

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



**BÁO CÁO ĐỒ ÁN**

Môn học: Kho dữ liệu và OLAP

Học kỳ I (2020-2021)

**PHÂN TÍCH DỮ LIỆU TAI NẠN GIAO THÔNG  
CỦA UK**

**Giáo viên hướng dẫn: Đỗ Thị Minh Phụng**

**Sinh viên thực hiện:**

Nguyễn Đình Văn Khoa - 14520426

Phạm Quốc Cường – 16520157

Lớp: IS217.L11

TP. Hồ Chí Minh, ngày 11 tháng 10 năm 2020

## LỜI CẢM ƠN

Lời đầu tiên, nhóm xin gửi lời cảm ơn chân thành đến quý Thầy Cô trường Đại học Công nghệ thông tin đặc biệt là quý Thầy Cô Khoa Hệ thống thông tin - những người đã dùng tri thức và tâm huyết của mình để truyền đạt cho chúng em vốn kiến thức vô cùng quý báu trong khoảng thời gian học tập tại trường. Những kiến thức mà Thầy Cô truyền đạt là bước đệm quan trọng giúp chúng em có thể hoàn thành đề tài tốt hơn.

Nhóm xin gửi lời cảm ơn đặc biệt chân thành tới Cô Đỗ Thị Minh Phụng - giảng viên môn Kho dữ liệu và OLAP đã tận tình giúp đỡ, trực tiếp chỉ bảo, hướng dẫn nhóm trong suốt quá trình làm đồ án môn học. Nhờ đó, chúng em đã tiếp thu được nhiều kiến thức bổ ích trong việc vận dụng cũng như kỹ năng làm đồ án.

Trải qua thời gian một học kỳ thực hiện đề tài. Với sự hướng dẫn tận tình cùng những đóng góp quý báu của Cô và các bạn giúp nhóm hoàn thành tốt hơn báo cáo môn học của mình. Bên cạnh việc vận dụng những kiến thức được học trên lớp đồng thời kết hợp với việc học hỏi và tìm hiểu những kiến thức mới. Từ đó, nhóm đã vận dụng tối đa những gì đã tiếp thu được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, không tránh khỏi những sai sót. Dó đó, rất mong nhận được những sự góp ý từ Cô nhằm giúp nhóm hoàn thiện những kiến thức đã học tập và cũng là hành trang để nhóm thực hiện tiếp các đề tài khác trong tương lai.

Xin chân thành cảm ơn quý Thầy Cô và các bạn!

*Nhóm sinh viên thực hiện*

# Mục Lục

CHƯƠNG 1: MÔ TẢ DỮ LIỆU.....	6
1. Nguồn Dữ Liệu.....	6
2. Mô tả các cột của nguồn dữ liệu .....	6
CHƯƠNG 2: LUẬC ĐỒ KHO DỮ LIỆU .....	8
1. Sơ đồ cơ sở dữ liệu .....	8
2. Mô tả các bảng Dimension.....	9
3. Mô tả bảng Fact .....	10
4. Danh sách các câu truy vấn .....	12
CHƯƠNG 3: QUÁ TRÌNH SSIS.....	13
1. Khởi tạo Database.....	13
1.1. Khởi tạo Database .....	13
1.2. Tạo kết nối đến Database gốc.....	14
2. Quá trình SSIS .....	15
2.1. Làm sạch dữ liệu .....	16
2.2. Tạo các bảng Dimension.....	22
2.3. Tạo bảng Fact .....	31
2.4. Tạo các khóa ngoại .....	38
2.5. Chạy project .....	40
CHƯƠNG 4: XỬ LÝ PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN TỪ KHO DỮ LIỆU(SSAS) .....	41
1. Thiết lập project SSAS.....	41
1.1. Khởi tạo Data source .....	41
1.2. Tạo Data source view .....	44
1.3. Tạo Cube .....	46
1.4. Chính sửa server phục vụ cho việc deloy.....	50
1.5. Thiết kế các bản Dimension .....	51
2. Thực hiện truy vấn .....	53
2.1. Xếp loại số lượng các vụ tai nạn của từng loại nguy hiểm trên đường với loại đường One way street (Đường 1 chiều).....	53
2.2. Cho biết tổng số nạn nhân tử vong trong các điều kiện ánh sáng nhưng trừ Daylight: Street light present. ....	54

2.3. Danh sách loại đường số nạn nhân hơn 20000 hoặc số phương tiện gây tai nạn nhỏ hơn 20000 .....	56
2.4. Thống kê số lượng phương tiện gây tai nạn theo tháng trong năm 2007 theo từng điều kiện ánh sáng.....	58
2.5. Đếm tổng số tai nạn theo mỗi loại kiểm soát ở những nơi giao nhau trong năm 2009 rồi sắp xếp giảm dần.....	58
2.6. Theo từng năm, chọn ra top 3 loại thời tiết gây tai nạn ít nhất.....	60
2.7. Thống kê số lượng vụ tai nạn có điều kiện ánh sáng với tên điều kiện ánh sáng có chuỗi "present" trong tên.....	62
2.8. Các loại đường tại nơi xảy ra tai nạn vừa nằm trong top 3 số vụ tai nạn và top 3 số phương tiện gây tai nạn.....	64
2.9. Xác định 10 cao tốc có số nạn nhân nhiều nhất.....	65
2.10. Tìm số lượng vụ tai nạn ít nhất theo năm công bố của loại đường Single carriageway ....	66
2.11. Xác định top 3 điều kiện đặc biệt có nhiều tai nạn nhất ngoại trừ loại điều kiện đặc biệt Unknown.....	68
2.12. Cao tốc có số tai nạn ít nhất có loại đường là Dual carriageway theo mỗi năm.....	69
2.13. Thống kê số lượng nạn nhân trong các vụ tai nạn ở điều kiện ánh sáng có tên có 4 kí tự đầu là "Dark". .....	71
2.14. Thống kê số lượng các vụ tai nạn trong các điều kiện mặt đường sao cho tổng ít nhất là 10000 với ít loại mặt đường nhất có thể. ....	73
2.15. Với các vụ tai nạn xảy ra vào các điều kiện thời tiết khi không cảnh sát trực, xuất ra "Thương vong lớn" nếu số người thương vong hơn 2000 ngược lại xuất ra "Thương vong nhỏ". ...	76
<b>CHƯƠNG 5: QUÁ TRÌNH SSRS .....</b>	<b>78</b>
1. Tạo báo cáo với SSRS Project.....	78
1.1. Tạo mới một project.....	78
1.2. Kết nối.....	78
1.3. Thiết kế .....	83
1.4. Thiết kế Report dạng biểu đồ .....	86
1.5. Tạo các Report .....	88
2. Tạo báo cáo với Power BI.....	89
2.1. Kết nối nguồn dữ liệu .....	89
2.2. Tạo mới một table .....	92
2.3. Tạo các report .....	95
<b>CHƯƠNG 6: QUÁ TRÌNH DATA MINING .....</b>	<b>96</b>
1. Quá trình khai phá dữ liệu .....	96

2.	Tạo báo cáo với SSRS Project.....	104
2.1.	Thuật toán cây quyết định.....	104
2.2.	Thuật toán Naïve bayes .....	105
2.3.	Thuật toán gom cụm .....	108
2.4.	So sánh các mô hình khai phá dữ liệu .....	109

# CHƯƠNG 1: MÔ TẢ DỮ LIỆU

## 1. Nguồn Dữ Liệu

Tên nguồn dữ liệu: Dữ liệu tai nạn ở vương quốc Anh từ 2005 đến 2012

Giới thiệu: Nhằm mục đích xem xét và đánh giá các điều kiện, thời gian ảnh hưởng đến các vụ tai nạn như thế nào.

Số cột: 33 cột

Số dòng: 1042129 dòng

Dữ liệu thu thập trong: 2005 tới 2012

Kho dữ liệu đề xuất theo hướng chủ đề: Tai nạn đường bộ

Nguồn dữ liệu: [https://www.kaggle.com/daveianhickey/2000-16-traffic-flow-england-scotland-wales?fbclid=IwAR0aelR8qUhP74jVQ-gZGYEKzsy4uR\\_CsfS\\_Mt-YH8\\_Ob3nNm4GP\\_t62Fgo](https://www.kaggle.com/daveianhickey/2000-16-traffic-flow-england-scotland-wales?fbclid=IwAR0aelR8qUhP74jVQ-gZGYEKzsy4uR_CsfS_Mt-YH8_Ob3nNm4GP_t62Fgo)

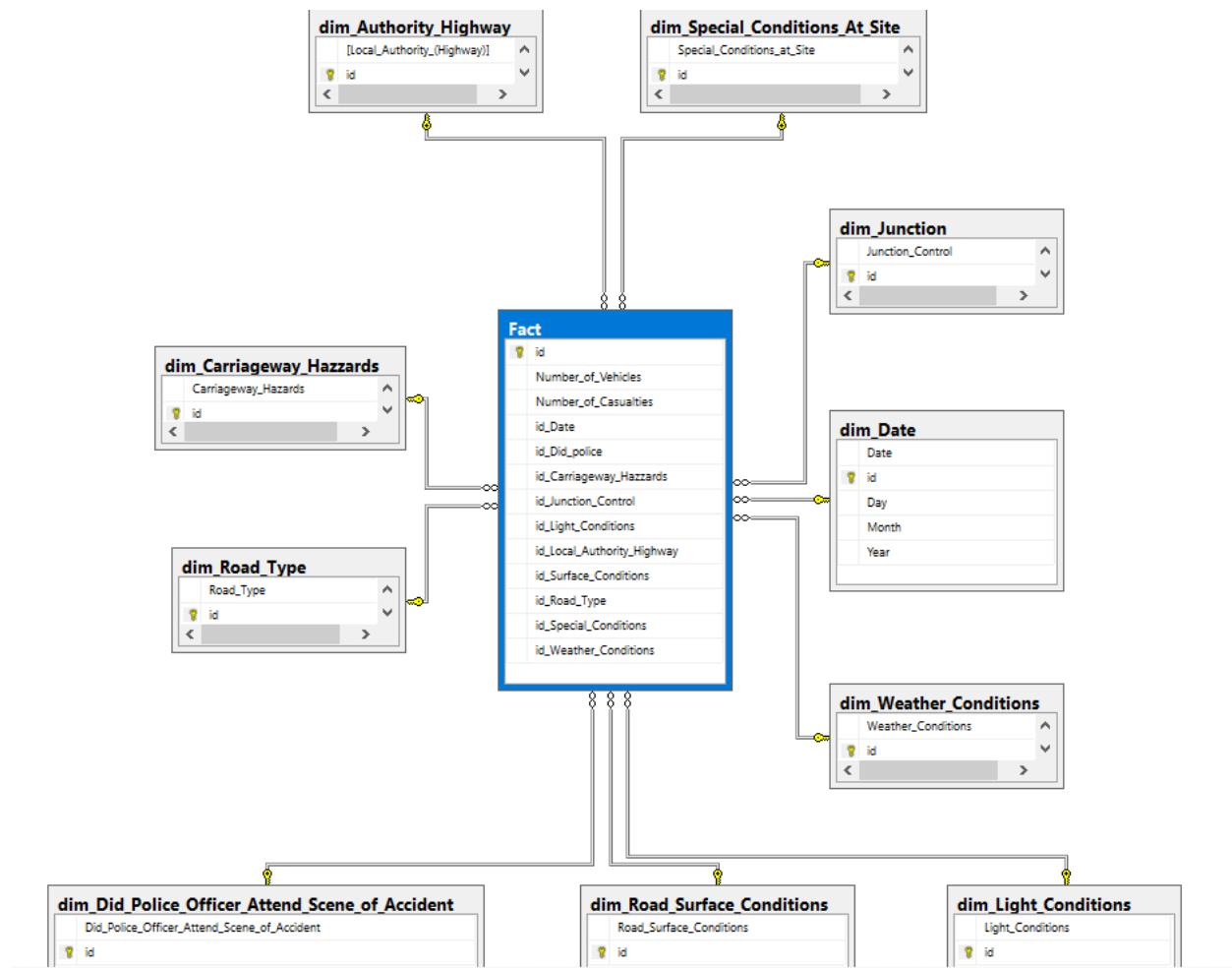
## 2. Mô tả các cột của nguồn dữ liệu

- Police\_Force: Id lực lượng cảnh sát quản lý.
- Accident\_Severity: Id mức độ nghiêm trọng.
- Number\_of\_Vehicles: Số phương tiện gây tai nạn.
- Number\_of\_Casualties: Số thương vong.
- Date: Ngày xảy ra tai nạn.
- Day\_of\_Week: Id thứ trong tuần.
- Local\_Authority\_District: Id chính quyền quận.
- Local\_Authority\_Highway: Mã cao tốc địa phương.
- \_1st\_Road\_Class: Id loại làn đường.
- \_1st\_Road\_Number: Số đường nội đô.
- Road\_Type: Tên loại đường.
- Speed\_limit: Tốc độ tối đa.
- Junction\_control: Thông tin kiểm soát của những nơi giao nhau.
- Pedestrian\_Crossing-Physical\_Facilities: Cơ sở vật chất cho người đi bộ.
- Light\_Conditions: Điều kiện ánh sáng.
- Weather\_Conditions: Điều kiện thời tiết.

- Road\_Surface\_Conditions: Điều kiện bề mặt đường.
- Special\_Conditions\_at\_Site: Điều kiện đặc biệt tại nơi xảy ra tai nạn.
- Carriageway\_Hazards: Tên nguy hiểm trên đường.
- Urban\_or\_Rural\_Area: Id khu vực thành thị hay nông thôn.
- Did\_Police\_Officer\_Attend\_Scene\_of\_Accident: Có cảnh sát trực trên tuyến đường không.

## CHƯƠNG 2: LUẬC ĐỒ KHO DỮ LIỆU

### 1. Sơ đồ cơ sở dữ liệu



## 2. Mô tả các bảng Dimension

STT	Tên bảng	Ý nghĩa	Thuộc tính	Kiểu dữ liệu	Giải thích
1	Dim_Special_Condition_at_Site	Điều kiện đặc biệt tại nơi xảy ra tai nạn	Id Name	Int Varchar(50)	Mã điều kiện đặc biệt
					Tên điều kiện
2	Dim_Junction_Control	Thông tin kiểm soát của những nơi giao nhau	Id Name	Int Varchar(50)	Mã thông tin
					Mô tả thông tin
3	Dim_Local_Authority_Highway	Cao tốc địa phương	Id Name	Int Varchar(50)	Mã cao tốc
					Tên cao tốc
4	Dim_Did_Police_Office_r_Attend_Scene_of_Accident	Có sảnh sát trực trên tuyến đường không	Id Name	Varchar(50) Varchar(50)	Mã trạng thái
					Mô tả trạng thái
5	Dim_Carriageway_Hazards	Nguy hiểm trên đường	Id Name	Int Varchar(50)	Mã nguy hiểm
					Tên nguy hiểm
6	Dim_Road_Type	Loại đường	Id Name	Int Varchar(50)	Mã loại đường
					Tên loại đường
7	Dim_Road_Surface_Conditions	Điều kiện mặt đường	Id Name	Int Varchar(50)	Mã điều kiện
					Mô tả điều kiện
8	Dim_Light		Id	Int	Mã điều kiện

	_Conditions	Điều kiện ánh sáng	Name	Varchar(50)	Mô tả điều kiện
9	Dim_Weath er _Conditions	Điều kiện thời tiết	Id	Int	Mã điều kiện
			Name	Varchar(50)	Mô tả điều kiện
10	Dim_Date	Thời gian xảy ra tai nạn	Id	Int	Mã thời gian
			Date	Date	Thời gian
			Year	Int	Năm
			Month	Int	Tháng
			Day	Int	Ngày

### 3. Mô tả bảng Fact

Tên bảng	Thuộc tính	Thuộc tính	Kiểu dữ liệu	Giải thích
Fact	Bình thường	Id	Int	Mã bảng
		Id_Date	Int	Mã thời gian
		Id_Did_Police_Scene	Int	Mã trạng thái xuất hiện của cảnh sát
		Id_Carriageway_Hazards	Int	Mã rủi ro tại nơi xảy ra tai nạn
		Id_Junction_Control	Int	Mã thông tin kiểm soát của giao lộ
		Id_Light_Conditions	Int	Mã điều kiện ánh sáng
		Id_Local_Authority_Highway	Int	Mã cao tốc địa phương
		Id_Surface_Condition	Int	Mã điều kiện mặt đường
		Id_Road_Type	Int	Mã loại đường

		Id_Special_Conditions	Int	Mã điều kiện đặc biệt
		Id_Weather_Conditions	Int	Mã điều kiện thời tiết
Độ đo		Number_of_Vehicles	Int	Số phương tiện gây tai nạn
		Number_of_Casualties	Int	Số nạn nhân

#### 4. Danh sách các câu truy vấn

**Câu 1:** Xếp loại số lượng các vụ tai nạn của từng loại nguy hiểm trên đường với loại đường One way street (Đường 1 chiều).

**Câu 2:** Cho biết tổng số nạn nhân tử vong trong các điều kiện ánh sáng nhưng trừ Daylight: Street light present. (Nonempty, Except)

**Câu 3:** Danh sách loại đường số nạn nhân hơn 1000 hoặc số phương tiện gây tai nạn nhỏ hơn 1000 (Union, Filter).

**Câu 4:** Thống kê số lượng phương tiện gây tai nạn theo tháng trong năm 2007 theo từng điều kiện ánh sáng.

**Câu 5:** Đếm tổng số tai nạn theo mỗi loại kiểm soát ở những nơi giao nhau trong năm 2009 rồi sắp xếp giảm dần.

**Câu 6:** Theo từng năm, chọn ra top 3 loại thời tiết gây tai nạn ít nhất.

**Câu 7:** Thống kê số lượng vụ tai nạn có điều kiện ánh sáng với tên điều kiện ánh sáng có chuỗi "present" trong tên.

**Câu 8:** Loại đường tại nơi xảy ra tai nạn vừa nằm trong top 3 số vụ tai nạn và top 3 số phương tiện gây tai nạn.

**Câu 9:** Xác định 10 cao tốc có số nạn nhân nhiều nhất.

**Câu 10:** Tìm số lượng vụ tai nạn ít nhất theo năm công bố của loại đường Single carriageway.

**Câu 11:** Xác định top 3 điều kiện đặc biệt có nhiều tai nạn nhất ngoại trừ loại điều kiện đặc biệt Unknown.

**Câu 12:** Cao tốc có số tai nạn ít nhất có loại đường là Dual carriageway theo mỗi năm.

**Câu 13:** Thống kê số lượng nạn nhân trong các vụ tai nạn ở điều kiện ánh sáng có tên có 4 kí tự đầu là "Dark" theo từng năm

**Câu 14:** Thống kê số lượng các vụ tai nạn trong các điều kiện mặt đường sao cho tổng ít nhất là 10000 với ít loại mặt đường nhất có thể.

**Câu 15:** Với các vụ tai nạn xảy ra vào các điều kiện thời tiết khi không có cảnh sát trực, xuất ra "Thương vong lớn" nếu số người thương vong hơn 2000 ngược lại xuất ra "Thương vong nhỏ".

## CHƯƠNG 3: QUÁ TRÌNH SSIS

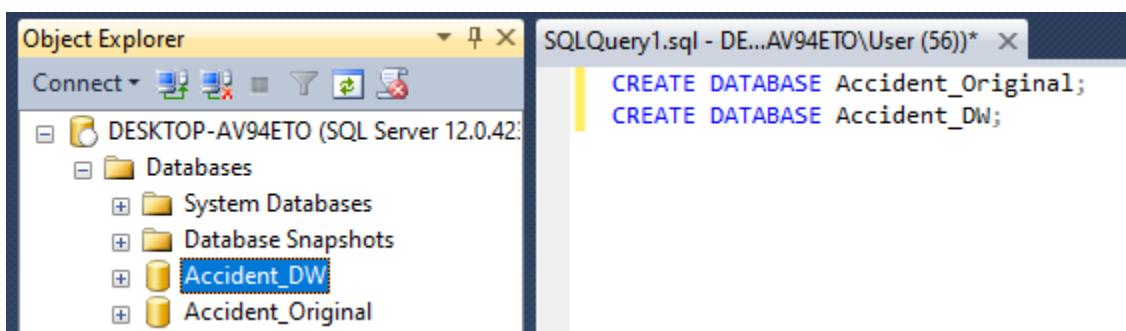
### 1. Khởi tạo Database

#### 1.1. Khởi tạo Database

Tạo các database chứa các dữ liệu

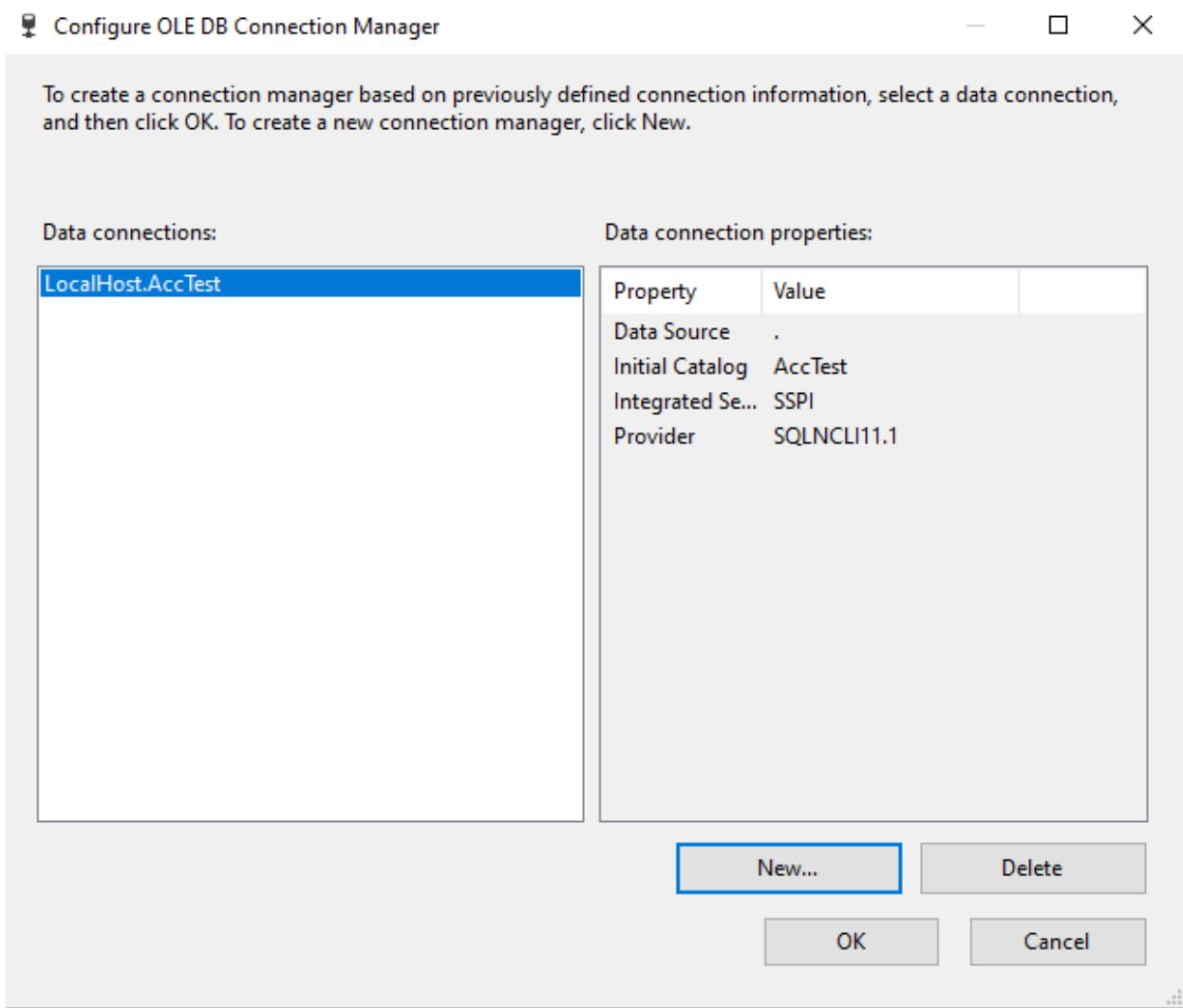
Cần tạo 2 csdl bao gồm:

1. Accident\_Original: Chứa dữ liệu đính từ excel vào.
2. Accident\_DW: Chứa dữ liệu của bảng Fact và các bảng DIM.

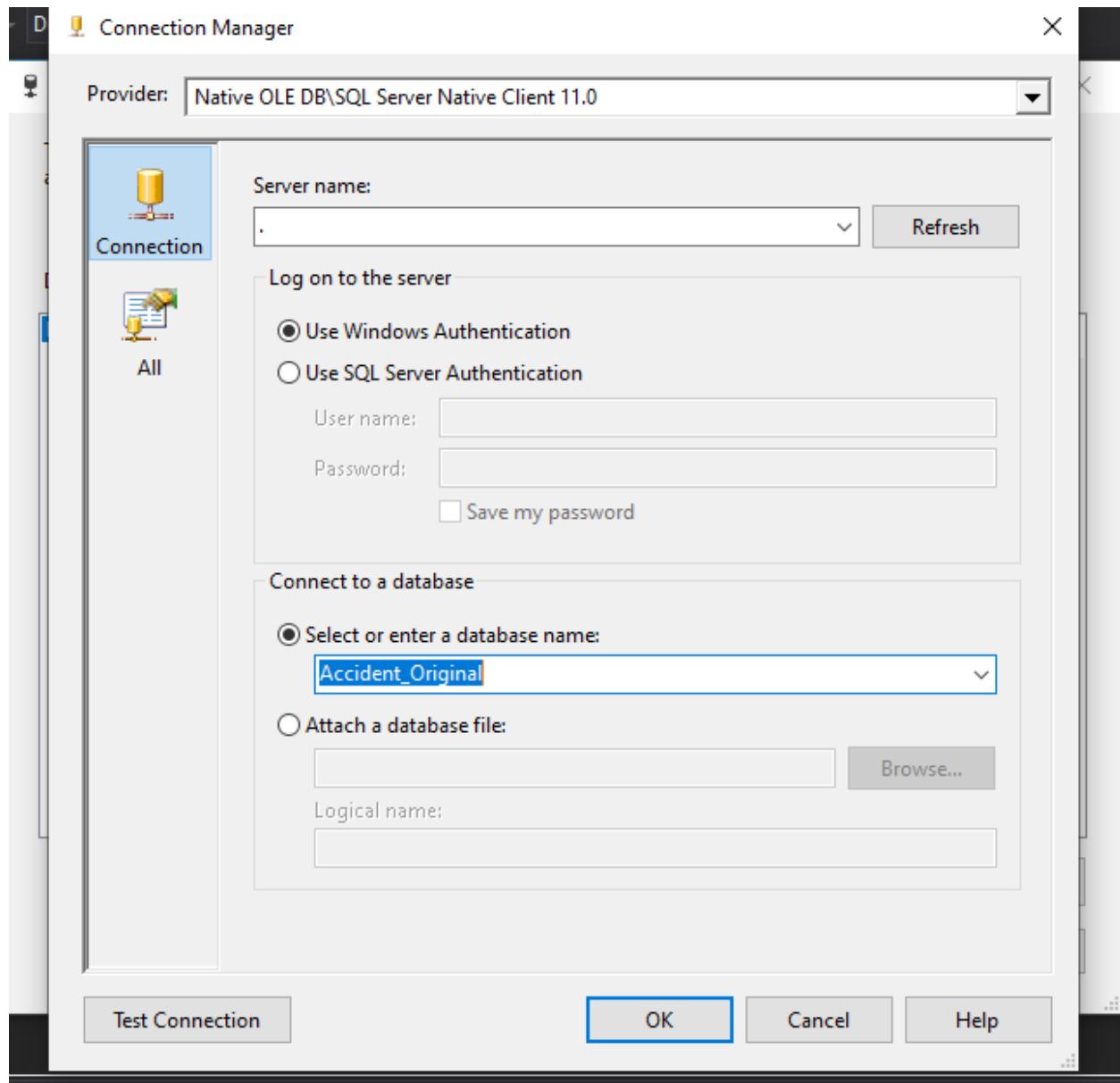


### 1.2. Tạo kết nối đến Database gốc

Trong cửa sổ Solution Explorer, click phải vào Connection Managers -> New Connection Manager-> Chọn new



Nhập server name( Chính là server name lúc connect trong sql server), chọn database đã tạo ở bước trên -> Nhấn Test Connection để kiểm tra kết nối -> Nhấn OK.

