

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
-----OOBOOKOO-----  
**BÁO CÁO ĐÒ ÁN**  
**Năm học: 2021 – 2022**



## **KHO DỮ LIỆU VÀ OLAP**

### **ĐỀ TÀI: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU VỀ BẢO HIỂM TAI NẠN XE**

**Giảng viên hướng dẫn: Ths. Nguyễn Thị Kim Phụng**

**Sinh viên thực hiện: Phạm Thảo Nhi – 19520815**

**Lớp: IS217.M21**

**TP. Hồ Chí Minh, tháng 06 năm 2022**

## MỤC LỤC

<b>NHẬN XÉT CỦA GIẢNG VIÊN</b> .....	5
<b>CHƯƠNG 1: MÔ TẢ DỮ LIỆU</b> .....	6
1.1. Tổng quan về dữ liệu.....	6
1.2. Mô tả thuộc tính .....	6
1.2.1. Bảng Fact.....	6
1.2.2. Bảng dim_driver.....	6
1.2.3. Bảng dim_gender .....	7
1.2.4. Bảng dim_marital_status.....	7
1.2.5. Bảng dim_area .....	7
1.2.6. Bảng dim_living_status.....	7
1.2.7. Bảng dim_time .....	7
1.2.8. Bảng dim_vehicle.....	7
<b>CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU – QUÁ TRÌNH SSIS</b> .....	8
2.1. Tạo mới một database .....	8
2.2. Tạo project mới với SSIS .....	8
<b>CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN _ QUÁ TRÌNH SSAS.</b> 13	13
3.1. Cấu hình project .....	13
3.1.1. Tạo datasource .....	13
3.1.2. Tạo datasource views.....	15
3.1.3. Tạo Cubes .....	16
3.1.4. Phân cấp dim_time.....	18
3.1.5. Process cube.....	19
3.2. Thực thi truy vấn .....	20
3.2.1. Roll up.....	20
3.2.2. Drill down .....	21
3.2.3. Slice and Dice .....	22
3.2.3.1. Đếm số báo cáo theo ngày với loại xe là Compact.....	22
3.2.3.2. Thống kê số lượng báo cáo với giới tính của tài xế là nữ theo từng năm .....	23
3.2.3.3. Thống kê số báo cáo theo từng tình trạng hôn nhân của tài xế tháng 5 và tháng 7 theo từng năm.....	23
3.2.3.5. Thống kê số lượng báo cáo với giới tính của tài xế là nữ và sở hữu nhà ở theo từng năm .....	24
3.2.4. Pivot .....	25
3.2.4.1. Thống kê số báo cáo được ghi nhận theo năm và loại xe .....	25
3.2.4.2. Thống kê số báo cáo theo zip code và giới tính.....	26
3.2.4.3. Thống kê số báo cáo được ghi nhận theo giới tính, loại xe và tình trạng nhà ở.....	29

3.3. Ngôn ngữ MDX .....	30
3.3.1. Đếm số báo cáo theo từng giới tính .....	30
3.3.2. Đếm số báo cáo theo từng tháng, năm .....	30
3.3.3. Đếm số báo cáo theo từng năm, từng quý .....	31
3.3.4. Đếm số báo cáo theo từng giới tính và loại xe là Large .....	31
3.3.5. Tìm tuổi cao nhất của các tài xế được claim bảo hiểm theo từng loại xe .....	32
3.3.6. Thống kê tổng thu nhập của những tài xế được claim bảo hiểm theo tình trạng nhà ở trong năm 2015 .....	32
3.3.7. Thống kê số báo cáo lớn hơn 2000 theo tình trạng hôn nhân và loại xe .....	33
3.3.8. Thống kê số báo cáo theo zip code và giới tính của tài xế với loại xe Compact .....	33
3.3.9. Thống kê tổng số lần claim bảo hiểm trước đó của những báo cáo trong năm 2015 và giới tính tài xế là nam .....	34
3.3.10. Thống kê top 3 báo cáo có tài xế có thu nhập cao nhất trong năm 2016 .....	34
3.3.11. Lọc những ngày có số báo cáo trên 30 .....	35
3.3.12. Thống kê số báo cáo gian lận theo ngày .....	35
3.3.13. Thống kê năm có nhiều báo cáo gian lận nhất .....	36
3.3.14. Thống kê top 3 báo cáo có số lần claim bảo hiểm trước đó cao nhất trong năm 2015 .....	37
3.3.15. Thống kê tình trạng hôn nhân có số báo cáo gian lận nhiều nhất trong quý 3 năm 2016 .....	37
3.3.16. Pivot: Thống kê số báo cáo theo ngày và theo tình trạng nhà ở .....	38
3.3.17. Pivot: Thống kê số báo cáo theo giới tính và theo tình trạng hôn nhân .....	39
3.3.18. Pivot: Thống kê số báo cáo gian lận theo loại xe và theo zip code .....	40
<b>CHƯƠNG 4: KHAI PHÁ DỮ LIỆU – DATAMINING .....</b>	<b>41</b>
4.1. Mining Structure .....	41
4.2. Mining Model.....	45
<b>CHƯƠNG 5: POWER BI .....</b>	<b>48</b>
5.1. Tạo Project Power BI .....	48
5.2. Tạo báo cáo dựa trên giới tính của tài xế .....	49
5.3. Tạo báo cáo dựa trên tình trạng hôn nhân của tài xế .....	49
5.4. Thống kê số báo cáo theo năm và theo giới tính.....	49
5.5. Thống kê số báo cáo gian lận theo từng năm .....	50
5.6. Thống kê số báo cáo theo từng loại xe.....	50
5.7. Thống kê số báo cáo và số báo cáo gian lận theo từng giới tính .....	50
5.8. Thống kê số báo cáo và số báo cáo gian lận theo tình trạng nhà ở .....	51

5.9. Thống kê số báo cáo theo từng tháng, năm.....	51
5.10. Thống kê số báo cáo theo từng quý, năm.....	51
5.11. Pivot: Thống kê số báo cáo theo từng năm và tình trạng nhà ở.....	52
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>53</b>

## NHẬN XÉT CỦA GIẢNG VIÊN

---

---

---

---

## CHƯƠNG 1: MÔ TẢ DỮ LIỆU

### 1.1. Tổng quan về dữ liệu

- Bộ dữ liệu được thu thập từ kaggle (link bộ dữ liệu: <https://www.kaggle.com/datasets/surekharamireddy/fraudulent-claim-on-cars-physical-damage>)

- Bộ dữ liệu có 24 thuộc tính và 17996 dòng.
- Bộ dữ liệu bao gồm các thuộc tính chi tiết của một báo cáo claim bảo hiểm xe, gồm có: Chỉ số an toàn, Tuổi của tài xế, Giới tính, Tình trạng hôn nhân, Ngày claim bảo hiểm, Loại xe, Tuổi của xe, Giá xe, ...

### 1.2. Mô tả thuộc tính

#### 1.2.1. Bảng Fact

Tên thuộc tính	Mô tả
<b>Claim_number</b>	Khóa chính
<b>Claim_date</b>	Ngày claim
<b>Age_of_driver</b>	Tuổi của tài xế
<b>Safety_rating</b>	Chỉ số an toàn
<b>Annual_income</b>	Thu nhập
<b>Past_num_of_claims</b>	Số lần claim trước đó
<b>Claim_est_payout</b>	Số tiền dự tính chi trả
<b>Age_of_vehicle</b>	Tuổi của xe
<b>Vehicle_category</b>	Loại xe
<b>Vehicle_price</b>	Giá xe
<b>Vehicle_weight</b>	Cân nặng xe
<b>Fraud</b>	Báo cáo gian lận hay không?

#### 1.2.2. Bảng dim\_driver

Tên thuộc tính	Mô tả
<b>Claim_number</b>	Khóa chính
<b>gender</b>	Giới tính
<b>Marital_status</b>	Tình trạng hôn nhân
<b>Living_status</b>	Tình trạng nhà ở

<b>Zip_code</b>	Khu vực
-----------------	---------

#### 1.2.3. Bảng dim\_gender

Tên thuộc tính	Mô tả
<b>gender</b>	Giới tính

#### 1.2.4. Bảng dim\_marital\_status

Tên thuộc tính	Mô tả
<b>Marital_status</b>	Tình trạng hôn nhân

#### 1.2.5. Bảng dim\_area

Tên thuộc tính	Mô tả
<b>Zip_code</b>	Khu vực

#### 1.2.6. Bảng dim\_living\_status

Tên thuộc tính	Mô tả
<b>Living_status</b>	Tình trạng nhà ở

#### 1.2.7. Bảng dim\_time

Tên thuộc tính	Mô tả
<b>Claim_date</b>	Ngày claim
<b>Claim_month</b>	Tháng claim
<b>Claim_year</b>	Năm claim
<b>Claim_quarter</b>	Quý claim

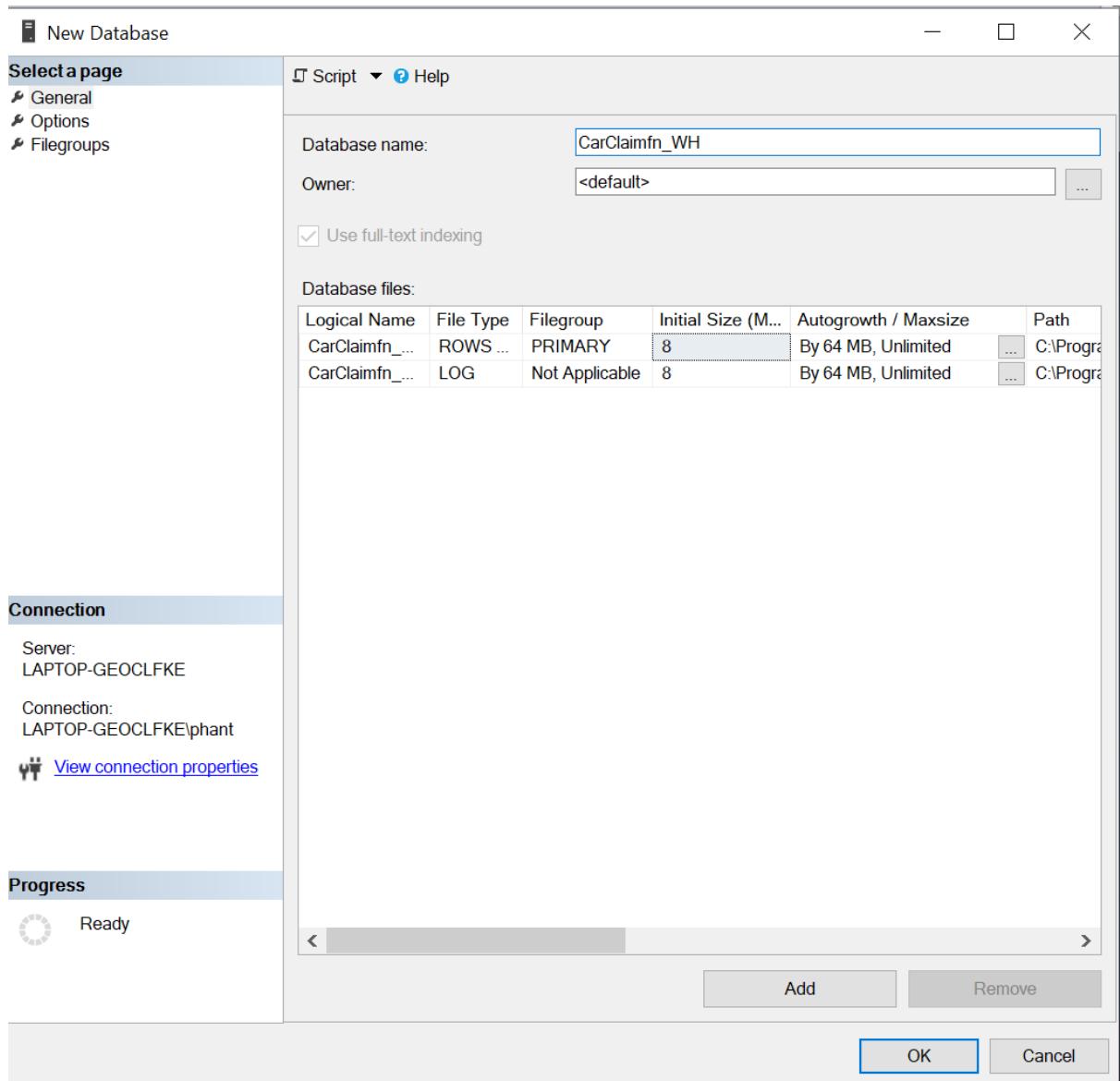
#### 1.2.8. Bảng dim\_vehicle

Tên thuộc tính	Mô tả
<b>Vehicle_category</b>	Loại xe

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

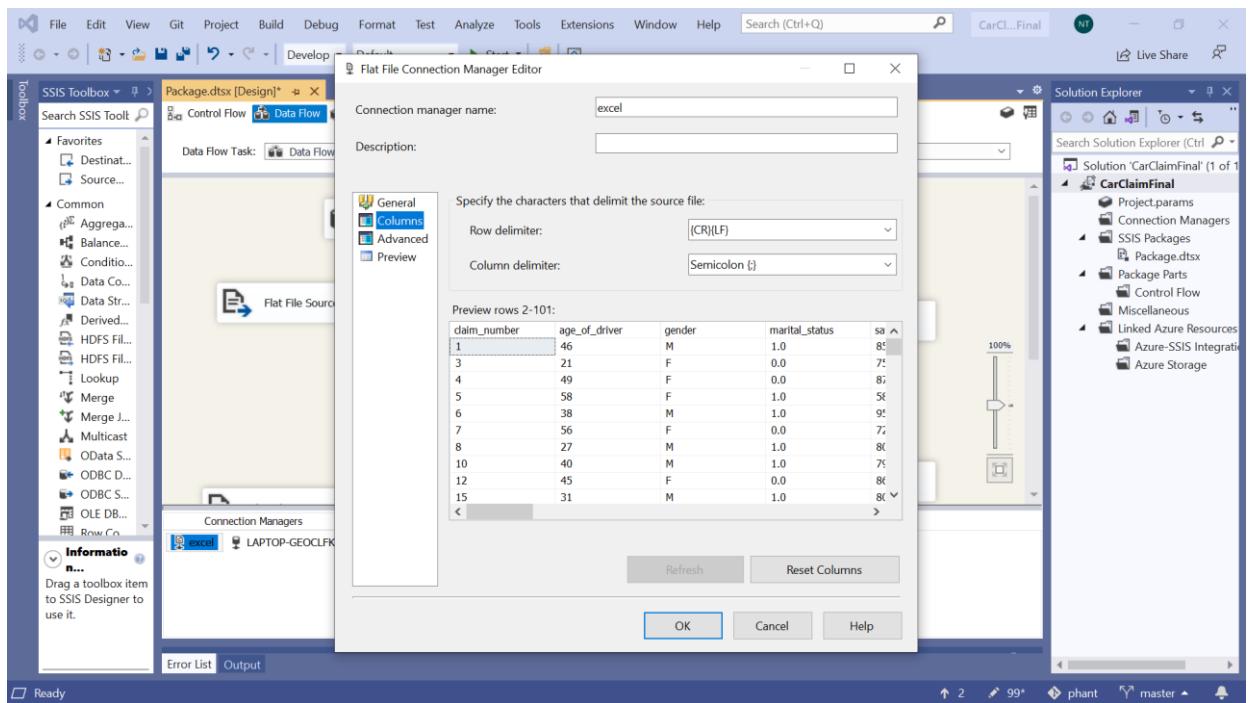
### 2.1. Tạo mới một database

- Thực hiện tạo một database với SQL Server để lưu data với tên là CarClaimfn\_WH.

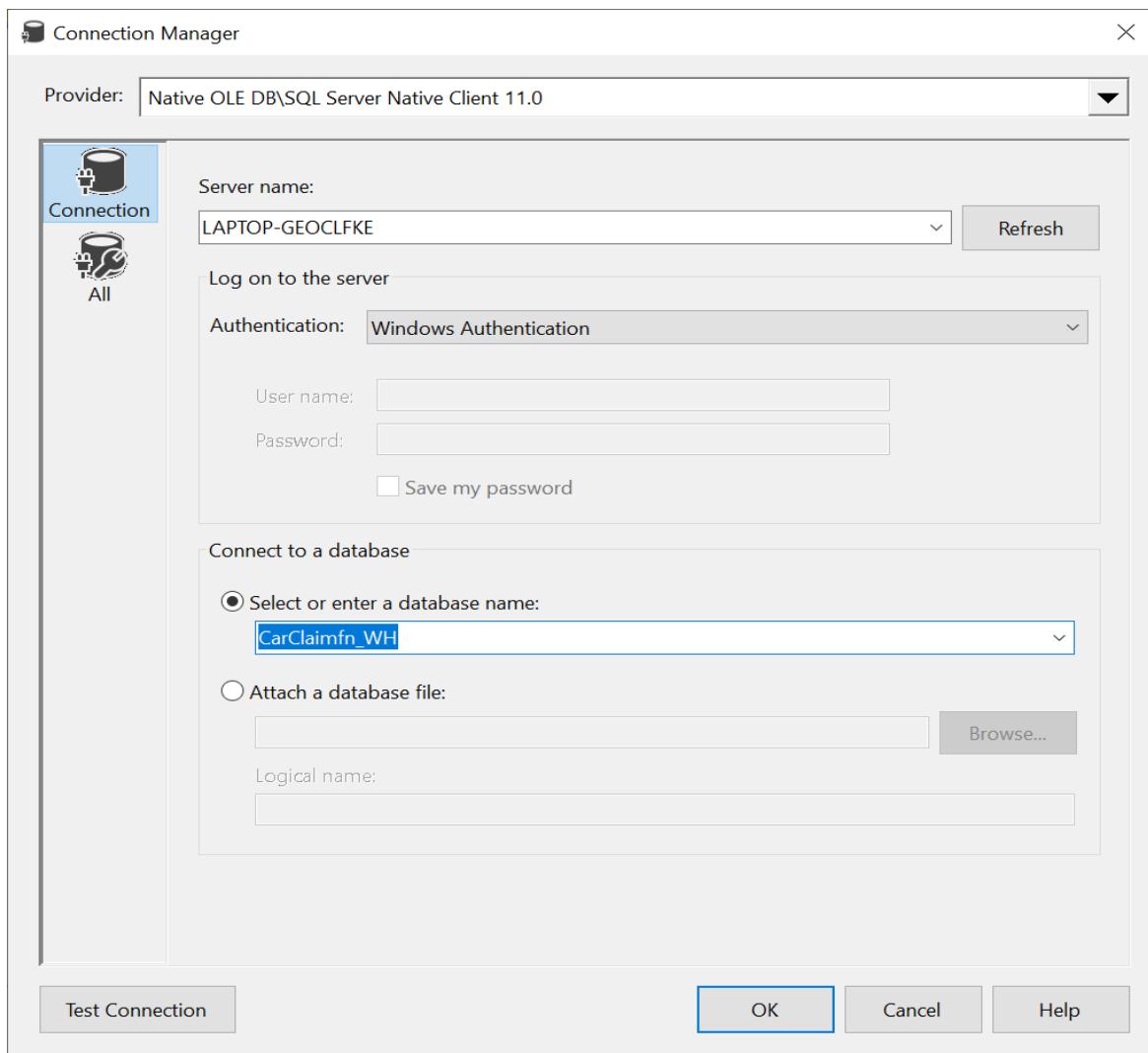


### 2.2. Tạo project mới với SSIS

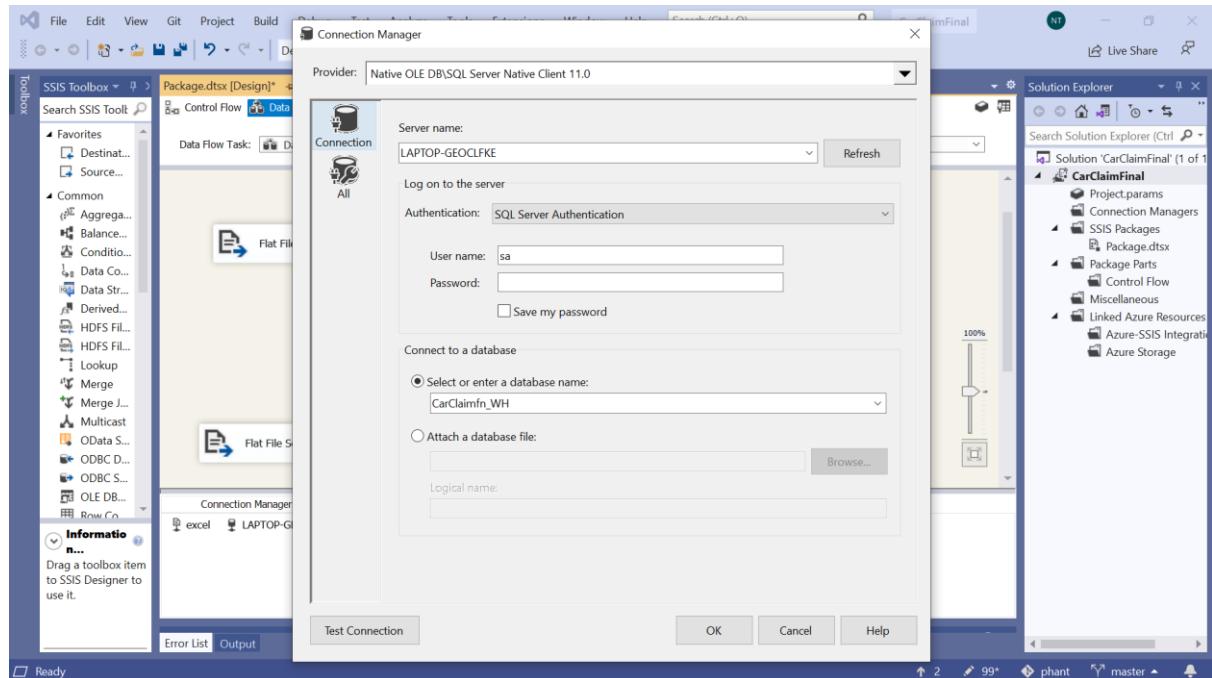
- Vào Microsoft Visual Studio 2019: Tạo Integration Services Project.
- Import dữ liệu với Flat file source.



