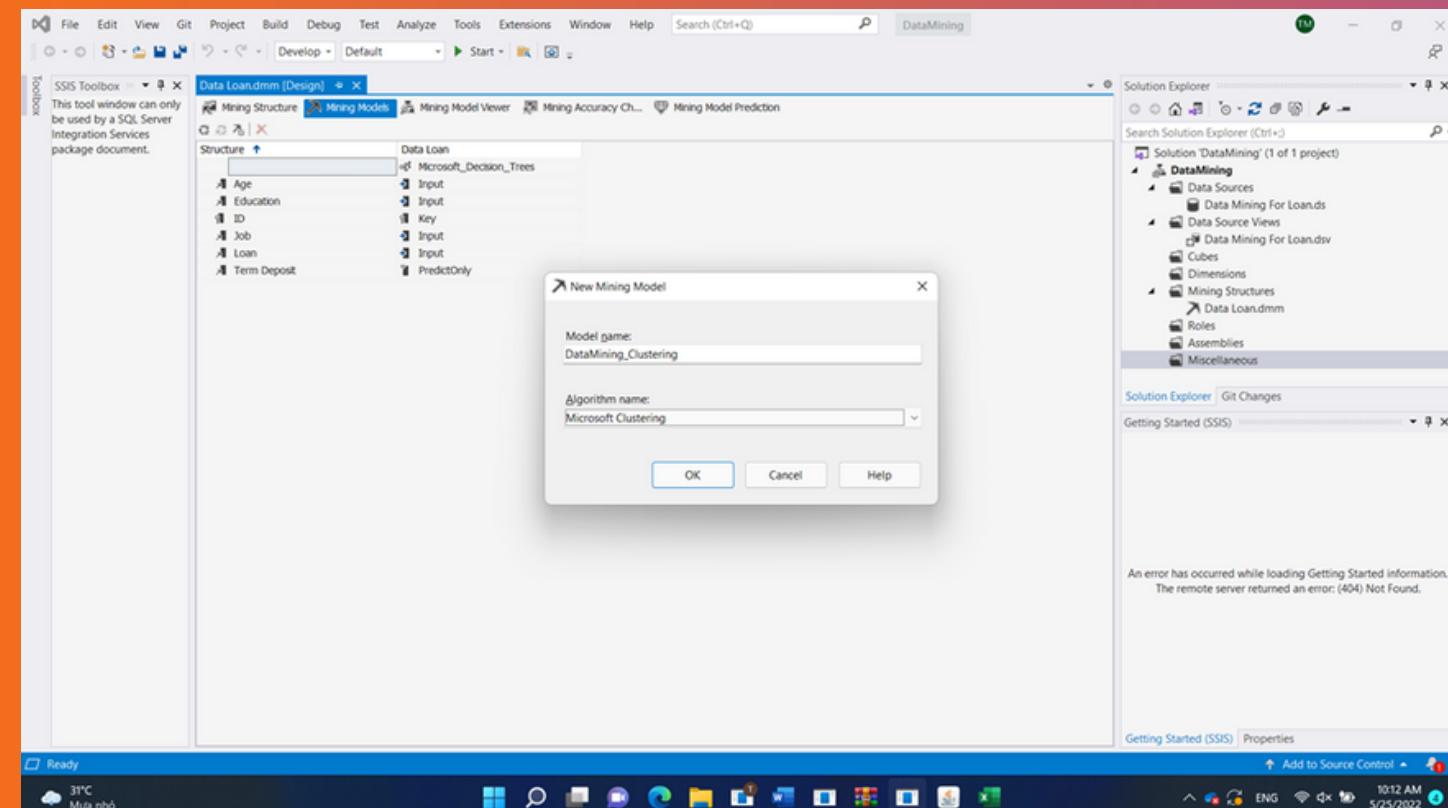


Công việc job not = 'blue-collar'
chiếm 92.05%

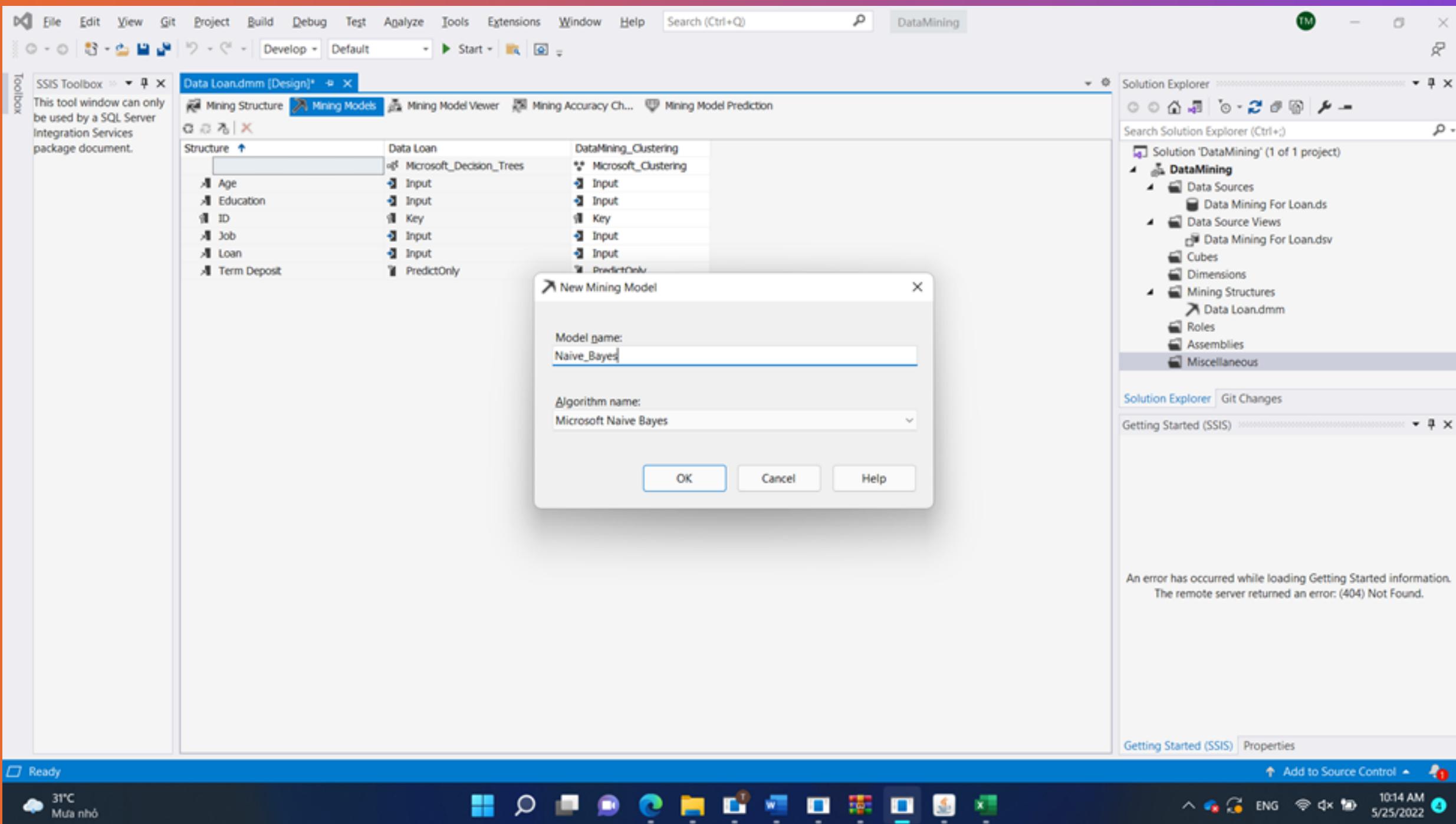
. Nhận xét kết quả sau khi mining dữ liệu

Những khách hàng có độ tuổi từ 33.4 đến 56.5 có mức giáo dục là đại học và có khoản vay với công việc là 'blue-collar' thì có khả năng khách hàng chưa đăng ký một khoản tiền có kỳ hạn