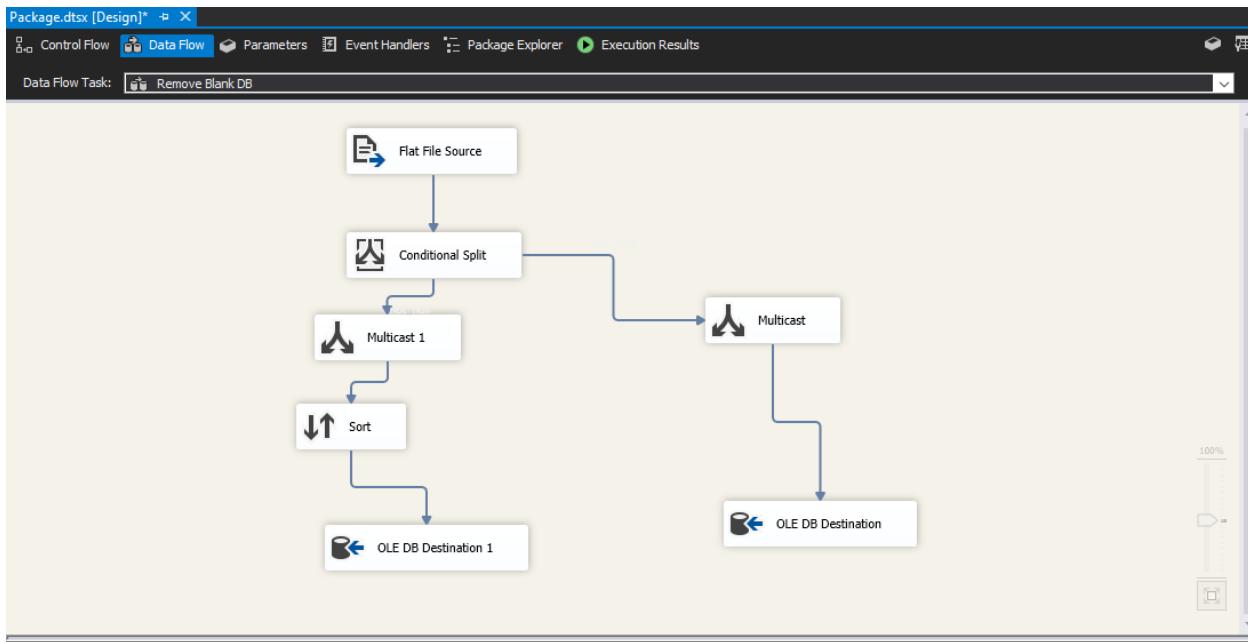


## 2. Quá trình SSIS

## 2.1. Làm sạch dữ liệu

Mục đích: Loại bỏ các dòng có dữ liệu trống

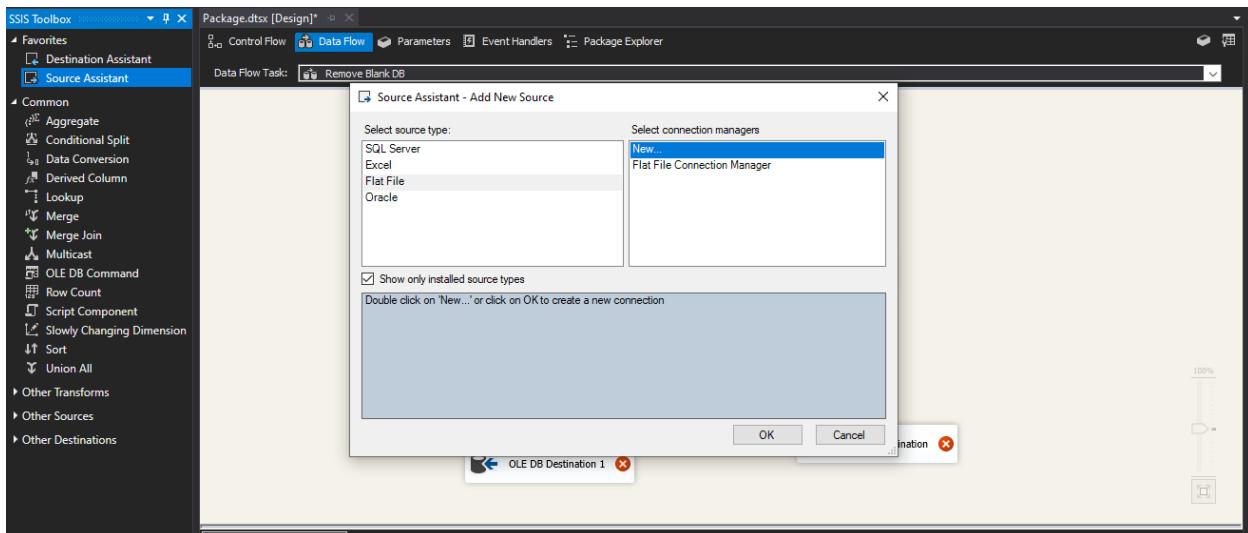
Cấu trúc:



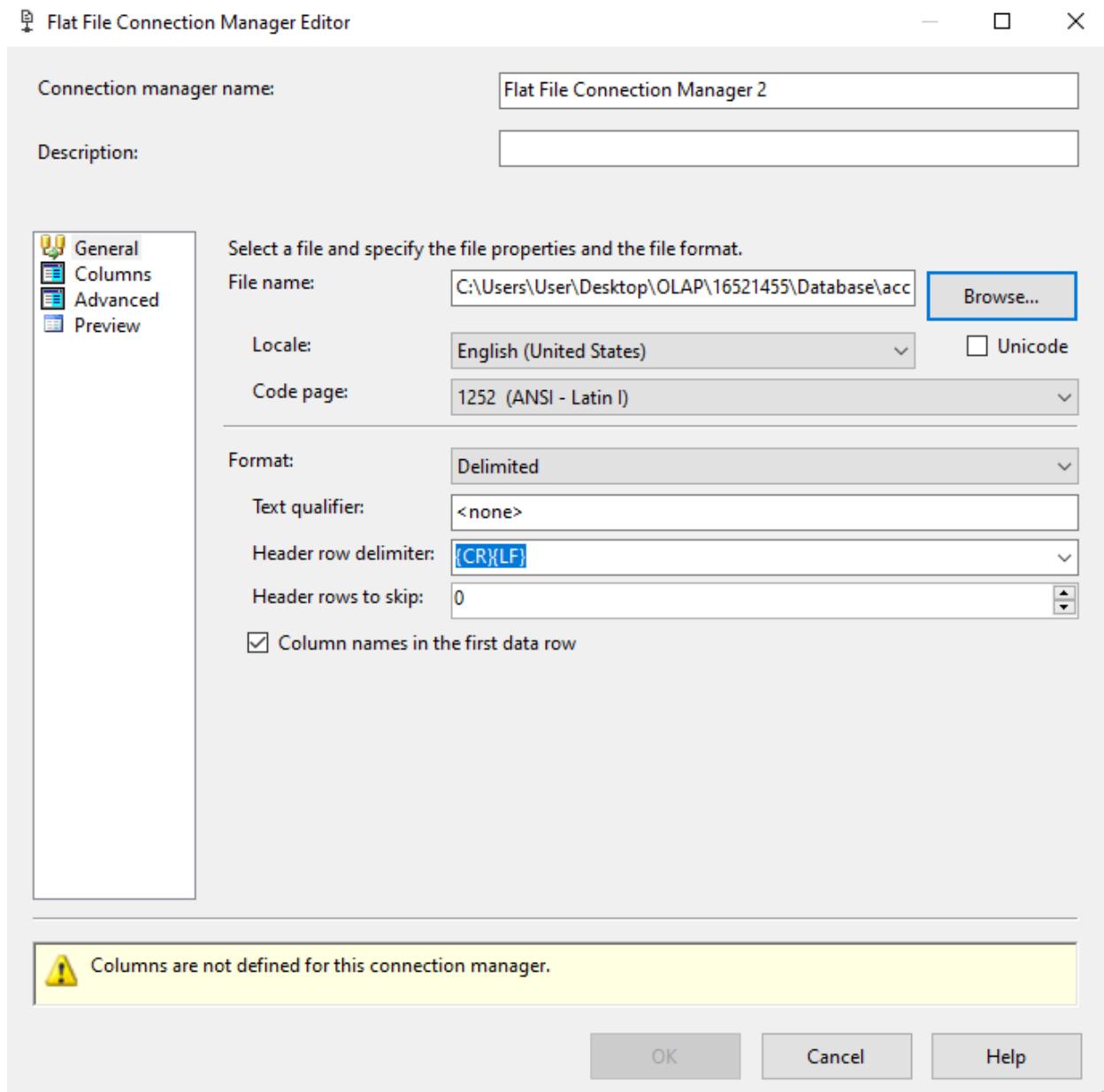
Chi tiết quá trình làm sạch dữ liệu:

- Flat file source

Dùng để đỗ dữ liệu từ file Excel vào. Ở tab SSIS ToolBox, ta nhấp chọn Source Assistant, sẽ có một hộp thoại hiện lên. Ta nhấn vào tab Flat File và chọn New.



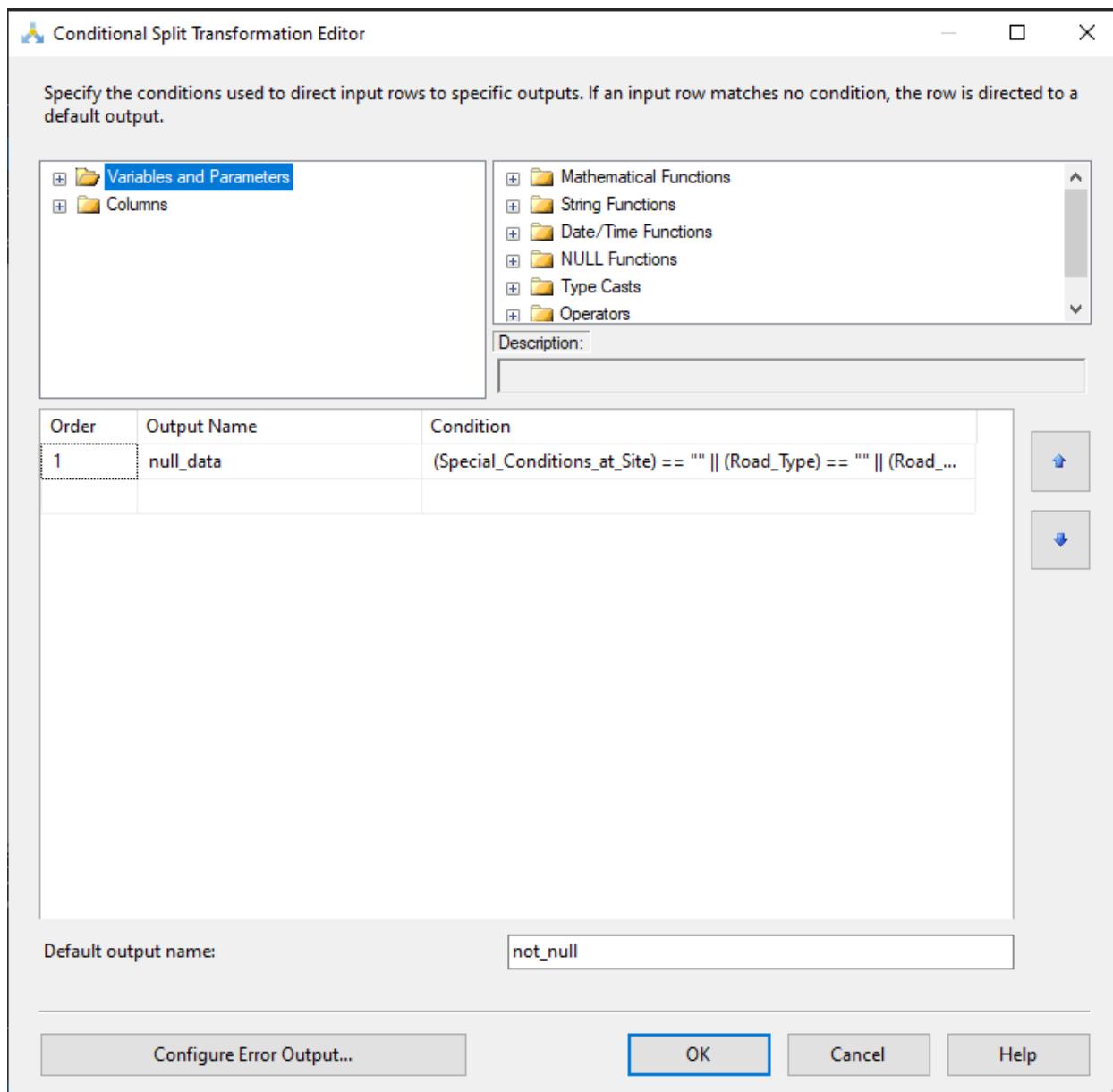
Tại hộp thoại vừa hiện lên, ta nhấn vào Browser chọn đường dẫn tới File excel. Sau đó nhấn Ok.



- Conditional Split

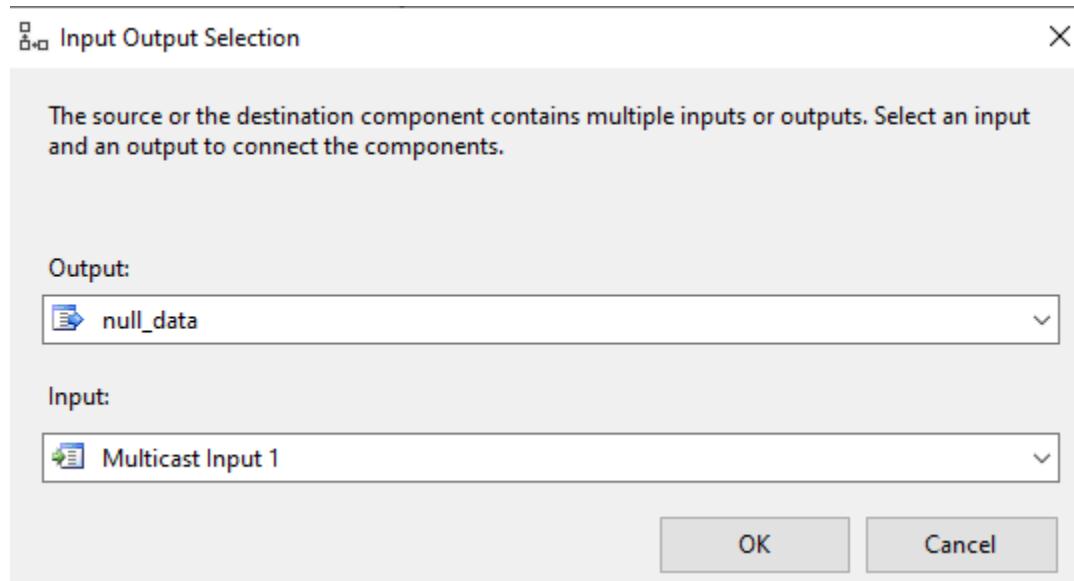
Dùng để kiểm tra dữ liệu có bị trống hay không. Tương tự Assistant Source ta nhập chọn condition split. Khi ấy một hộp thoại sẽ hiện lên. Ta cần thiết lập các mục

- Output Name: Tên output của dữ liệu
- Condition: Kiểm tra thuộc có giá trị trống hay không.
- Default output name: Tên mặc định của output (Dữ liệu không thỏa các điều kiện đã khai báo trong condition)



- Multicast

Ta nhấp chọn Multicast hai lần để tạo ra 2 multicast. Khi kéo liên kết từ Condition Split đến Multicast. Một hộp thoại sẽ hiện lên, nó yêu cầu người dùng nhập vào output. Đối với Multicast hướng tới database null thì ta chọn output là null data, còn multicast hướng tới database clean thì ta chọn output là not null.



- Sort

Mục đích: Sắp xếp data theo một thuộc tính nào đó trước khi đổ dữ liệu vào Database clean.

Sort Transformation Editor

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

Available Input Columns		
	Name	Pass Through
<input type="checkbox"/>	Name	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Did_Police_Officer_Attend_...	<input checked="" type="checkbox"/>
<input type="checkbox"/>	LSOA_of_Accident_Location	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Year	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Accident_Severity	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Number_of_Vehicles	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Police_Force	<input type="checkbox"/>

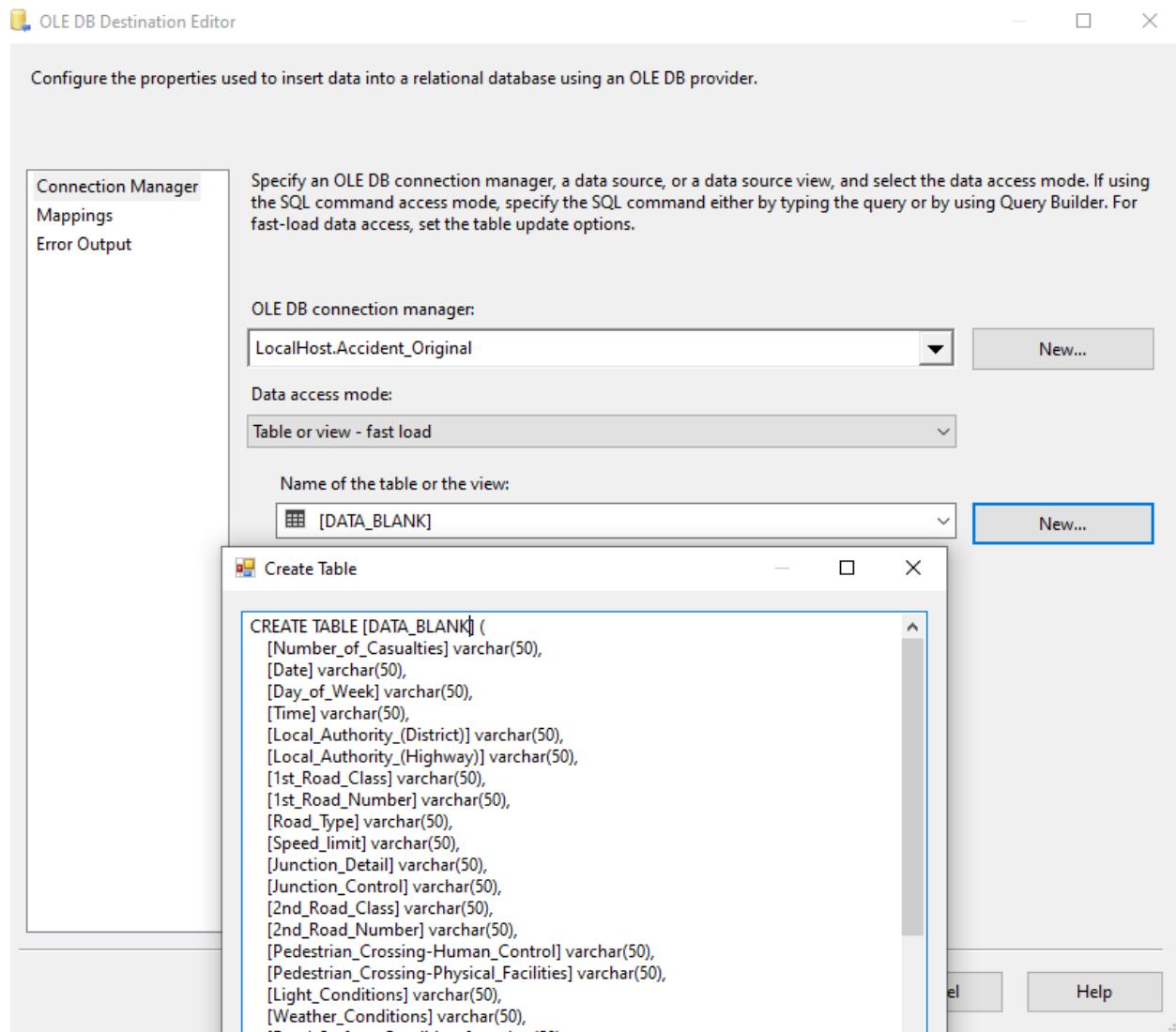
Input Column	Output Alias	Sort Type	Sort Order	Compa
Police_Force	Police_Force	ascending	1	

Remove rows with duplicate sort values

OK Cancel Help

### - OLE BD Destination

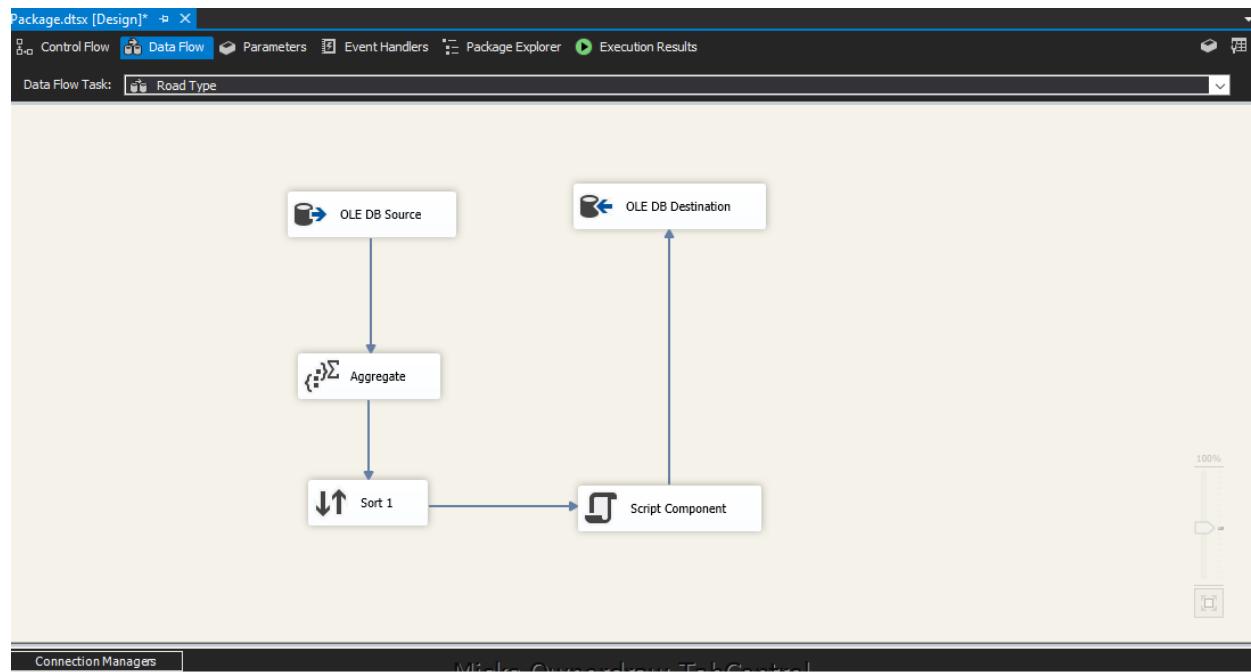
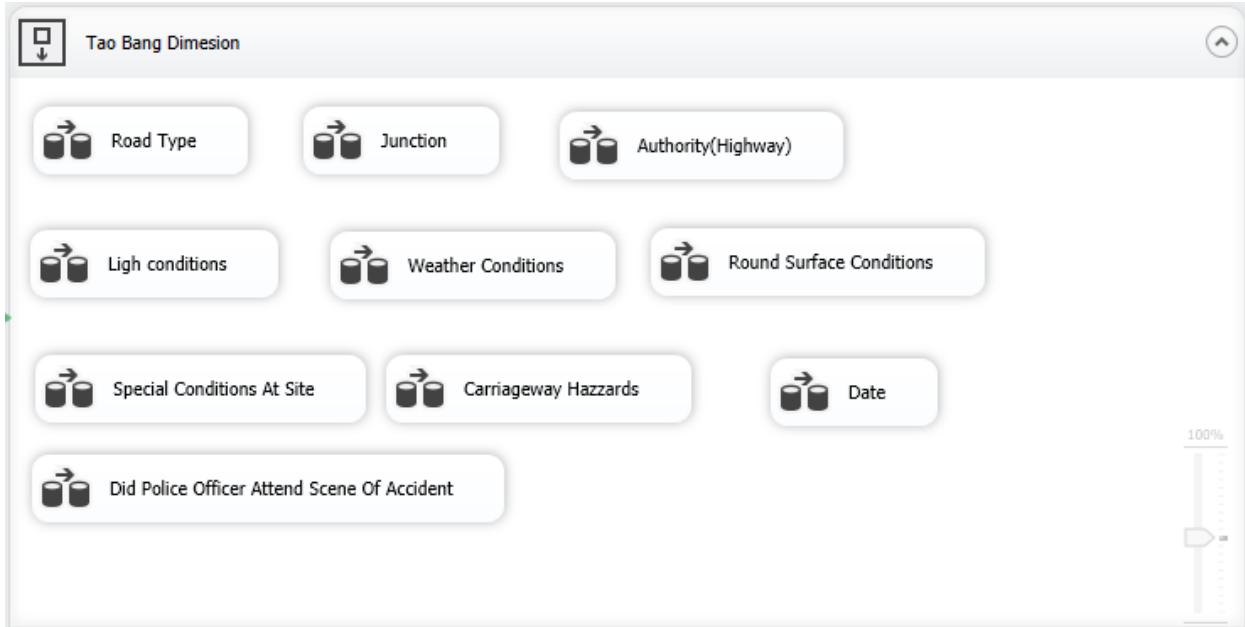
Ở tab SSIS ToolBox, ta nhập chọn Destination Assistant hai lần để tạo ra hai destination database. Một cái đỗ dữ liệu trống và một dùng để đỗ dữ liệu sạch. Sẽ có một hộp thoại hiện lên, ta cần nhấn New và đặt tên cho Database kêt bên dòng Create Table. Sau đó kiểm tra các trường dữ liệu của bảng rồi nhấn OK.



## 2.2. Tạo các bảng Dimension

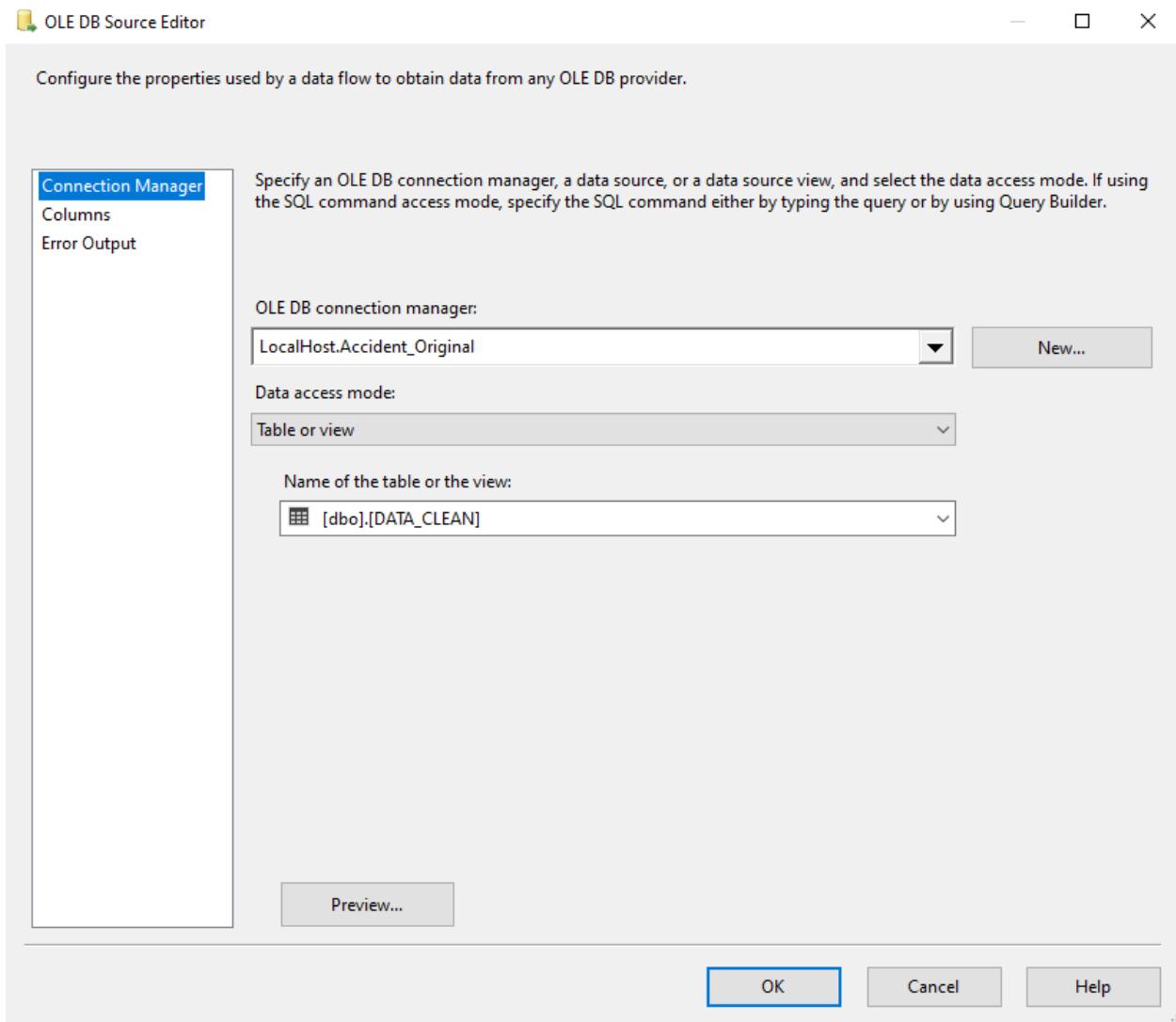
Mục đích: Dùng Sequence Container để chạy song song các Flow Task tạo bảng Dimension.

Cấu trúc các Flow Task thông thường:



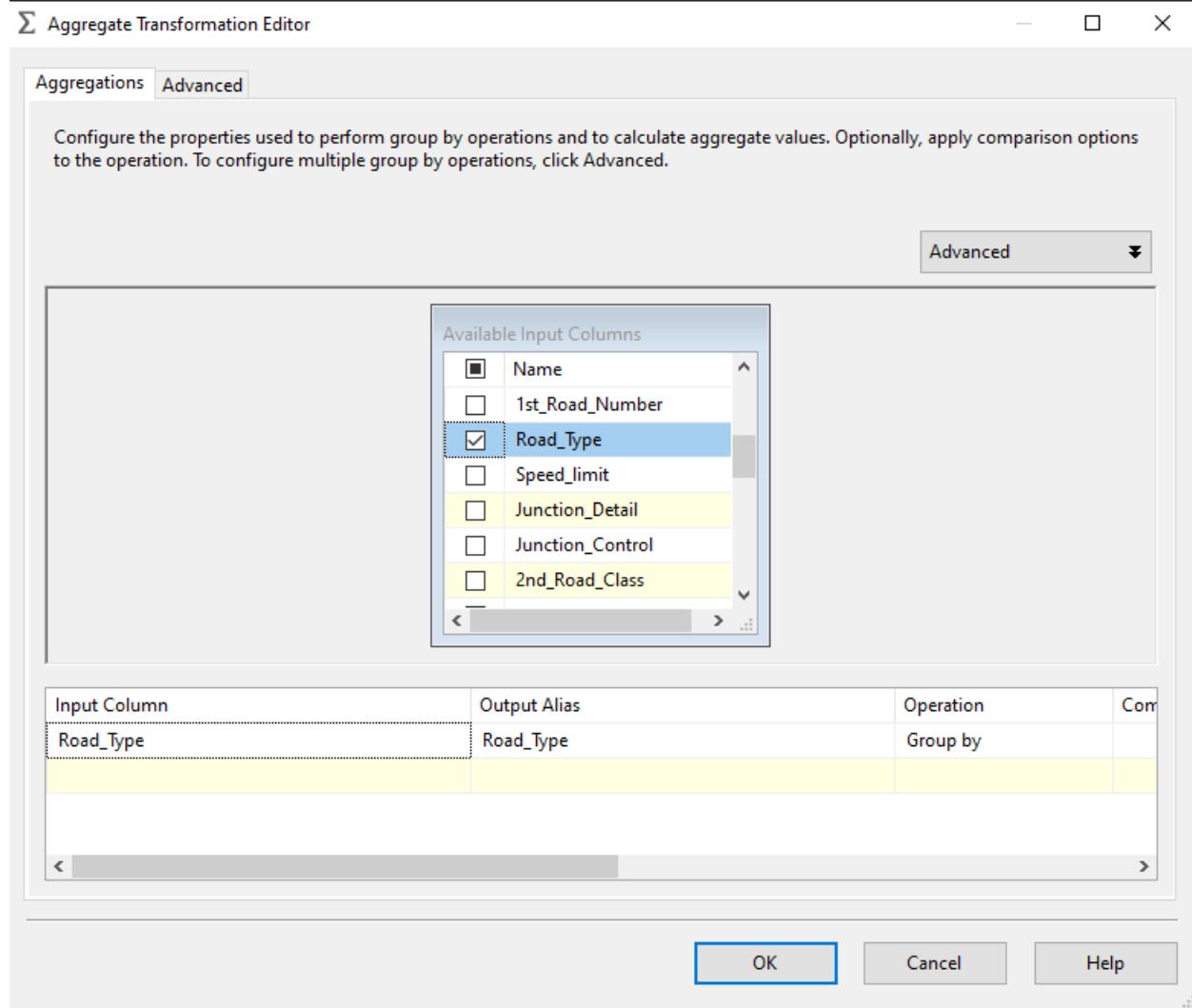
- OLE DB Source

Mục đích: Lấy dữ liệu đã làm sạch từ bảng Data\_clean



- Aggregate

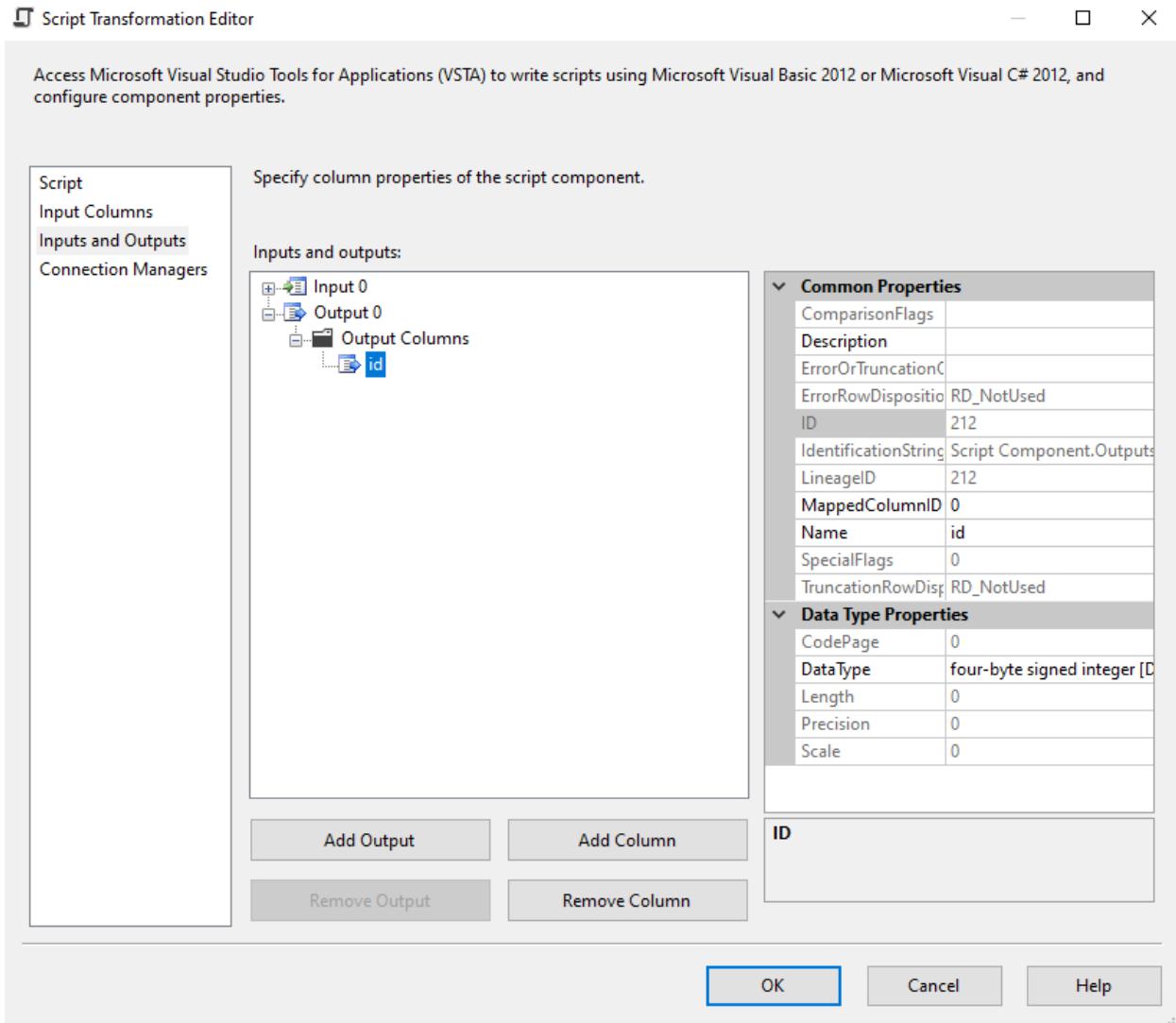
Mục đích: Group by theo thuộc tính được tick chọn.