- Thêm tên server để chọn database ware house.



- Tạo thêm OLE DB Destination kết nối với database, và nối Flat file Source với OLE DB để đổ dữ liệu vào.



- Tạo connection và các bảng dim.
- Sort các dữ liệu trùng lặp.

**Sort Transformation Editor**

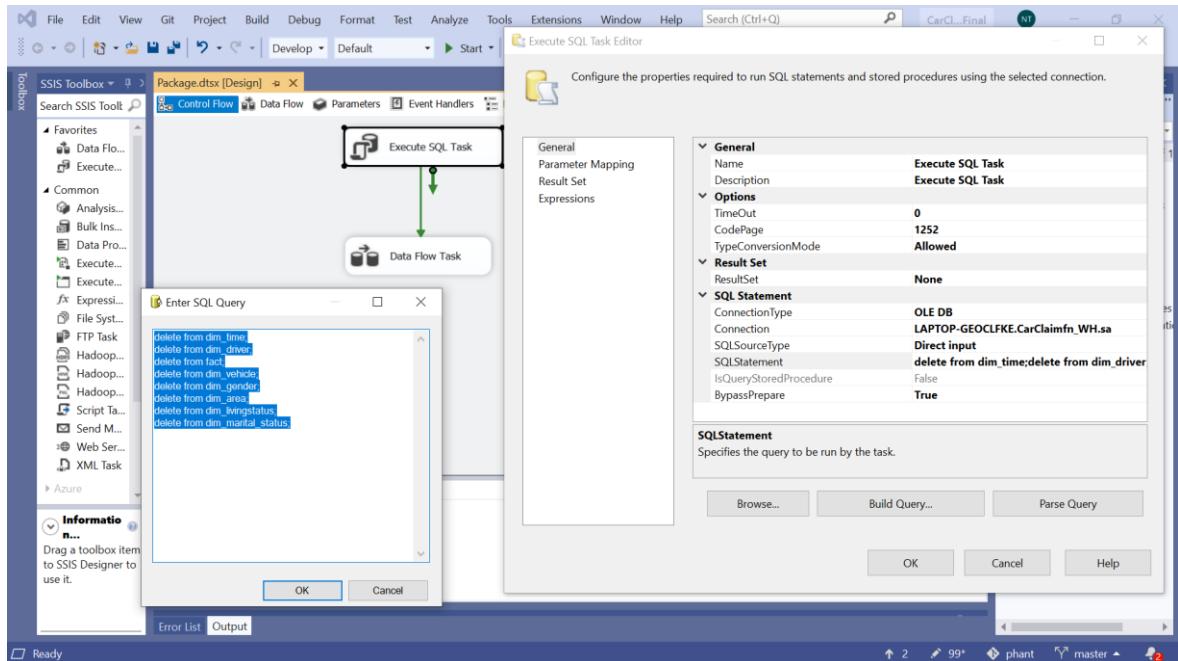
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Input Column	Output Alias	Sort Type	Sort Order	Com
vehicle_category	vehicle_category	ascending	1	

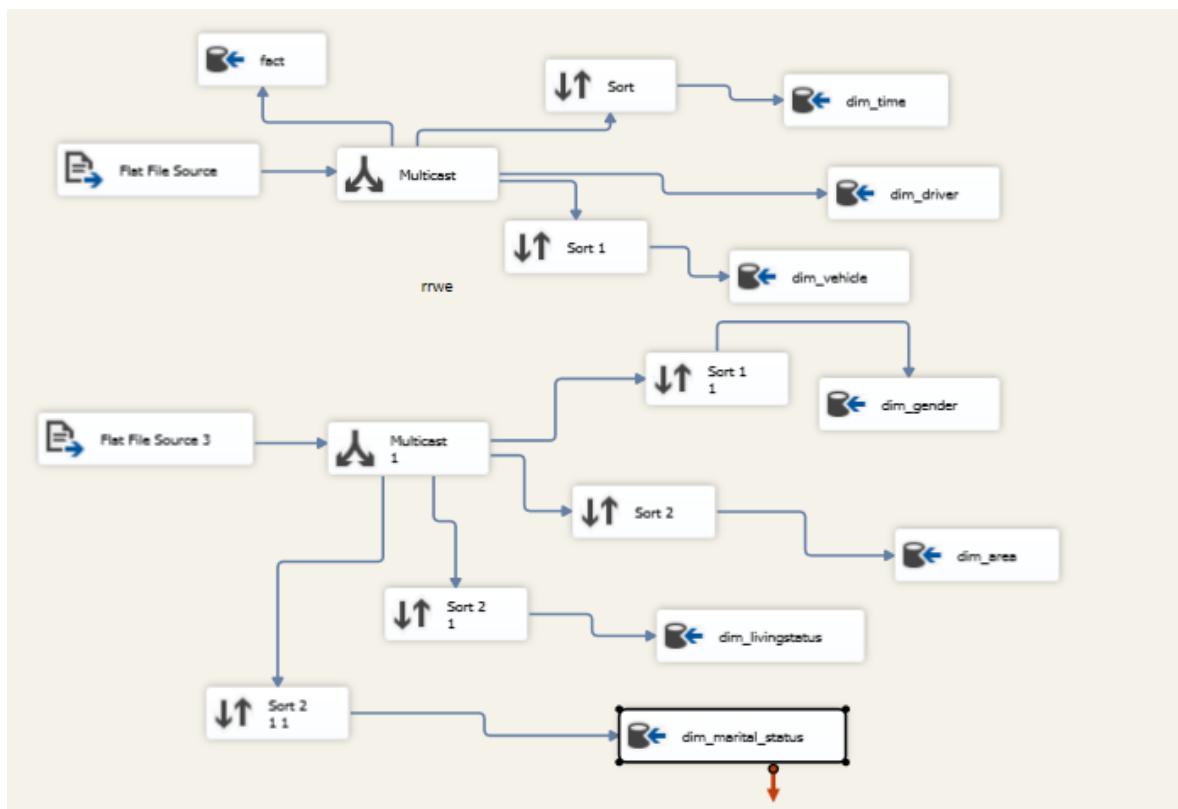
Remove rows with duplicate sort values

OK Cancel Help

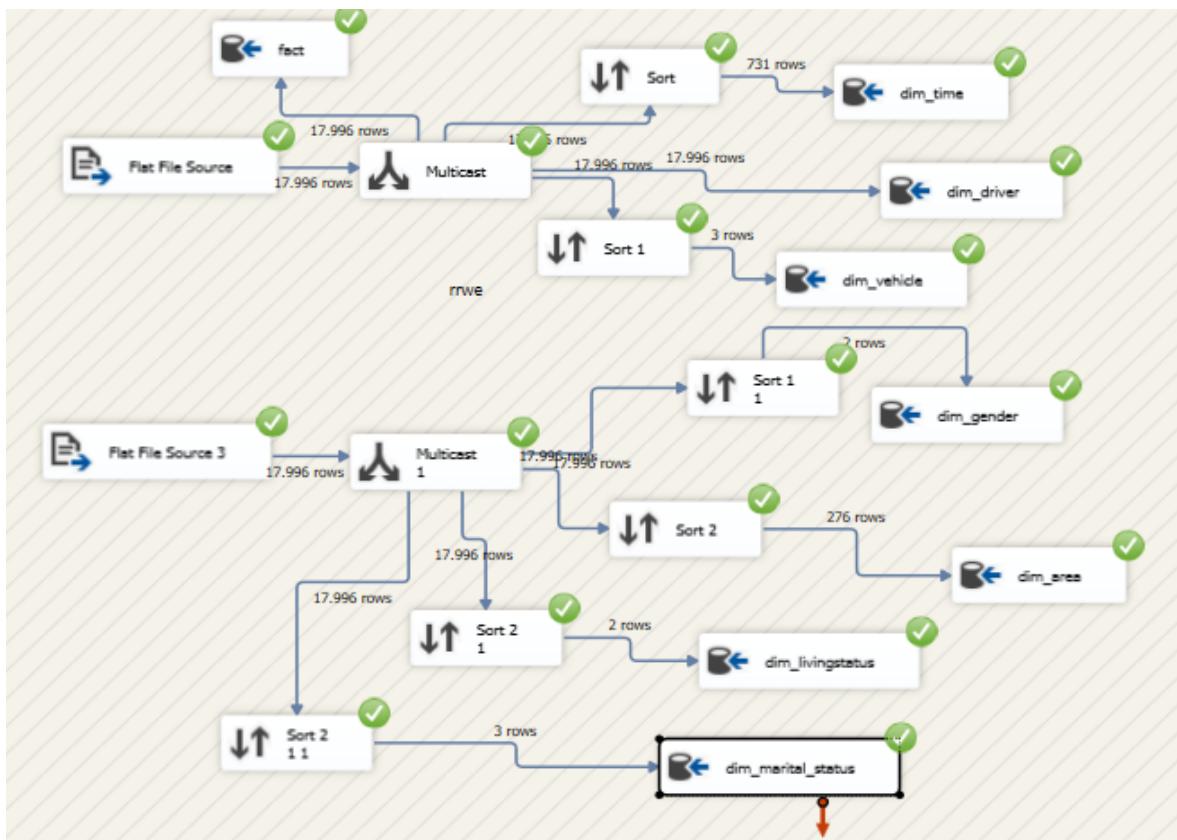
- Execute SQL Task để tránh việc dữ liệu bị trùng lặp khi đã nhiều lần.



- Sau khi hoàn tất ta được



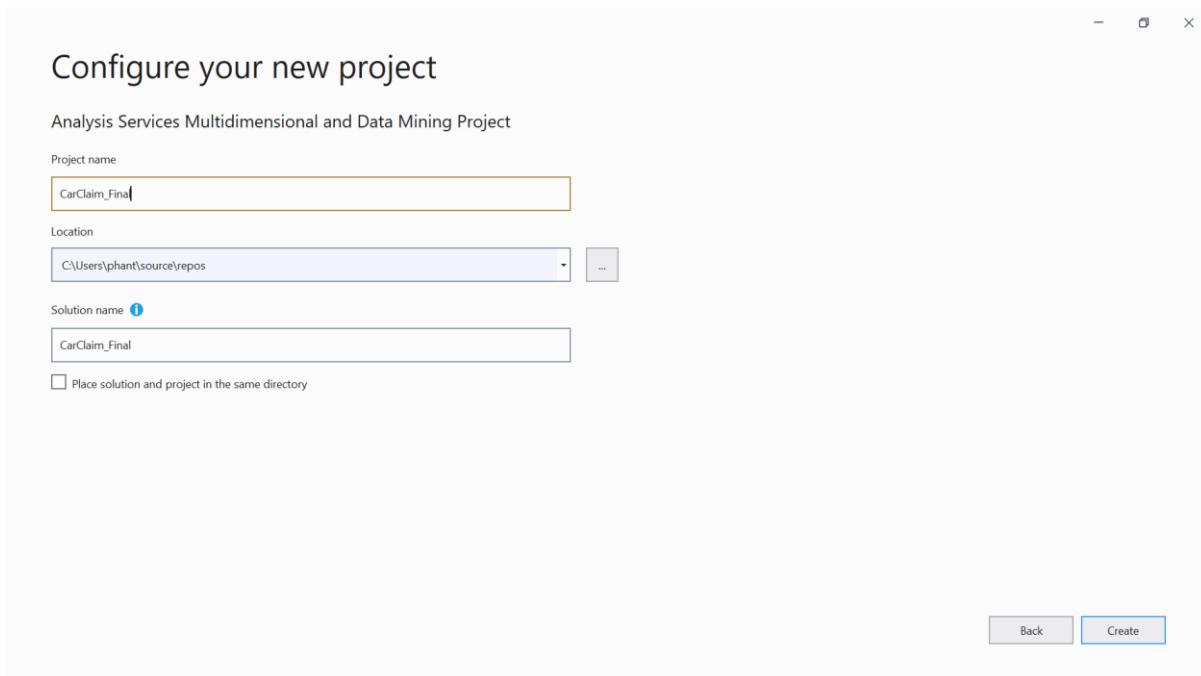
- Đồ dữ liệu thành công



## CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN \_ QUÁ TRÌNH SSAS

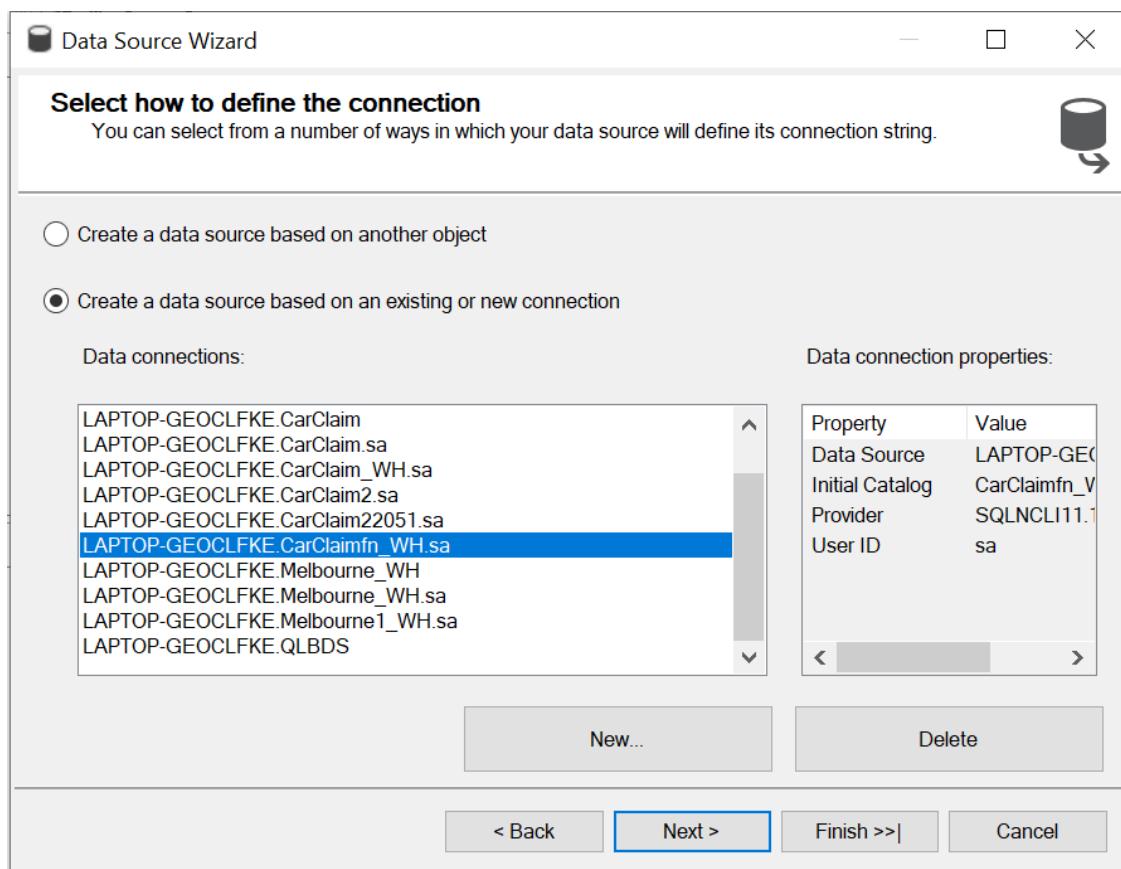
### 3.1. Cấu hình project

Tạo một project Analysis Services với tên CarClaim\_Final.

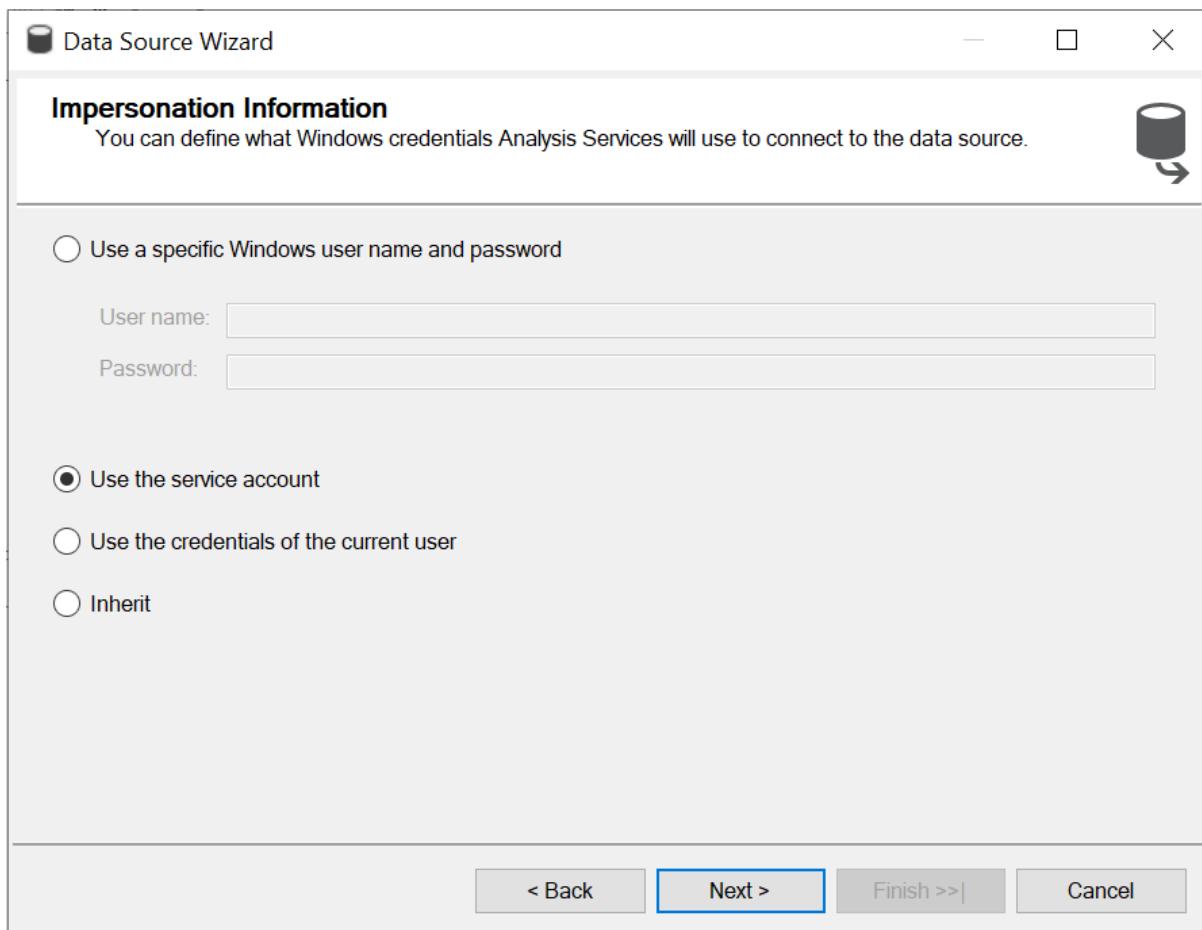


#### 3.1.1. Tạo datasource

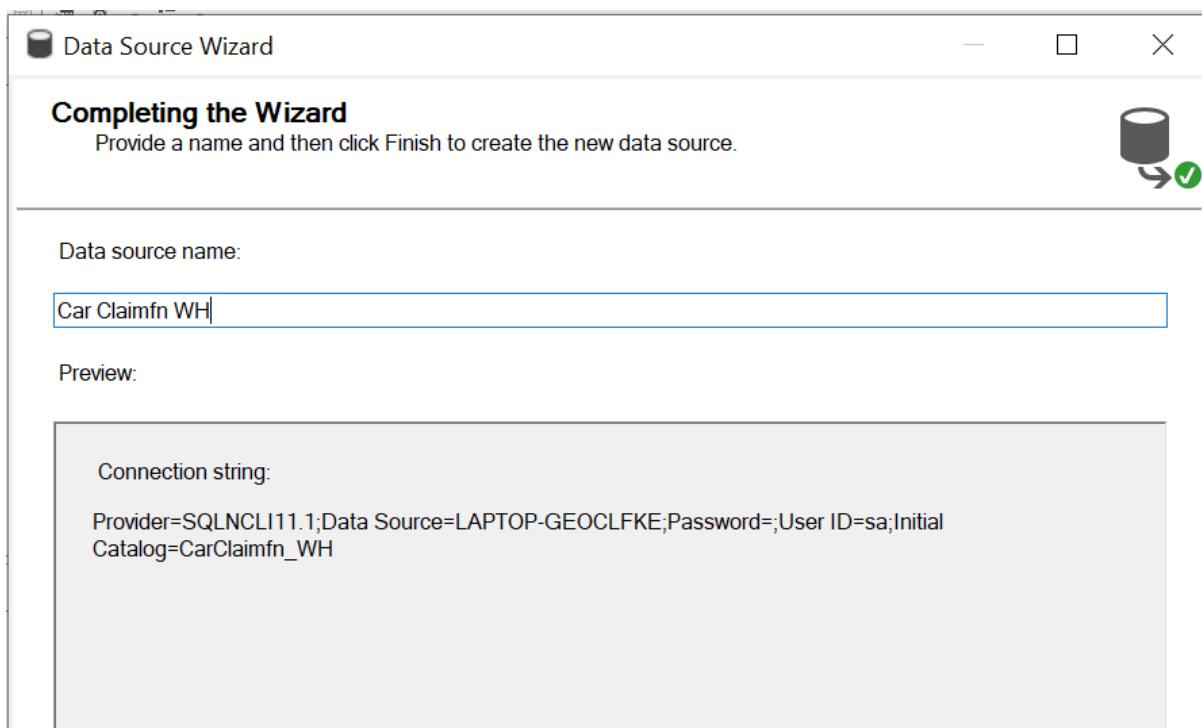
- Tạo datasource dựa trên connection có sẵn.