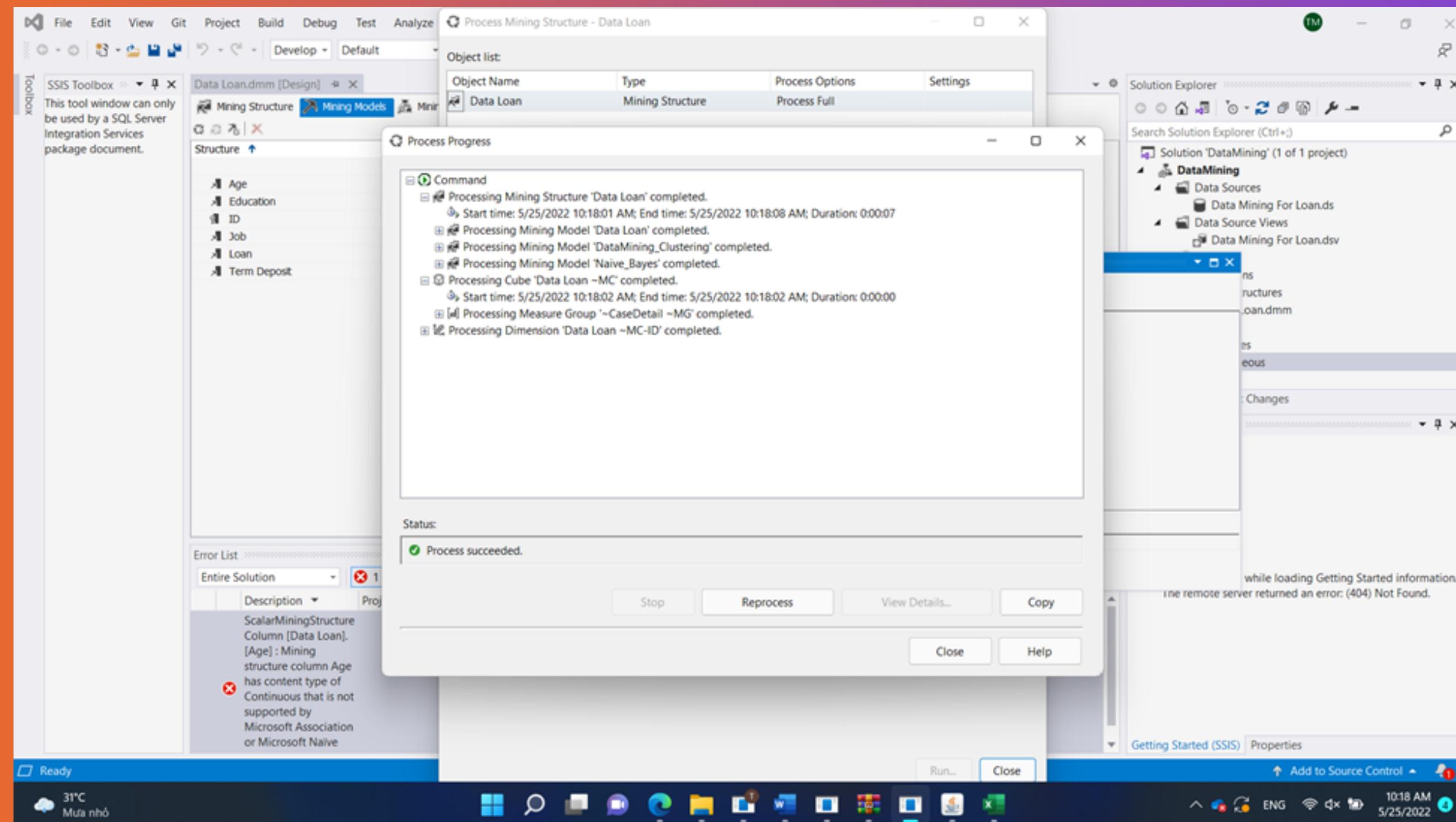
Đánh giá các thuật toán bằng Mining Accurary Chart