- Script Component

Mục đích: Thêm trường id tự tăng cho bảng Dimension mới.

Tạo trường mới: Chọn (Inputs and Outputs) -> Chọn (Output) -> Chọn (Add Column) -> Đổi tên cột thành id -> Chọn (OK).



Chọn (Script) -> Chọn (Edit Script) -> Khai báo biến count = 1 -> Trong hàm Input0\_ProcessInputRow Thêm vào 2 dòng lệnh

- Row.id = count;
- count++;

The screenshot shows a code editor window with the following details:

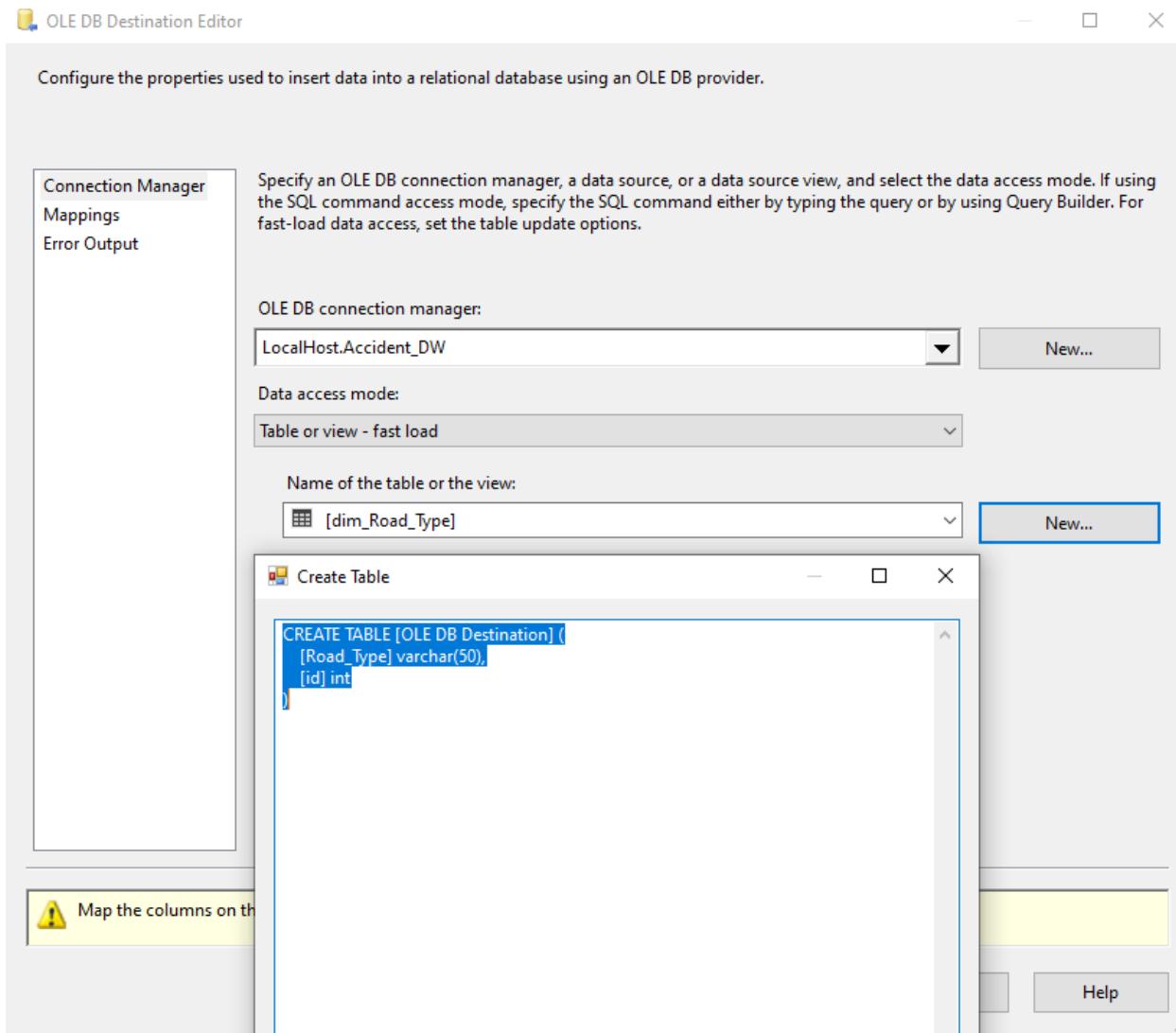
- Title Bar:** main.cs
- Tab Bar:** ScriptMain
- Code Area:**

```
/// You can delete this method if you don't need to do anything here.
/// </summary>
public override void PostExecute()
{
    base.PostExecute();
    /*
     * Add your code here
     */
}

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

- OLE DB Destination

Mục đích: Đỗ dữ liệu vào bảng mới tạo

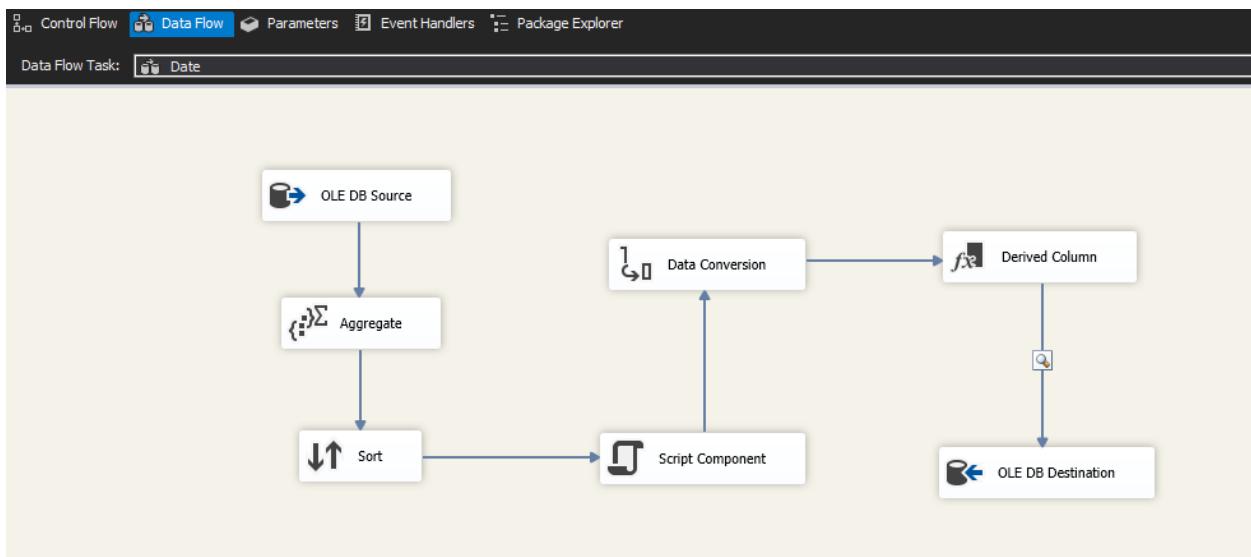


Tương tự lúc đổ dữ liệu vào Data\_clean.

Ngoài ra ta có thể nhán vào new để tạo bảng mới trong database ở SQL Server. Sau đó ta nhán vào mapping để kiểm tra dữ liệu xem đúng chưa.

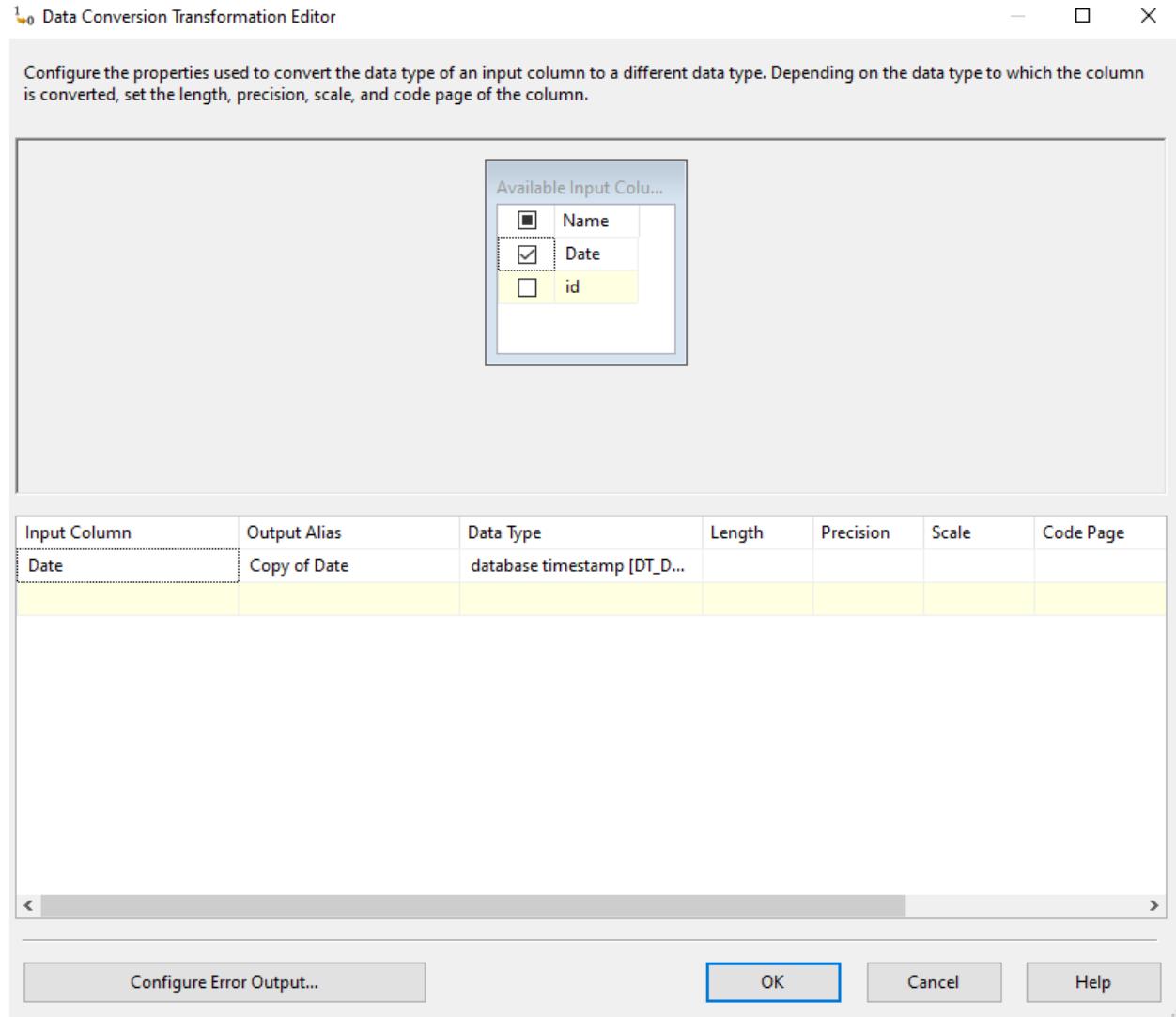
**Lưu ý:** Các bảng chiêu khác được thiết lập với cấu trúc tương tự như trên. Riêng bảng chiêu ngày được thiết lập với cấu trúc sau.

Cấu trúc xử lý của bảng chiều Date:



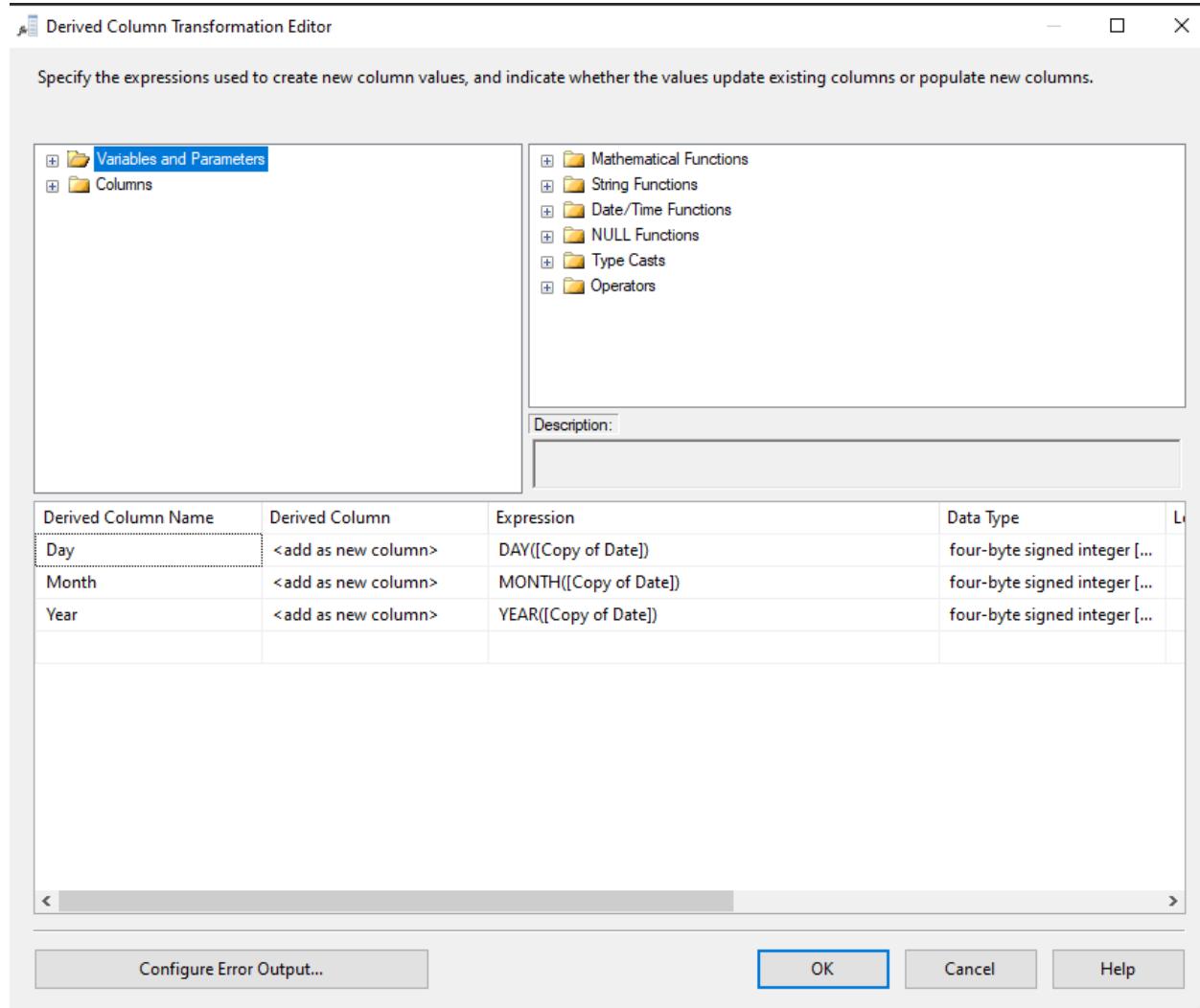
OLE DB Source, Aggregate, Sort, Script Component: Giống như trên

- Data Conversion: Chuyển thuộc tính ngày tháng sang kiểu dữ liệu database timestamp



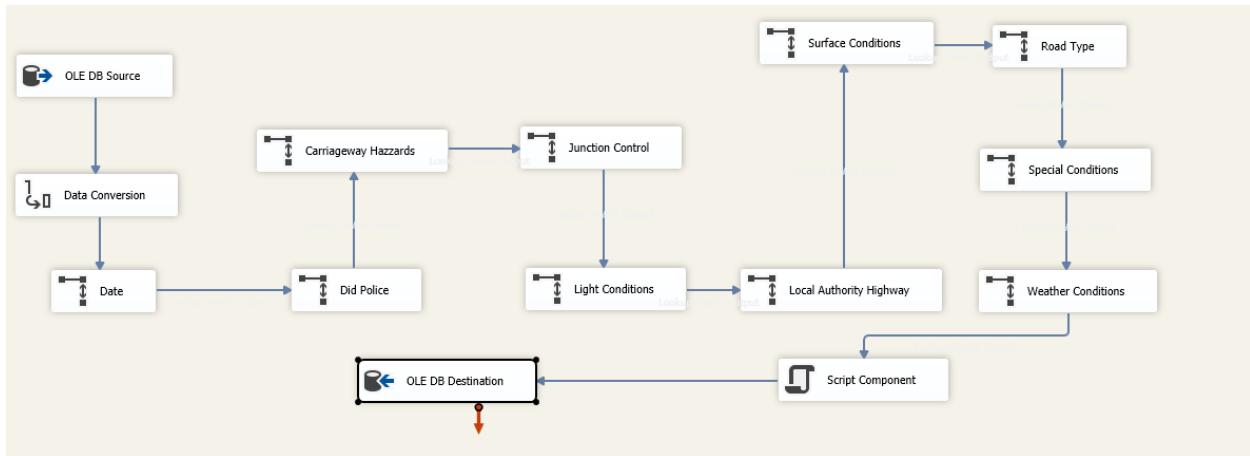
- Derived Column:

Mục đích: Tách dữ liệu thành từng thuộc tính ngày, tháng, năm.



### 2.3. Tạo bảng Fact

Cấu trúc:



OLE DB Source: Tương tự như các bảng Dimension.

- Data Conversion

Mục đích: Chuyển các thuộc tính độ đo sang kiểu dữ liệu số.

## 1 Data Conversion Transformation Editor

— □ ×

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Number_of_Casualties	Copy of Number_of_Cas...	single-byte signed integer [...]				
Date	Copy of Date	database timestamp [DT_D...				
Day_of_Week	Copy of Day_of_Week	string [DT_STR]	50			1252 (ANSI -
Local_Authority_(District)	Copy of Local_Authority...	string [DT_STR]	50			1252 (ANSI -
Local_Authority_(Highw...	Copy of Local_Authority...	string [DT_STR]	50			1252 (ANSI -
1st_Road_Class	Copy of 1st_Road_Class	four-byte signed integer [D...				
1st_Road_Number	Copy of 1st_Road_Num...	string [DT_STR]	50			1252 (ANSI -
Road_Type	Copy of Road_Type	string [DT_STR]	50			1252 (ANSI -
Speed_limit	Copy of Speed_limit	four-byte signed integer [D...				
Junction_Control	Copy of Junction_Control	string [DT_STR]	50			1252 (ANSI -
Police_Force	Copy of Police_Force	string [DT_STR]	50			1252 (ANSI -
Number_of_Vehicles	Copy of Number_of_Vehicles	string [DT_STR]	50			1252 (ANSI -
Accident_Severity	Copy of Accident_Severity	string [DT_STR]	50			1252 (ANSI -
1st_Road_Class	Copy of 1st_Road_Class	four-byte signed integer [D...				

**Available Input Columns**

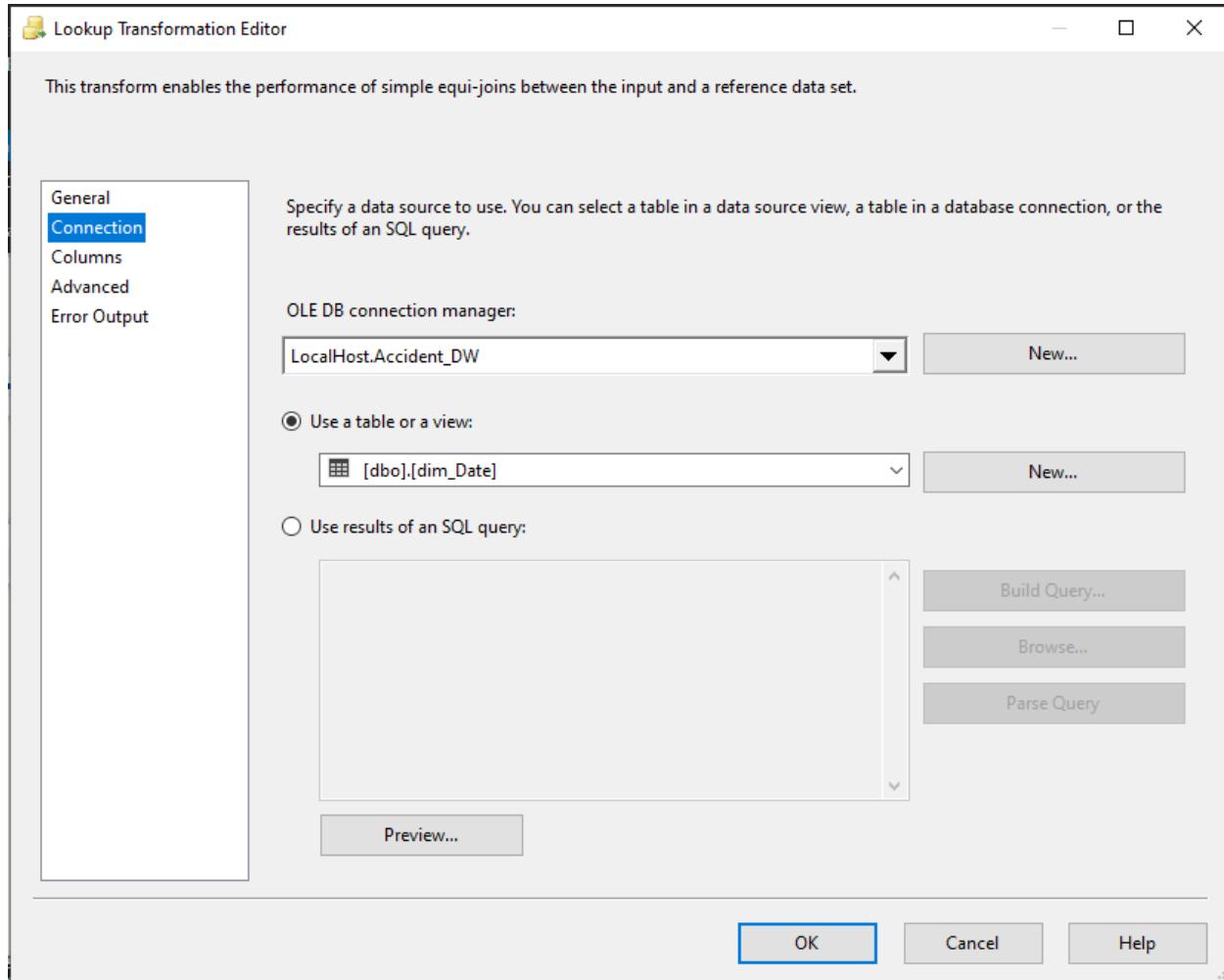
- Name
- Accident\_Severity
- Police\_Force
- Number\_of\_Vehicles
- Number\_of\_Casualties
- Day\_of\_Week
- Local\_Authority\_(District)
- 1st\_Road\_Class

Configure Error Output...      OK      Cancel      Help

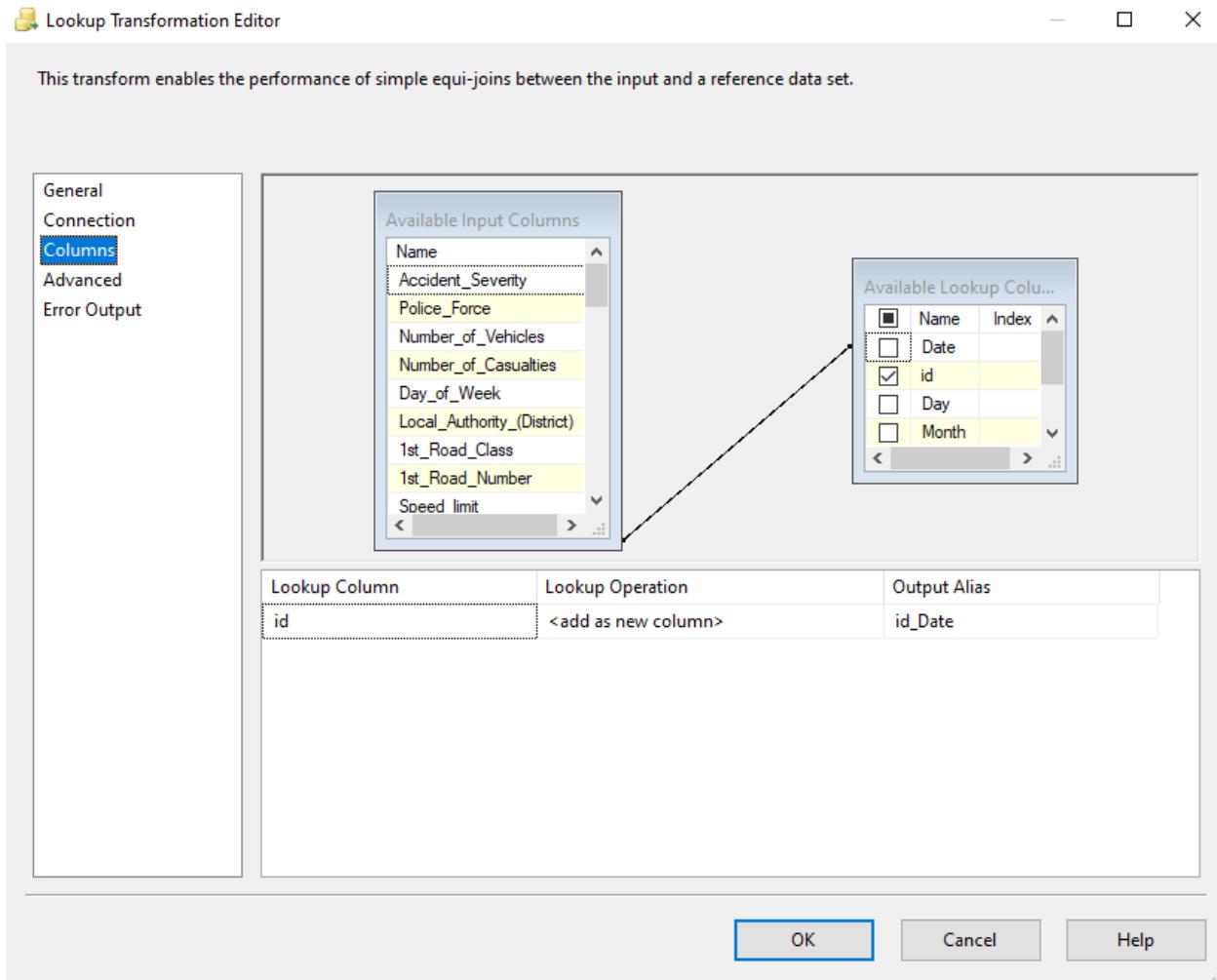
- Các Lookup tới các bảng Dimension

Mục đích: Lấy Id của các bảng Dimension bằng cách kết với data gốc tới các bảng Dimension.

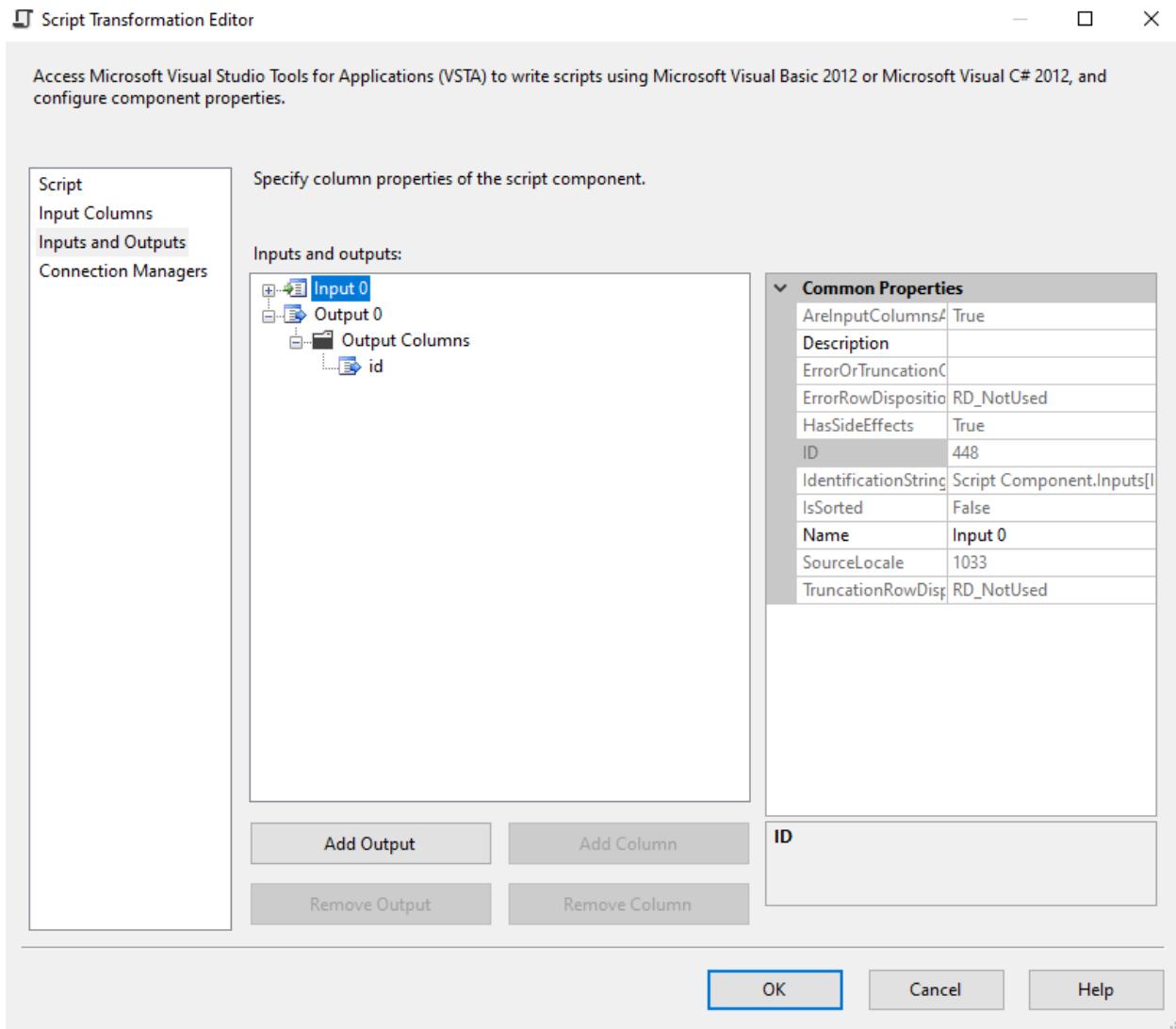
Chọn connection và chọn bảng cần kết



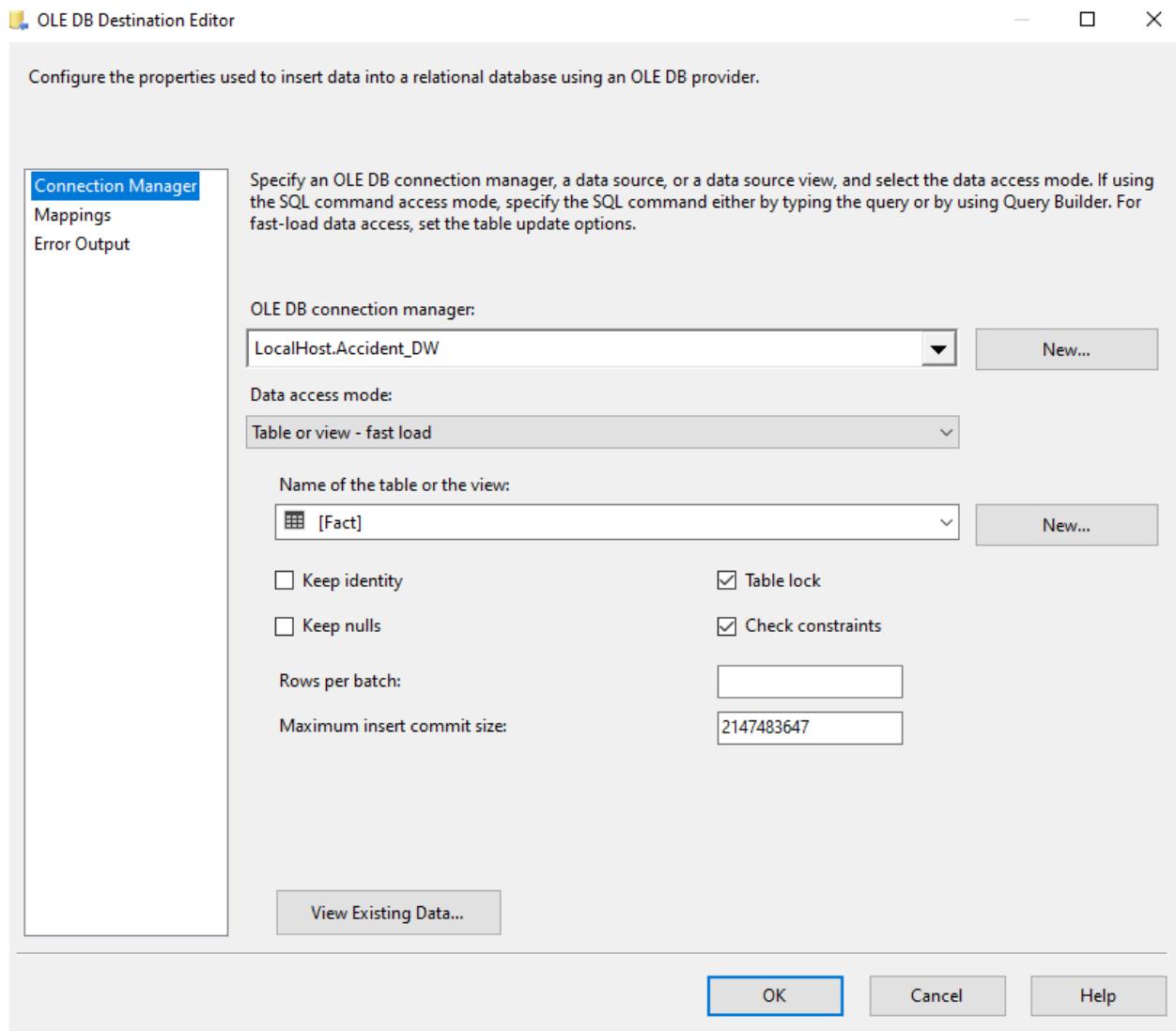
Kéo nối bảng từ trái sang phải sau đó đổi tên giá trị ở Output Alias cho phù hợp

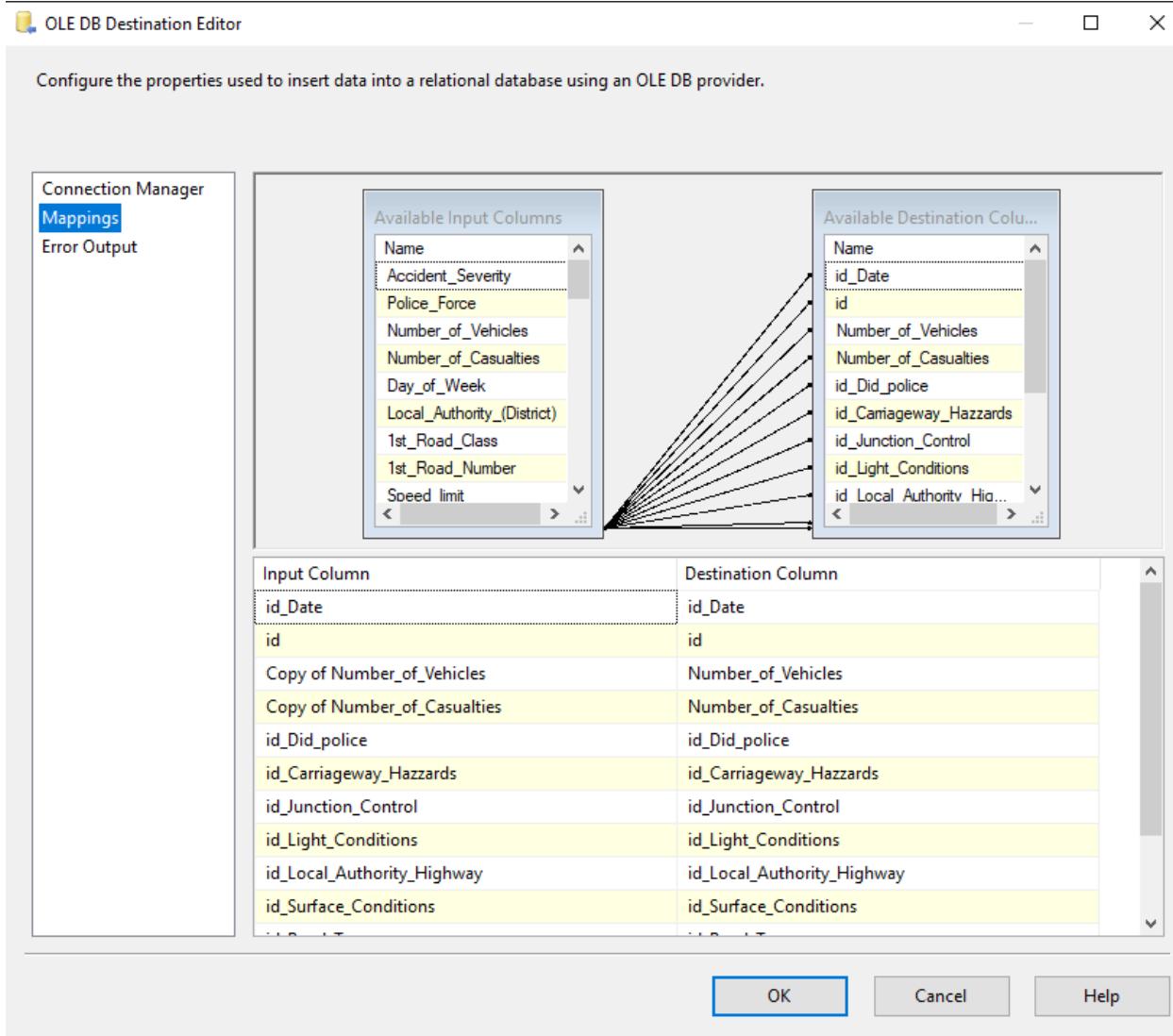


- Script Component: Tương tự các bảng Dimension



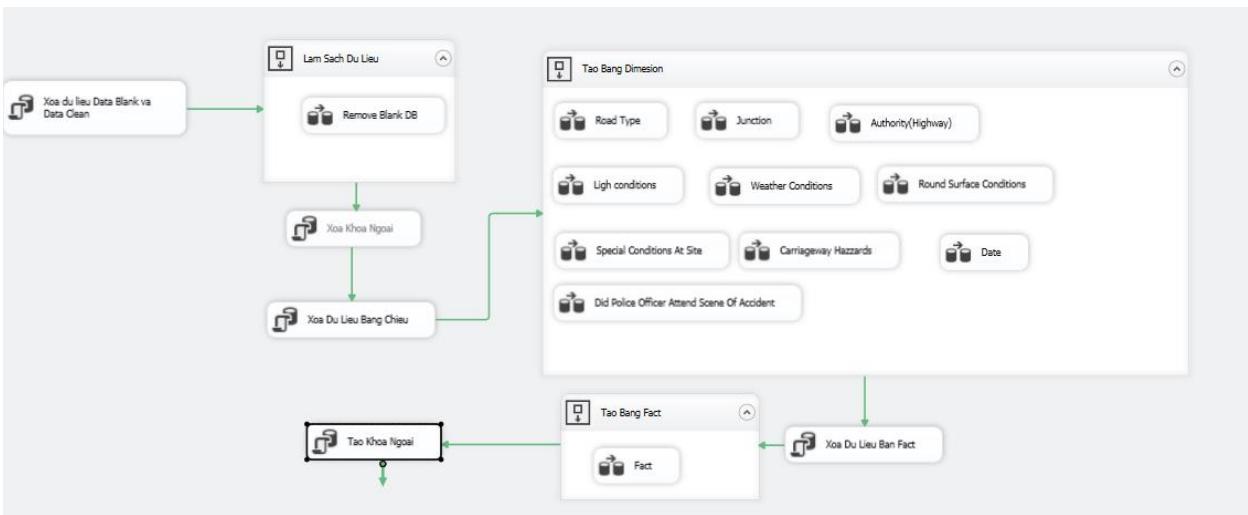
- OLE DB Destination: Giống các bảng Dimension bình thường nhưng lưu ý thao tác mapping dữ liệu (Ưu tiên chọn dữ liệu sau khi đã convert để mapping).

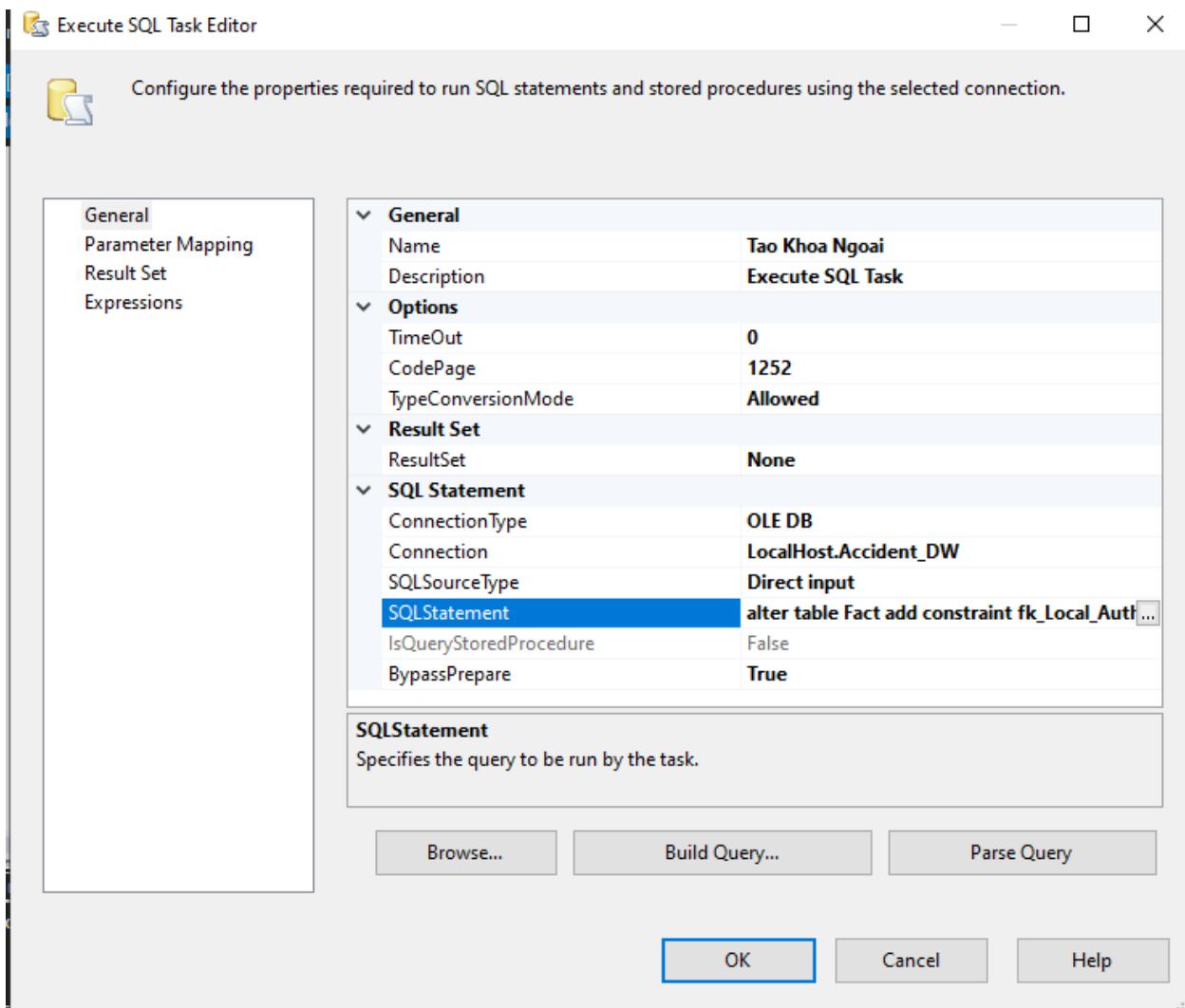




## 2.4. Tạo các khóa ngoại

Dùng Execute SQL Task trong SISS ToolBox





Chọn mục SQLStatement và ghi những dòng lệnh SQL bên dưới đây, để tạo khóa ngoại.