- Sử dụng service account.

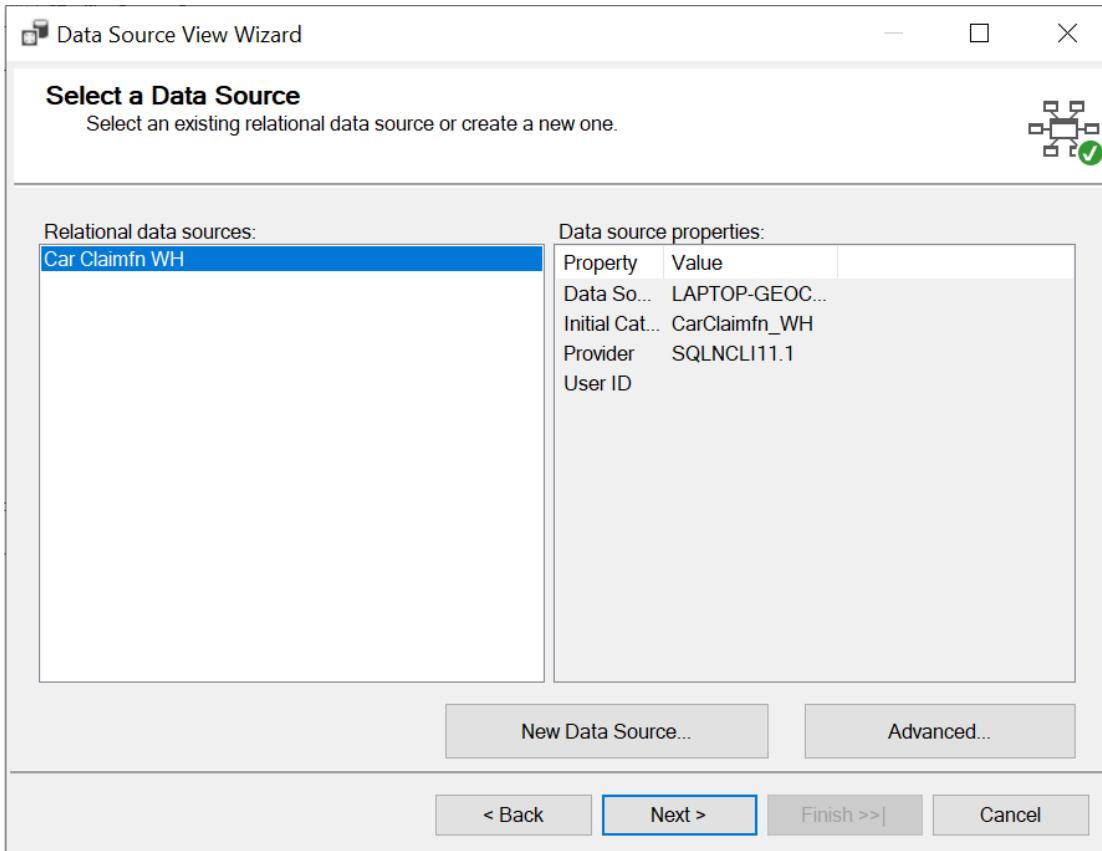


- Hoàn tất tạo datasource.

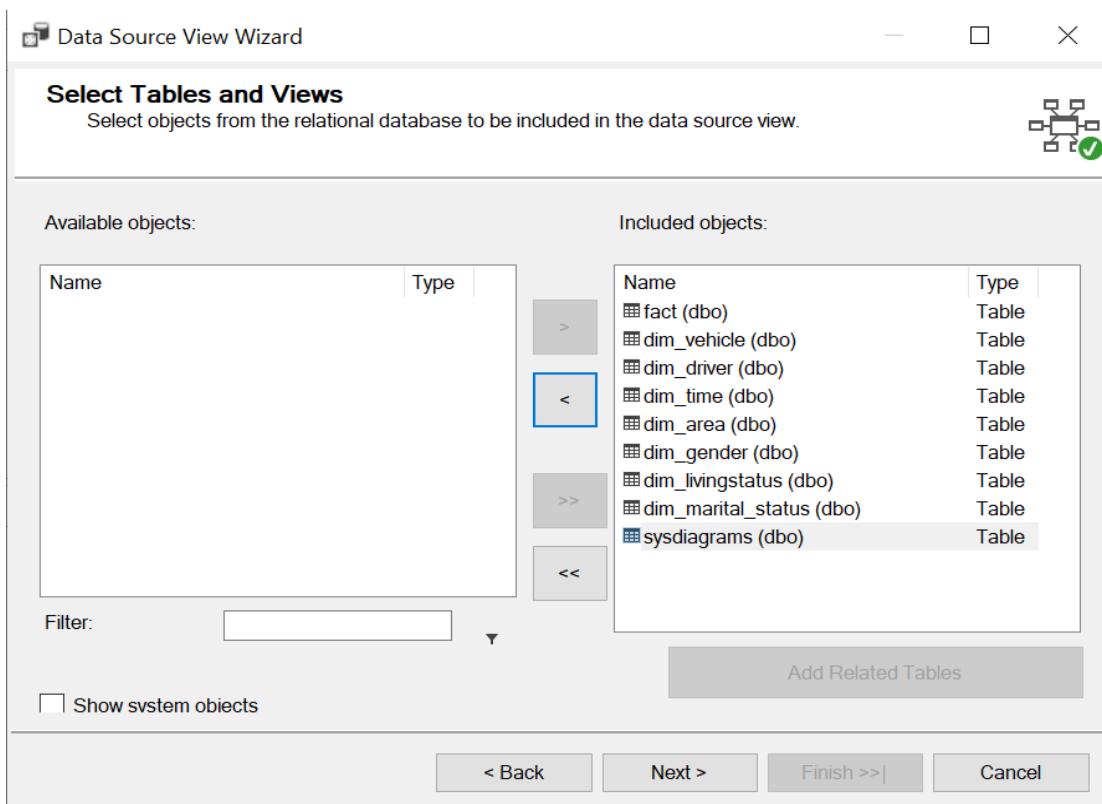


### 3.1.2. Tạo datasource views

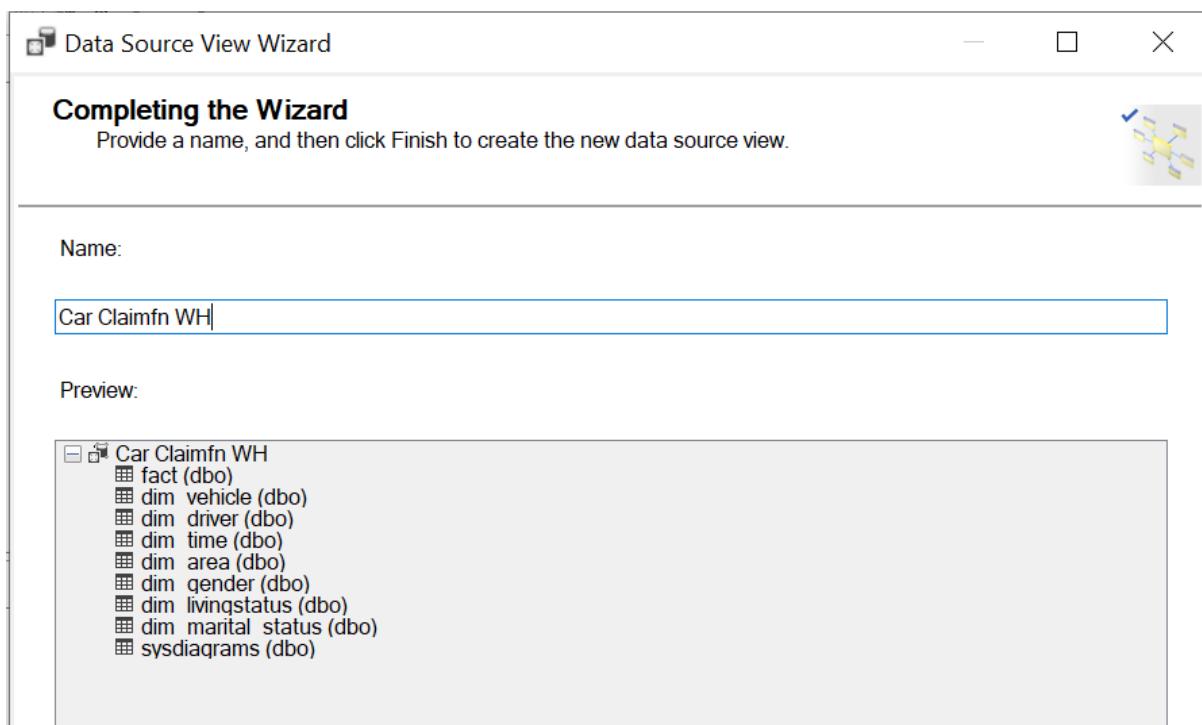
- Chọn datasource CarClaimfn\_WH vừa tạo.



- Chọn các bảng dim và fact.

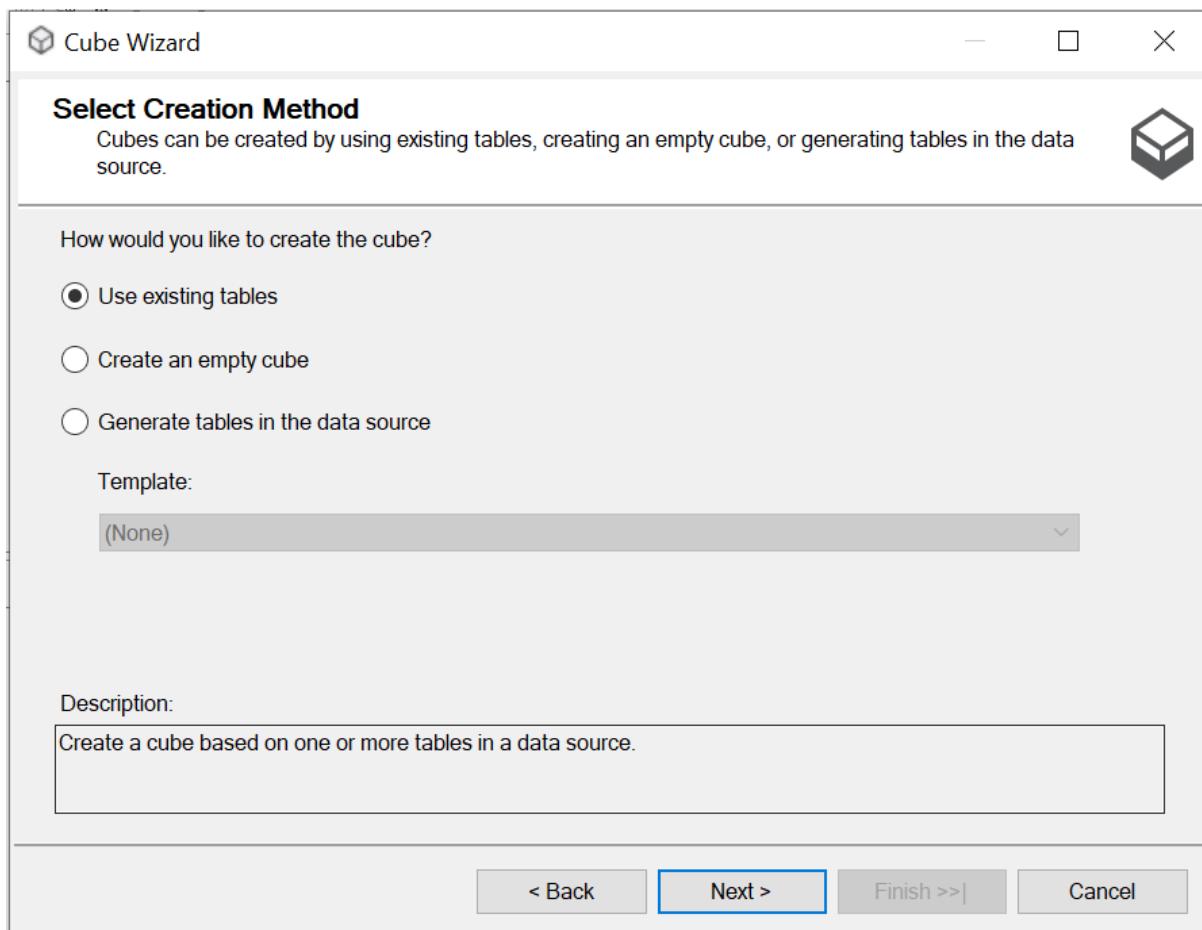


- Hoàn tất tạo datasource view.