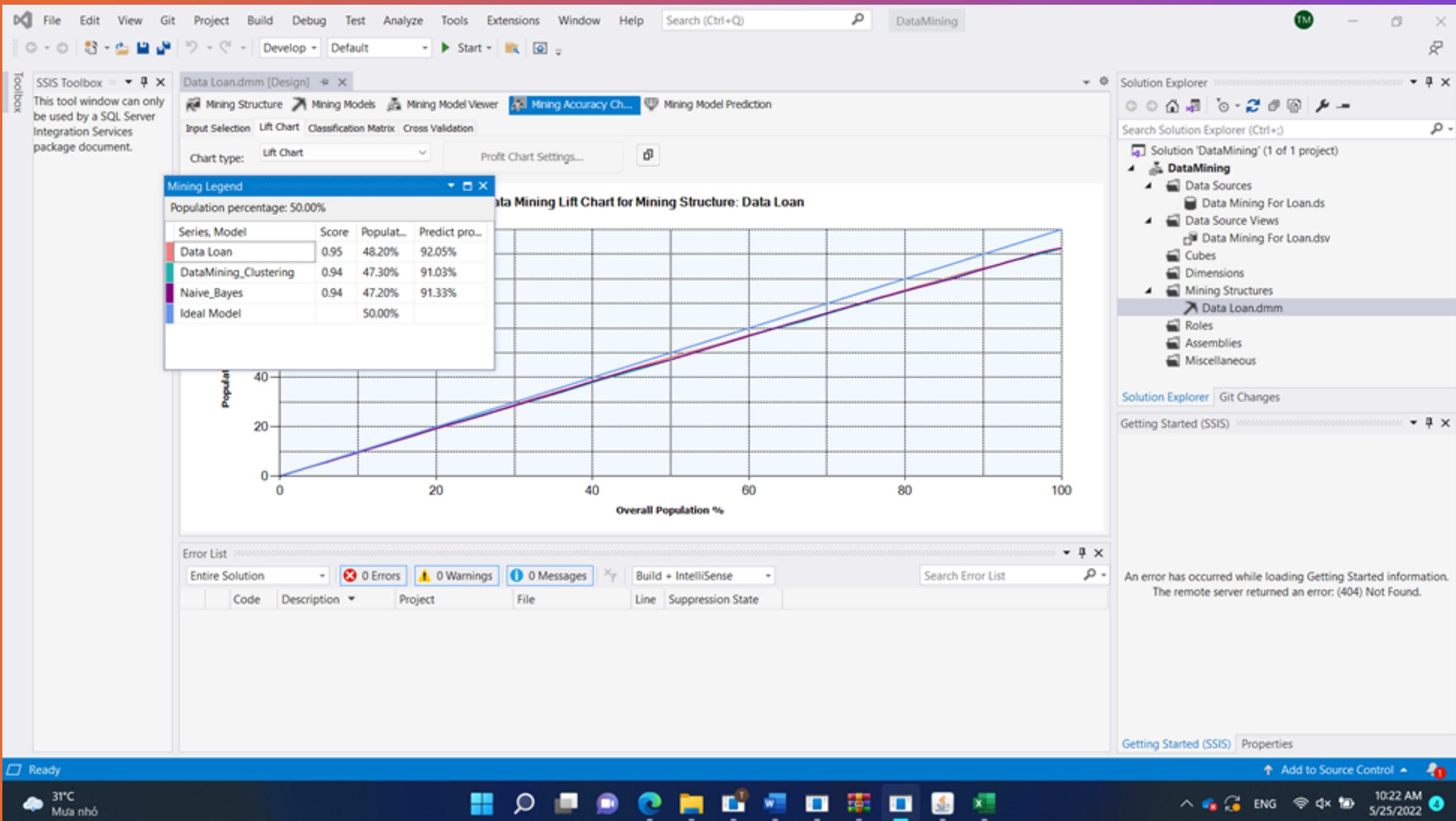
Click phải chuột vào vùng trống chọn New Mining Model
Chọn thuật toán là Microsoft Clustering



Tiếp tục tạo thêm thuật toán Mining Microsoft Naïve Bayes



Tiến hành Process



So sánh và đánh giá các thuật toán bằng đồ thị Lift
Để xem đồ thị Lift ta vào trong tab Mining Accuracy Chart chọn Lift Chart

Thuật toán Decision Tree có phần trăm chính xác 48.2% với số điểm 0.95

Thuật toán Clustering có phần trăm chính xác là 47.3% với số điểm 0.94

Thuật toán NaiveBayes có phần trăm chính xác là 47.2% với số điểm 0.94

=> Ta ưu tiên chọn thuật toán Decision Tree





STUDIO SHODWE

Thank You

SLIDE PRESENTATIONS DESIGN