```

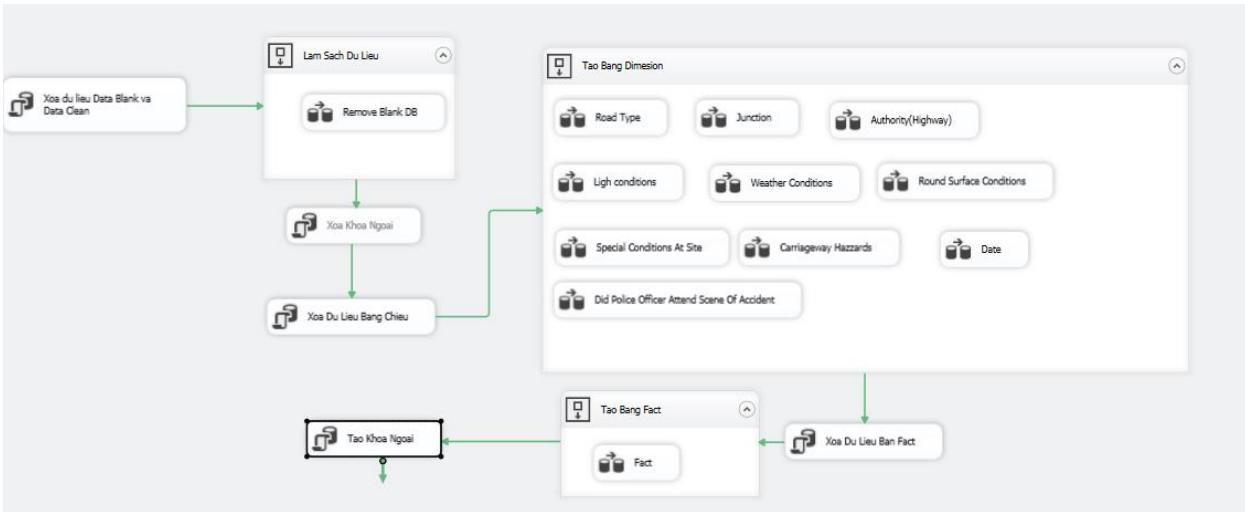
alter table Fact add constraint fk_Local_Authority_Highway foreign key (id_Local_Authority_Highway) references dim_Authority_Highway(id)
GO
alter table Fact add constraint fk_Date foreign key (id_Date) references dim_Date(id)
GO
alter table Fact add constraint fk_Did_Police_Sence foreign key (id_Did_Police) references dim_Did_Police_Officer_Attend_Scene_of_Accident(id)
GO
alter Table Fact add constraint fk_Camageway foreign key (id_Camageway_Hazzards) references dim_Camageway_Hazzards(id)
GO
alter Table Fact add constraint fk_Junction_Control foreign key (id_Junction_Control) references dim_Junction(id)
GO
alter Table Fact add constraint fk_Light_Conditions foreign key (id_Light_Conditions) references dim_Light_Conditions(id)
GO
alter Table Fact add constraint fk_Pedestrian_Crossing foreign key (id_Pedestrian_Crossing) references dim_Pedestrian_Crossing_Physical_Facilities(id)
GO
alter Table Fact add constraint fk_Surface_Conditions foreign key (id_Surface_Conditions) references dim_Road_Surface_Conditions(id)
GO
alter Table Fact add constraint fk_Road_Type foreign key (id_Road_Type) references dim_Road_Type(id)
GO
alter Table Fact add constraint fk_Special_Conditions foreign key (id_Special_Conditions) references dim_Special_Conditions_at_Site(id)
GO
alter Table Fact add constraint fk_Weather_Conditions foreign key (id_Weather_Conditions) references dim_Weather_Conditions(id)

```

## 2.5. Chạy project

Để có thể Execute toàn bộ Project cho lần sau, cần bổ sung các task sau để tránh xung đột ràng buộc:

- Xóa dữ liệu bảng DataNull và DataClean:  
Connection: Chọn kết nối tới database chứa dữ liệu gốc  
SQL Statement: Gõ lệnh sql xóa 2 bảng Data Null và Data Clean
- Xóa khóa ngoại:  
Connection: Chọn kết nối tới database chứa bảng Fact và Dimension  
SQL Statement: Gõ lệnh sql xóa các ràng buộc khóa ngoại
- Xóa dữ liệu các bảng Dimension:  
Connection: Chọn kết nối tới database chứa bảng Fact và Dimension  
SQL Statement: Gõ lệnh sql xóa dữ liệu các bảng Dimension.
- Xóa dữ liệu bảng Fact:  
Connection: Chọn kết nối tới database chứa bảng Fact và Dimension  
SQL Statement: Gõ lệnh sql xóa dữ liệu bảng Fact.

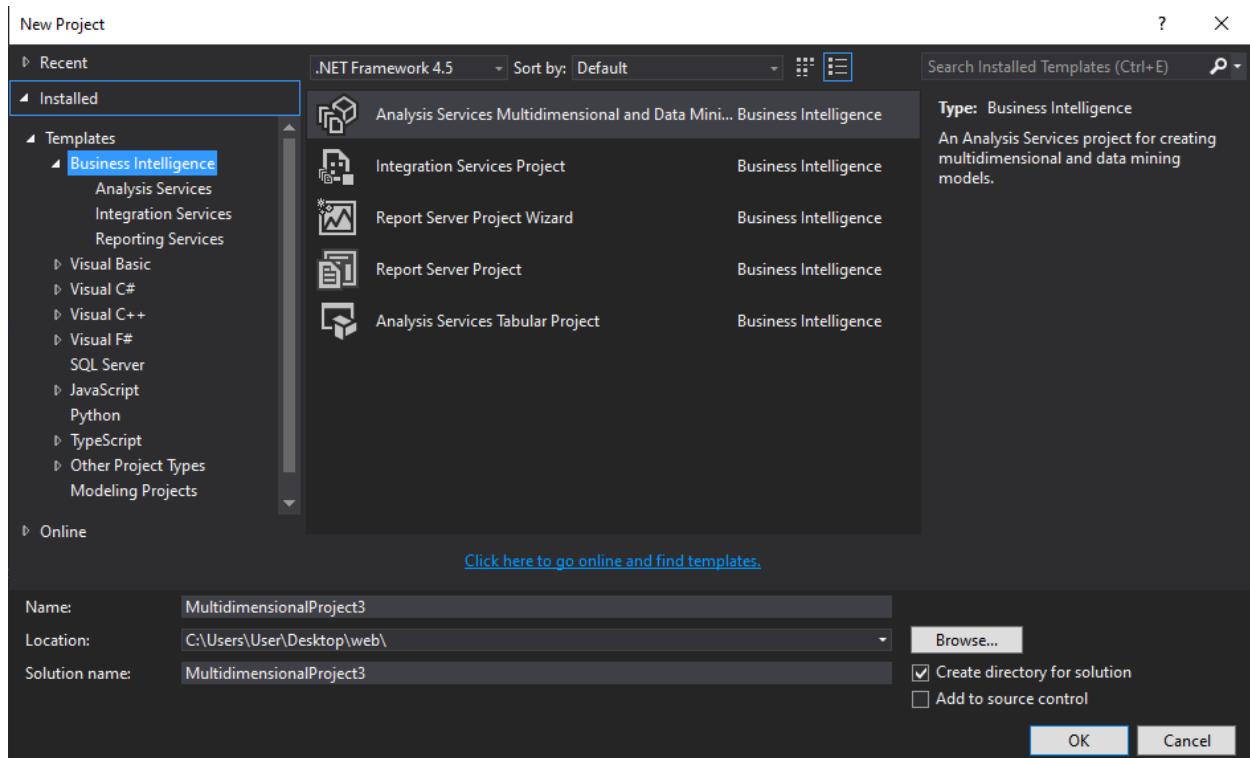


## CHƯƠNG 4: XỬ LÝ PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN TỪ KHO DỮ LIỆU(SSAS)

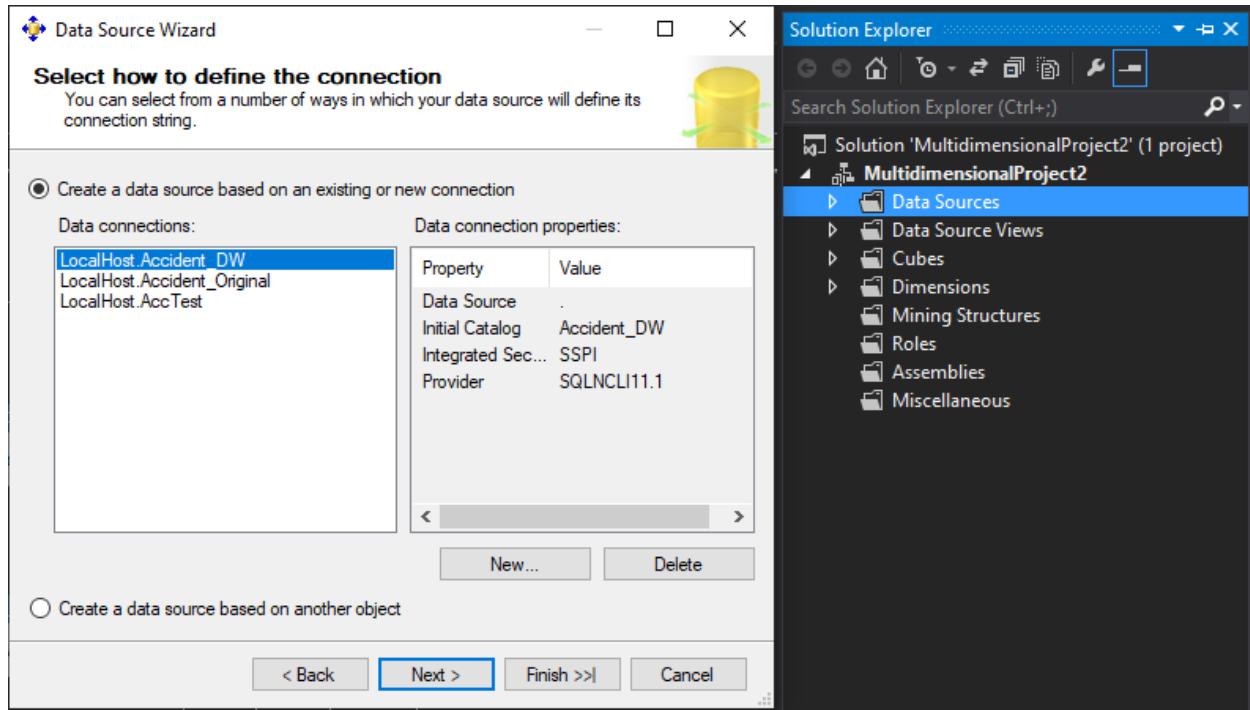
### 1. Thiết lập project SSAS

#### 1.1. Khởi tạo Data source

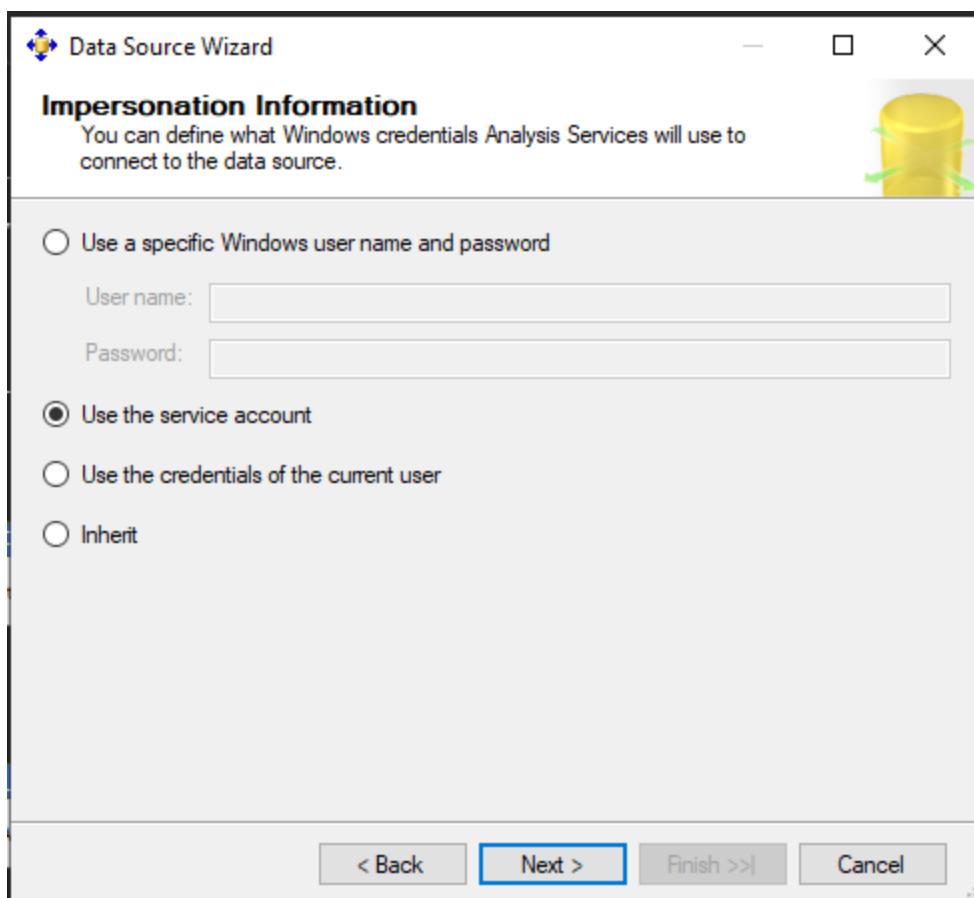
- Tạo project Analysis Services Multidimensional and Data Mining:



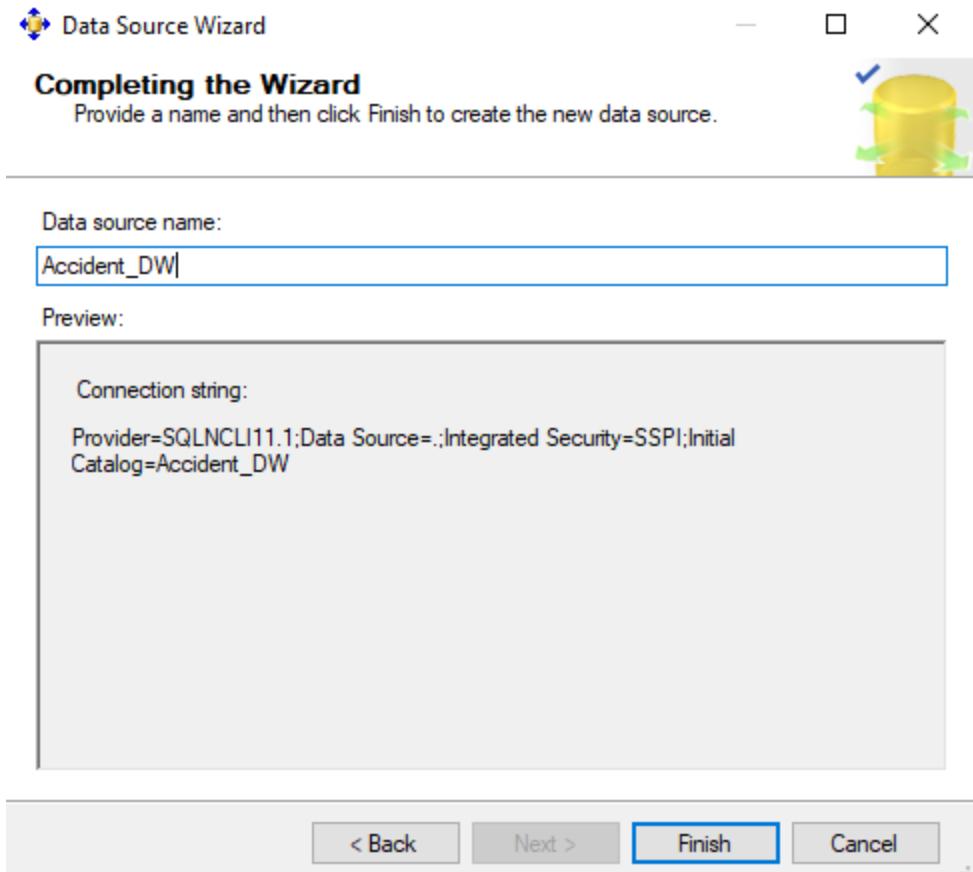
- Click phải vào Data Sources-> New Data Source -> Xuất hiện hộp thoại -> Nhấn Next



- Chọn “Create a data source based on an existing or new connection” -> Chọn Data connections hoặc nhấn “New” nếu muốn tạo connection mới. -> Nhấn Next



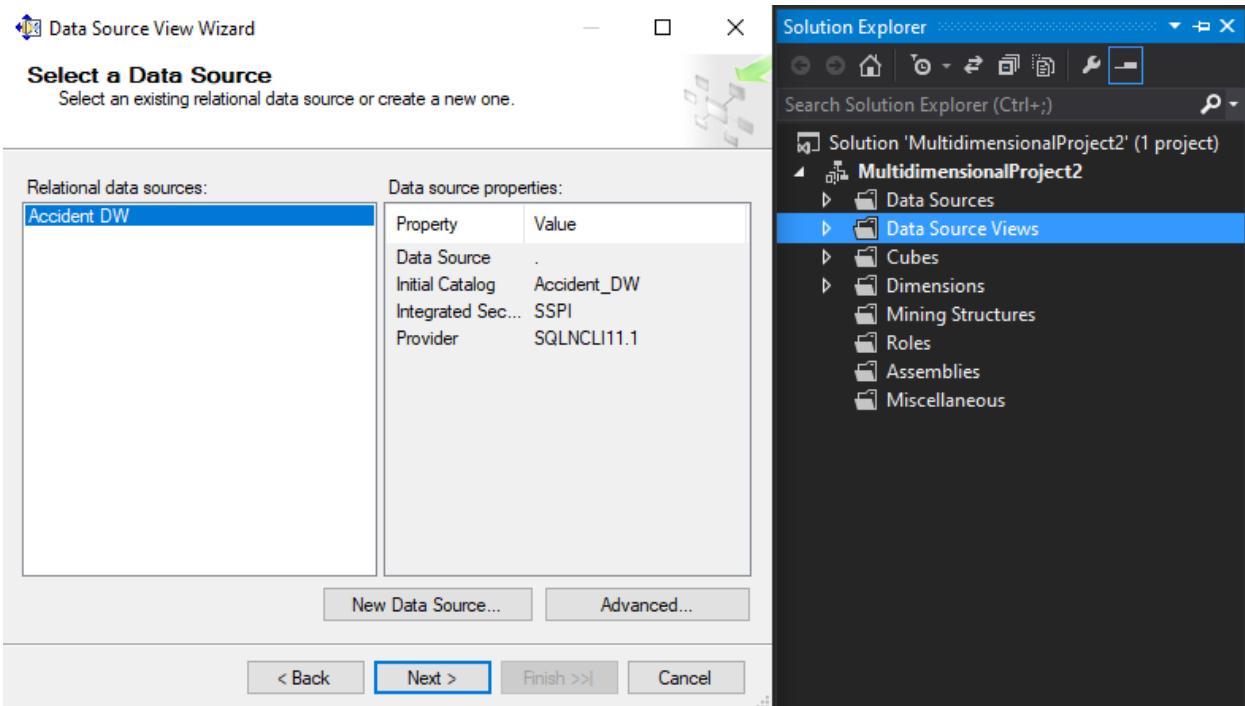
- Chọn “Use the service account” -> Next.



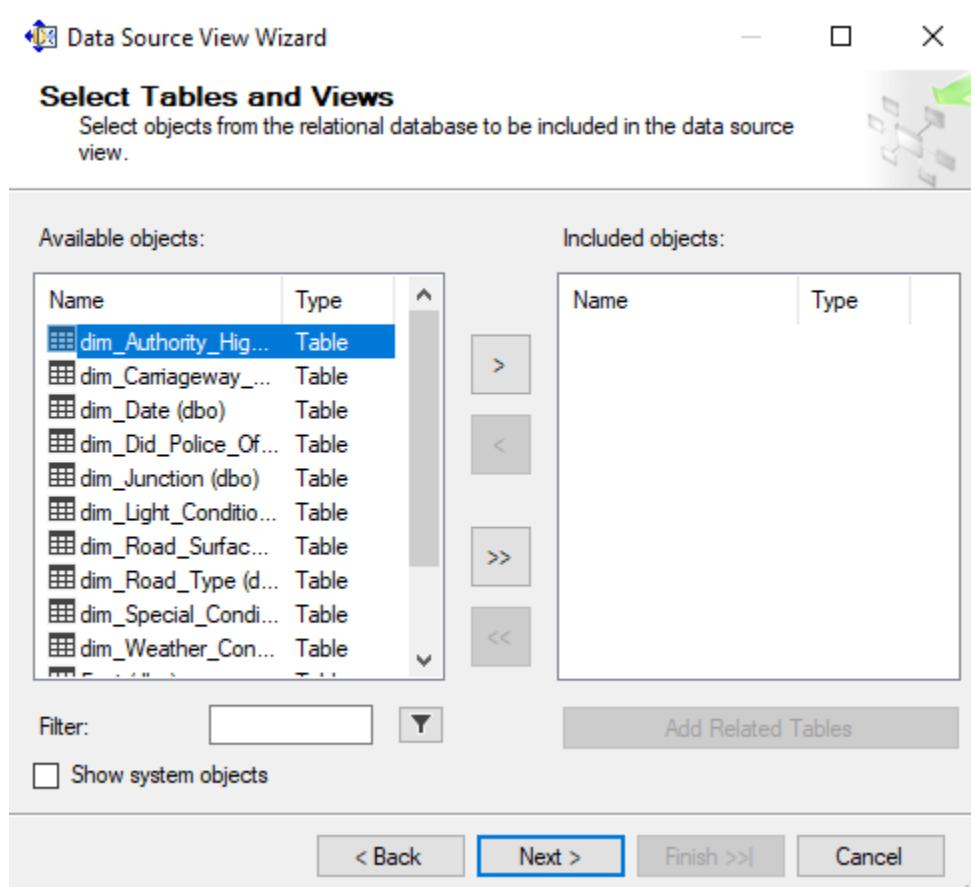
- Đặt tên cho Data souce -> Finish

#### 1.2. Tạo Data source view

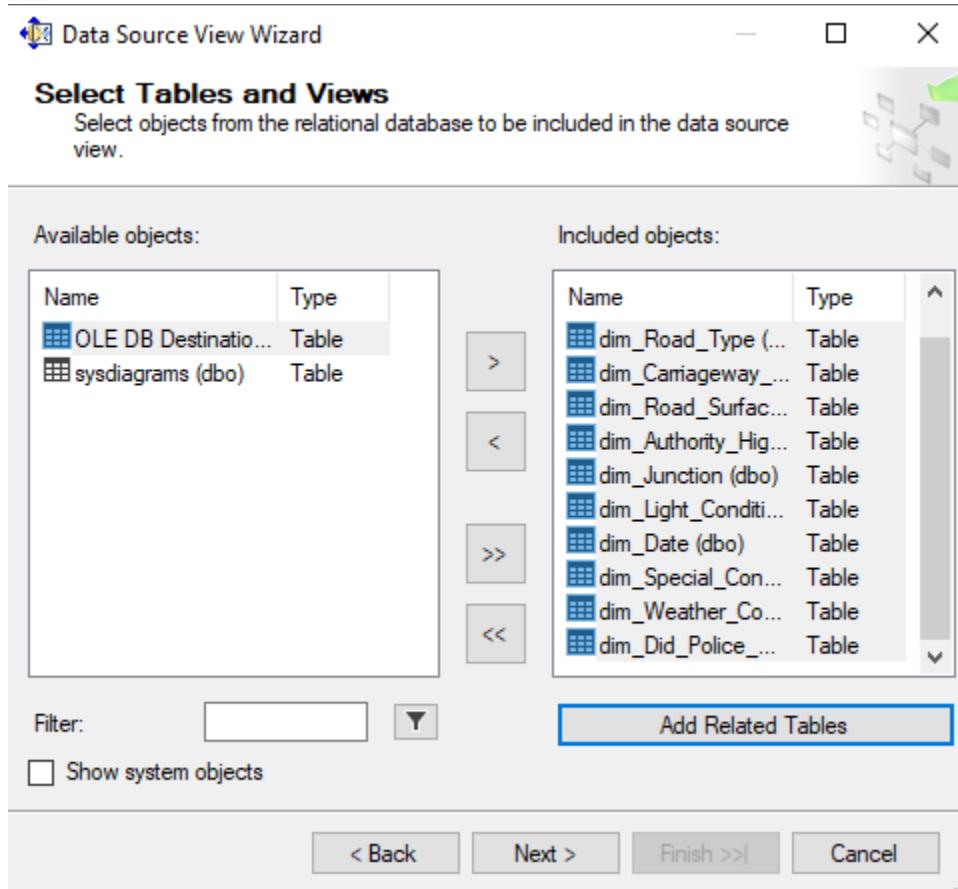
- Click phải Data source views -> New Data Source View-> Xuất hiện hộp thoại ->Next



- Chọn data source -> Next.

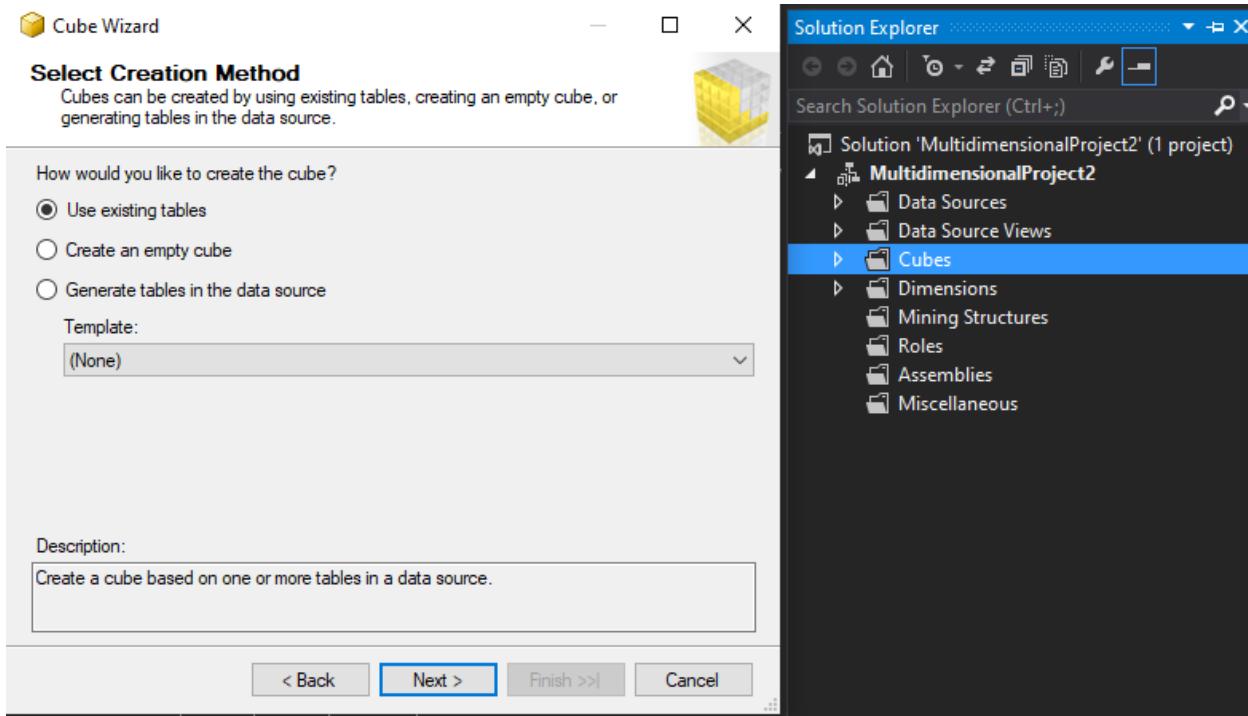


- Chọn bảng Fact -> Chọn Add Related Table -> Finish

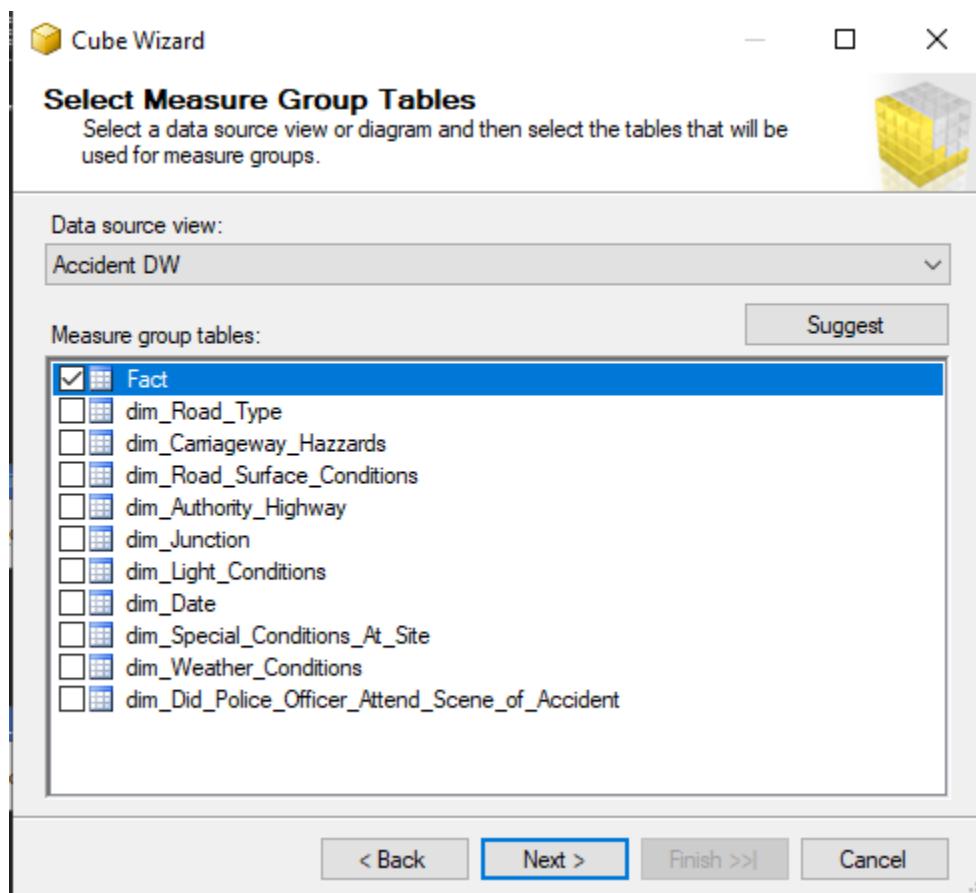


### 1.3. Tạo Cube

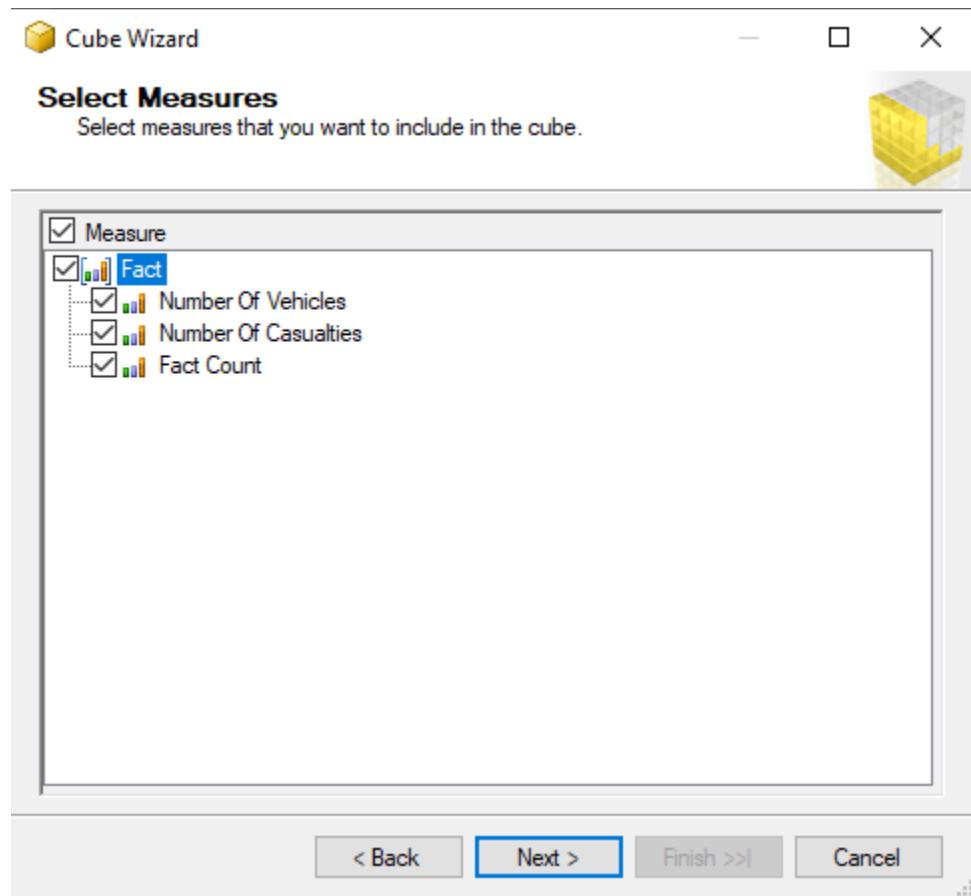
- Click phải Cubes -> New cube -> Xuất hiện hộp thoại -> nhấn Next



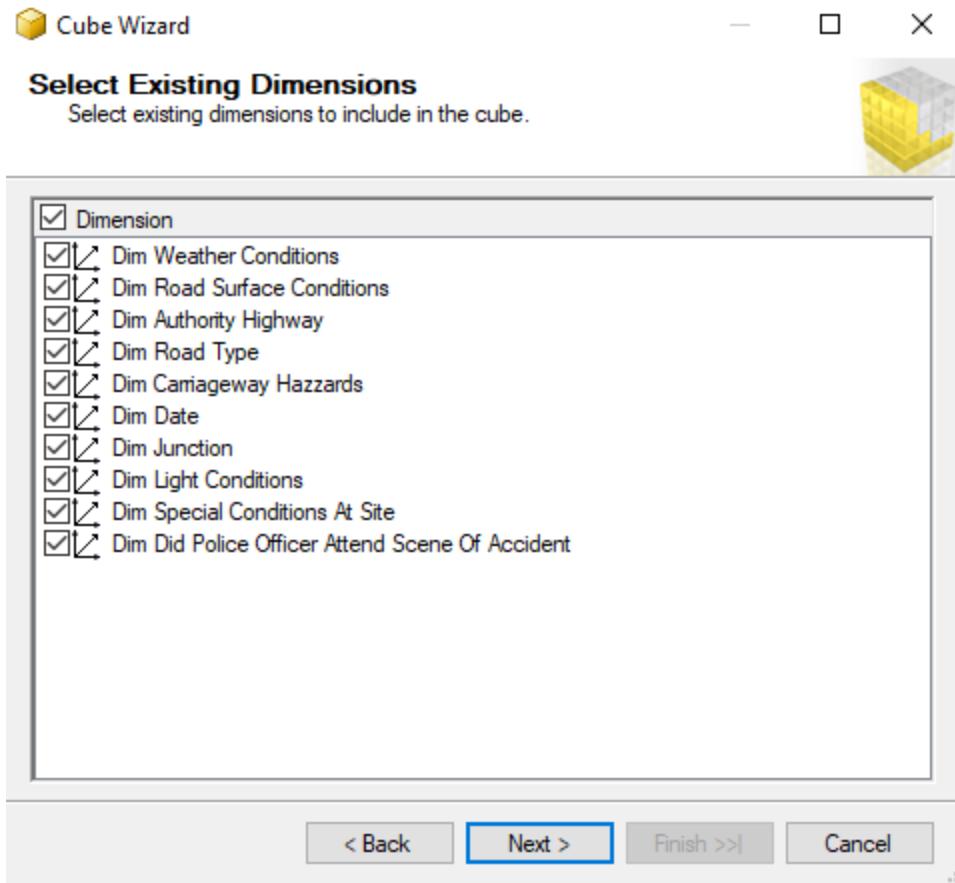
- Chọn “Use existing tables” -> Next



- Chọn bảng Fact -> Next



- Chọn các thuộc tính độ đo -> Next



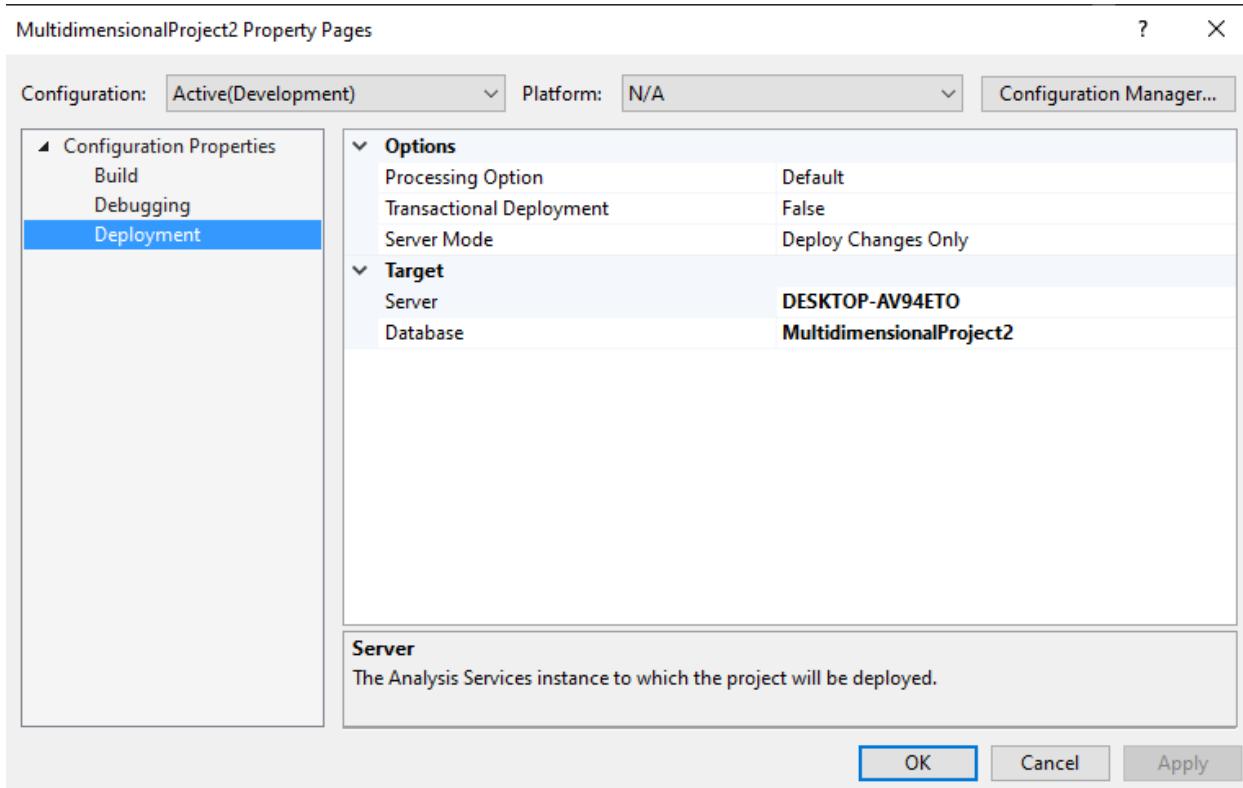
- Chọn tất cả các bảng dimensions -> Next



- Đặt tên cho cube -> Finish

#### 1.4. Chính sửa server phục vụ cho việc deloy

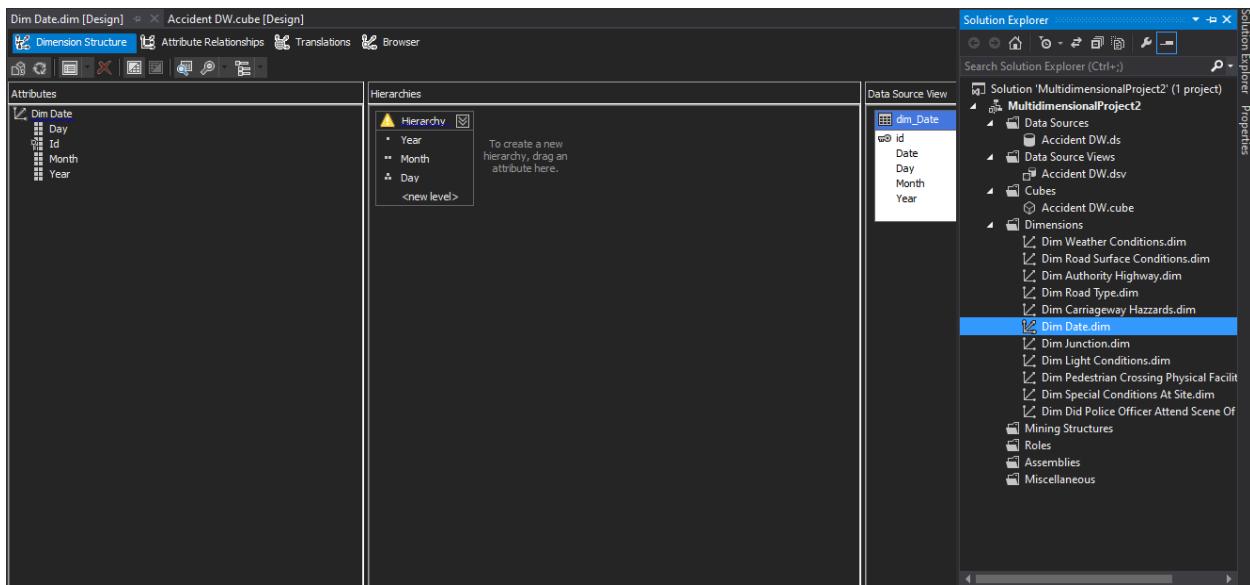
- Click phải vào tên project -> Properties -> Chọn mục Deployment -> sửa lại tên server( chính là server đăng nhập trên SQL Server). -> OK



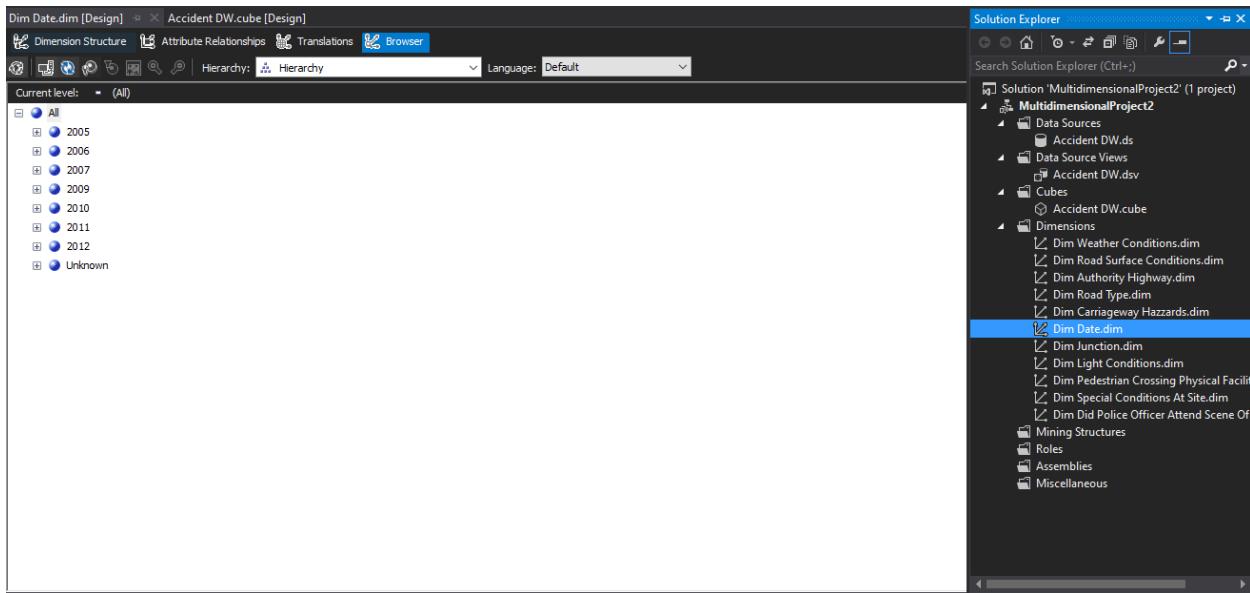
- Click phải vào tên Project -> Deploy.

#### 1.5. Thiết kế các bản Dimension

- Chọn bảng dimemsion trong mục Dimensions.
- Thực hiện kéo thả các thuộc tính (cần cho quá trình truy vấn dữ liệu) trong khung Data source view vào khung Attributes.
- Có thể thay đổi tên thuộc tính hoặc một số property khác trong cửa sổ Properties
- Kéo các thuộc tính cần tạo phân cấp từ khung Attributes sang Hierarchies



- Kiểm tra các trường dữ liệu của bảng dimension trong tab Browser:



- Thiết lập tương tự cho toàn bộ các bảng dimension khác.

## 2. Thực hiện truy vấn

2.1. Xếp loại số lượng các vụ tai nạn của từng loại nguy hiểm trên đường với loại đường One way street (Đường 1 chiều).

- Truy vấn theo BI:

The screenshot shows the configuration of a calculated measure in a BI tool. The left pane displays a list of measures, with 'Cau\_1' selected. The right pane shows the properties of 'Cau\_1': Parent Properties (Measures), Expression (containing a case statement for accident counts), and Additional Properties (Format string: True, Visible: True, Non-empty behavior: Fact, Associated measure group: Fact). Below this is a 'Calculation Tools' section with tabs for Metadata, Functions, and Templates.

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Type	Road Type	Equal	{ One way street }
<Select dimension>			

Carriageway Hazards	Cau_1
Any animal (except ...)	IT
Dislodged vehicle lo...	IT
Involvement with pr...	IT
None	NHIEU
Other object in carri...	BINH THUONG
Pedestrian in carriag...	IT

- Truy vấn theo Pivot Excel

- Truy vấn theo MDX

```

//Cau 1 Xếp loại số lượng các vụ tai nạn của từng loại nguy hiểm trên đường với loại đường One way street (Đường 1 chiều).
with member [Measures].Result as
case
    when [Measures].[Fact Count] > 200 then 'NHIEU'
    when [Measures].[Fact Count] > 100 then 'BINH THUONG'
    when [Measures].[Fact Count] > 0 then 'IT'
end
select [Measures].[Result] on 0,
non empty [Dim Carriageway Hazards].[Carriageway Hazards].children on 1
from [Accident DW]
where [Dim Road Type].[Road Type].&[One way street]

```

Carriageway Hazards	Cau_1
Any animal (except a ridden horse)	IT
Dislodged vehicle load in carriageway	IT
Involvement with previous accident	IT
None	NHIEU
Other object in carriageway	BINH THUONG
Pedestrian in carriageway (not injured)	IT

2.2. Cho biết tổng số nạn nhân tử vong trong các điều kiện ánh sáng nhưng trừ Daylight: Street light present.

- Truy vấn theo BI

The screenshot shows the Microsoft Analysis Services (SSAS) Management Studio interface. On the left, the 'Script Organizer' pane lists a script with several lines containing the expression '[Cau\_2]'. The 'Expression' pane on the right shows the expression: `NONEMPTY(except([Dim Light Conditions].[Light Conditions].[Light Conditions], [Dim Light Conditions].[Light Conditions].[Light Conditions].&[Daylight: Street light present]))`. The 'Additional Properties' section indicates the type is 'Dynamic'. The 'Measure Group' pane on the far left shows the 'Accident DW' measure group with various dimensions like 'Measures', 'Fact', and 'KPIs'. The top navigation bar includes tabs for 'Cube Structure', 'Dimension Usage', 'Calculations', 'KPIs', 'Actions', 'Partitions', 'Aggregations', 'Perspectives', 'Translations', and 'Browser'. The 'Language' dropdown is set to 'Default'.