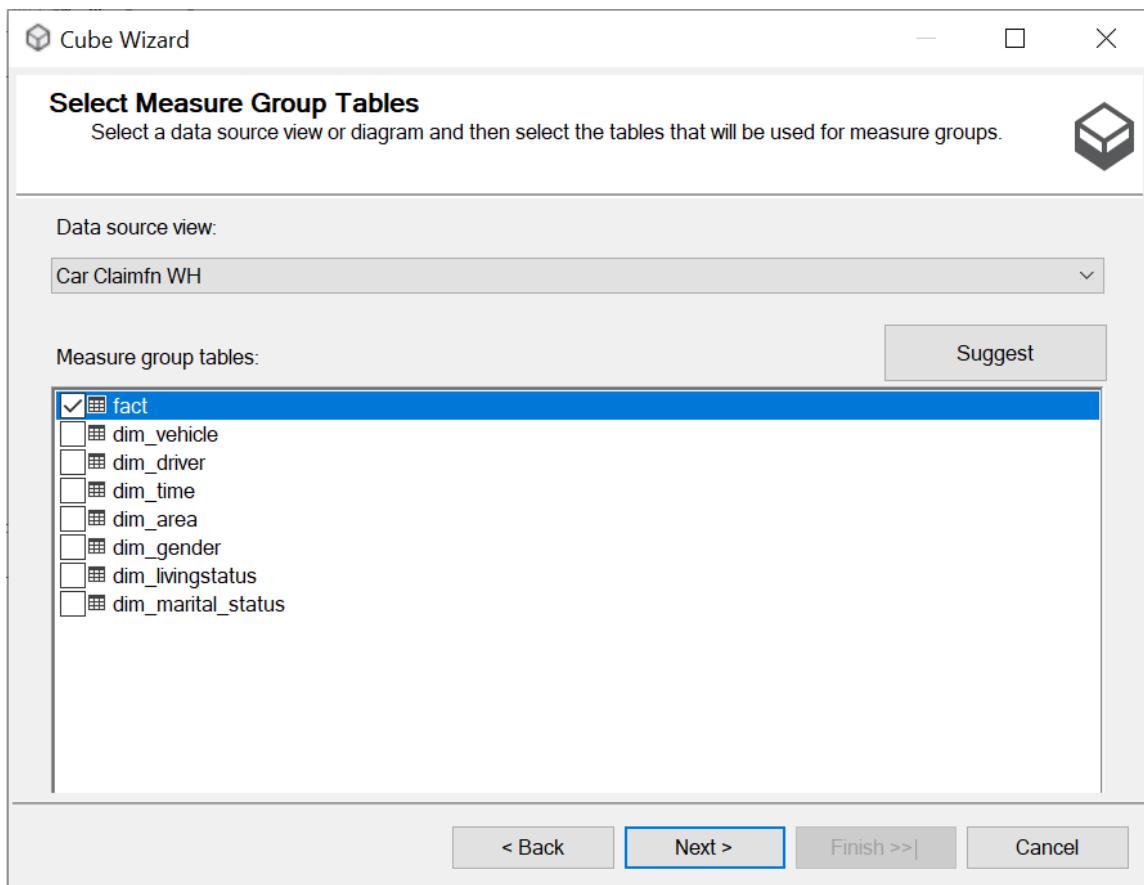


### 3.1.3. Tạo Cubes

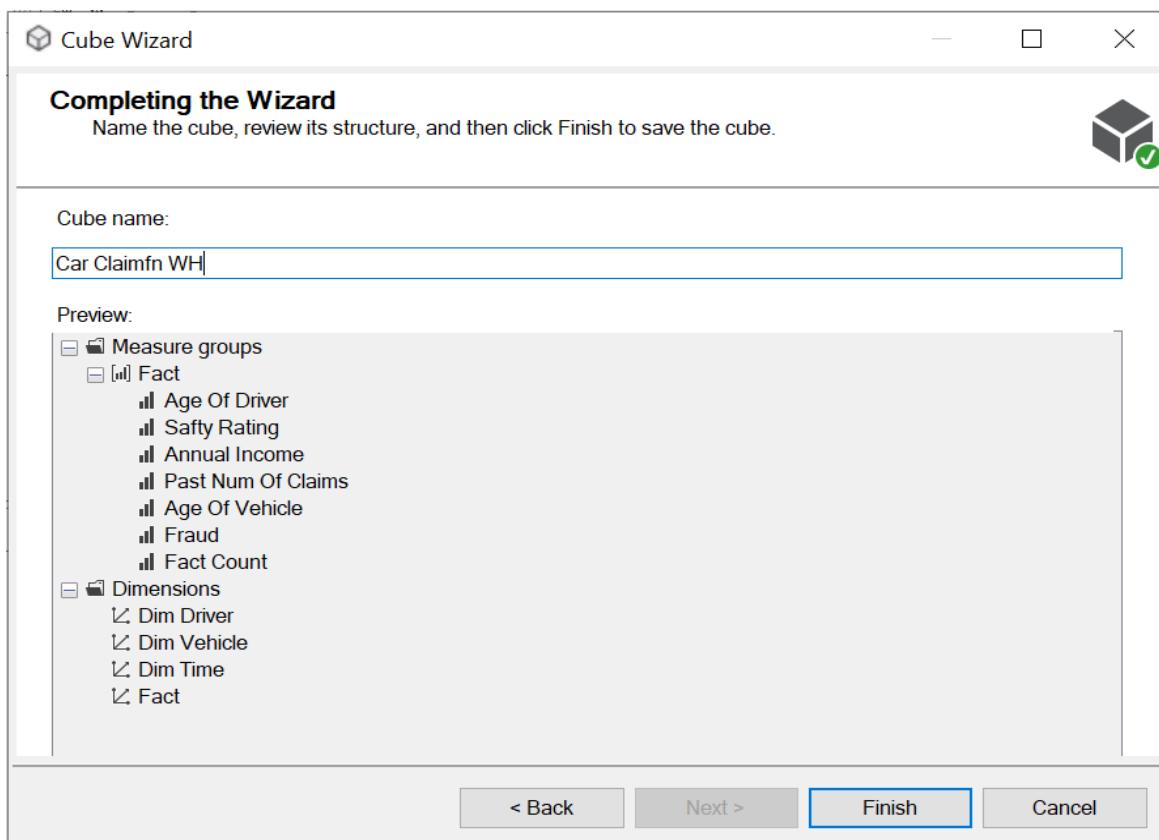
- Sử dụng các table có sẵn



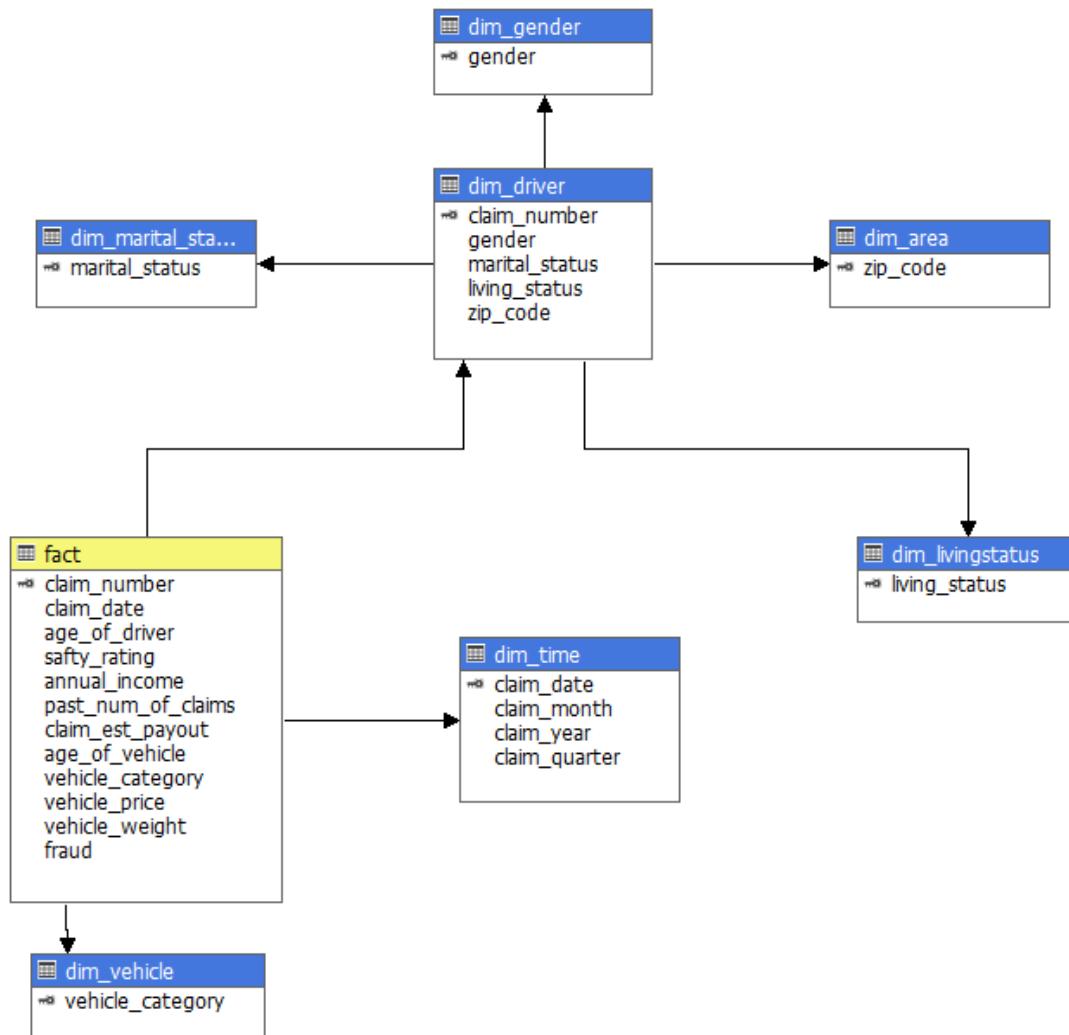
- Chọn measure group table



- Hoàn tất tạo cubes

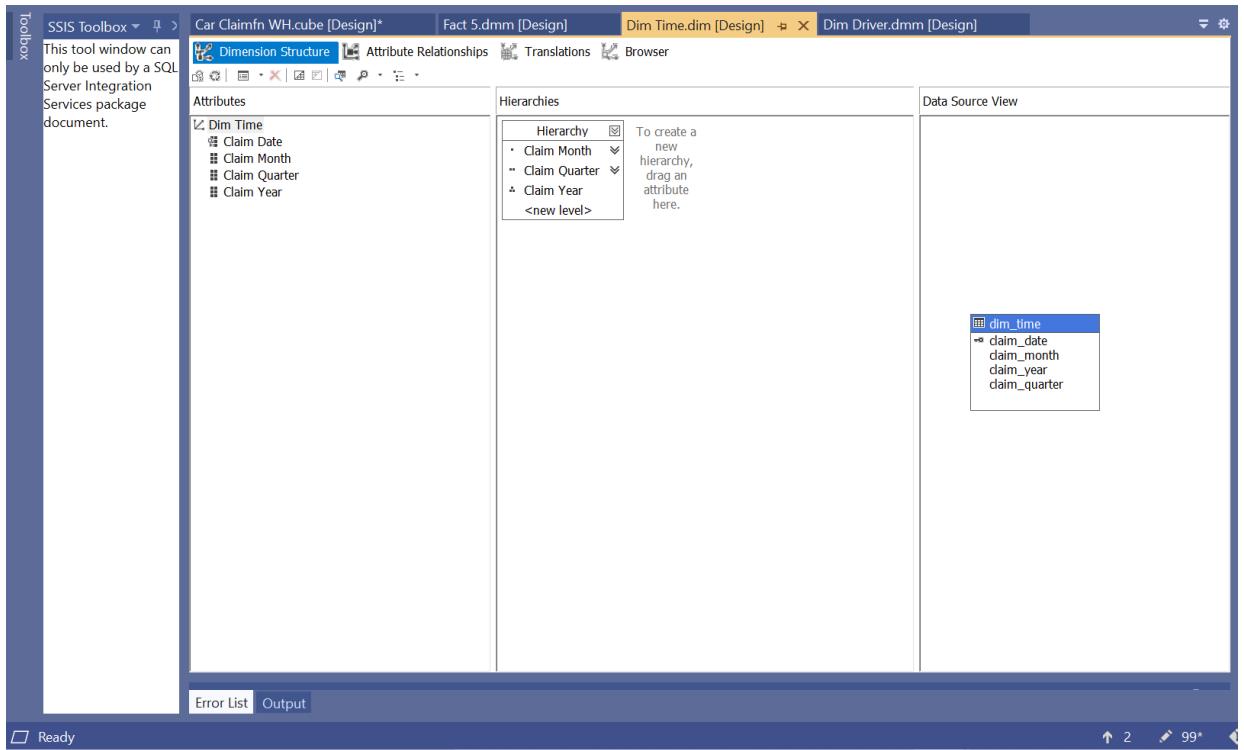


Sau khi tạo thành công ta được sơ đồ hình bông tuyêt:



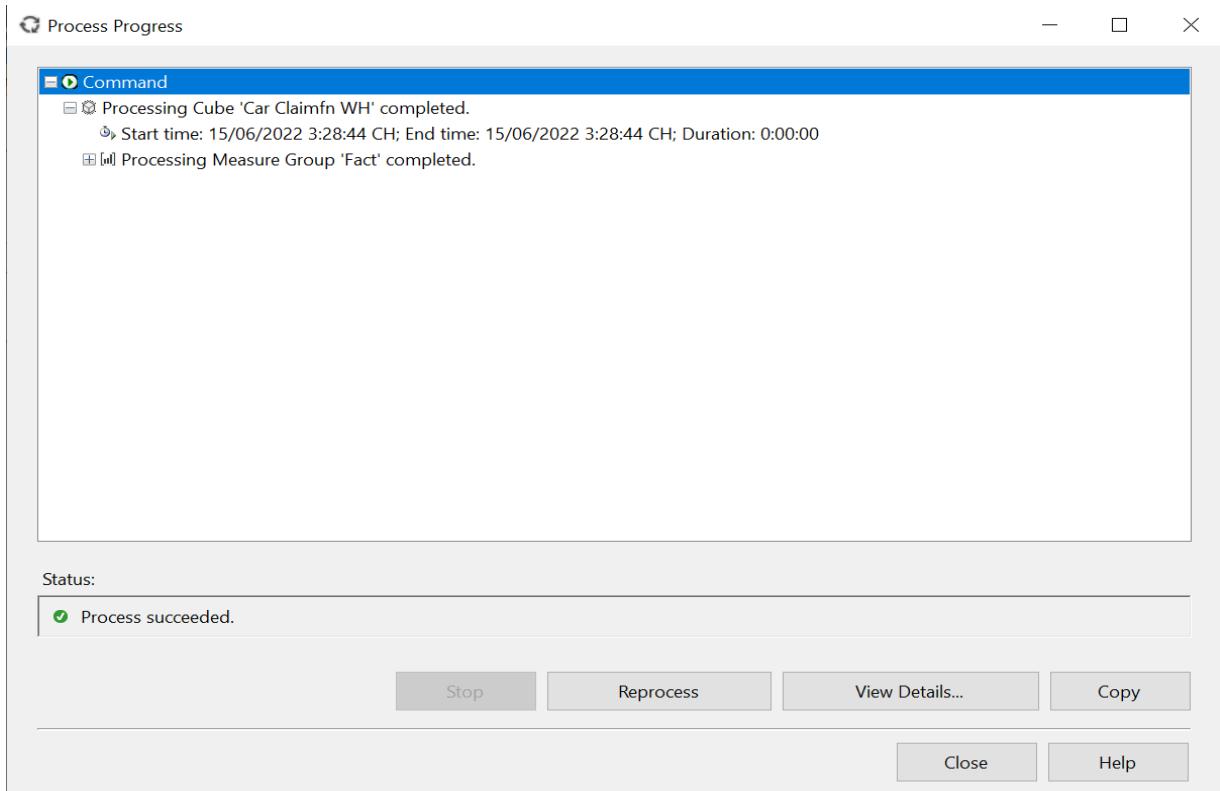
### 3.1.4. Phân cấp dim\_time

Kéo thả các thuộc tính Claim\_date, Claim\_month, Claim\_quarter, Claim\_year để thực hiện phân cấp.



### 3.1.5. Process cube

- Click chuột phải vào Car Claimfn WH.cube sau đó chọn Process, xuất hiện thông báo ta chọn Yes.
- Chọn full màn hình Process, chọn nút Run, kết quả chạy thành công như hình sau.



### 3.2. Thực thi truy vấn

#### 3.2.1. Roll up

Đếm số báo cáo claim bảo hiểm theo từng tháng, năm.

The screenshot shows the SSAS MDX query editor interface. At the top, there is a toolbar with various icons. Below the toolbar is a header row with columns for Dimension, Hierarchy, Operator, and Filter. The Dimension column contains the text '<Select dimension>'. The table below has two columns: 'Claim Month' and 'Fact Count'. The data rows are:

Claim Month	Fact Count
1	1606
10	1474
11	1444
12	1484
2	1439
3	1556
4	1423
5	1502
6	1456
7	1532
8	1595
9	1487

The screenshot shows the SSAS cube design interface for 'Car Claimfn WH.cube [Design]'. The top navigation bar includes 'Cube Structure', 'Dimension Usage', 'Calculations', 'KPIs', 'Actions', and a language dropdown set to 'Default'. The main area has a toolbar at the top with icons for 'Edit as Text', 'Import...', 'MDX', and others. On the left is a tree view of the cube structure under 'Car Claimfn WH' and 'Metadata'. The right side shows a table with 'Dimension' and 'Hierarchy' columns, both set to '<Select dimension>'. The table below has two columns: 'Claim Year' and 'Fact Count'. The data rows are:

Claim Year	Fact Count
2015	9111
2016	8885

### 3.2.2. Drill down

Đếm số báo cáo claim bảo hiểm theo từng tháng, từng ngày.

Car Claimfn WH.cube [Design] X

Cube Structure Dimension Usage Calculations KPIs Actions Partitions  
Language: Default

Edit as Text Import... MDX MDX View Tools Options Help

Dimension Hierarchy Operator

<Select dimension>

Measure Group:

<All>

- Fraud
- Maximum Sa
- Past Num Of
- Safty Rating
- KPIs
- Dim Driver
- Dim Time
  - Claim Date
    - Members
      - Claim Date
    - Hierarchy
  - Claim Month
  - Claim Quarter
  - Claim Year
- Dim Vehicle
- Fact

Calculated Members

Claim Date	Fact Count
2015-01-01	29
2015-01-02	24
2015-01-03	34
2015-01-04	23
2015-01-05	17
2015-01-06	27
2015-01-07	32
2015-01-08	19
2015-01-09	35
2015-01-10	28
2015-01-11	20
2015-01-12	27
2015-01-13	28
2015-01-14	25
2015-01-15	33
2015-01-16	19
2015-01-17	27
2015-01-18	36
2015-01-19	18
2015-01-20	27
2015-01-21	25
2015-01-22	30

MDX ▾

The screenshot shows a BI tool interface with a toolbar at the top containing various icons. Below the toolbar is a header row with four columns: 'Dimension', 'Hierarchy', 'Operator', and 'Filter'. The 'Dimension' column has a dropdown menu with the placeholder '<Select dimension>'. The 'Hierarchy' column contains a tree view of dimension members. The 'Operator' and 'Filter' columns are currently empty.

Dimension	Hierarchy	Operator	Filter
<Select dimension>			
<b>Claim Month</b>	<b>Fact Count</b>		
1	1606		
10	1474		
11	1444		
12	1484		
2	1439		
3	1556		
4	1423		
5	1502		
6	1456		
7	1532		
8	1595		
9	1487		

### 3.2.3. Slice and Dice

#### 3.2.3.1. Đếm số báo cáo theo ngày với loại xe là Compact

Car Claimfn WH.cube [Design] ▾

The screenshot shows the 'Car Claimfn WH.cube [Design]' interface. The top navigation bar includes 'Cube Structure', 'Dimension Usage', 'Calculations', 'KPIs', 'Actions', 'Partitions', 'Aggregations', 'Perspectives', 'Translations', and 'Browser'. The 'Browser' tab is selected.

The main area features a toolbar with icons for 'Edit as Text', 'Import...', 'MDX', and others. Below the toolbar is a table with columns: 'Dimension', 'Hierarchy', 'Operator', and 'Filter Expression'. The 'Dimension' column is set to 'Dim Vehicle', 'Hierarchy' is 'Vehicle Category', 'Operator' is 'Equal', and 'Filter Expression' is '{ Compact }'.

On the left, there is a navigation pane with sections for 'Car Claimfn WH', 'Metadata', 'Search Model', 'Measure Group', and 'Calculated Members'. The 'Measure Group' section is expanded, showing 'Fact' and its sub-members: 'Maximum Safty Rating', 'Past Num Of', 'Safty Rating', 'KPIs', 'Dim Driver', 'Dim Time', 'Claim Date', 'Members', 'Claim Month', 'Claim Quarter', 'Claim Year', 'Hierarchy', 'Dim Vehicle', 'Vehicle Category', and 'Fact'.

The main table displays data for 'Claim Date', 'Vehicle Category', and 'Fact Count' from January 1st to January 22nd, filtered by 'Compact' vehicle category.

Claim Date	Vehicle Category	Fact Count
2015-01-01	Compact	13
2015-01-02	Compact	4
2015-01-03	Compact	9
2015-01-04	Compact	7
2015-01-05	Compact	8
2015-01-06	Compact	8
2015-01-07	Compact	12
2015-01-08	Compact	6
2015-01-09	Compact	14
2015-01-10	Compact	9
2015-01-11	Compact	7
2015-01-12	Compact	12
2015-01-13	Compact	10
2015-01-14	Compact	5
2015-01-15	Compact	13
2015-01-16	Compact	4
2015-01-17	Compact	10
2015-01-18	Compact	9
2015-01-19	Compact	4
2015-01-20	Compact	9
2015-01-21	Compact	8
2015-01-22	Compact	11

### 3.2.3.2. Thống kê số lượng báo cáo với giới tính của tài xế là nữ theo từng năm

The screenshot shows the SSAS Cube Design interface for the "Car Claimfn WH.cube". The top navigation bar includes "Cube Structure", "Dimension Usage", "Calculations", "KPIs", "Actions", "Partitions", and "Aggregates". The language is set to "Default". The main area displays a query in MDX:

```
SELECT
    [Dim Driver].[Gender].<Select dimension>
    ON COLUMNS,
    [Claim Year].[Year].MEMBERS
    ON ROWS
    FROM [Fact Count]
```

The results table shows the count of facts for each gender and year:

Claim Year	Gender	Fact Count
2015	F	4341
2016	F	4243

The left pane shows the cube's metadata, including measure groups like "Maximum Sales", "Past Num Of", and "Safty Rating", and dimensions like "Dim Driver" which includes "Claim Number", "Gender", "Living Status", "Marital Status", and "Zip Code".

### 3.2.3.3. Thống kê số báo cáo theo từng tình trạng hôn nhân của tài xế tháng 5 và tháng 7 theo từng năm

The screenshot shows the SSAS Cube Design interface for the "Car Claimfn WH.cube". The top navigation bar includes "Cube Structure", "Dimension Usage", "Calculations", "KPIs", "Actions", "Partitions", and "Aggregates". The language is set to "Default". The main area displays a query in MDX:

```
SELECT
    [Dim Time].[Claim Month].<Select dimension>
    ON COLUMNS,
    [Dim Time].[Month].MEMBERS
    ON ROWS
    FROM [Fact Count]
```

The results table shows the count of facts for each marital status and month:

Marital Status	Claim Month	Claim Year	Fact Count
0.0	5	2015	110
0.0	5	2016	116
0.0	7	2015	101
0.0	7	2016	92
1.0	5	2015	253
1.0	5	2016	260
1.0	7	2015	239
1.0	7	2016	238

The left pane shows the cube's metadata, including measure groups like "Claim Number", "Gender", "Living Status", "Marital Status", and "Zip Code", and dimensions like "Dim Time" which includes "Claim Date", "Members", "Claim Month", "Claim Quarter", "Claim Year", and "Hierarchy".

### 3.2.3.4. Thống kê số lượng báo cáo gian lận vào tháng 9 từng năm

The screenshot shows the SSAS Design Studio interface with the title bar "Car Claimfn WH.cube [Design]". The ribbon menu includes "Cube Structure", "Dimension Usage", "Calculations", "KPIs", "Actions", "Partitions", and "Aggr". The language is set to "Default". The main area has tabs "Edit as Text" and "Import...". The MDX dropdown is set to "MDX". Below these are various toolbar icons. On the left, there's a tree view of the cube structure under "Car Claimfn WH" and "Metadata". A search model pane is also visible. The right side shows a dimension usage grid and a table preview.

Dimension	Hierarchy	Operator
Dim Time	Claim Month	Equal
<Select dimension>		

Claim Month	Claim Year	Fraud
9	2015	56
9	2016	63

### 3.2.3.5. Thống kê số lượng báo cáo với giới tính của tài xế là nữ và sở hữu nhà ở theo từng năm

The screenshot shows the SSAS Design Studio interface with the title bar "Car Claimfn WH.cube [Design]". The ribbon menu includes "Cube Structure", "Dimension Usage", "Calculations", "KPIs", "Actions", "Partitions", "Aggregations", "Perspectives", "Translations", and "P". The language is set to "Default". The main area has tabs "Edit as Text" and "Import...". The MDX dropdown is set to "MDX". Below these are various toolbar icons. On the left, there's a tree view of the cube structure under "Car Claimfn WH" and "Metadata". A search model pane is also visible. The right side shows a dimension usage grid and a table preview.