### - Truy vấn theo Pivot Excel

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields ribbon on the right is open, showing the 'Σ Fact' category expanded with fields like '16', 'Cau\_1', 'Cau\_10', 'Cau\_15', and 'Fact Count'. The 'Fact Count' field is checked. The PivotTable itself displays data from the 'Fact' dimension, grouped by 'Light Conditions' and 'Hierarchy'. The data is as follows:

Light Conditions	Hierarchy	Operator	Filter Expression
Dim Light Conditions	Light Conditions	In	Cau_2
			<Select dimension>
Darkness: No ...	Fact Count		12827
Darkness: Stre... unknown			4965
Darkness: Stre... present and lit			136384
Darkness: Stre... present but unlit			2543

- Truy vấn theo MDX

```
--Cau2
select [Measures].[Fact Count] on columns,
NONEMPTY(except([Dim Light Conditions].[Light Conditions].[Light Conditions],
[Dim Light Conditions].[Light Conditions].[Light Conditions].&[Daylight: Street light present]
)) on rows
from [Accident DW]
```

	Fact Count
Darkness: No street lighting	12827
Darkness: Street lighting unknown	4965
Darkness: Street lights present and lit	136384
Darkness: Street lights present but unlit	2543

2.3. Danh sách loại đường số nạn nhân hơn 20000 hoặc số phương tiện gây tai nạn nhỏ hơn 20000 .

- Truy vấn theo BI

**Script Organizer**

- 1 CALCULATE
- 2 Cau\_1
- 3 (-.) [Cau\_2]
- 4 (-.) [Cau\_3]
- 5 (-.) [Cau\_5]
- 6 (-.) [Cau\_6]
- 7 (-.) [Cau\_7]
- 8 (-.) [Cau\_8]
- 9 (-.) [Cau\_9]
- 10 [Cau\_10]
- 11 (-.) [Cau\_11]

**Dimension Tools**

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Type	Road Type	In	Cau_3
<Select dimension>			

Road Type	Number Of Casualties	Number Of Vehicles
Dual carriageway	94768	129548
Roundabout	85859	128993
Single carriageway	610147	854193
Slip road	11321	15604
Unknown	4031	5616

- Truy vấn theo Pivot Excel

The screenshot shows the Microsoft Power BI Data Editor interface. On the left is a pivot table with rows labeled from 3 to 27. The columns are "Row Labels", "Number Of Casualties", and "Number Of Vehicles". The data includes:

Row Labels	Number Of Casualties	Number Of Vehicles
4 Dual carriageway	94768	129548
5 Roundabout	85859	128993
6 Single carriageway	610147	854193
7 Slip road	11321	15604
8 Unknown	4031	5616
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		

The configuration pane on the right shows fields under "Fact" and "Values" sections, with "Number Of Casualties" and "Number Of Vehicles" selected.

- Truy vấn theo MDX

The screenshot shows Microsoft SQL Server Management Studio (SSMS) with an MDX query window and a results grid.

The MDX query is:

```
//câu 3
SELECT {[Measures].[Number Of Casualties],
          [Measures].[Number Of Vehicles]} ON COLUMNS,
NONEMPTY(
    union(Filter([Dim Road Type].[Road Type].[Road Type],[Measures].[Number Of Casualties]>20000),
         Filter([Dim Road Type].[Road Type].[Road Type],[Measures].[Number Of Vehicles]<20000))) ON ROWS
from [Accident DW]
```

The results grid displays the same data as the Power BI table:

	Number Of Casualties	Number Of Vehicles
Dual carriageway	94768	129548
Roundabout	85859	128993
Single carriageway	610147	854193
Slip road	11321	15604
Unknown	4031	5616

2.4. Thống kê số lượng phương tiện gây tai nạn theo tháng trong năm 2007 theo từng điều kiện ánh sáng.

- Truy vấn theo MDX

```
//câu 4 Thống kê số lượng phương tiện gây tai nạn theo tháng trong năm 2007 theo từng điều kiện ánh sáng.
SELECT [Measures].[Number Of Vehicles] ON 0
,NONEMPTY(DRILLDOWNLEVEL(
CROSSJOIN([Dim Date].[Hierarchy].[Year].&[2007],[Dim Light Conditions].[Light Conditions].[Light Conditions]))) ON 1
FROM [Accident DW]
```

100 % <

		Number Of Vehicles
2007	Darkness: No street lighting	4215
2007	Darkness: Street lighting unknown	1685
2007	Darkness: Street lights present and lit	42121
2007	Darkness: Street lights present but unlit	758
2007	Daylight: Street light present	150149
1	Darkness: No street lighting	705
1	Darkness: Street lighting unknown	275
1	Darkness: Street lights present and lit	6136
1	Darkness: Street lights present but unlit	82
1	Daylight: Street light present	9239
10	Darkness: No street lighting	381
10	Darkness: Street lighting unknown	175
10	Darkness: Street lights present and lit	4087
10	Darkness: Street lights present but unlit	78
10	Daylight: Street light present	12271
11	Darkness: No street lighting	720
11	Darkness: Street lighting unknown	220
11	Darkness: Street lights present and lit	6625
11	Darkness: Street lights present but unlit	110
11	Daylight: Street light present	10216

2.5. Đếm tổng số tai nạn theo mỗi loại kiểm soát ở những nơi giao nhau trong năm 2009 rồi sắp xếp giảm dần.

- Truy vấn theo BI

The screenshot shows the Microsoft Analysis Services Script Organizer interface. On the left, there's a tree view of scripts and calculated members. In the center, a script editor window is open with the following details:

- Name:** [Cau\_5]
- Expression:** ORDER([Dim Junction].[Junction Control].children,[MEASURES].[Fact Count],desc)
- Additional Properties:**
  - Type: Dynamic
  - Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Date	Year	Equal	{ 2009 }
Dim Junction	Junction Control	In	Cau_5
<Select dimension>			

Junction Control	Fact Count
Authorised per...	249
Automatic traffi...	16749
Giveaway or unc...	78938
Stop Sign	942

- Truy vấn theo Pivot Excel

Year	2009
Row Labels	Fact Count
Authorised person	249
Automatic traffic signal	16749
Giveaway or uncontrolled	78938
Stop Sign	942

- Truy vấn theo MDX

```
//câu 5
SELECT [Measures].[Fact Count] ON COLUMNS,
NON EMPTY ORDER([Dim Junction].[Junction Control].children,[Measures].[Fact Count],desc) ON ROWS
FROM [Accident DW]
WHERE [Dim Date].[Year].&[2009]
```

100 % <

	Fact Count
Giveaway or uncontrolled	78938
Automatic traffic signal	16749
Stop Sign	942
Authorised person	249

2.6. Theo từng năm, chọn ra top 3 loại thời tiết gây tai nạn ít nhất.

- Truy vấn theo BI

Name: [Cau\_6]

Expression:

```
BOTTOMCOUNT([Dim Weather Conditions].[Weather Conditions].CHILDREN,3,[Measures].[Fact Count])
```

Additional Properties:

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Weather Conditions	Weather Conditions	In	Cau_6
<Select dimension>			

Year	Weather Conditions	Fact Count
2005	Fog or mist	566
2005	Snowing with high winds	91
2006	Fog or mist	660
2006	Snowing with high winds	37
2007	Fog or mist	394
2007	Snowing with high winds	43
2009	Fog or mist	368
2009	Snowing with high winds	100
2010	Fog or mist	433
2010	Snowing with high winds	137
2011	Fog or mist	317
2011	Snowing with high winds	19
2012	Fog or mist	4
2012	Snowing with high winds	1

- Truy vấn theo Pivot Excel

PivotTable Fields

Choose fields to add to report:

Σ Fact

- 16
- Cau\_1
- Cau\_10
- Cau\_15
- Fact Count
- Number Of Casualties
- Number Of Vehicles

Drag fields between areas below:

Filters	Columns
Rows	Σ Values
Year	Fact Count
Cau_6	

- Truy vấn theo MDX

The screenshot shows the SSMS interface with an MDX query in the query editor and a results grid below it.