Dimension	Hierarchy	Operator	Filter Expression
Dim Driver	Gender	Equal	{ F }
Dim Driver	Living Status	Equal	{ Own }

Claim Year	Gender	Living Status	Fact Count
2015	F	Own	2423
2016	F	Own	2408

### 3.2.4. Pivot

#### 3.2.4.1. Thống kê số báo cáo được ghi nhận theo năm và loại xe

Truy vấn theo SSAS

Vehicle Category	Claim Year	Fact Count
Compact	2015	1431
Compact	2016	1473
Large	2015	1441
Large	2016	1355
Medium	2015	1467
Medium	2016	1415

- Dòng: Claim\_year, cột: Vehicle\_category

	A	B	C	D	E	F
1	Fact Count	Column Labels				
2	Row Labels	Compact	Large	Medium	Grand Total	
3	2015		3040	3020	3051	9111
4	2016		3021	2966	2898	8885
5	Grand Total		6061	5986	5949	17996
6						

- Dòng: Vehicle\_category, cột: Claim\_year

A	B	C	D	E
1	Fact Count	Column Labels		
2	Row Labels	2015	2016	Grand Total
3	Compact	3040	3021	6061
4	Large	3020	2966	5986
5	Medium	3051	2898	5949
6	Grand Total	9111	8885	17996
7				
8				
9				

### 3.2.4.2. Thống kê số báo cáo theo zip code và giới tính

Truy vấn theo SSAS

Car Claimfn WH.cube [Design] X

Cube Structure Dimension Usage Calculations KPIs Actions Partitions

Language: Default

Edit as Text Import... MDX

**Dimension** **Hierarchy** **Operator**  
**Dim Driver** **Gender** **Equal**  
**<Select dimension>**

Zip Code	Gender	Fact Count
0	F	11
0	M	12
15001	F	141
15001	M	142
15003	F	36
15003	M	51
15004	F	24
15004	M	16
15005	F	25
15005	M	15
15006	F	35
15006	M	21
15007	F	22
15007	M	24
15009	F	103
15009	M	105
15010	F	73
15010	M	86
15012	F	123
15012	M	153
15014	F	22
15014	M	22

Safty Rating  
KPIs  
Dim Driver: Claim Number, Gender, Living Status, Marital Status, Zip Code  
Dim Time: Claim Date (Members: Claim Date), Claim Month, Claim Quarter, Claim Year, Hierarchy  
Dim Vehicle

Calculated Members

- Dòng: Zip code, cột: Gender

	A	B	C	D	E
1	Fact Count	Column Labels			
2	Row Labels	F	M	Grand Total	
3	0		11	12	23
4	15001		141	142	283
5	15003		36	51	87
6	15004		24	16	40
7	15005		25	15	40
8	15006		35	21	56
9	15007		22	24	46
10	15009		103	105	208
11	15010		73	86	159
12	15012		123	153	276
13	15014		22	22	44
14	15015		22	23	45
15	15017		21	23	44
16	15018		16	30	46
17	15019		24	24	48
18	15020		21	30	51
19	15021		75	104	179
20	15022		55	81	136
21	15024		22	27	49
22	15025		123	113	236
23	15026		21	21	42
	15027		21	21	42

- Dòng: Gender, Cột: Zip code

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Fact Count	Column Labels																
2	Row Labels	0	15001	15003	15004	15005	15006	15007	15009	15010	15012	15014	15015	15017	15018	15019	15020	15021
3	F		11	141	36	24	25	35	22	103	73	123	22	22	21	16	24	21
4	M		12	142	51	16	15	21	24	105	86	153	22	23	23	30	24	30
5	Grand Total		23	283	87	40	40	56	46	208	159	276	44	45	44	46	48	51
6																		
7																		
8																		
9																		

### 3.2.4.3. Thống kê số báo cáo được ghi nhận theo giới tính, loại xe và tình trạng nhà ở

Truy vấn theo SSAS

Gender	Living Status	Vehicle Category	Fact Count
F	Own	Compact	1688
F	Own	Large	1551
F	Own	Medium	1592
F	Rent	Compact	1216
F	Rent	Large	1247
F	Rent	Medium	1290
M	Own	Compact	1704
M	Own	Large	1787
M	Own	Medium	1647
M	Rent	Compact	1453
M	Rent	Large	1401
M	Rent	Medium	1420

- Dòng: Gender, Cột: Living\_status và Vehicle\_category

	A	B	C	D	E	F	G	H	I
1	Fact Count	Column Labels							
2		Own	Rent				Grand Total		
3	Row Labels	Compact	Large	Medium	Compact	Large	Medium		
4	F		1688	1551	1592	1216	1247	1290	8584
5	M		1704	1787	1647	1453	1401	1420	9412
6	Grand Total		3392	3338	3239	2669	2648	2710	17996
7									
8									
9									
10									

- Dòng: Gender và Living\_status, cột: Vehicle\_category

	A	B	C	D	E	F
1	Fact Count	Column Labels				
2	Row Labels	Compact	Large	Medium	Grand Total	
3						
4	Own	1688	1551	1592	4831	
5	Rent	1216	1247	1290	3753	
6						
7						
8	Own	1704	1787	1647	5138	
9	Rent	1453	1401	1420	4274	
9	Grand Total	6061	5986	5949	17996	
10						
11						

### 3.3. Ngôn ngữ MDX

#### 3.3.1. Đếm số báo cáo theo từng giới tính

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

CarClaim\_Final Execute

-- Đếm số báo cáo theo từng giới tính

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

Messages Results

Fact Count	
F	8584
M	9412

#### 3.3.2. Đếm số báo cáo theo từng tháng, năm

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Time].[Claim Month].CHILDREN, [Dim Time].[Claim Year].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Đếm số báo cáo theo từng tháng, năm
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Time].[Claim Month].CHILDREN, [Dim Time].[Claim Year].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

		Fact Count
1	2015	814
1	2016	792
10	2015	746
10	2016	728
11	2015	723
11	2016	721
12	2015	782
12	2016	701
2	2015	724
2	2016	715
3	2015	789
3	2016	767
4	2015	699
4	2016	723
5	2015	765
5	2016	737
6	2015	754
6	2016	702
7	2015	770
7	2016	762
8	2015	821
8	2016	774

Query executed successfully.

### 3.3.3. Đếm số báo cáo theo từng năm, từng quý

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Time].[Claim Year].CHILDREN, [Dim Time].[Claim Quarter].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Đếm số báo cáo theo từng năm, từng quý
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Time].[Claim Year].CHILDREN, [Dim Time].[Claim Quarter].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

		Fact Count
2015	1	2327
2015	2	2218
2015	3	2315
2015	4	2251
2016	1	2274
2016	2	2162
2016	3	2299
2016	4	2150

### 3.3.4. Đếm số báo cáo theo từng giới tính và loại xe là Large

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
WHERE [Dim Vehicle].[Vehicle Category].&[Large];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Đếm số báo cáo theo từng giới tính và loại xe là Large
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
WHERE [dim Vehicle].[Vehicle Category].&[Large];
```

Results:

	Fact Count
F	2798
M	3188

### 3.3.5. Tìm tuổi cao nhất của các tài xế được claim bảo hiểm theo từng loại xe

```
SELECT NON EMPTY [Measures].[Age Of Driver] ON COLUMNS,
NON EMPTY([Dim Vehicle].[Vehicle Category].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Tim tuoi cao nhat cua cac tai xe theo tung loai xe duoc claim bao hiem.
SELECT NON EMPTY [Measures].[Age Of Driver] ON COLUMNS,
NON EMPTY([Dim Vehicle].[Vehicle Category].CHILDREN) ON ROWS
FROM [Car Claimfn WH];
```

Results:

	Age Of Driver
Compact	95
Large	113
Medium	133

### 3.3.6. Thống kê tổng thu nhập của những tài xế được claim bảo hiểm theo tình trạng nhà ở trong năm 2015

```
SELECT NON EMPTY [Measures].[Annual Income] ON COLUMNS,
NON EMPTY([Dim Driver].[Living Status].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Time].[Claim Year].&[2015]);
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Thống kê tổng thu nhập của những tài xe được claim bảo hiểm theo tình trạng hôn nhân
SELECT NON EMPTY [Measures].[Annual Income] ON COLUMNS,
NON EMPTY([Dim Driver].[Living Status].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Time].[Claim Year].&[2015]);
```

Results

Own	Rent
190530599	149846151

### 3.3.7. Thống kê số báo cáo lớn hơn 2000 theo tình trạng hôn nhân và loại xe

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Marital Status].CHILDREN, [Dim Vehicle].[Vehicle Category].CHILDREN)
HAVING [Measures].[Fact Count] > 2000 ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Thống kê số báo cáo lớn hơn 2000 theo tình trạng hôn nhân và loại xe.
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Marital Status].CHILDREN, [Dim Vehicle].[Vehicle Category].CHILDREN)
HAVING [Measures].[Fact Count] > 2000 ON ROWS
FROM [Car Claimfn WH];
```

Results

Vehicle Category	Fact Count
Compact	4267
Large	4271
Medium	4284

### 3.3.8. Thống kê số báo cáo theo zip code và giới tính của tài xe với loại xe Compact

```
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Zip code].CHILDREN, [Dim Driver].[Gender].CHILDREN)
HAVING [Measures].[Fact Count] ON ROWS
FROM [Car Claimfn WH]
WHERE ({[Dim Vehicle].[Vehicle Category].&[Compact]});
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Thống kê so bao cao theo zip code và giới tính của tài xe được claim bảo hiểm với số lần claim trước đó
SELECT NON EMPTY [Measures].[Fact Count] ON COLUMNS,
NON EMPTY([Dim Driver].[Zip code].CHILDREN, [Dim Driver].[Gender].CHILDREN)
HAVING [Measures].[Fact Count] ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Vehicle].[Vehicle Category].&[Compact]);
```

	Fact Count
0 F	4
0 M	5
15001 F	49
15001 M	53
15003 F	10
15003 M	12
15004 F	9
15004 M	4
15005 F	7
15005 M	3
15006 F	9
15006 M	9
15007 F	6
15007 M	7
15009 F	35
15009 M	40
15010 F	36
15010 M	29
15012 F	44
15012 M	45

Query executed successfully.

### 3.3.9. Thống kê tổng số lần claim bảo hiểm trước đó của những báo cáo trong năm 2015 và giới tính tài xe là nam

```
SELECT NON EMPTY [Measures].[Past Num of Claims] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Thống kê tổng số lần claim bảo hiểm trước đó của những báo cáo trong năm 2015 và giới tính tài xe là nam
SELECT NON EMPTY [Measures].[Past Num of Claims] ON COLUMNS,
NON EMPTY([Dim Driver].[Gender].CHILDREN) ON ROWS
FROM [Car Claimfn WH]
```

	Past Num Of Claims
F	4261
M	4826

### 3.3.10. Thống kê top 3 báo cáo có tài xe có thu nhập cao nhất trong năm 2016

```
SELECT NON EMPTY [Measures].[Annual Income] ON COLUMNS,
NON EMPTY TOPCOUNT ([Fact].[Claim number].CHILDREN,3,[Measures].[Annual Income]) ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Time].[Claim Year].&[2016]);
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phantom)\* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query MDX DAX EML DAT Execute

CarClaim\_Final

MDXQuery2.mdx ...P-GEOCLFKE\phantom\*

Cube: Car Claimfn WH

Metadata Functions

Search Model

Measure Group: <All>

Car Claimfn WH

Measures

Fact

Age Of Driver

Age Of Vehicle

Annual Income

Fact Count

Fraud

Maximum Safety Rating

```
-- Thống kê top 3 báo cáo có tài xế có thu nhập cao nhất trong năm 2016
SELECT NON EMPTY [Measures].[Annual Income] ON COLUMNS,
NON EMPTY TOPCOUNT ([Fact].[Claim number].CHILDREN,3,[Measures].[Annual Income]) ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Time].[Claim Year].&[2016]);
```

121 %

Messages Results

	Annual Income
1096	46066
3383	45960
10891	45842

### **3.3.11. Lọc những ngày có số báo cáo trên 30**

```
SELECT NON EMPTY [Measures].[Fact count] ON COLUMNS,  
NON EMPTY([Dim time].[Claim Date].CHILDREN)  
HAVING [Measures].[Fact count] > 30 ON ROWS  
FROM [Car Claimfn WH];
```

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phantom)\* - Microsoft SQL Server Management Studio". The left pane displays the cube structure for "Car Claimfn WH", including dimensions like Fact, KPIs, and Dim Driver, and measures like Fact Count. The main pane shows the execution results of an MDX query:

```
-- Lọc những ngày có số lượng báo cáo trên 30
SELECT NON EMPTY [Measures].[Fact count] ON COLUMNS,
NON EMPTY([Dim time].[Claim Date].CHILDREN )
HAVING [Measures].[Fact count] > 30 ON ROWS
FROM [Car Claimfn WH];
```

Fact Date	Fact Count
2015-01-03	34
2015-01-07	32
2015-01-09	35
2015-01-15	33
2015-01-18	36
2015-01-23	36
2015-01-25	31
2015-02-03	31
2015-02-04	32
2015-02-06	31
2015-02-13	35
2015-02-22	34
2015-02-23	32
2015-03-01	32
2015-03-03	32
2015-03-05	32
2015-03-16	32
2015-03-21	34
2015-04-20	37
2015-05-03	34
2015-05-06	31
2015-05-13	32

### **3.3.12. Thống kê số báo cáo gian lận theo ngày**

```
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,  
NON EMPTY([Dim time].[Claim Date].CHILDREN) on rows  
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant) - Microsoft SQL Server Management Studio

```
-- Thông kê số báo cáo gian lận theo ngày
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,
NON EMPTY([Dim time].[Claim Date].CHILDREN ) on rows
FROM [Car Claimfn WH];
```

The screenshot shows the Microsoft SQL Server Management Studio interface. The left pane displays the cube structure for 'Car Claimfn WH' with various dimensions like Dim Driver, Dim Time, and Fact, and measures like Fraud, Maximum Safety Rating, and Past Num Of Claims. The main pane contains the MDX query. The results pane on the right shows a table titled 'Fraud' with data from January 1, 2015, to January 23, 2015, with values ranging from 1 to 10.

Date	Fraud
2015-01-01	3
2015-01-02	6
2015-01-03	3
2015-01-04	8
2015-01-05	4
2015-01-06	4
2015-01-07	7
2015-01-08	3
2015-01-09	3
2015-01-10	6
2015-01-11	6
2015-01-12	7
2015-01-13	6
2015-01-14	1
2015-01-15	4
2015-01-16	2
2015-01-17	3
2015-01-18	4
2015-01-19	5
2015-01-20	1
2015-01-21	10
2015-01-22	3
2015-01-23	3

Query executed successfully.

### 3.3.13. Thông kê năm có nhiều báo cáo gian lận nhất

```
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,
NON EMPTY TOPCOUNT ([Dim time].[Claim year].[Claim year],1,[Measures].[Fraud]) ON ROWS
FROM [Car Claimfn WH];
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant) - Microsoft SQL Server Management Studio

```
-- Thông kê năm có nhiều báo cáo gian lận nhất
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,
NON EMPTY TOPCOUNT ([Dim time].[Claim year].[Claim year],1,[Measures].[Fraud]) ON ROWS
FROM [Car Claimfn WH];
```

The screenshot shows the Microsoft SQL Server Management Studio interface. The left pane displays the cube structure for 'Car Claimfn WH'. The main pane contains the MDX query. The results pane on the right shows a table titled 'Fraud' with one row for the year 2016, indicating 1522 fraudulent claims.

Year	Fraud
2016	1522

Query executed successfully.

### 3.3.14. Thống kê top 3 báo cáo có số lần claim bảo hiểm trước đó cao nhất trong năm 2015

```
SELECT NON EMPTY [Measures].[Past Num of Claims] ON COLUMNS,  
NON EMPTY TOPCOUNT ([Fact].[Claim  
number].CHILDREN,3,[Measures].[Past Num of Claims]) ON ROWS  
FROM [Car Claimfn WH]  
WHERE ([Dim Time].[Claim Year].&[2015]);
```

The screenshot shows the Microsoft SQL Server Management Studio interface with the following details:

- Title Bar:** MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phantom)\* - Microsoft SQL Server Management Studio
- Toolbar:** Standard toolbar with various icons for file operations, queries, and database management.
- Query Editor:** The main area contains the MDX query:

```
-- Thống kê top 3 báo cáo có số lần claim bảo hiểm trước đó cao nhất trong năm 2015  
SELECT NON EMPTY [Measures].[Past Num of Claims] ON COLUMNS,  
NON EMPTY TOPCOUNT ([Fact].[Claim number].CHILDREN,3,[Measures].[Past Num of Claims]) ON ROWS  
FROM [Car Claimfn WH]  
WHERE ([Dim Time].[Claim Year].&[2015]);
```
- Results Window:** Shows the execution results:

Past Num of Claims
14581
15832
18555
- Object Explorer:** On the left, it shows the cube structure:
  - Cube: Car Claimfn WH
  - Metadata & Functions
  - Search Model
  - Measure Group: <All>
  - Fact: Age Of Driver, Age Of Vehicle, Annual Income, Fact Count, Fraud, Maximum Safety Rating, Past Num Of Claims, Safety Rating
  - KPIs
  - Dim Driver: Claim Number, Gender, Living Status, Marital Status, Zip Code
  - Dim Time: Claim Date, Claim Month, Claim Quarter, Claim Year, Hierarchy
  - Dim Vehicle
  - Fact: Claim Number
- Status Bar:** Query executed successfully. Laptop-GEOCLFKE | LAPTOP-GEOCLFKE\phantom | CarClaim\_Final | 00:00:01

### 3.3.15. Thống kê tình trạng hôn nhân có số báo cáo gian lận nhiều nhất trong quý 3 năm 2016

```
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,  
NON EMPTY TOPCOUNT([Dim Driver].[Marital Status].CHILDREN,1,  
[Measures].[Fraud]) ON  
ROWS  
FROM [Car Claimfn WH]  
WHERE ([Dim Time].[Claim Year].&[2016], [Dim Time].[Claim Quarter].&[3]);
```

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Thông kê tình trạng hôn nhân có số báo cáo gian lận nhiều nhất vào quý 3 năm 2016
SELECT NON EMPTY [Measures].[Fraud] ON COLUMNS,
NON EMPTY TOPCOUNT([Dim Driver].[Marital Status].CHILDREN,1, [Measures].[Fraud]) ON ROWS
FROM [Car Claimfn WH]
WHERE ([Dim Time].[Claim Year].&[2016], [Dim Time].[Claim Quarter].&[3]);
```

Query executed successfully.

### 3.3.16. Pivot: Thông kê số báo cáo theo ngày và theo tình trạng nhà ở

**SELECT NON EMPTY ([Measures].[Fact Count], [Dim Driver].[Living Status].CHILDREN) ON COLUMNS,**

**NON EMPTY [Dim Time].[Claim Date].CHILDREN ON ROWS**

**FROM [Car Claimfn WH];**

MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio

```
-- Pivot: Thông kê số báo cáo theo ngày và theo tình trạng nhà ở
SELECT NON EMPTY ([Measures].[Fact Count], [Dim Driver].[Living Status].CHILDREN) ON COLUMNS,
NON EMPTY [Dim Time].[Claim Date].CHILDREN ON ROWS
FROM [Car Claimfn WH];
```

	Fact Count	Fact Count	Fact Count
F	2	2499	6083
M	3	2670	6739

Query executed successfully.

### 3.3.17. Pivot: Thống kê số báo cáo theo giới tính và theo tình trạng hôn nhân

**SELECT NON EMPTY ([Measures].[Fact Count], [Dim Driver].[Marital status].  
CHILDREN) ON COLUMNS,**

**NON EMPTY [Dim Driver].[Gender].CHILDREN ON ROWS**

**FROM [Car Claimfn WH];**

	Fact Count	Fact Count
F	0.0	10
M	2499	6739

**SELECT NON EMPTY ([Measures].[Fact Count], [Dim**

**Driver].[Gender].CHILDREN) ON COLUMNS,**

**NON EMPTY [Dim Driver].[Marital status].CHILDREN ON ROWS**

**FROM [Car Claimfn WH];**

	Fact Count	Fact Count
F	0.0	1.0
M	2	3

### 3.3.18. Pivot: Thống kê số báo cáo gian lận theo loại xe và theo zip code

**SELECT NON EMPTY ([Measures].[Fact Count], [Dim Driver].[zip code].  
CHILDREN) ON COLUMNS,**

**NON EMPTY [Dim vehicle].[vehicle category].CHILDREN ON ROWS**

**FROM [Car Claimfn WH];**

The screenshot shows the Microsoft SQL Server Management Studio interface with the following details:

- Title Bar:** MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio
- Toolbar:** Standard toolbar with various icons for file operations, queries, and database management.
- Query Editor:** Contains the MDX query:
 

```
-- Pivot: Thống kê số báo cáo gian lận theo loại xe và zip code
SELECT NON EMPTY ([Measures].[Fact Count], [Dim Driver].[zip code].CHILDREN) ON COLUMNS,
NON EMPTY [Dim vehicle].[vehicle category].CHILDREN ON ROWS
FROM [Car Claimfn WH];
```
- Results Grid:** Displays a pivot table with columns for Fact Count (0, 15001, 15003, 15004, 15005, 15006, 15007, 15009, 15010, 15012, 15014, 15015, 15017, 15018) and rows for Compact, Large, and Medium categories. The data is summarized as follows:
 

	Fact Count											
Compact	9	102	22	13	10	18	13	75	65	89	18	14
Large	7	81	30	17	15	19	14	63	51	95	9	15
Medium	7	100	35	10	15	19	19	70	43	92	17	16
- Status Bar:** Shows "Query executed successfully." and other session details like "Laptop-GEOCLFKE" and "00:00:01".

**SELECT NON EMPTY ([Measures].[Fact Count], [Dim vehicle].[vehicle  
category].CHILDREN) ON COLUMNS,**

**NON EMPTY [Dim Driver].[zip code].CHILDREN ON ROWS**

**FROM [Car Claimfn WH];**

The screenshot shows the Microsoft SQL Server Management Studio interface with the following details:

- Title Bar:** MDXQuery2.mdx - LAPTOP-GEOCLFKE.CarClaim\_Final (LAPTOP-GEOCLFKE\phant)\* - Microsoft SQL Server Management Studio
- Toolbar:** Standard toolbar with various icons for file operations, queries, and database management.
- Query Editor:** Contains the MDX query:
 

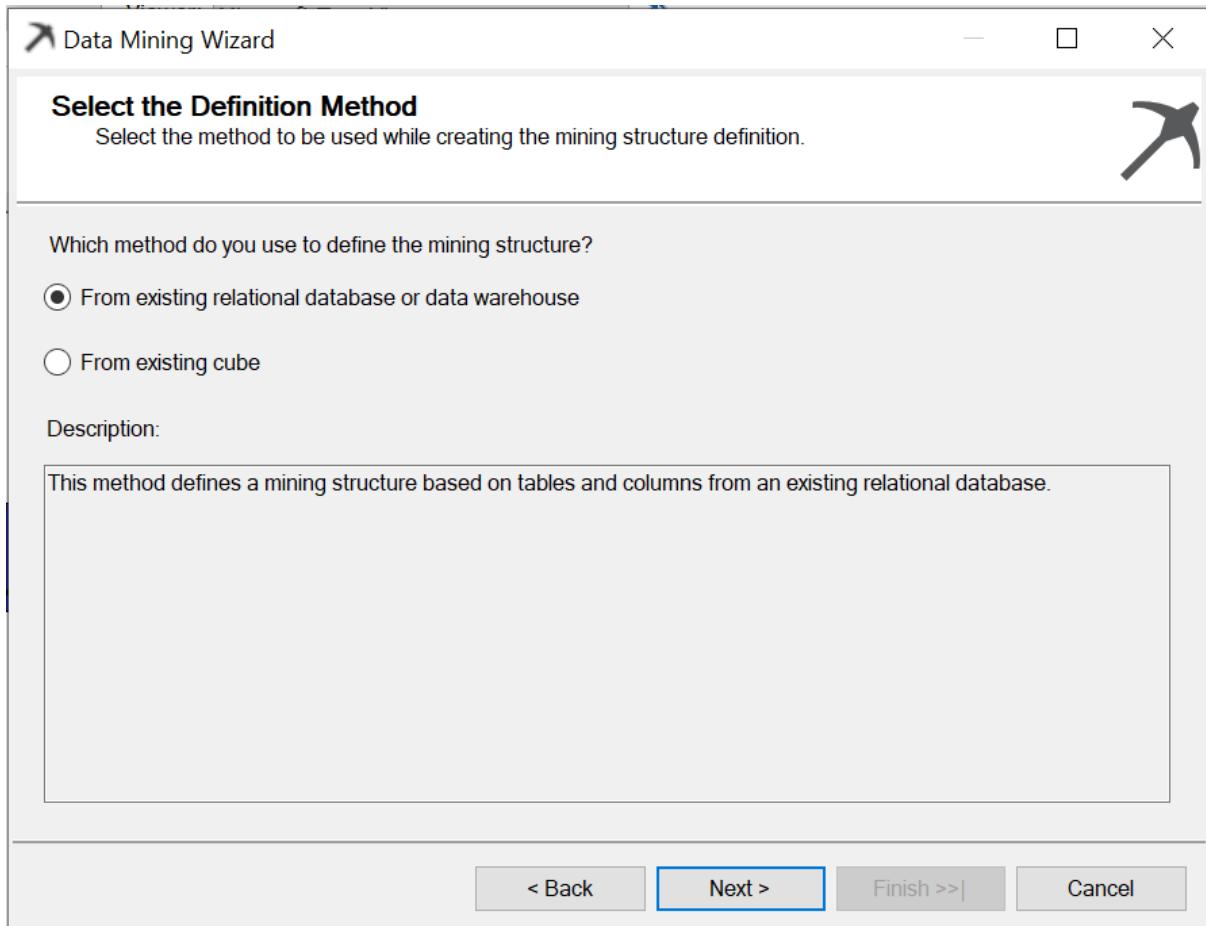
```
SELECT NON EMPTY ([Measures].[Fact Count], [Dim vehicle].[vehicle category].CHILDREN) ON COLUMNS,
NON EMPTY [Dim Driver].[zip code].CHILDREN ON ROWS
FROM [Car Claimfn WH];
```
- Results Grid:** Displays a pivot table with columns for Fact Count (0, 15001, 15003, 15004, 15005, 15006, 15007, 15009, 15010, 15012, 15014, 15015, 15017, 15018) and rows for Compact, Large, and Medium categories. The data is summarized as follows:
 

	Fact Count	Fact Count	Fact Count
0	9	7	7
15001	102	81	100
15003	22	30	35
15004	13	17	10
15005	10	15	15
15006	18	19	19
15007	13	14	19
15009	75	63	70
15010	65	51	43
15012	89	95	92
15014	18	9	17
15015	14	15	16
15017	15	15	14
15018	17	16	13
15019	18	20	10
15020	16	24	11
- Status Bar:** Shows "Query executed successfully." and other session details like "Laptop-GEOCLFKE" and "00:00:01".

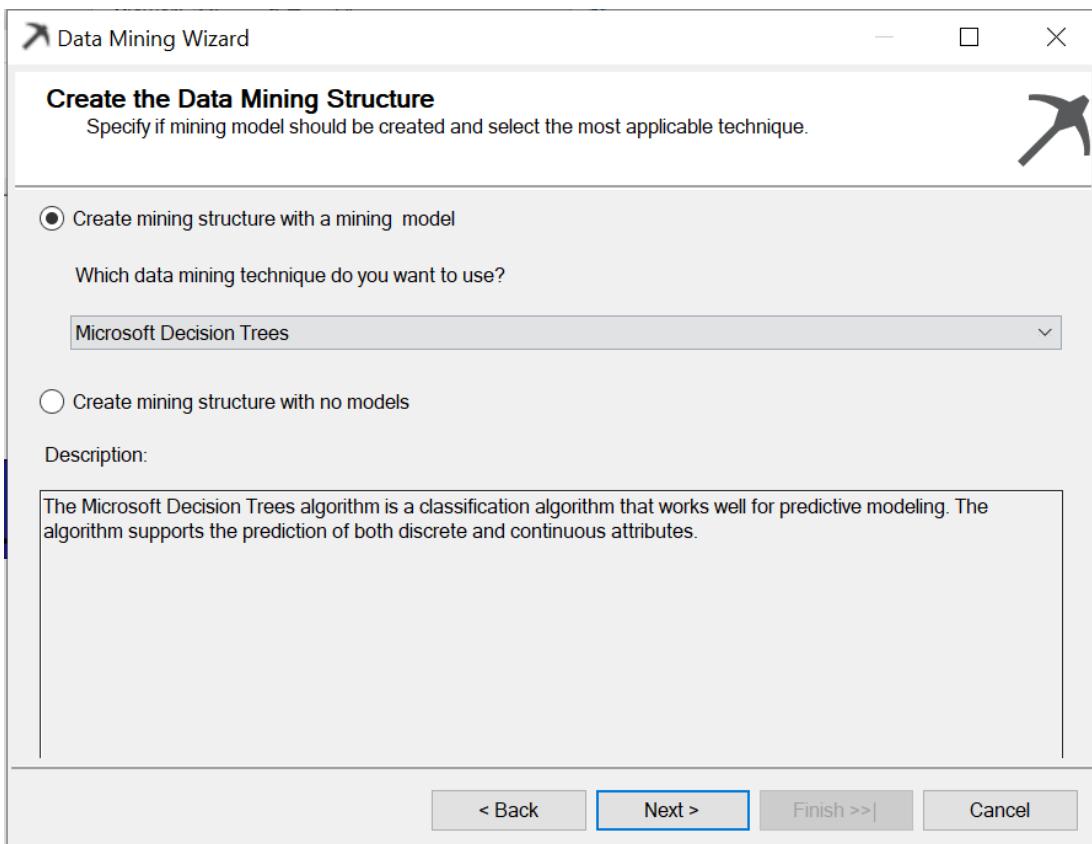
## CHƯƠNG 4: KHAI PHÁ DỮ LIỆU – DATAMINING

### 4.1. Mining Structure

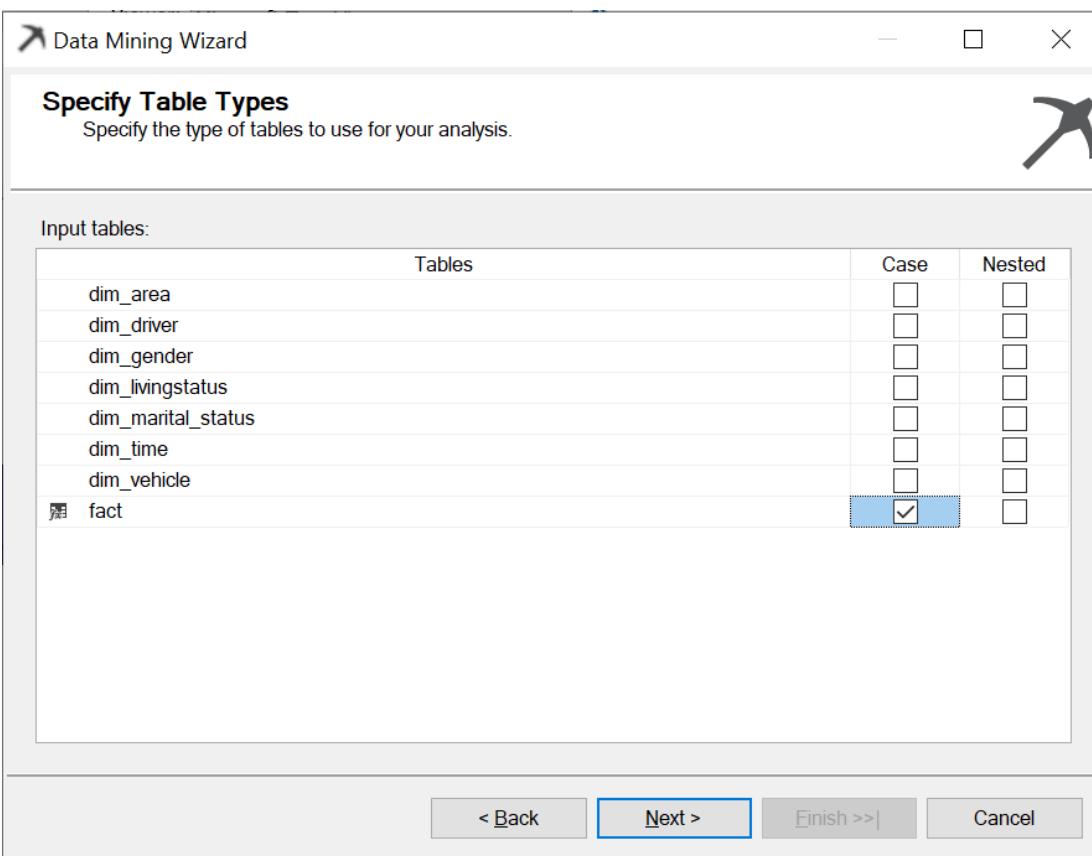
- Chọn method database có sẵn.



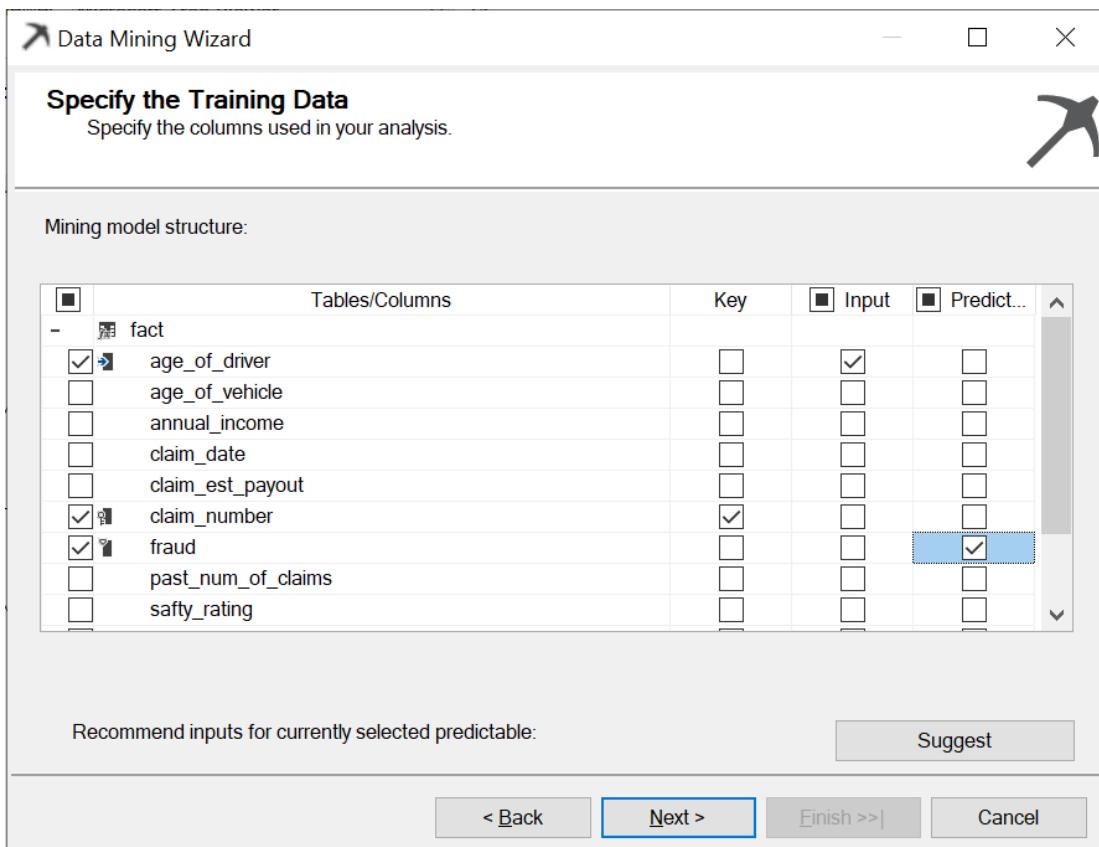
- Chọn datamining technique là Microsoft Decision Trees.



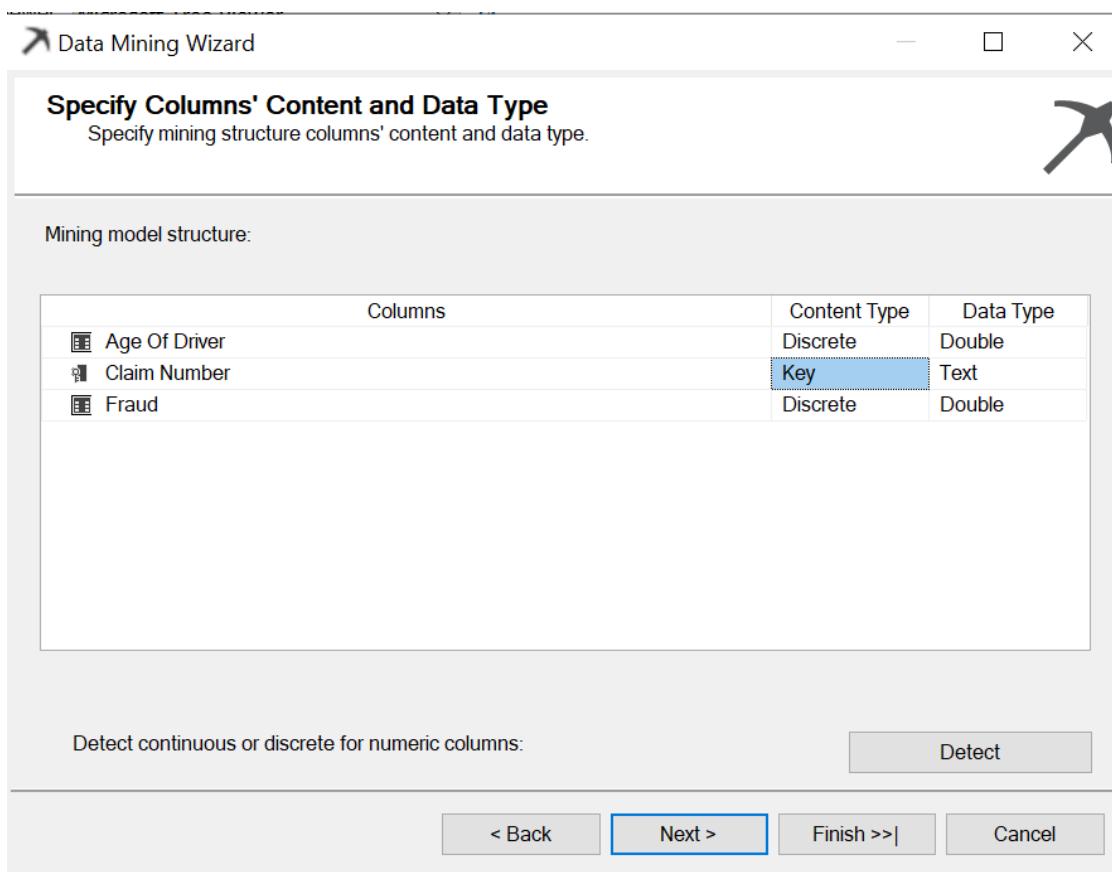
- Chọn table



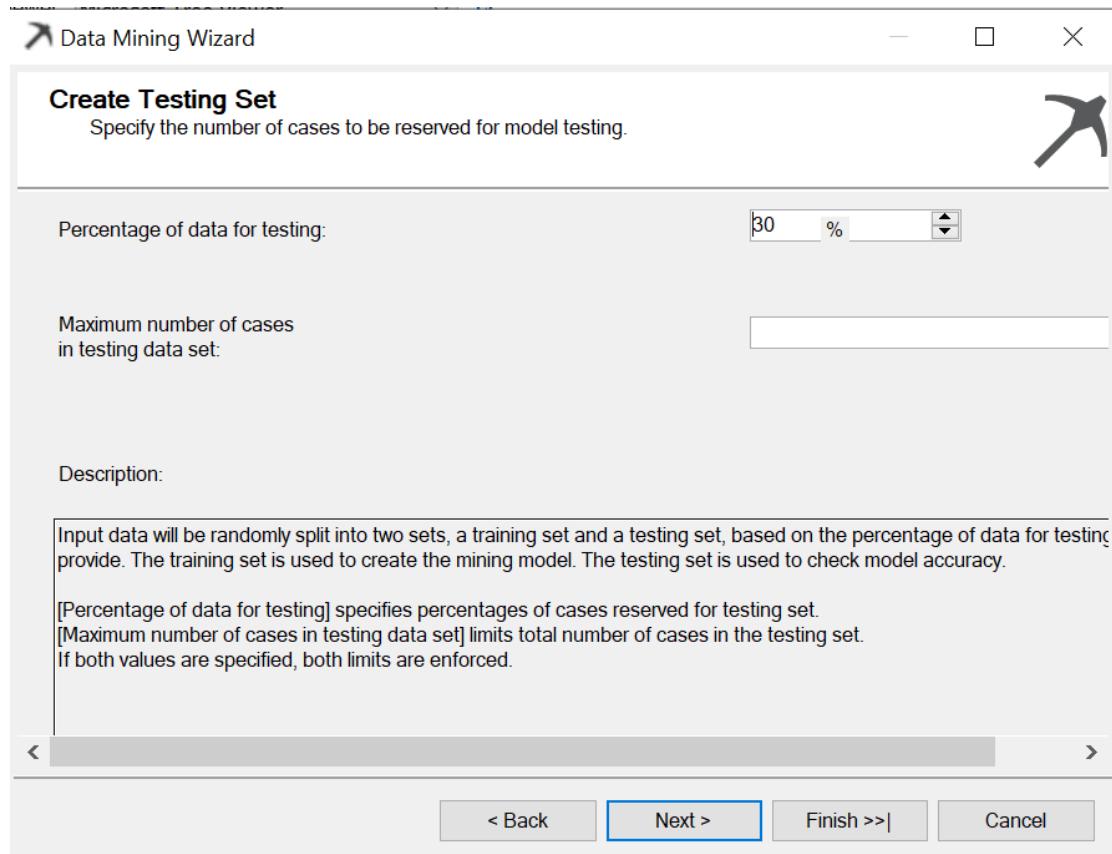
- Chọn thuộc tính input và prediction.