```
//câu 6
SELECT [Measures].[Fact Count] ON 0,
NON EMPTY CROSSJOIN([Dim Date].[Year].[Year],
BOTTOMCOUNT([Dim Weather Conditions].[Weather Conditions].CHILDREN,3,[Measures].[Fact Count])) ON 1
FROM [Accident DW]
```

The results grid displays the following data:

		Fact Count
2005	Snowing with high winds	91
2005	Fog or mist	566
2006	Snowing with high winds	37
2006	Fog or mist	660
2007	Snowing with high winds	43
2007	Fog or mist	394
2009	Snowing with high winds	100
2009	Fog or mist	368
2010	Snowing with high winds	137
2010	Fog or mist	433
2011	Snowing with high winds	19
2011	Fog or mist	317
2012	Snowing with high winds	1
2012	Fog or mist	4

2.7. Thống kê số lượng vụ tai nạn có điều kiện ánh sáng với tên điều kiện ánh sáng có chuỗi "present" trong tên.

- Truy vấn theo BI

The screenshot shows the 'Properties' dialog box for a calculated member named '[Cau\_7]'.

**Name:** [Cau\_7]

**Expression:**

```
FILTER([Dim Light Conditions].[Light Conditions].[Light Conditions],
Instr([Dim Light Conditions].[Light Conditions].CURRENTMEMBER.Name,"present")>0)
```

**Additional Properties:**

- Type: Dynamic
- Display folder: (empty)

Screenshot of the SSAS Dimension Filter dialog and a table showing the count of street light conditions.

**Dimension Filter Dialog:**

Dimension	Hierarchy	Operator	Filter Expression
Dim Light Conditions	Light Conditions	In	Cau_7
<Select dimension>			

**Table Data:**

Light Conditions	Fact Count
Darkness: Street lights present and lit	136384
Darkness: Street lights present but unlit	2543
Daylight: Street light present	458287

- Truy vấn theo Pivot Excel

Screenshot of a Microsoft Excel PivotTable setup.

**PivotTable Fields Panel:**

- Choose fields to add to report:
- Search:
- Dim Light Conditions:
  - Hierarchy
  - More Fields
  - Sets
    - Cau\_11
    - Cau\_13
    - Cau\_2
    - Cau\_7

**PivotTable Data View:**

Row Labels	Fact Count
Darkness: Street lights present and lit	136384
Darkness: Street lights present but unlit	2543
Daylight: Street light present	458287

- Truy vấn theo MDX

Screenshot of the SSMS MDX Query window and its results.

**MDX Query:**

```
//câu 7
SELECT [Measures].[Fact Count] ON 0,
NON EMPTY FILTER([Dim Light Conditions].[Light Conditions].[Light Conditions],
Instr([Dim Light Conditions].[Light Conditions].CURRENTMEMBER.Name,"present")>0) ON 1
FROM [Accident DW]
```

**Results:**

	Fact Count
Darkness: Street lights present and lit	136384
Darkness: Street lights present but unlit	2543
Daylight: Street light present	458287

2.8. Các loại đường tại nơi xảy ra tai nạn vừa nằm trong top 3 số vụ tai nạn và top 3 số phương tiện gây tai nạn.

- Truy vấn theo BI

Name: [Cau\_8]

Expression:

```
INTERSECT(
TOPCOUNT([Dim Road Type].[Road Type].CHILDREN,3,[Measures].[Fact Count]),
TOPCOUNT([Dim Road Type].[Road Type].CHILDREN,3,[Measures].[Number Of Vehicles]))
```

Additional Properties:

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Type	Road Type	In	Cau_8
<Select dimension>			

Road Type	Fact Count	Number Of Vehicles
Dual carriageway	66049	129548
Roundabout	67785	128993
Single carriageway	455929	854193

- Truy vấn theo Pivot Excel

Row Labels	Fact Count	Number Of Vehicles
4 Dual carriageway	66049	129548
5 Roundabout	67785	128993
6 Single carriageway	455929	854193
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		

Choose fields to add to report:

Σ Fact

Fact Count

Number Of Vehicles

Drag fields between areas below:

Filters

Values

Rows

Cau\_8

Fact Count

Number Of Vehicles

- Truy vấn theo MDX

```
//câu 8 Loại đường tại nơi xảy ra tai nạn vừa nằm trong top 3 số vụ tai nạn và top 3 số phương tiện gây tai nạn.
SELECT {[Measures].[Fact Count],[Measures].[Number Of Vehicles]} ON 0,
INTERSECT(
TOPCOUNT([Dim Road Type].[Road Type].CHILDREN,3,[Measures].[Fact Count]),
TOPCOUNT([Dim Road Type].[Road Type].CHILDREN,3,[Measures].[Number Of Vehicles])) ON 1
FROM [Accident DW]
```

100 % <

	Fact Count	Number Of Vehicles
Single carriageway	455929	854193
Roundabout	67785	128993
Dual carriageway	66049	129548

## 2.9. Xác định 10 cao tốc có số nạn nhân nhiều nhất.

- Truy vấn theo BI

Name:

Expression:

```
SUBSET(ORDER([Dim Authority Highway].[Local Authority_Highway].CHILDREN,[Measures].[Number Of Casualties],DESC),0,10)
```

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Authority Highway	Local Authority_Highway	In	Cau_9
<Select dimension>			

Local Authority_Highway	Number Of Casualties
E08000003	12172
E08000025	19804
E08000035	12380
E10000012	14943
E10000014	14669
E10000015	13994
E10000016	18723
E10000017	20898
E10000028	12960
E10000030	19615

- Truy vấn theo Pivot Excel

- Truy vấn theo MDX

```
//câu 9
SELECT [Measures].[Number Of Casualties] ON COLUMNS,
SUBSET(ORDER([Dim Authority Highway].[Local Authority_Highway].CHILDREN,[Measures].[Number Of Casualties],DESC),0,10)
ON ROWS
FROM [Accident DW]
```

	Number Of Casualties
E10000017	20898
E08000025	19804
E10000030	19615
E10000016	18723
E10000012	14943
E10000014	14669
E10000015	13994
E10000028	12960
E08000035	12380
E08000003	12172

- 2.10. Tìm số lượng vụ tai nạn ít nhất theo năm công bố của loại đường Single carriageway
- Truy vấn theo BI

Name: [Cau\_10]

Parent Properties

Parent hierarchy: Measures

Parent member: Change

Expression

```
MIN([Dim Date].[Year].CHILDREN,[Measures].[Fact Count])
```

Additional Properties

Format string:

Visible: True

Non-empty behavior:

Associated measure group: (Undefined)

Display folder:

Color Expressions

Font Expressions

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Type	Road Type	Equal	{ Single carriageway }
<Select dimension>			

Road Type	Cau_10
Single carriageway	1653

- Truy vấn theo Pivot Excel

2

3 Row Labels [Cau\_10]

4 Single carriageway 1653

5 Grand Total 1653

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

Choose fields to add to report:

Search

Hierarchy

More Fields

Sets

Dim Road Type

Road Type

Filters

Rows

Road Type

Values

Cau\_10

- Truy vấn theo MDX

```
//câu 10
WITH MEMBER MEASURES.[MIN CASUALTIES] AS MIN(
[Dim Date].[Year].CHILDREN,[Measures].[Fact Count])
SELECT MEASURES.[MIN CASUALTIES] ON 0,
NONEMPTY([Dim Road Type].[Road Type].&[Single carriageway]) ON 1
FROM [Accident DW]
```

The screenshot shows the SSMS MDX query editor. The query window contains the provided MDX code. Below it, the results pane shows a single row:

	MIN CASUALTIES
Single carriageway	1653

2.11. Xác định top 3 điều kiện đặc biệt có nhiều tai nạn nhất ngoại trừ loại điều kiện đặc biệt Unknown.

- Truy vấn theo BI

The screenshot shows the Analysis Services Properties dialog for a calculated member named [Cau\_11].

**Name:** [Cau\_11]

**Expression:**

```
TOPCOUNT(
[Dim Special Conditions At Site].[Special Conditions At Site].CHILDREN
- EXISTS([Dim Special Conditions At Site].[Special Conditions At Site].CHILDREN,
[Dim Special Conditions At Site].[Special Conditions At Site].&[Unknown]),
,3,[Measures].[Fact Count])
```

**Additional Properties:**

- Type: Dynamic
- Display folder: (empty)

**Dimension Properties:**

Dimension	Hierarchy	Operator	Filter Expression
Dim Special Conditions At Site	Special Conditions At Site	In	Cau_11
<Select dimension>			

**Result Table:**

Special Conditions At Site	Fact Count
None	602480
Oil or diesel	1650
Roadworks	5938

- Truy vấn theo Pivot Excel

The screenshot shows the Microsoft Power BI Data Editor interface. On the left is a Pivot Table with rows labeled 'Row Labels' and 'Fact Count'. The data includes:

Row Labels	Fact Count
None	602480
Oil or diesel	1650
Roadworks	5938

To the right is the 'Choose fields to add to report:' pane, which contains sections for Fact (with 'Fact Count' checked) and Dim Authority Highway. Below it are 'Filters' and 'Columns' sections, and a 'Rows' section set to 'Cau 11' with 'Fact Count' as the value.

- Truy vấn theo MDX

```
//câu 11 Xác định top 3 điều kiện đặc biệt có nhiều tai nạn nhất ngoại trừ loại điều kiện đặc biệt Unknown.
SELECT [Measures].[Fact Count] ON 0,
TOPCOUNT
[Dim Special Conditions At Site].[Special Conditions At Site].CHILDREN
- EXISTS([Dim Special Conditions At Site].[Special Conditions At Site].CHILDREN,
[Dim Special Conditions At Site].[Special Conditions At Site].[Special Conditions At Site].&[Unknown])
,3,[Measures].[Fact Count] ON 1
FROM [Accident DW]
```

The screenshot shows a BI tool's results grid. The title bar says '100 %'. The grid has two columns: 'Messages' and 'Results'. The 'Results' tab is selected and displays the following data:

	Fact Count
None	602480
Roadworks	5938
Oil or diesel	1650

2.12. Cao tốc có số tai nạn ít nhất có loại đường là Dual carriageway theo mỗi năm.

- Truy vấn theo BI

Name: [Cau\_12]

Expression

```
TOPCOUNT([Dim Authority Highway].[Local Authority_Highway].CHILDREN,1,[Measures].[Fact Count])
```

Additional Properties

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Type	Road Type	Equal	{ Dual carriageway }
Dim Authority Highway	Local Authority_Highway	In	Cau_12
<Select dimension>			

Year	Local Authority_Highway	Fact Count
2005	E08000025	630
2006	E08000025	570
2007	E08000025	567
2009	E08000025	415
2010	E08000025	396
2011	E08000025	331

### - Truy vấn theo Pivot Excel

PivotTable Fields

Choose fields to add to report:

Search

Σ Fact

- 16
- Cau\_1
- Cau\_10
- Cau\_15
- Fact Count
- Number Of Casualties
- Number Of Vehicles

Drag fields between areas below:

Filters

Road Type

Rows

Cau\_12

Year

Columns

Values

Fact Count

- Truy vấn theo MDX

```
//câu 12
SELECT [Measures].[Fact Count] ON 0,
NON EMPTY CROSSJOIN([Dim Date].[Year].[Year],
TOPCOUNT([Dim Authority Highway].[Local Authority_ Highway].CHILDREN,1,[Measures].[Fact Count])) ON 1
FROM [Accident DW]
WHERE [Dim Road Type].[Road Type]&[Dual carriageway]

//câu 13
100 % <


|      |           | Fact Count |
|------|-----------|------------|
| 2005 | E08000025 | 630        |
| 2006 | E08000025 | 570        |
| 2007 | E08000025 | 567        |
| 2009 | E08000025 | 415        |
| 2010 | E08000025 | 396        |
| 2011 | E08000025 | 331        |


```

2.13. Thống kê số lượng nạn nhân trong các vụ tai nạn ở điều kiện ánh sáng có tên có 4 kí tự đầu là “Dark”.

- Truy vấn theo BI

Name: [Cau\_13]

Expression:

```
FILTER([Dim Light Conditions].[Light Conditions].[Light Conditions],
LEFT([Dim Light Conditions].[Light Conditions].CURRENTMEMBER.Name,4)='Dark')
```

Additional Properties:

Type: Dynamic

Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Light Conditions	Light Conditions	In	Cau_13
<Select dimension>			

Year	Light Conditions	Number Of Casualties
2005	Darkness: No street lighting	4245
2005	Darkness: Street lighting unknown	1005
2005	Darkness: Street lights present and lit	35780
2005	Darkness: Street lights present but unlit	787
2006	Darkness: No street lighting	3855
2006	Darkness: Street lighting unknown	931
2006	Darkness: Street lights present and lit	35128
2006	Darkness: Street lights present but unlit	594
2007	Darkness: No street lighting	3631
2007	Darkness: Street lighting unknown	1220
2007	Darkness: Street lights present and lit	32005
2007	Darkness: Street lights present but unlit	550
2009	Darkness: No street lighting	3143
2009	Darkness: Street lighting unknown	1171
2009	Darkness: Street lights present and lit	29457
2009	Darkness: Street lights present but unlit	499
2010	Darkness: No street lighting	2757
2010	Darkness: Street lighting unknown	931

- Truy vấn theo Pivot Excel

The screenshot shows the Microsoft Power BI Data view. On the left is a PivotTable with 'Number Of Casualties' in the Values field, 'Year' in the Rows field, and 'Light Conditions' in the Columns field. The table structure is as follows:

Light Conditions	2005	2006	2007	2009	2010	2011	2012
Darkness: No street lighting	4245	3855	3631	3143	2757	2809	1
Darkness: Street lighting unknown	1005	931	1220	1171	931	1212	12
Darkness: Street lights present and lit	35780	35128	32005	29457	26365	26686	749

On the right side of the interface, there are several panes: 'Choose fields to add to report', 'Filters', and 'Columns'. The 'Choose fields to add to report' pane shows fields from the Fact table, including 'Fact', 'Cau\_1', 'Cau\_10', 'Cau\_15', 'Fact Count', 'Number Of Casualties' (which is checked), and 'Number Of Vehicles'. The 'Filters' pane shows 'Rows' set to 'Cau\_13' and 'Year'. The 'Columns' pane shows 'Values' set to 'Number Of Casualties'.

- Truy vấn theo MDX

```
//câu 13
SELECT [Measures].[Number Of Casualties] ON 0,
NONEMPTYCROSSJOIN([Dim Date].[Year].CURRENTMEMBER.CHILDREN,
FILTER([Dim Light Conditions].[Light Conditions].[Light Conditions],
LEFT([Dim Light Conditions].[Light Conditions].CURRENTMEMBER.Name,4)='Dark')) ON 1
FROM [Accident DW]
```

100 % <

		Number Of Casualties
2005	Darkness: No street lighting	4245
2005	Darkness: Street lighting unknown	1005
2005	Darkness: Street lights present and lit	35780
2005	Darkness: Street lights present but unlit	787
2006	Darkness: No street lighting	3855
2006	Darkness: Street lighting unknown	931
2006	Darkness: Street lights present and lit	35128
2006	Darkness: Street lights present but unlit	594
2007	Darkness: No street lighting	3631
2007	Darkness: Street lighting unknown	1220
2007	Darkness: Street lights present and lit	32005
2007	Darkness: Street lights present but unlit	550
2009	Darkness: No street lighting	3143
2009	Darkness: Street lighting unknown	1171
2009	Darkness: Street lights present and lit	29457
2009	Darkness: Street lights present but unlit	499
2010	Darkness: No street lighting	2757
2010	Darkness: Street lighting unknown	931

2.14. Thống kê số lượng các vụ tai nạn trong các điều kiện mặt đường sao cho tổng ít nhất là 10000 với ít loại mặt đường nhất có thể.

- Truy vấn theo BI

Name: [Cau\_14]

Expression: BOTTOMSUM([Dim Road Surface Conditions].[Road Surface Conditions].CHILDREN,10000,[Measures].[Fact Count])

Additional Properties:

- Type: Dynamic
- Display folder:

Dimension	Hierarchy	Operator	Filter Expression
Dim Road Surface Conditions	Road Surface Conditions	In	Cau_14
<Select dimension>			

Road Surface Conditions	Fact Count
Flood (Over 3cm of water)	268
Frost/Ice	8757
Snow	3239

### - Truy vấn theo Pivot Excel

Choose fields to add to report:

Search

Fact

- 16
- Cau\_1
- Cau\_10
- Cau\_15
- Fact Count
- Number Of Casualties
- Number Of Vehicles

Drag fields between areas below:

Filters

Columns

Rows: Cau\_14

Values: Fact Count

### - Truy vấn theo MDX

```
FROM [Accident DW]

//câu 14
SELECT [Measures].[Fact Count] ON 0,
BOTTOMSUM([Dim Road Surface Conditions].[Road Surface Conditions].CHILDREN,10000,[Measures].[Fact Count]) ON 1
FROM [Accident DW]

//câu 15
```

00 % <

	Fact Count
Unknown	(null)
Flood (Over 3cm of water)	268
Snow	3239
Frost/Ice	8757

2.15. Với các vụ tai nạn xảy ra vào các điều kiện thời tiết khi không cảnh sát trực, xuất ra "Thương vong lớn" nếu số người thương vong hơn 2000 ngược lại xuất ra "Thương vong nhỏ".

- Truy vấn theo BI

Name: [Cau\_15]

**Parent Properties**

Parent hierarchy: Measures

Parent member:  Change

**Expression**

```
IIf([Measures].[Number Of Casualties] > 2000, "Thương vong lớn", "Thương vong nhỏ")
```

**Additional Properties**

Format string:

Visible: True

Non-empty behavior:

Associated measure group: (Undefined)

Display folder:

**Color Expressions**

**Font Expressions**

Dimension	Hierarchy	Operator	Filter Expression
Dim Did Police Officer Attend Scene...	Did Police Officer Attend Scene ...	Equal	{ No }
<Select dimension>			

Weather Conditions	Cau 15
Fine with high winds	Thương vong nhỏ
Fine without high ...	Thương vong lớn
Fog or mist	Thương vong nhỏ
Other	Thương vong lớn
Raining with high w...	Thương vong nhỏ
Raining without hig...	Thương vong lớn
Snowing with high ...	Thương vong nhỏ
Snowing without hi...	Thương vong nhỏ
Unknown	Thương vong lớn
Unknown	Thương vong nhỏ

- Truy vấn theo Pivot Excel

Did Police Officer Attend Scene Of Accident	No
Row Labels	Cau 15
Fine with high winds	Thương vong nhỏ
Fine without high winds	Thương vong lớn
Fog or mist	Thương vong nhỏ
Other	Thương vong lớn
Raining with high winds	Thương vong nhỏ
Raining without high winds	Thương vong lớn
Snowing with high winds	Thương vong nhỏ
Snowing without high winds	Thương vong nhỏ
Unknown	Thương vong lớn
Unknown	Thương vong nhỏ
Grand Total	Thương vong lớn
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	

- Truy vấn theo MDX

```

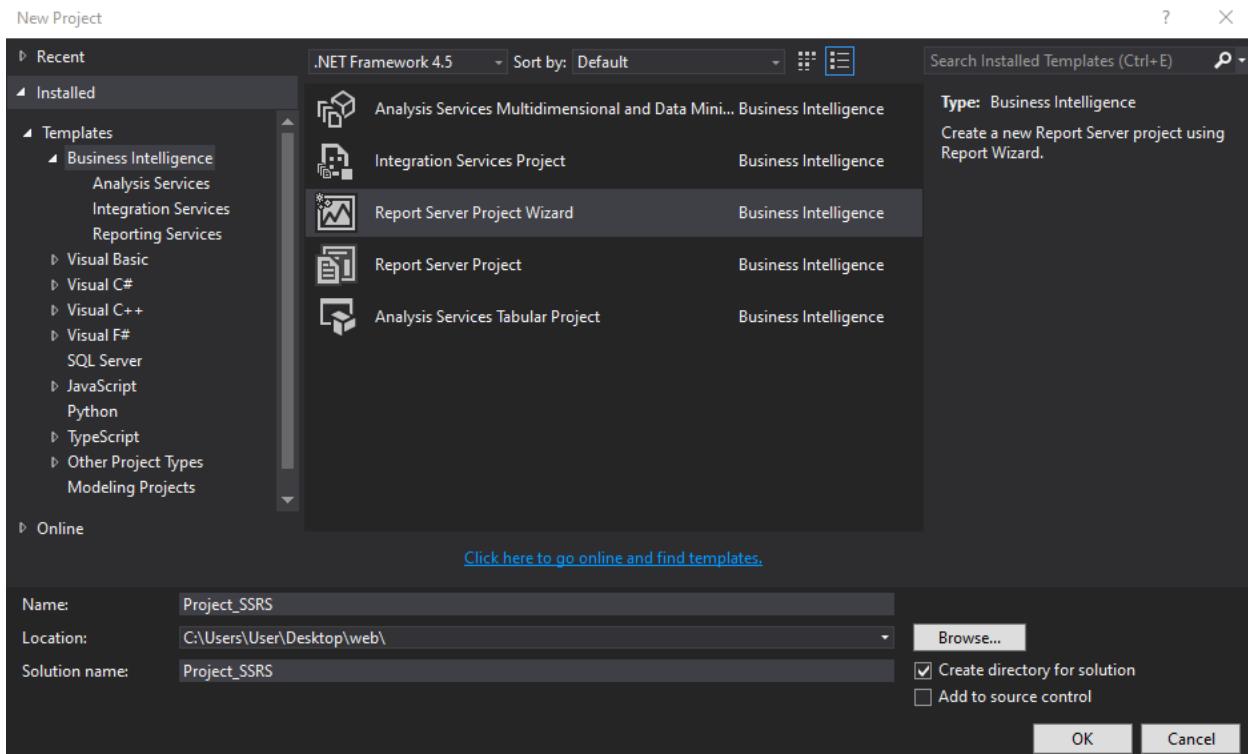
//câu 15 Với các vụ tai nạn xảy ra vào các điều kiện thời tiết khi không có cảnh sát trực, xuất ra "Thương vong lớn"
// nếu số người thương vong hơn 2000 ngược lại xuất ra "Thương vong nhỏ".
WITH MEMBER [Measures].[RESULT]
AS IIf([Measures].[Number Of Casualties] > 2000,"Thương vong lớn","Thương vong nhỏ")
SELECT NON EMPTY [MEASURES].[RESULT] ON COLUMNS,
      [Dim Weather Conditions].[Weather Conditions].children ON ROWS
FROM [Accident DW]
WHERE [Dim Did Police Officer Attend Scene Of Accident].[Did Police Officer Attend Scene Of Accident].&[No]
  
```

Messages	Results
100 %	
	RESULT
Fine with high winds	Thương vong nhỏ
Fine without high winds	Thương vong lớn
Fog or mist	Thương vong nhỏ
Other	Thương vong lớn
Raining with high winds	Thương vong nhỏ
Raining without high winds	Thương vong lớn
Snowing with high winds	Thương vong nhỏ
Snowing without high winds	Thương vong nhỏ
Unknown	Thương vong lớn
Unknown	Thương vong nhỏ

## CHƯƠNG 5: QUÁ TRÌNH SSRS

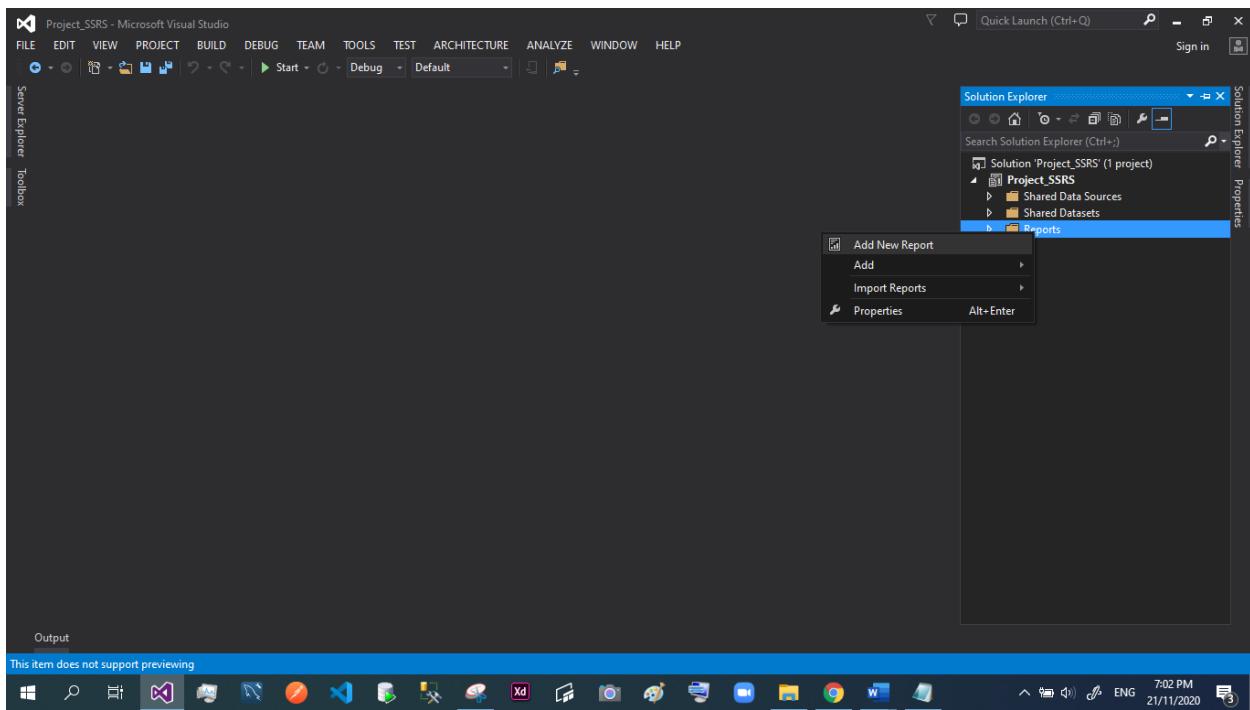
### 1. Tạo báo cáo với SSRS Project

#### 1.1. Tạo mới một project

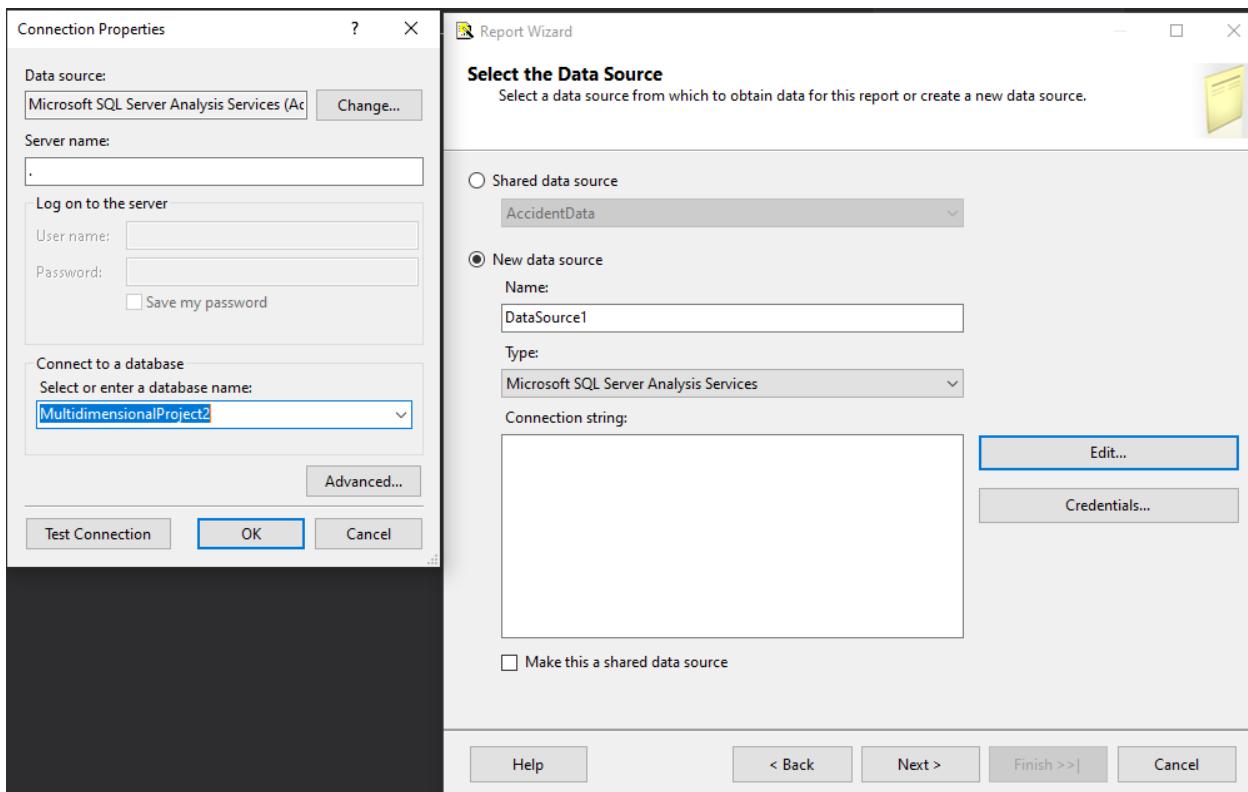


#### 1.2. Kết nối

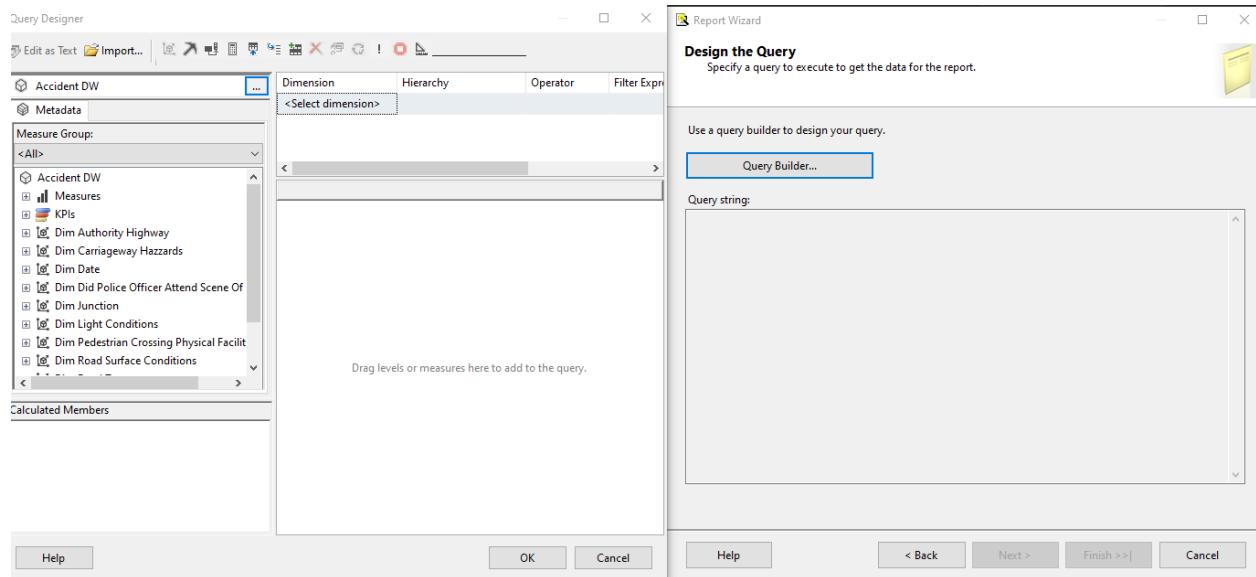
- Tại hộp thoại Solution Explorer, r-click vào Report và Add New Report



- Thiết lập tương tự như lúc truy vấn bằng công cụ BI ở mục SSAS lưu ý rằng ta cần chọn Microsoft SQL Server Analysis Services ở mục Type. Sau đó nhấn Next



- Nhấn chọn Query Builder sau đó chọn những thuộc tính, độ đo cần thiết cho báo cáo. Sau đó nhấn OK và chọn Next



- Chọn mục Tabular và nhấn Next

## Report Wizard

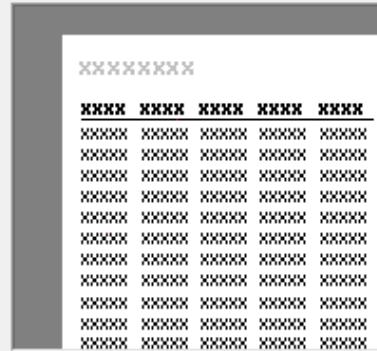
### Select the Report Type

Select the type of report that you want to create.



Tabular

Matrix



Help

< Back

Next >

Finish >>

Cancel

- Chuyển hết các thuộc tính từ Available fields sang mục Details sau đó chọn Next và cuối cùng là Finish



## Design the Table

Choose how to group the data in the table.

### Available fields:

Page>

Group>

Details>

< Remove

### Displayed fields:

Number\_Of\_Casualties  
Number\_Of\_Vehicles

< >



Help

< Back

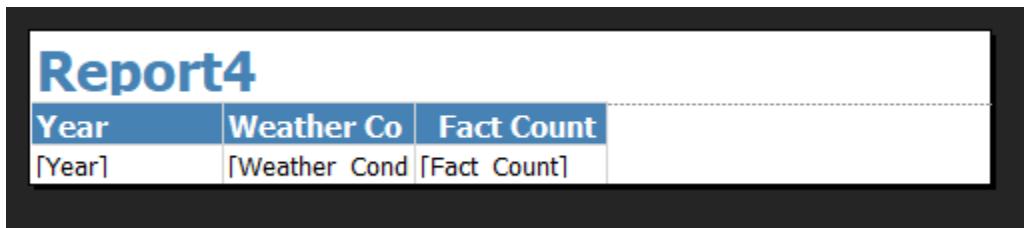
Next >

Finish >>

Cancel

### 1.3. Thiết kế

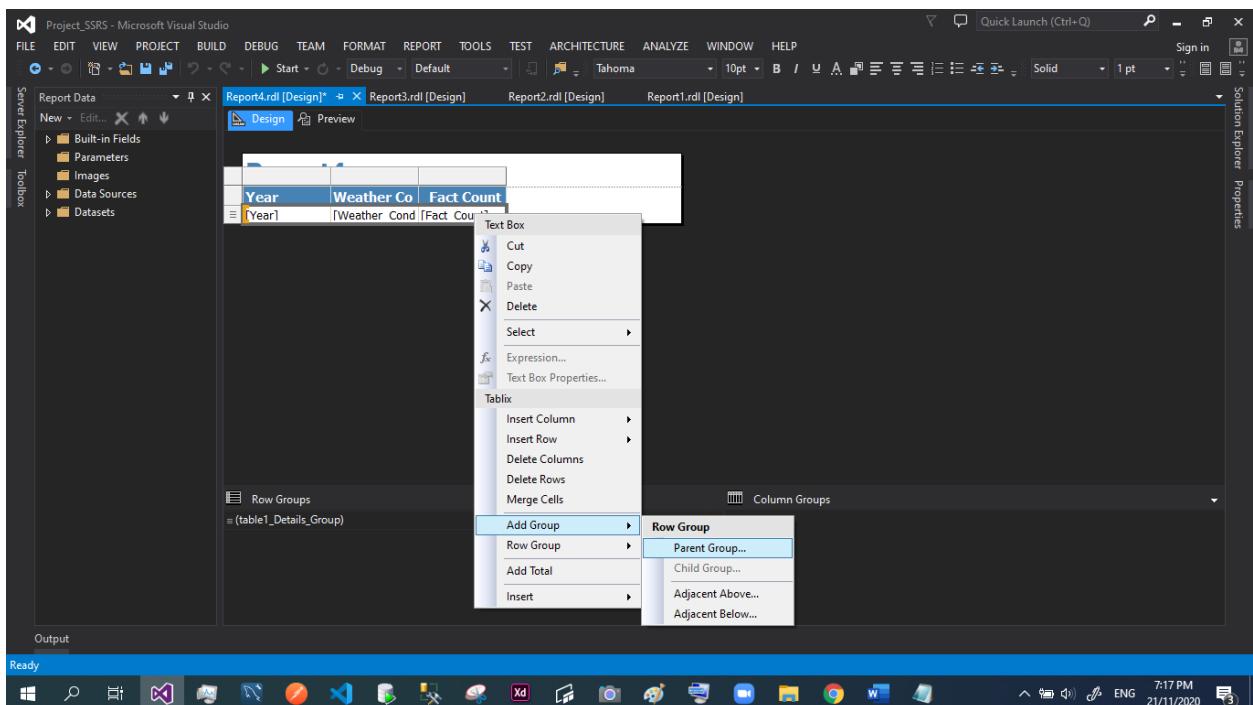
- Thiết kế cơ bản của một report



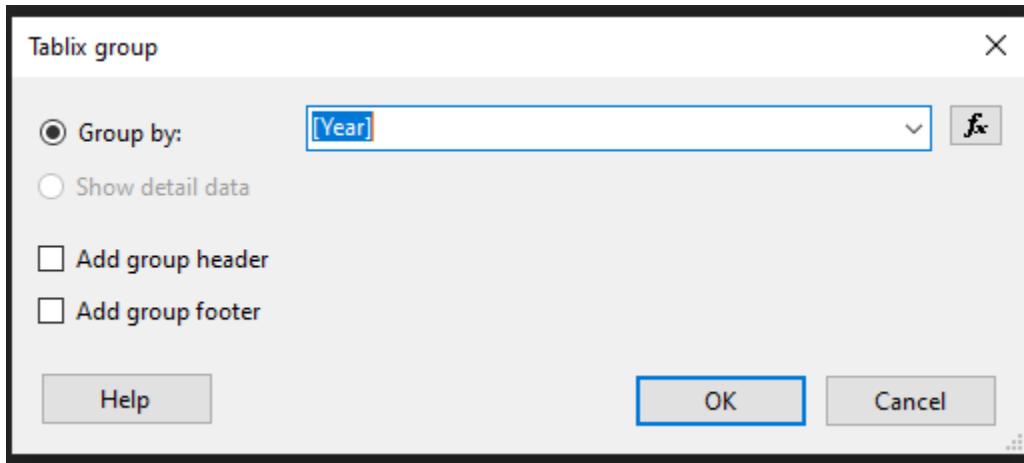
- Có thể chỉnh sửa bằng cách chọn đối tượng chỉnh sửa sau đó dùng các công cụ trên thanh công cụ để chỉnh sửa sao cho phù hợp với mong muốn.



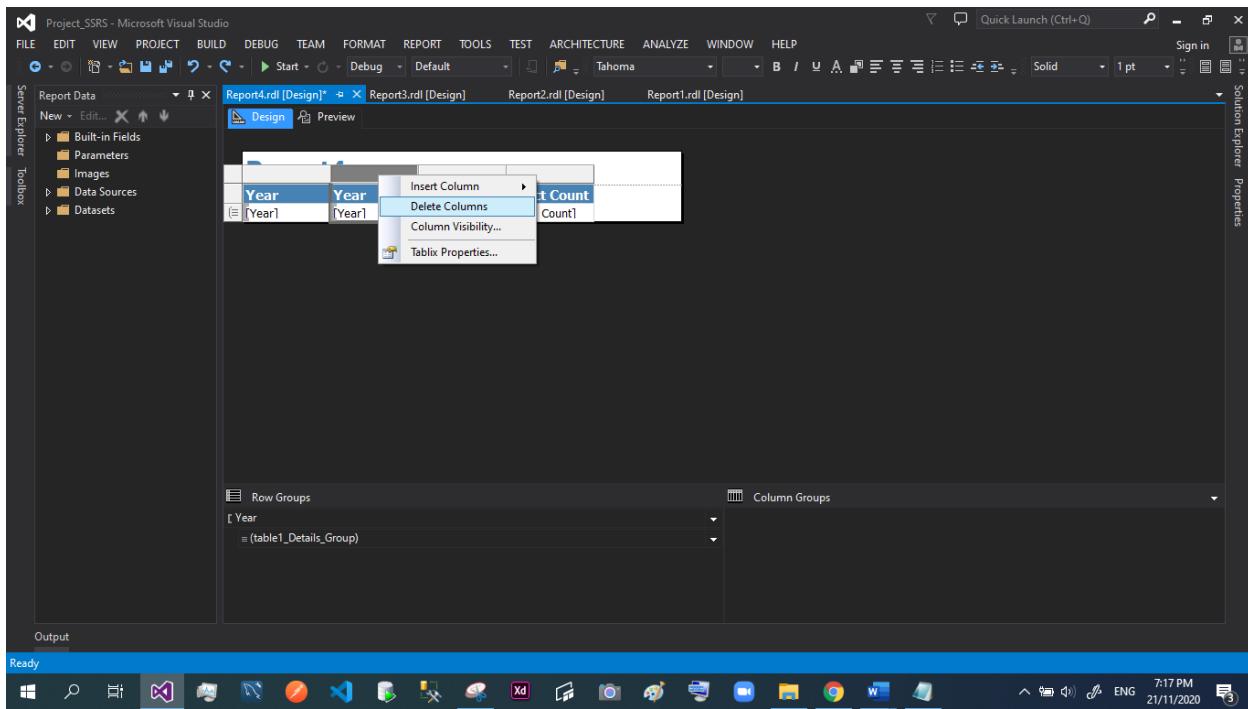
- Thao tác Add Group -> Parent Group dùng để gộp các row cùng giá trị lại với nhau.



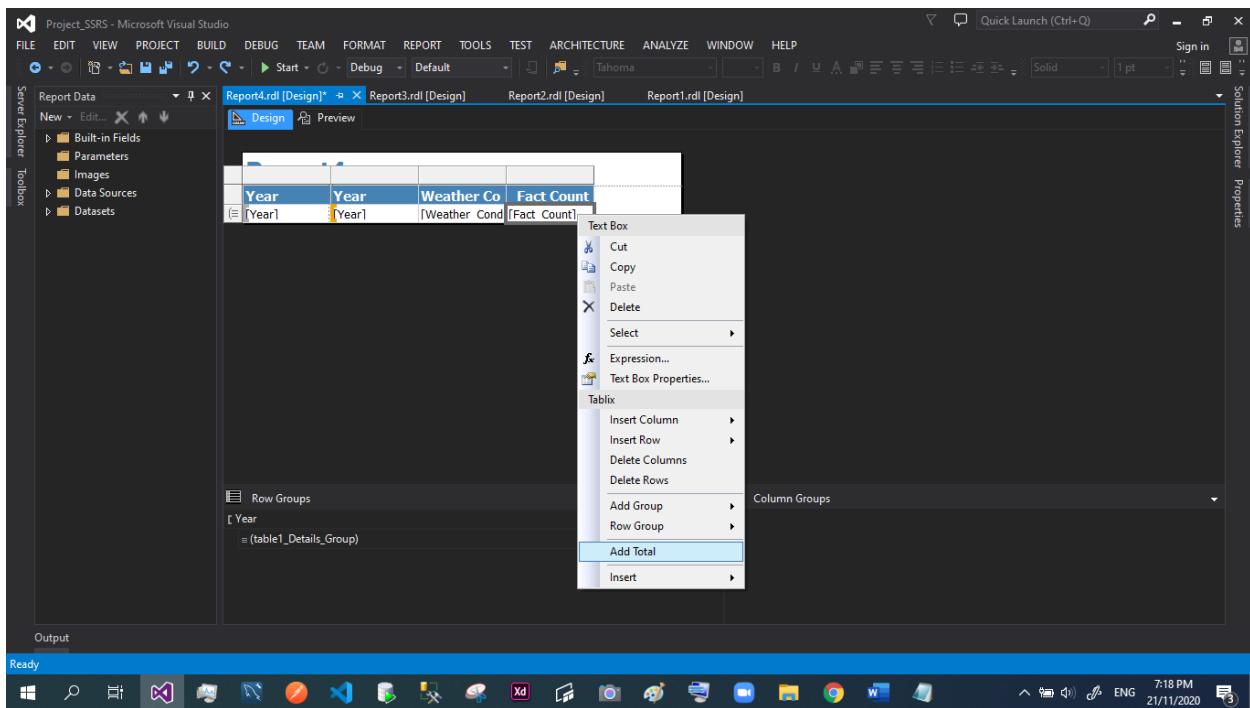
- Chọn Add group header hoặc Add group footer để xuất hiện dòng tổng (Cộng các độ đo trong group)



- Sau khi đã group xong thì nên xóa cột cũ trước khi group để đảm bảo tính thẩm mĩ



- Tính total và grand total của độ đo, ta có thể r-click vào ô thuộc tính cần tính total và nhấn Add total.



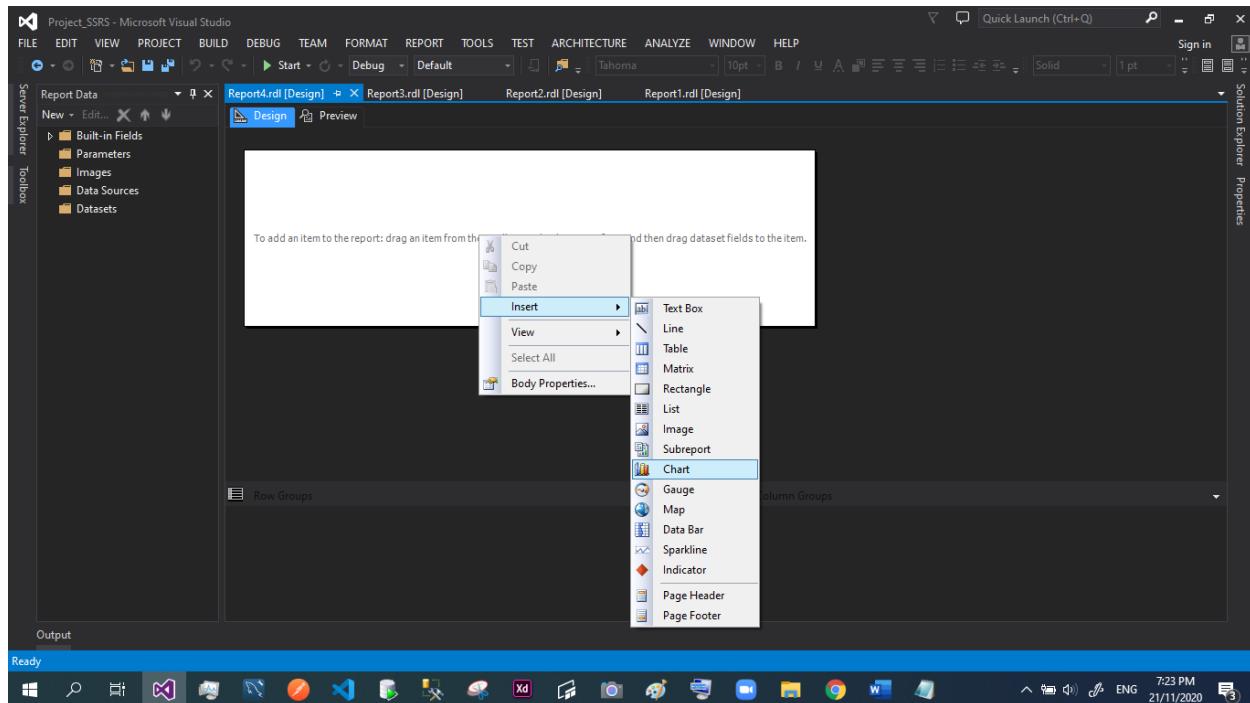
- Sau đó chọn Review để xem kết quả

Report4

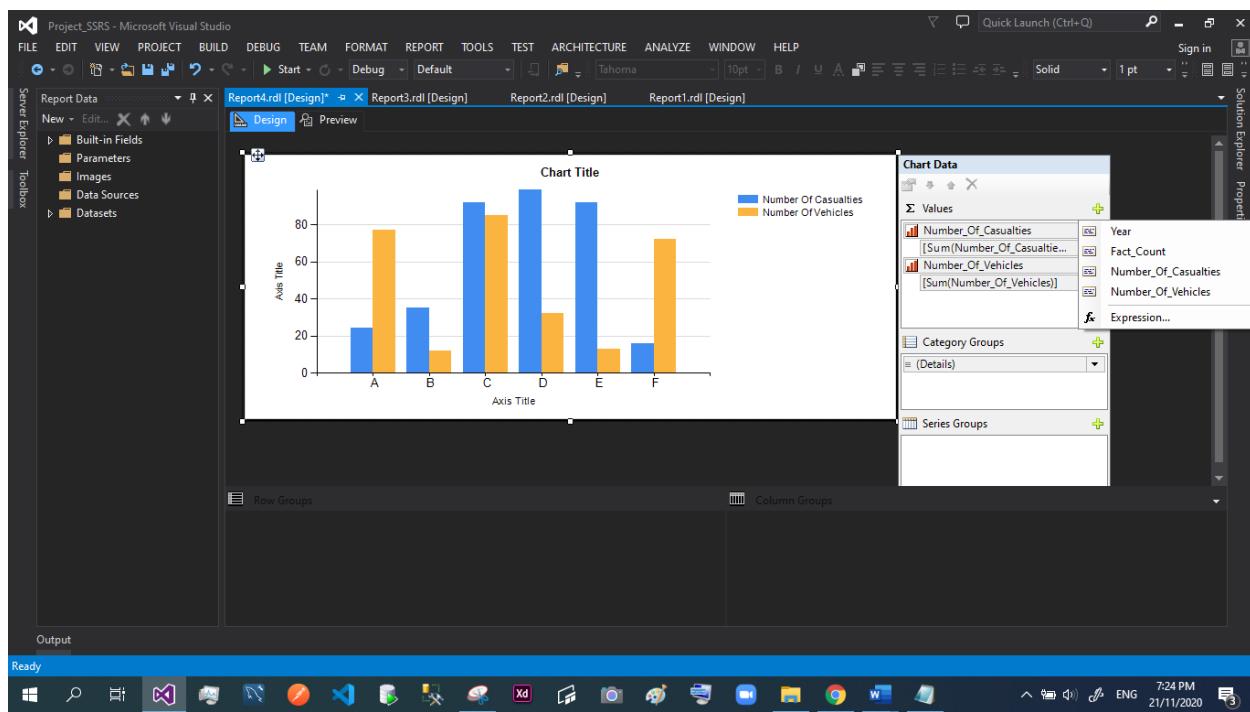
Year	Year	Weather Conditions	Fact Count
2005	2005	Fine with high winds	1308
	2005	Fine without high winds	93858
	2005	Fog or mist	566
	2005	Other	2920
	2005	Raining with high winds	1121
	2005	Raining without high winds	13048
	2005	Snowing with high winds	91
	2005	Snowing without high winds	803
	2005	Unknown	2033
			115748
2006	2006	Fine with high winds	1289
	2006	Fine without high winds	89221
	2006	Fog or mist	660

#### 1.4. Thiết kế Report dạng biểu đồ

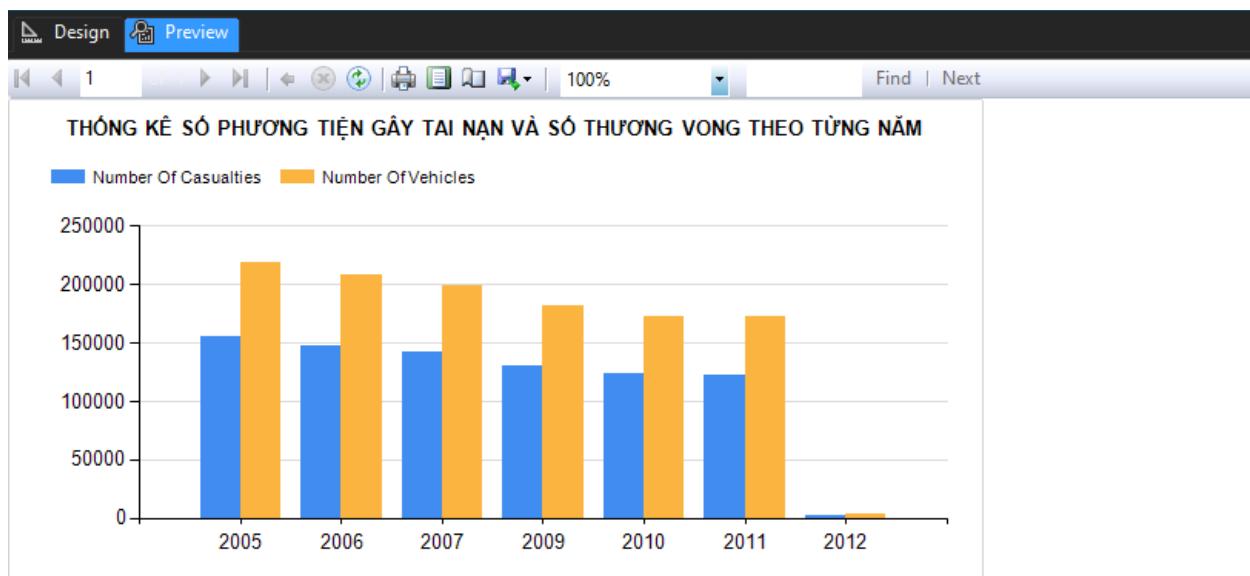
- Right click -> Insert-> Chart -> Chọn loại chart phù hợp



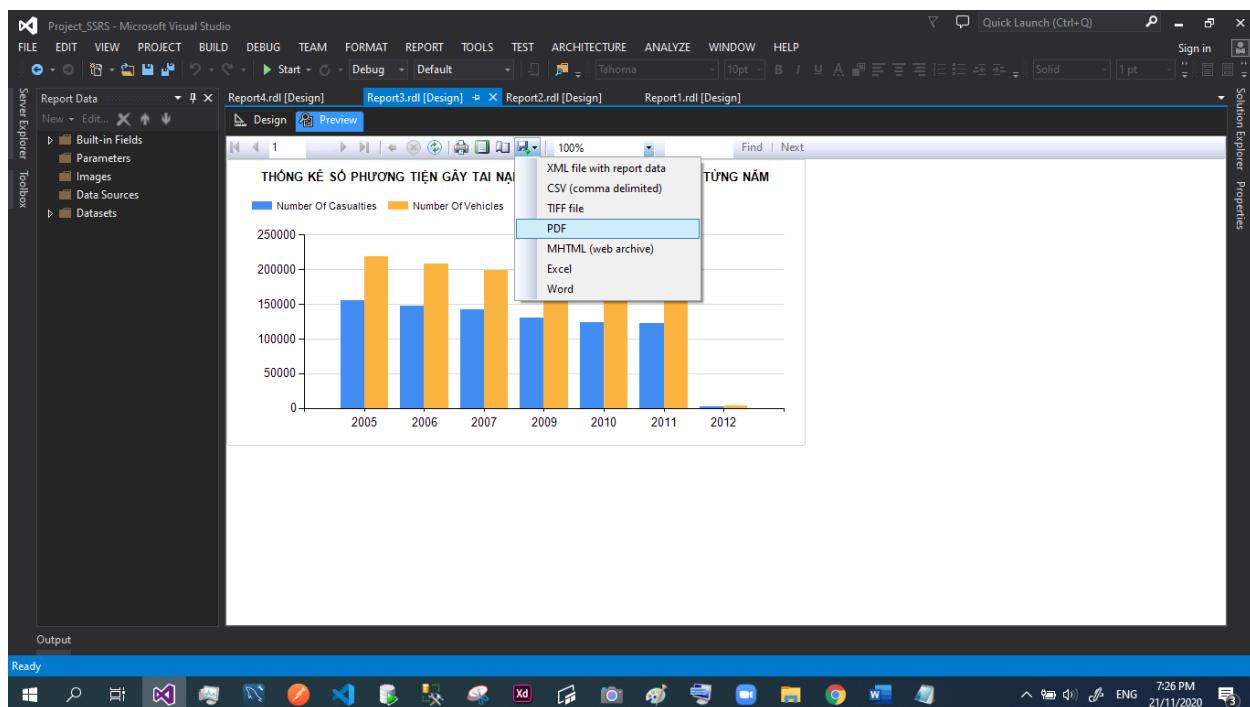
- Right click -> Click vào dấu + tại mục Value mục để thêm dữ liệu cho trực tung.
- Right click -> Click vào dấu + tại mục Category Group là dữ liệu trực hoành.



- Thiết kế các mục cần thiết rồi chọn Review để xem kết quả



- Có thể xuất báo cáo ra các định dạng



### 1.5. Tạo các Report

- Report 1: Thống kê số vụ tai nạn trong năm theo điều kiện thời tiết

**THỐNG KÊ SỐ VỤ TAI NẠN TRONG NĂM THEO ĐIỀU KIỆN THỜI TIẾT**

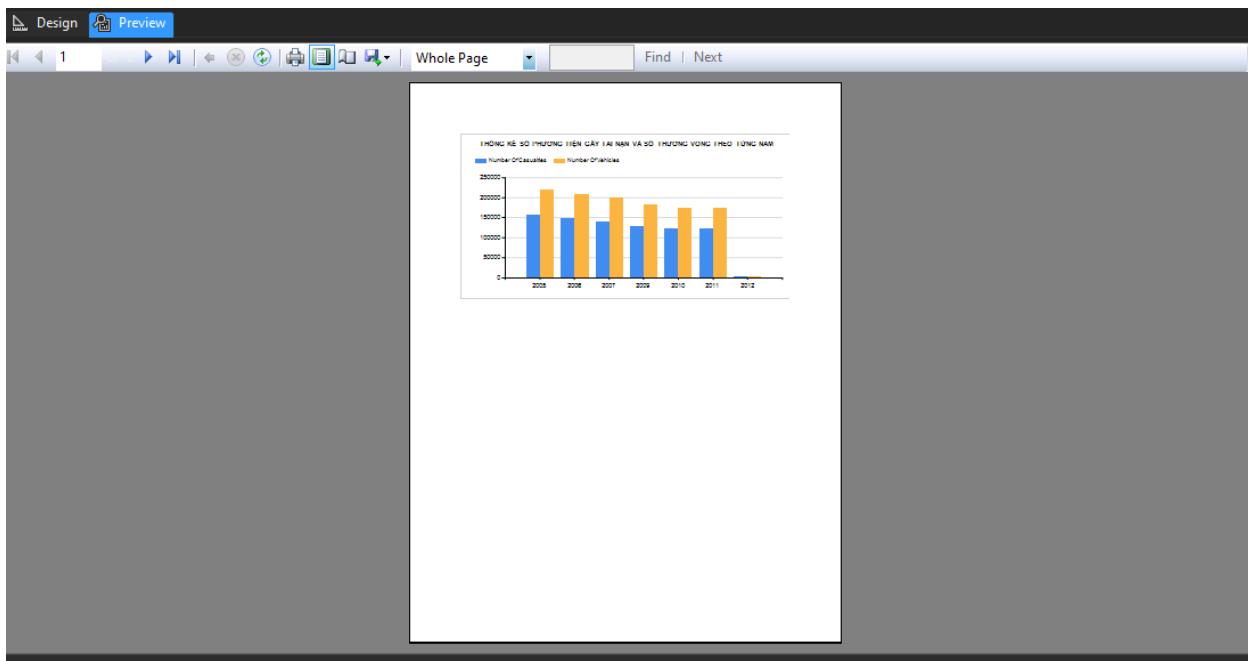
Year	Weather Conditions	Fact Count
2005	Pine with high winds	1308
	Pine without high winds	93858
	Rain or mist	856
	Other	2929
	Raining with high winds	1121
	Raining without high winds	13048
	Snowing with high winds	91
	Snowing without high winds	803
	Unknown	2033
	Wind only	13548
2006	Pine with high winds	1209
	Pine without high winds	95221
	Rain or mist	860
	Other	2320
	Raining with high winds	1436
	Raining without high winds	12683
	Snowing with high winds	37
	Snowing without high winds	296
	Unknown	1955
	Wind only	132277
2007	Pine with high winds	1141
	Pine without high winds	84477
	Rain or mist	394
	Other	2304
	Raining with high winds	1443
	Raining without high winds	12888
	Snowing with high winds	41
	Snowing without high winds	321
	Unknown	2265
	Wind only	105440
2008	Pine with high winds	1117
	Pine without high winds	77328
	Rain or mist	380
	Other	2370
	Raining with high winds	1236
	Wind only	130905

- Report 2 : Thống kê số phương tiện gây tai nạn và số thương vong theo điều kiện ánh sáng theo từng năm

**THỐNG KÊ SỐ PHƯƠNG TIỆN GÂY TAI NẠN VÀ SỐ THƯƠNG VONG THEO ĐIỀU KIỆN ÁNH SÁNG CỦA TỪNG NĂM**

Year	Light Conditions	Number Of Casualties	Number Of Vehicles
2005	Darkness: No street lighting	4248	4710
	Darkness: Street lighting unknown	1015	1452
	Darkness: Street lights present and lit	8070	4741
	Darkness: Street lights present but unlit	787	1597
	Daylight: Street light present	113647	137924
	Daylight: Street light absent	155454	219805
2006	Darkness: No street lighting	3655	4317
	Darkness: Street lighting unknown	932	1326
	Darkness: Street lights present and lit	8030	46030