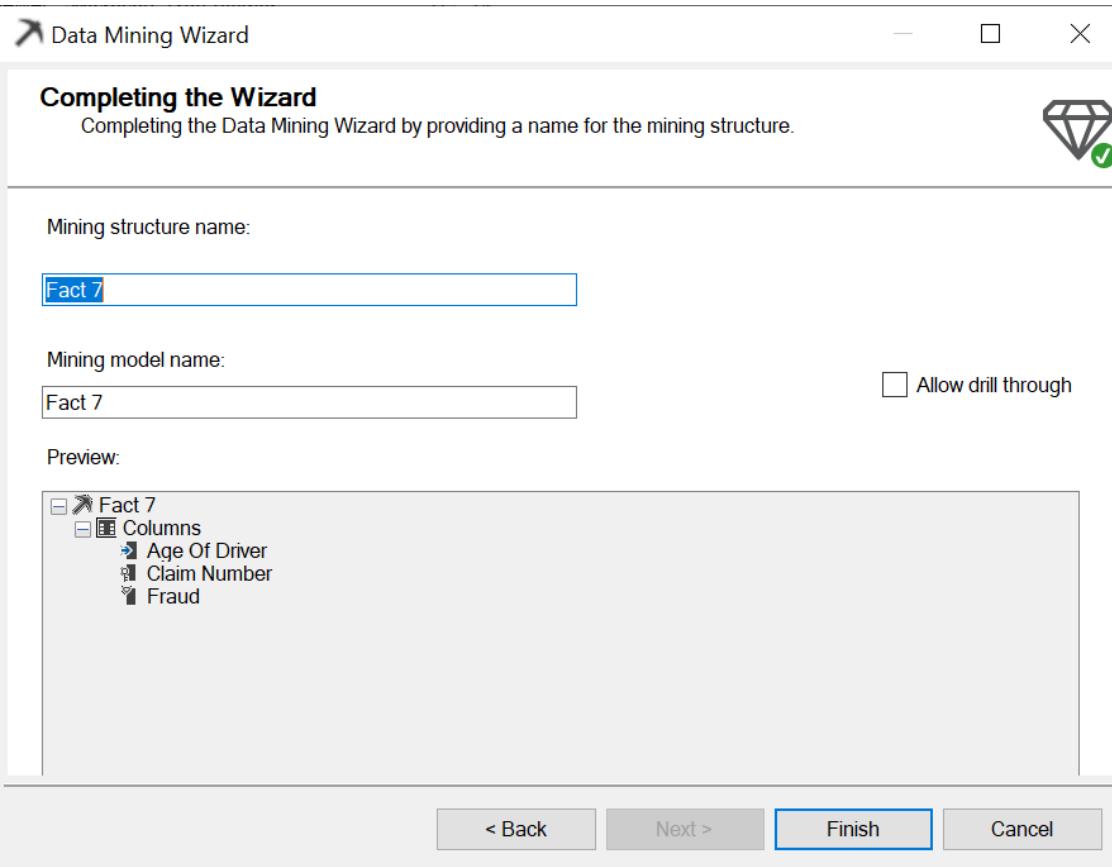
- Chọn datatype.



- Cài đặt 30% cho tập test và 70% cho tập train.



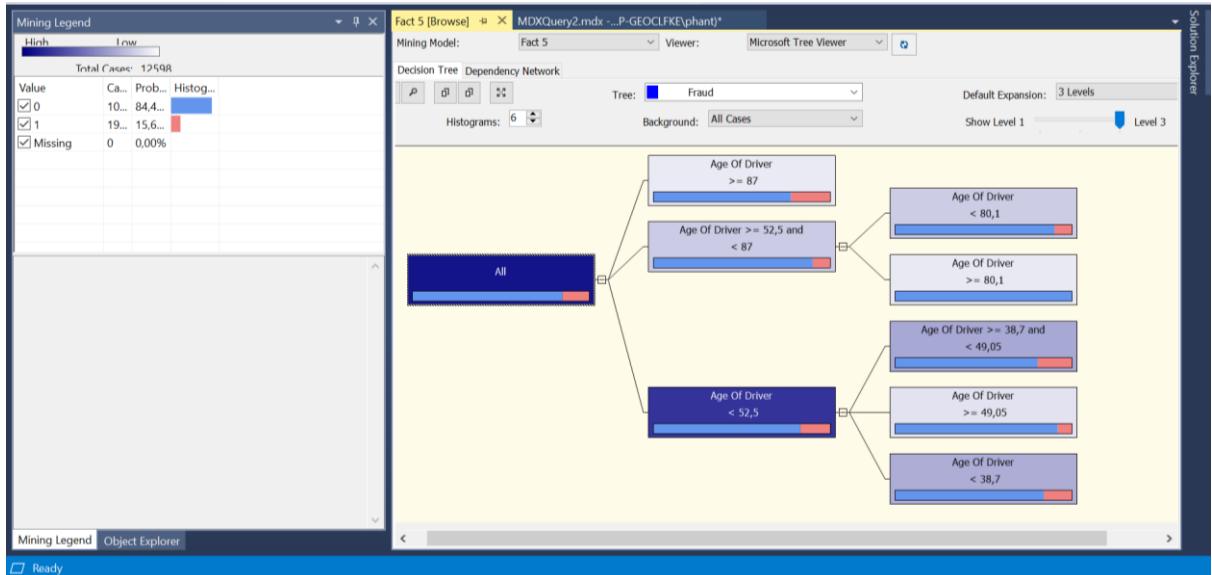
- Hoàn tất tạo một mining structure.



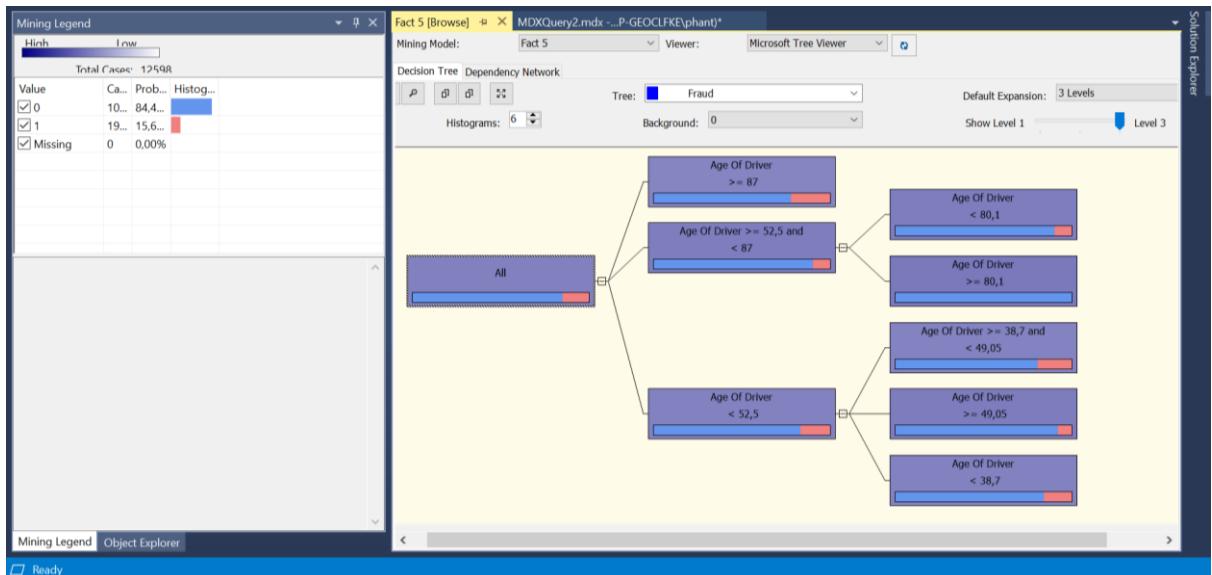
## 4.2. Mining Model

Sau khi thực hiện lần lượt các thao tác Process và Browse, chúng ta sẽ nhận được mô hình mining.

- Background: All cases (0, 1, Missing)

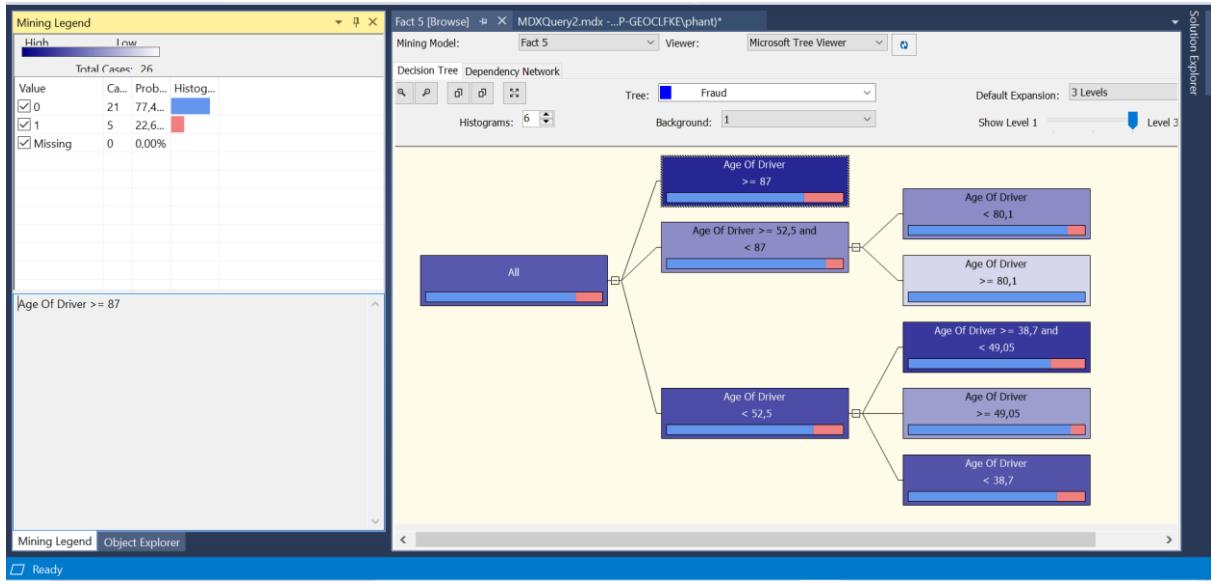


- Background: 0



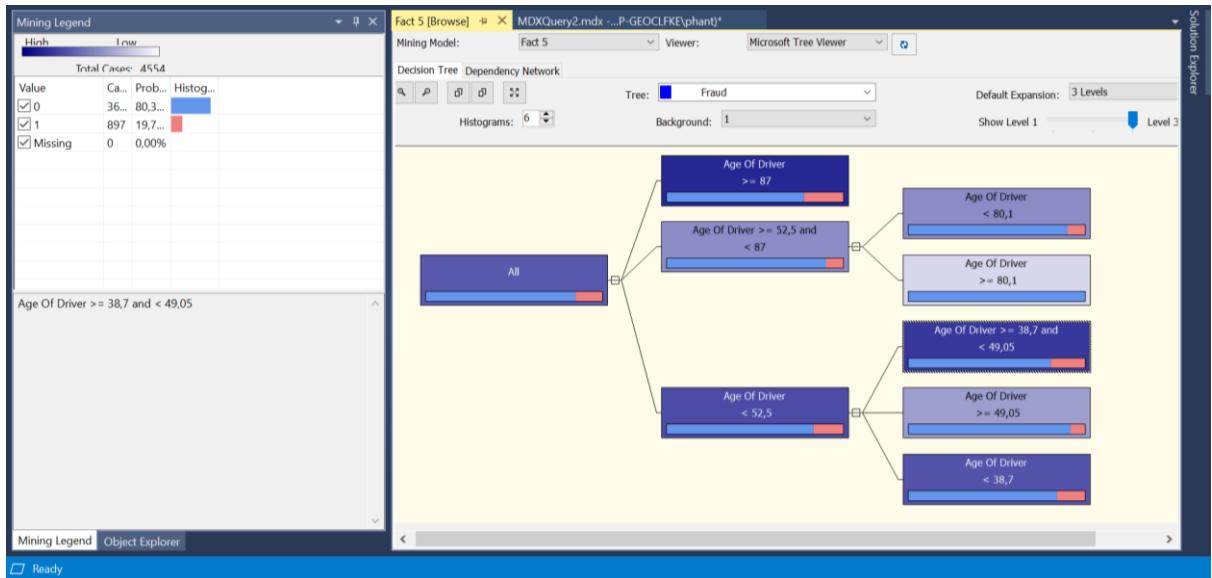
Đối với trường hợp này ta thấy luật không rõ ràng.

- Background: 1



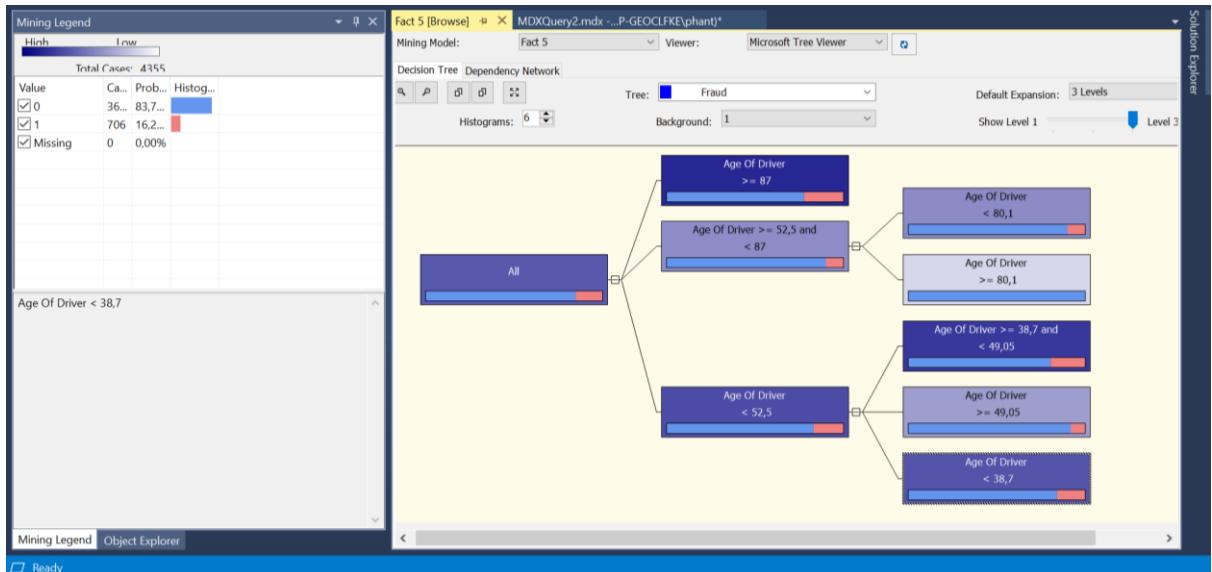
If Age of Driver  $\geq 87$  then Fraud = 1

⇒ Luật: “**Nếu tuổi của tài xế lớn hơn hoặc bằng 87 thì báo cáo là gian lận (77,4%)**”



If Age of Driver  $\geq 38,7$  and  $< 49,05$  then Fraud = 1

⇒ Luật: “**Nếu tuổi của tài xế lớn hơn hoặc bằng 38,7 và nhỏ hơn 49,05 thì báo cáo là gian lận (80,3%)**”



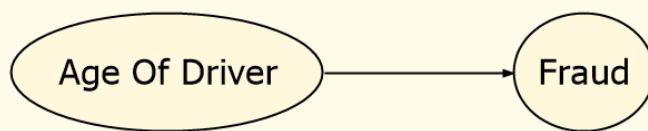
If Age of Driver  $< 38,7$  then Fraud = 1

⇒ Luật: “**Nếu tuổi của tài xế nhỏ hơn 38,7 thì báo cáo là gian lận (83,7%)**”

Từ 3 luật trên, ta có thể gom lại thành luật duy nhất với độ chính xác trên 80% (Trung bình độ chính xác của 3 luật) .

Luật: “**Nếu tuổi của tài xế lớn hơn hoặc bằng 87 hoặc nhỏ hơn 49,05 thì báo cáo là gian lận**”.

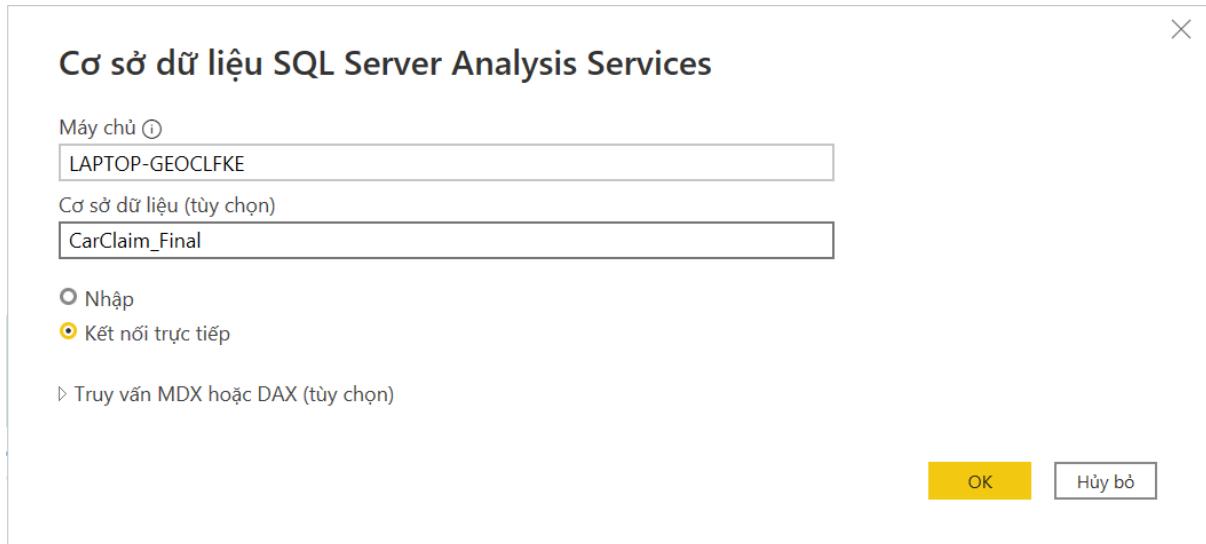
Các yếu tố ảnh hưởng tới Fraud:



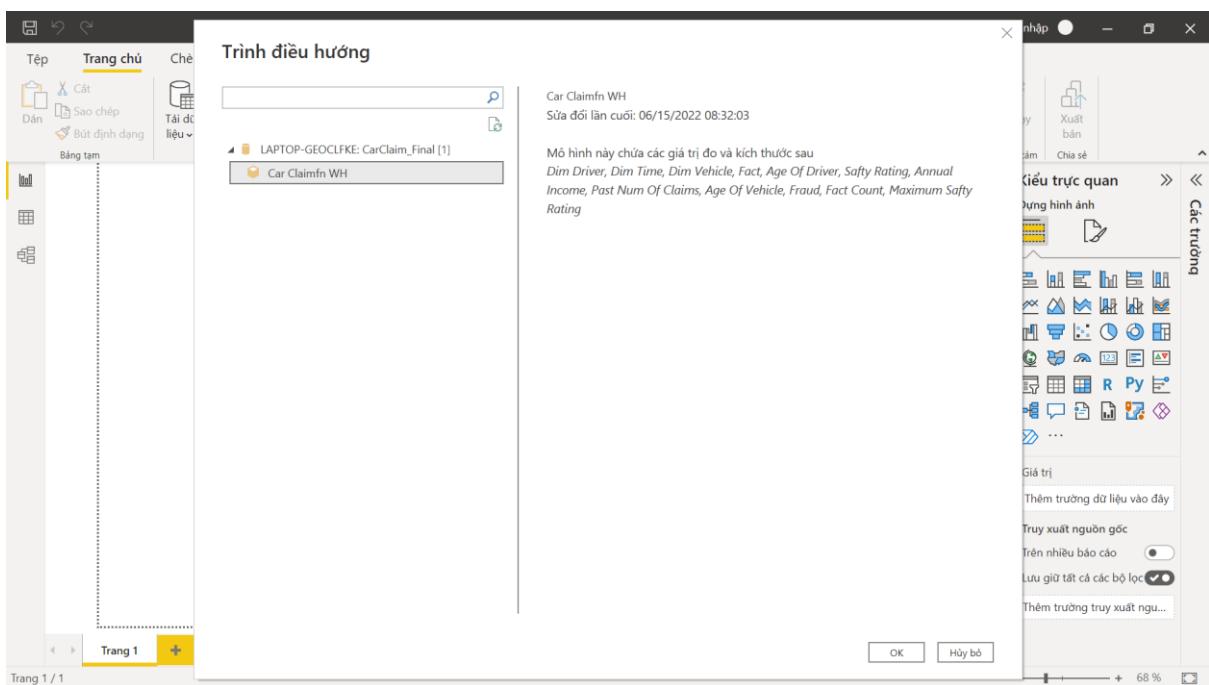
## CHƯƠNG 5: POWER BI

### 5.1. Tạo Project Power BI

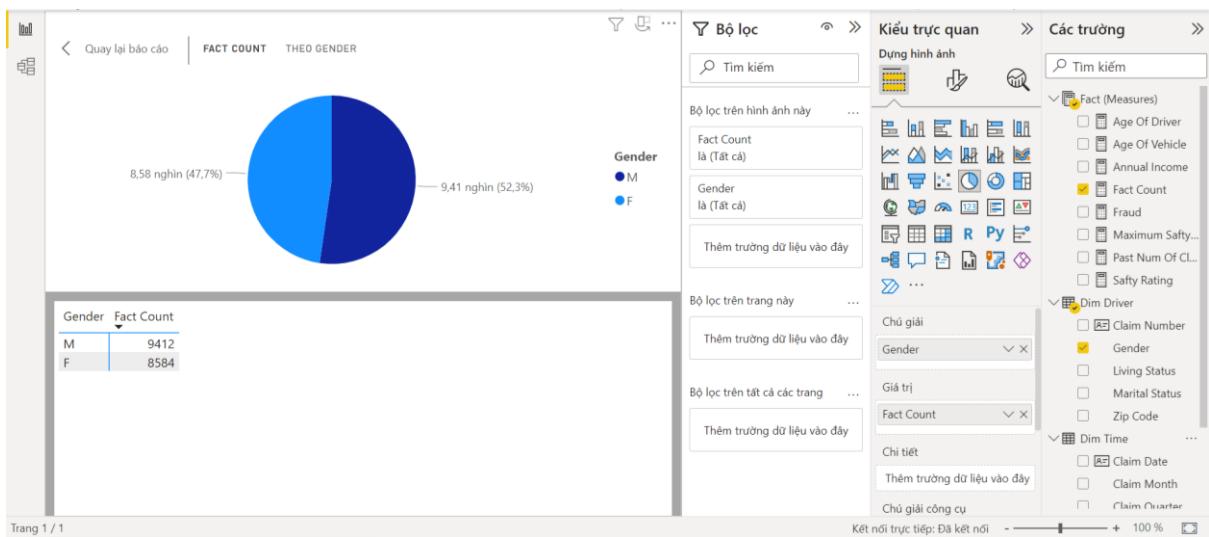
- Chọn Get data -> Analysis Services, điền tên máy chủ và database.



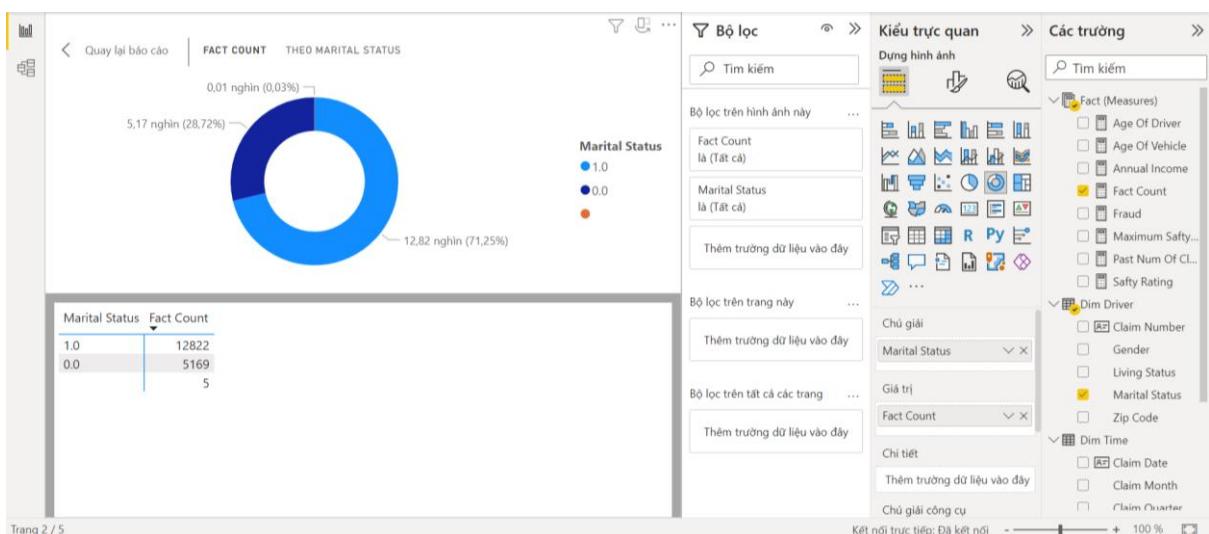
- Hoàn tất việc tạo project Power BI.



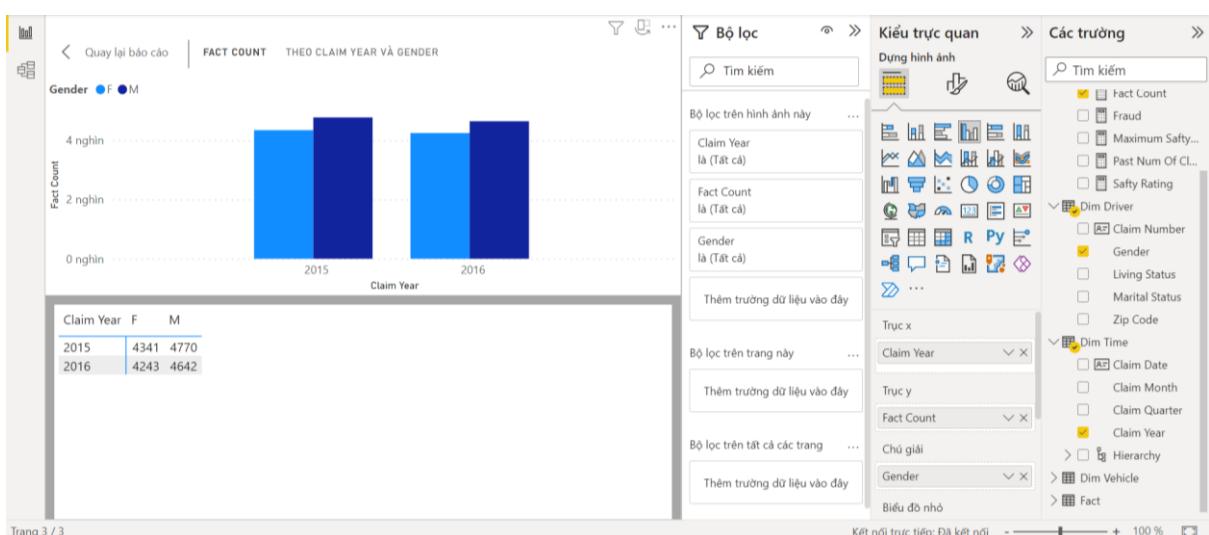
## 5.2. Tạo báo cáo dựa trên giới tính của tài xế



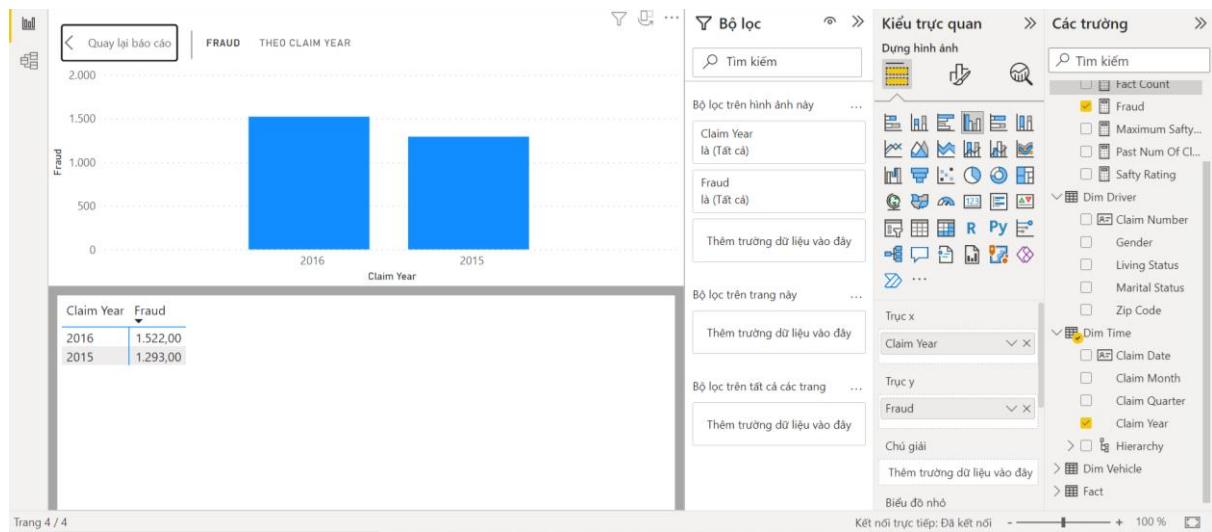
## 5.3. Tạo báo cáo dựa trên tình trạng hôn nhân của tài xế



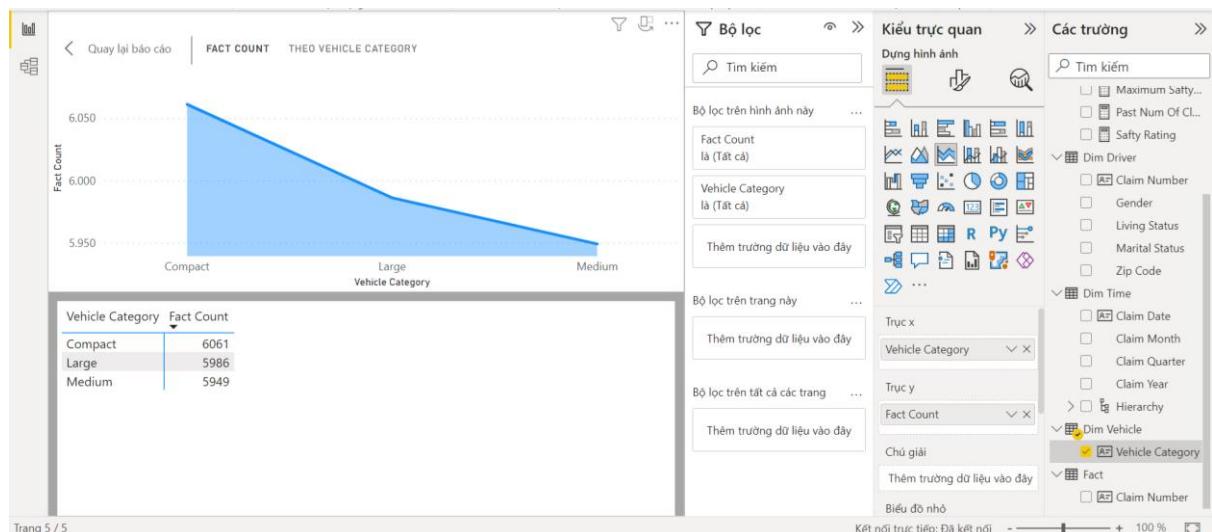
## 5.4. Thống kê số báo cáo theo năm và theo giới tính



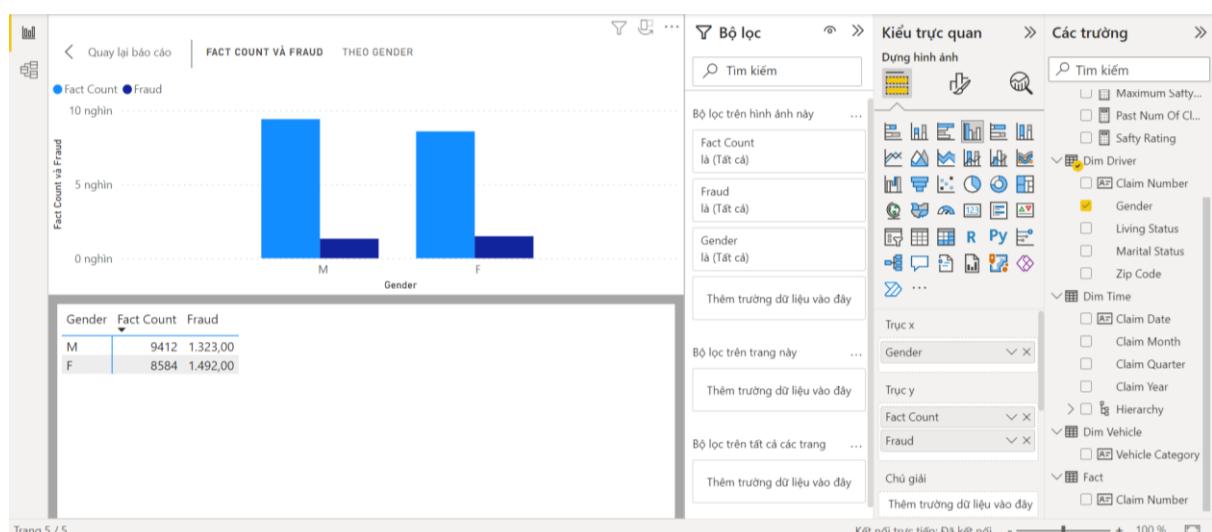
## 5.5. Thông kê số báo cáo gian lận theo từng năm



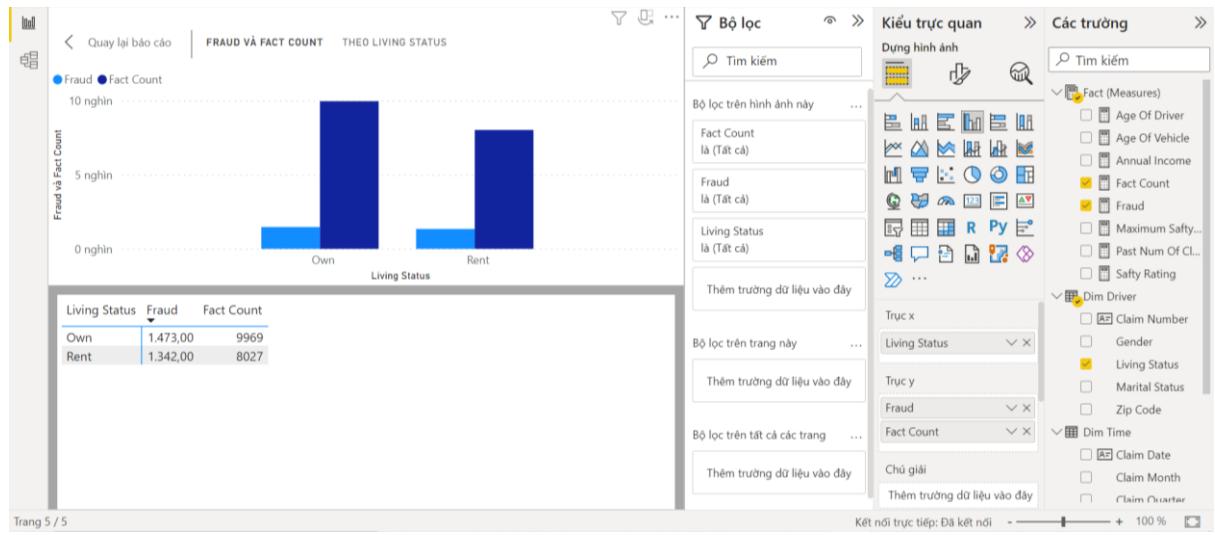
## 5.6. Thông kê số báo cáo theo từng loại xe



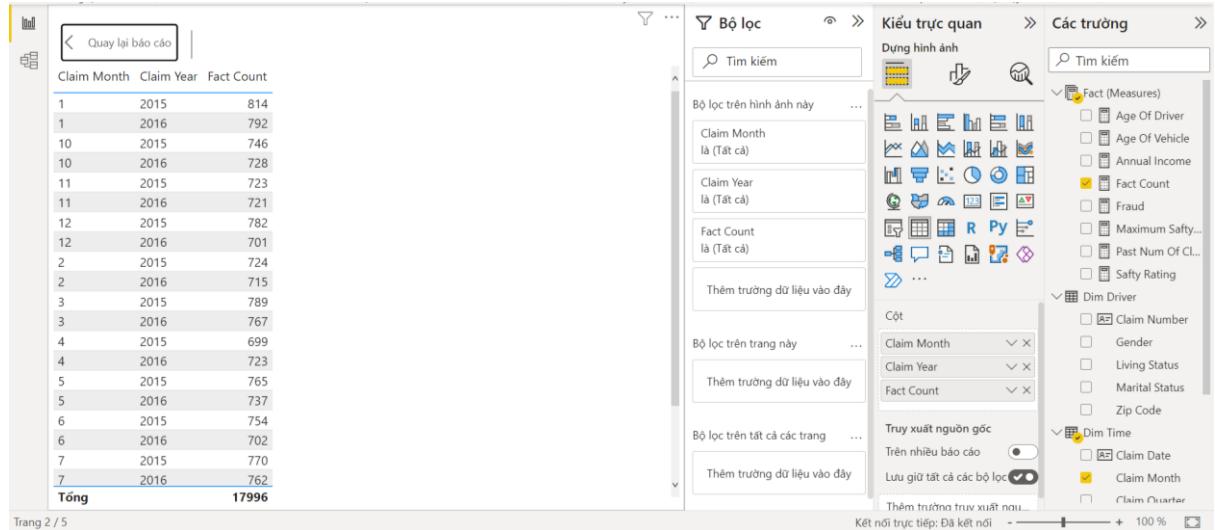
## 5.7. Thông kê số báo cáo và số báo cáo gian lận theo từng giới tính



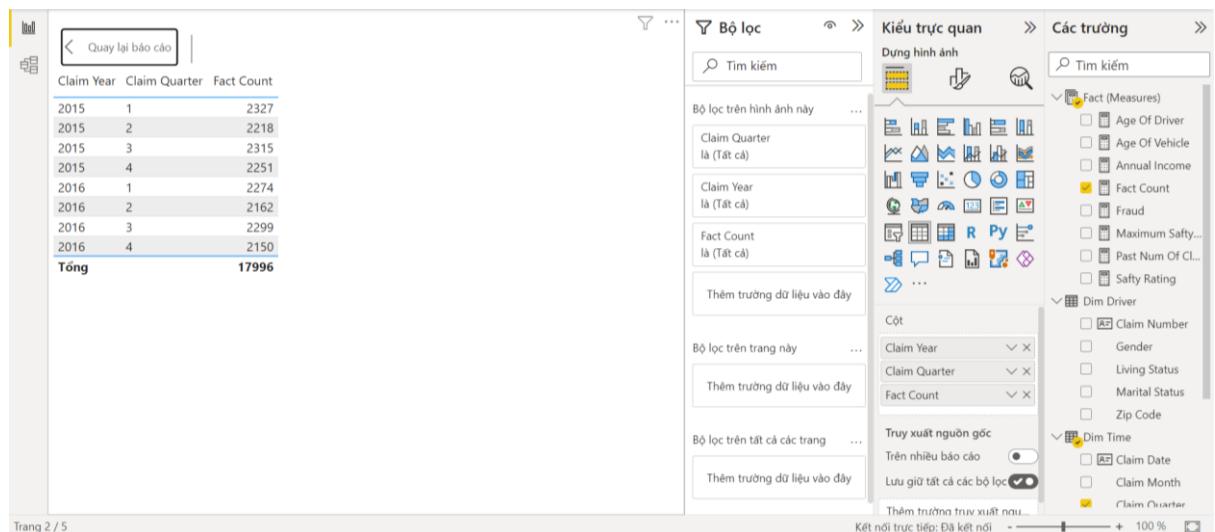
## 5.8. Thống kê số báo cáo và số báo cáo gian lận theo tình trạng nhà ở



## 5.9. Thống kê số báo cáo theo từng tháng, năm



## 5.10. Thống kê số báo cáo theo từng quý, năm



## 5.11. Pivot: Thông kê số báo cáo theo từng năm và tình trạng nhà ở

Two screenshots of a Power BI report interface showing pivot tables and filtering.

**Screenshot 1:** A pivot table showing the count of reports by year and ownership status. The table has columns: Claim Year, Own, Rent, and Tổng. Rows show data for 2015 and 2016, with a total row for Tổng.

	Own	Rent	Tổng
2015	5047	4064	<b>9111</b>
2016	4922	3963	<b>8885</b>
<b>Tổng</b>	<b>9969</b>	<b>8027</b>	<b>17996</b>

**Screenshot 2:** A pivot table showing the count of reports by year and living status. The table has columns: Living Status, 2015, 2016, and Tổng. Rows show data for Own and Rent, with a total row for Tổng.

	2015	2016	Tổng
Own	5047	4922	<b>9969</b>
Rent	4064	3963	<b>8027</b>
<b>Tổng</b>	<b>9111</b>	<b>8885</b>	<b>17996</b>

The interface includes a search bar, filter panes for 'Bộ lọc' (Filter), 'Kiểu trực quan' (Type of visualization), and 'Các trường' (Fields), and a 'Kết nối trực tiếp' (Direct connection) section at the bottom right.

## **TÀI LIỆU THAM KHẢO**

<https://www.techtarget.com/searchbusinessanalytics/definition/data-mining>  
<https://docs.microsoft.com/en-us/analysis-services/ssas-overview?view=sql-analysis-services-2022>  
<https://docs.microsoft.com/en-us/sql/integration-services/lesson-1-create-a-project-and-basic-package-with-ssis?view=sql-server-ver16>