	Darkness: Street lights present but unlit	554	815
	Daylight: Street light present	107594	159631
	Daylight: Street light absent	148012	208184
2007	Darkness: No street lighting	3631	4215
	Darkness: Street lighting unknown	1220	1685
	Darkness: Street lights present and lit	8078	42127
	Darkness: Street lights present but unlit	530	798
	Daylight: Street light present	104189	150149
	Daylight: Street light absent	141595	198328
2008	Darkness: No street lighting	3143	3551
	Darkness: Street lighting unknown	1171	1603
	Darkness: Street lights present and lit	24047	30197
	Darkness: Street lights present but unlit	490	865
	Daylight: Street light present	96439	137271
	Daylight: Street light absent	120709	162065
2009	Darkness: No street lighting	2757	3094
	Darkness: Street lighting unknown	921	1265
	Darkness: Street lights present and lit	26368	3483
	Darkness: Street lights present but unlit	398	528
	Daylight: Street light present	92756	132628
	Daylight: Street light absent	123204	172473
2010	Darkness: No street lighting	2859	3218
	Darkness: Street lighting unknown	1212	1667
	Darkness: Street lights present and lit	26686	35856

- Report 3: Thống kê số phương tiện gây tai nạn và số thương vong theo từng năm

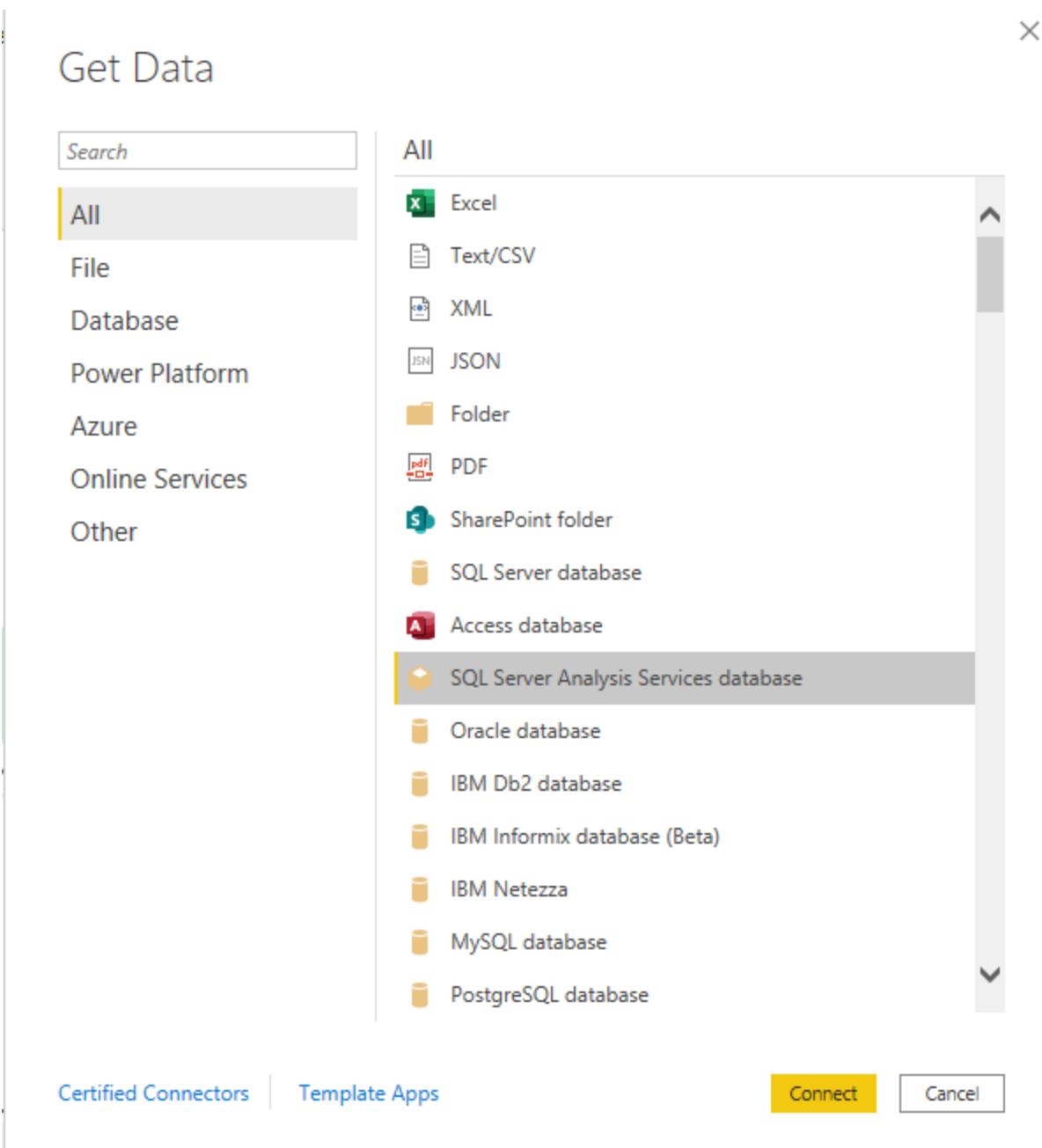


## 2. Tạo báo cáo với Power BI

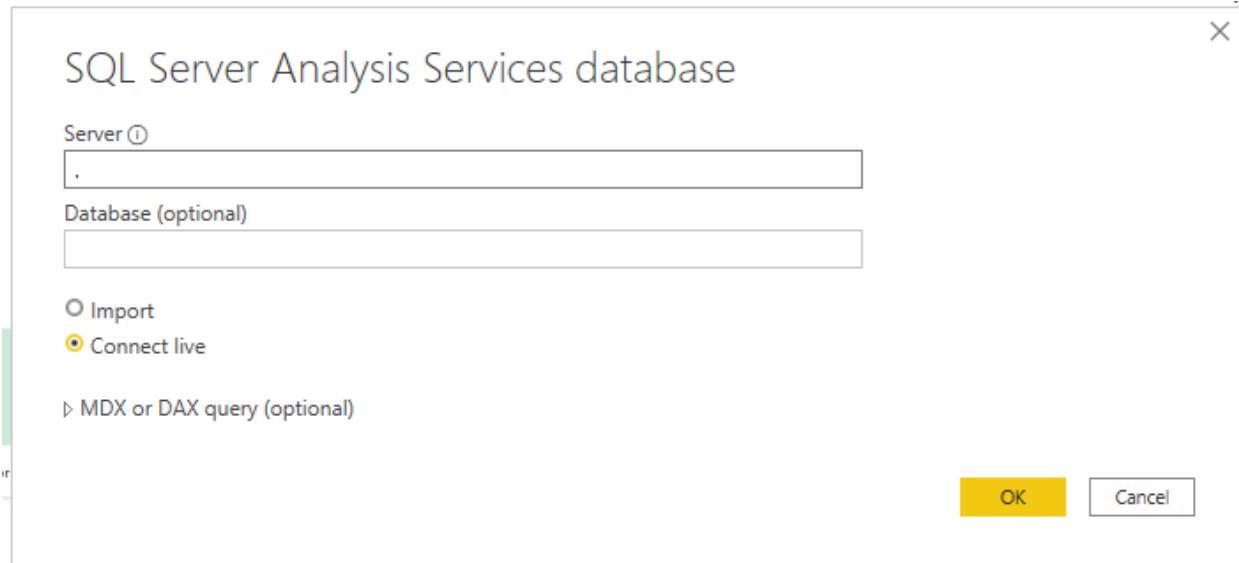
### 2.1. Kết nối nguồn dữ liệu

- Chọn Get Data -> Analysis Services

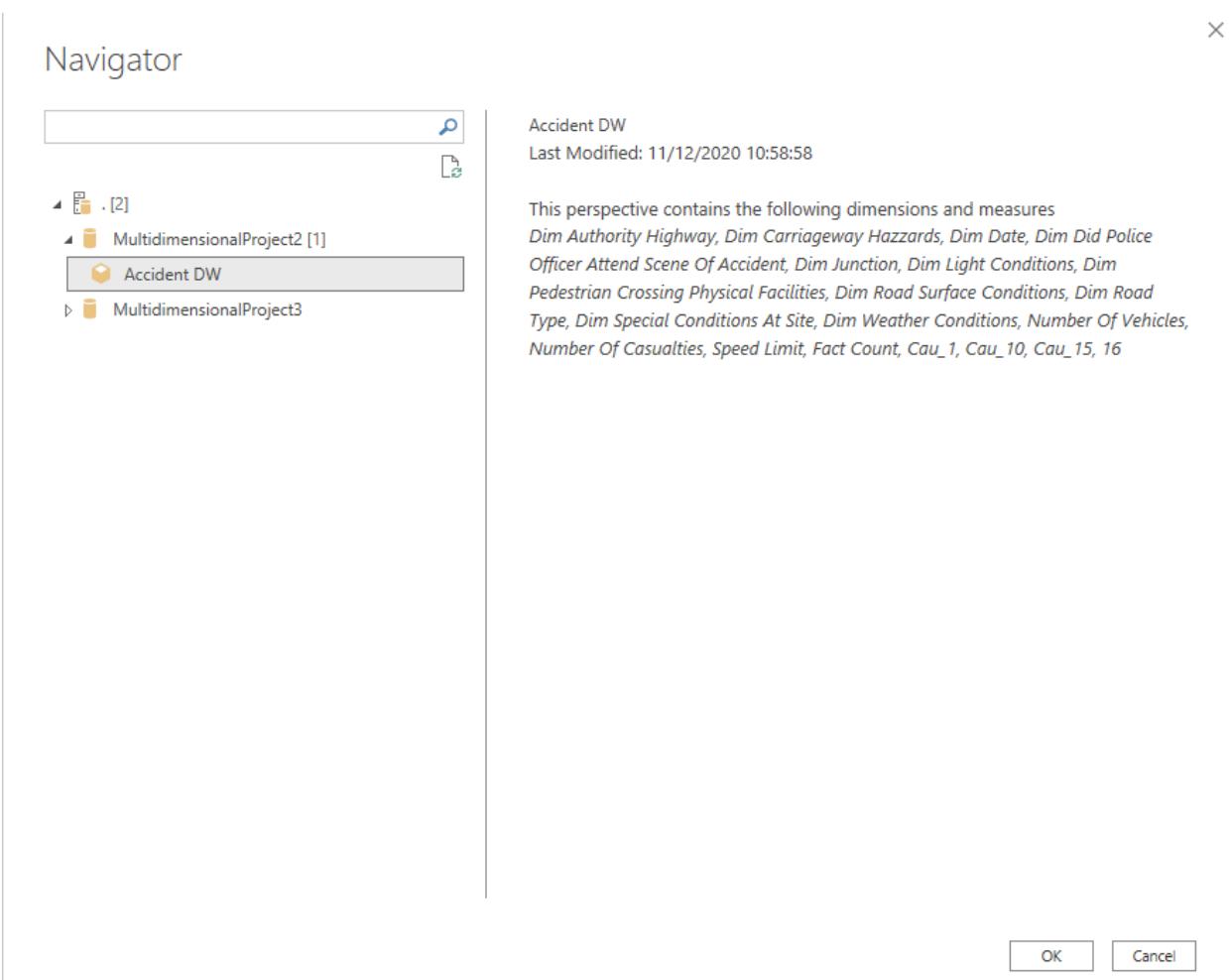
Name	Opened
tet2.pbix	22 minutes ago
tet.pbix	23 hours ago



- Điền tên server của mình và nhấn OK

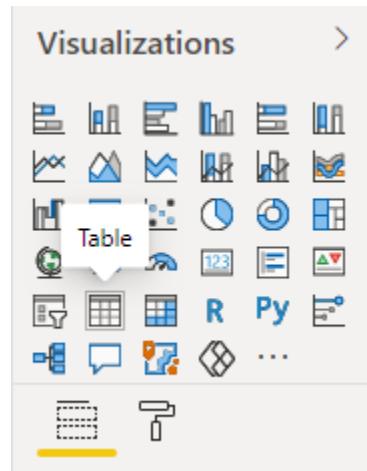


- Chọn Project SSAS cần tạo report và nhấn OK



## 2.2. Tạo mới một table

- Nhấn chọn Table trong hộp thoại Visualizations



- Chọn các trường thích hợp trong mục Field để thực hiện truy vấn

The screenshot shows the 'Fields' pane in Power BI. It contains a search bar and a list of fields categorized by table. The 'Fact' table has fields: Cau\_1, Fact Count (selected), Number Of ..., Number Of ..., and Speed Limit. The 'Measures' table is collapsed. The 'Dim Authority Hi...' and 'Dim Carriageway ...' tables are collapsed. The 'Dim Date' table is expanded, showing fields: Day, Hierarchy, Id, Month, and Year (selected). Below these are collapsed tables: 'Dim Did Police Of...' and 'Dim Junction'. At the bottom, it says 'Live connection: Connected'.

Table	Field	Status
Fact	Cau_1	
	Fact Count	Selected
	Number Of ...	
	Number Of ...	
	Speed Limit	
Measures		
Dim Authority Hi...		
Dim Carriageway ...		
Dim Date	Day	
	Hierarchy	
	Id	
	Month	
	Year	Selected
	Dim Did Police Of...	
Dim Junction		

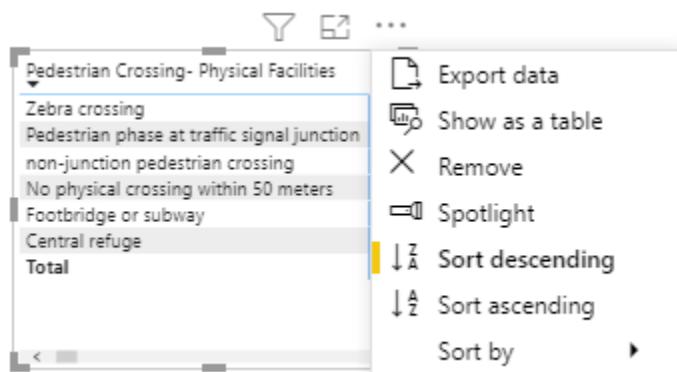
- Báo cáo kết quả hiển thị sau khi chọn các trường thích hợp

Back to report

Year	Pedestrian Crossing- Physical Facilities	Fact Count
2005	Central refuge	2424
2005	Footbridge or subway	365
2005	No physical crossing within 50 meters	92359
2005	non-junction pedestrian crossing	6661
2005	Pedestrian phase at traffic signal junction	10636
2005	Zebra crossing	3303
2006	Central refuge	2161
2006	Footbridge or subway	308
2006	No physical crossing within 50 meters	87617
2006	non-junction pedestrian crossing	6501
2006	Pedestrian phase at traffic signal junction	10441
2006	Zebra crossing	3149
2007	Central refuge	2228
2007	Footbridge or subway	358
2007	No physical crossing within 50 meters	84385
2007	non-junction pedestrian crossing	6340
2007	Pedestrian phase at traffic signal junction	9264
2007	Zebra crossing	2874
2009	Central refuge	2138
2009	Footbridge or subway	365
2009	No physical crossing within 50 meters	75104
<b>Total</b>		<b>615006</b>

of 1

- Click dấu 3 chấm ngang góc phải trên để có thể export report ra file định dạng .csv



### 2.3. Tạo các report

- Report 1: Thống kê số lượng vụ tai nạn theo từng loại điều kiện ánh sáng cho người tham gia giao thông theo từng năm

[Back to report](#)

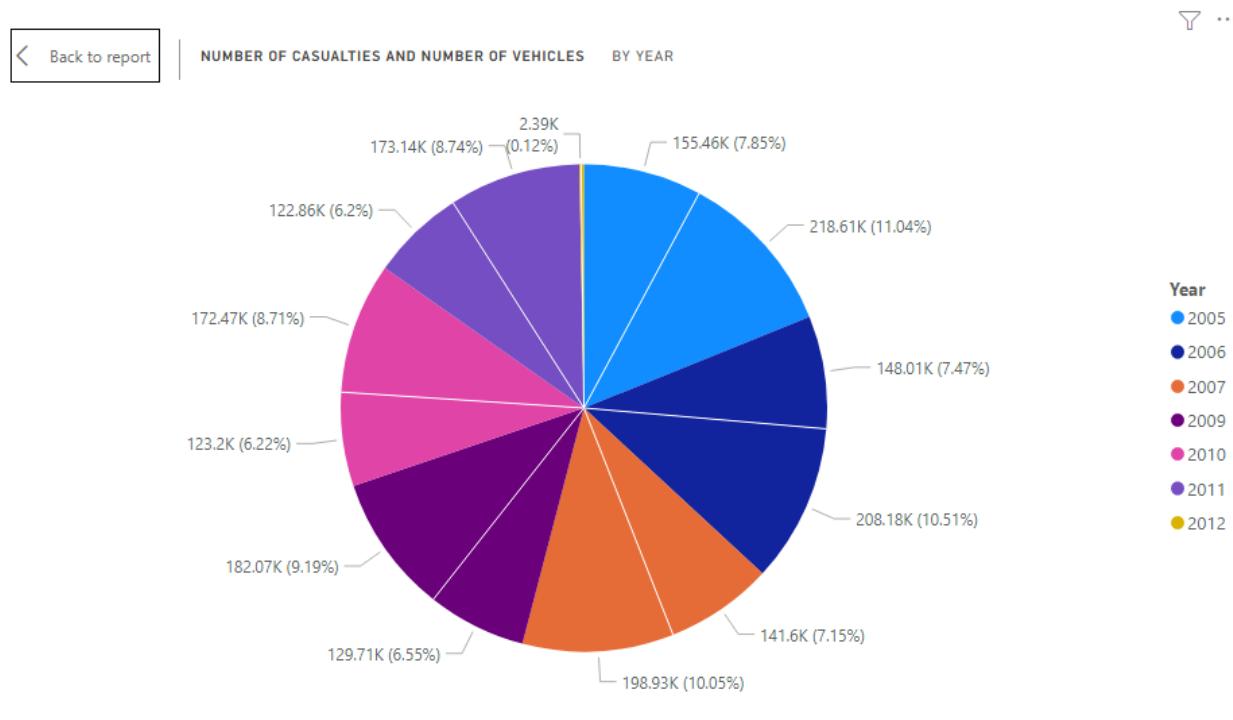
Light Conditions	2005	2006	2007	2009	2010	2011	2012	Total
Darkness: No street lighting	2592	2400	2313	1982	1734	1805	1	<b>12827</b>
Darkness: Street lighting unknown	813	718	922	886	707	910	9	<b>4965</b>
Darkness: Street lights present and lit	26086	25474	23268	21569	19430	19958	599	<b>136384</b>
Darkness: Street lights present but unlit	609	450	422	365	302	391	4	<b>2543</b>
Daylight: Street light present	85648	81135	78524	72076	69953	69545	1406	<b>458287</b>
<b>Total</b>	<b>115748</b>	<b>110177</b>	<b>105449</b>	<b>96878</b>	<b>92126</b>	<b>92609</b>	<b>2019</b>	<b>615006</b>

- Report 2: Thống kê số vụ tai nạn theo điều kiện thời tiết và điều kiện bề mặt đường

[Back to report](#)

Weather Conditions	Dry	Flood (Over 3cm of water)	Frost/Ice	Snow	Wet/Damp	Total
Fine with high winds	4402		4	72	5	2594 <b>7077</b>
Fine without high winds	422566		28	4969	539	70749 <b>498851</b>
Fog or mist	400			358	14	1970 <b>2742</b>
Other	2675		9	2201	233	8616 <b>13734</b>
Raining with high winds	74		92	32	12	6806 <b>7016</b>
Raining without high winds	470		120	181	75	69000 <b>69846</b>
Snowing with high winds	5		1	75	241	106 <b>428</b>
Snowing without high winds	46		1	728	2093	1166 <b>4034</b>
Unknown	9277		13	141	27	1820 <b>11278</b>
<b>Total</b>	<b>439915</b>		<b>268</b>	<b>8757</b>	<b>3239</b>	<b>162827</b> <b>615006</b>

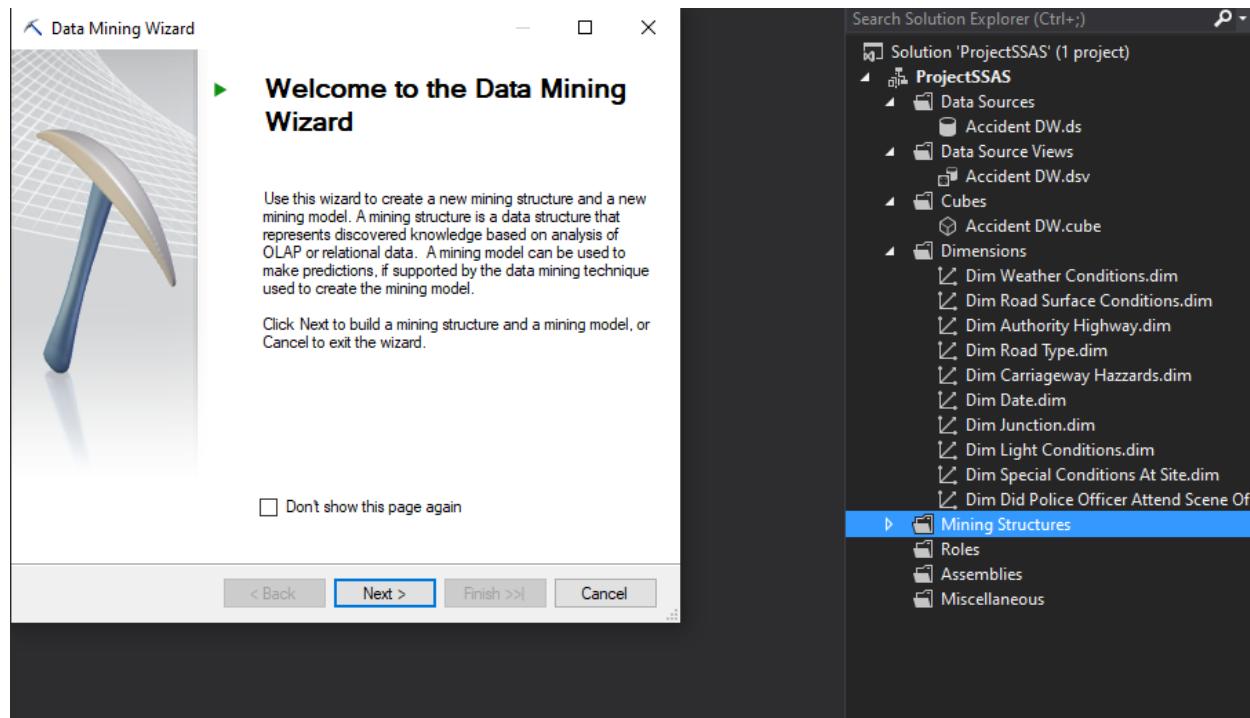
- Report 3: Thống kê số phương tiện gây tai nạn và số thương vong theo từng năm



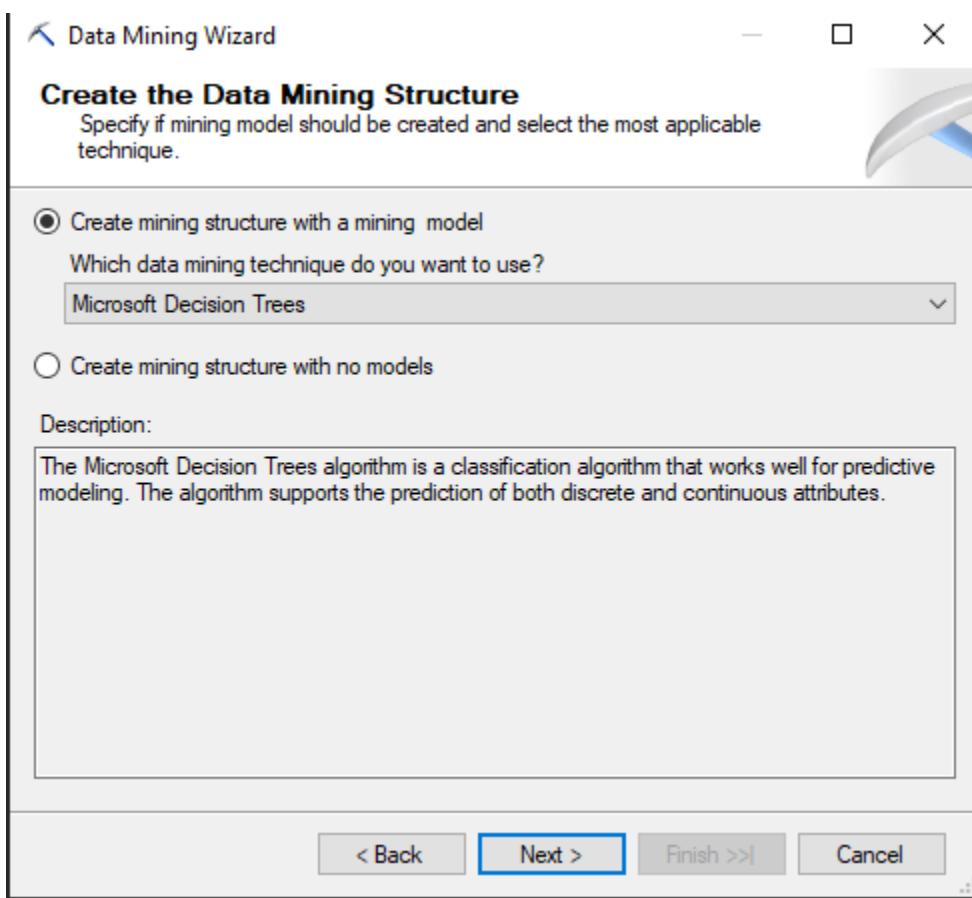
## CHƯƠNG 6: QUÁ TRÌNH DATA MINING

### 1. Quá trình khai phá dữ liệu

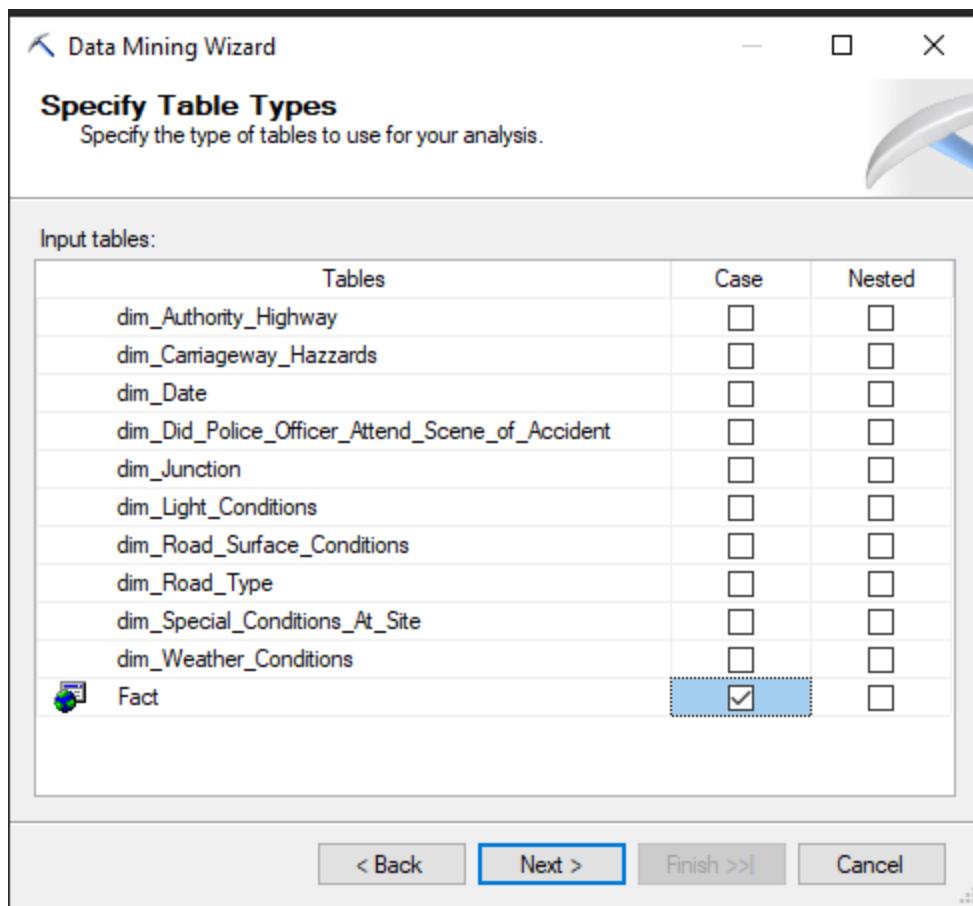
- Trong project SSAS, r-click vào Mining Structures và tạo mới file



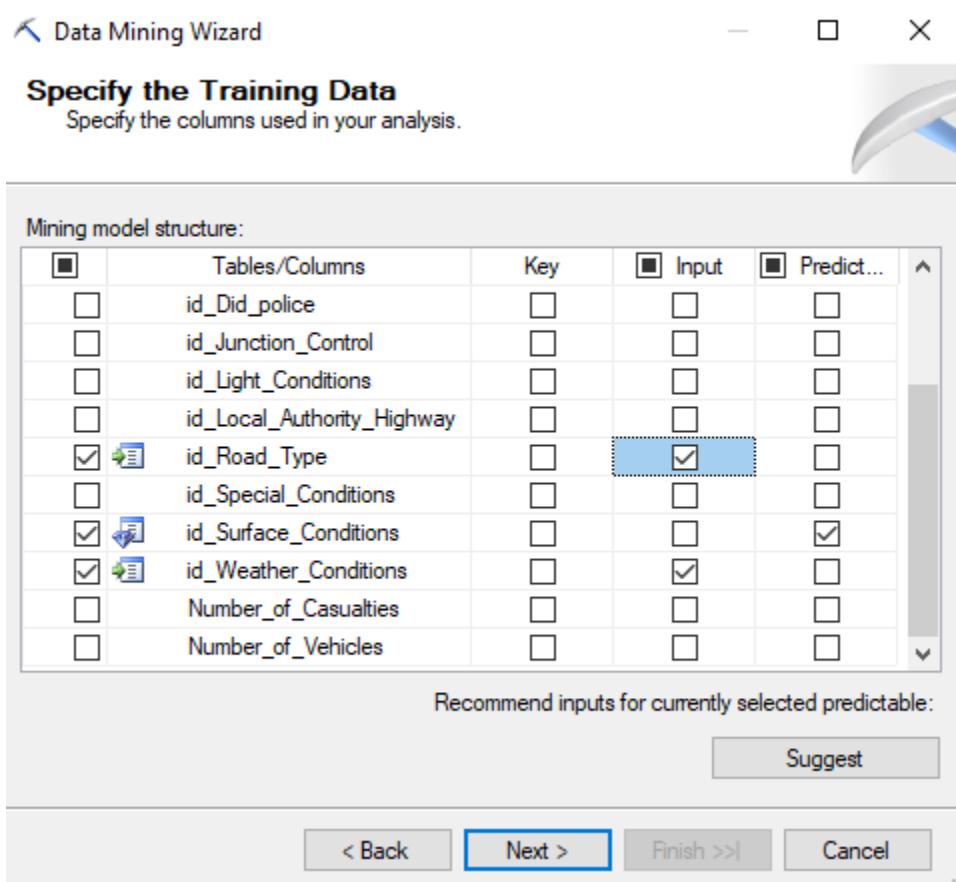
- Chọn cây quyết định- Microsoft Decision Trees để thực hiện mining.



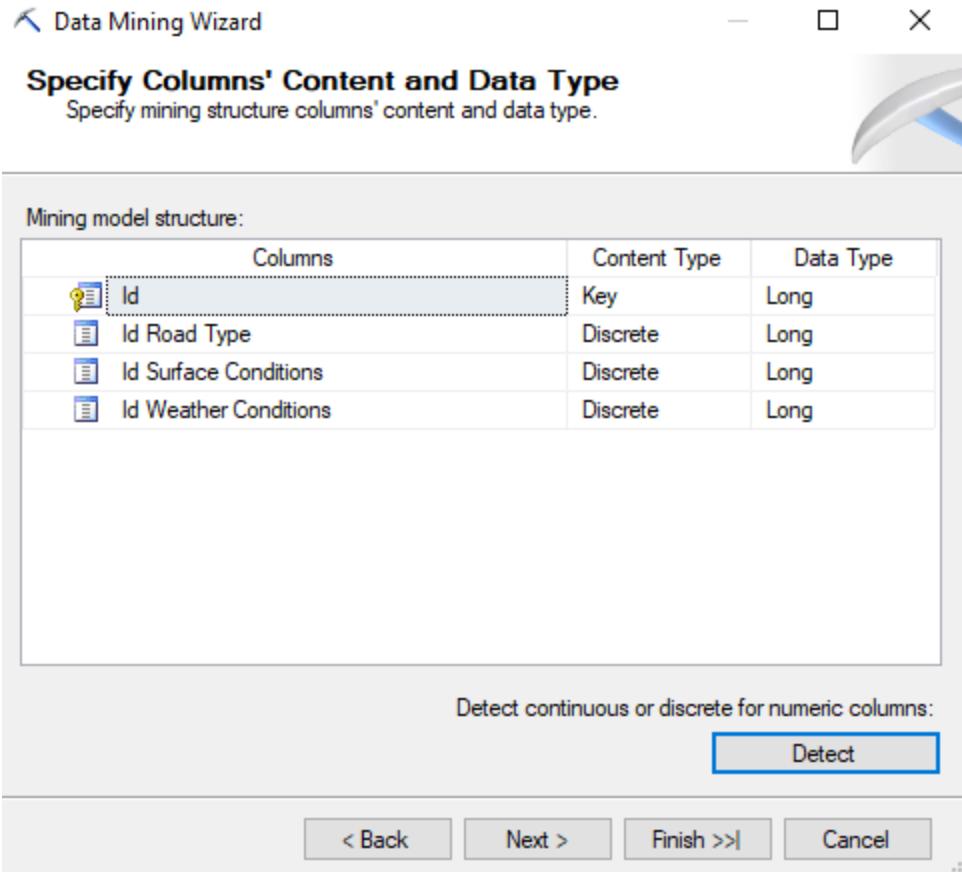
- Chọn bảng Fact để thực hiện mining



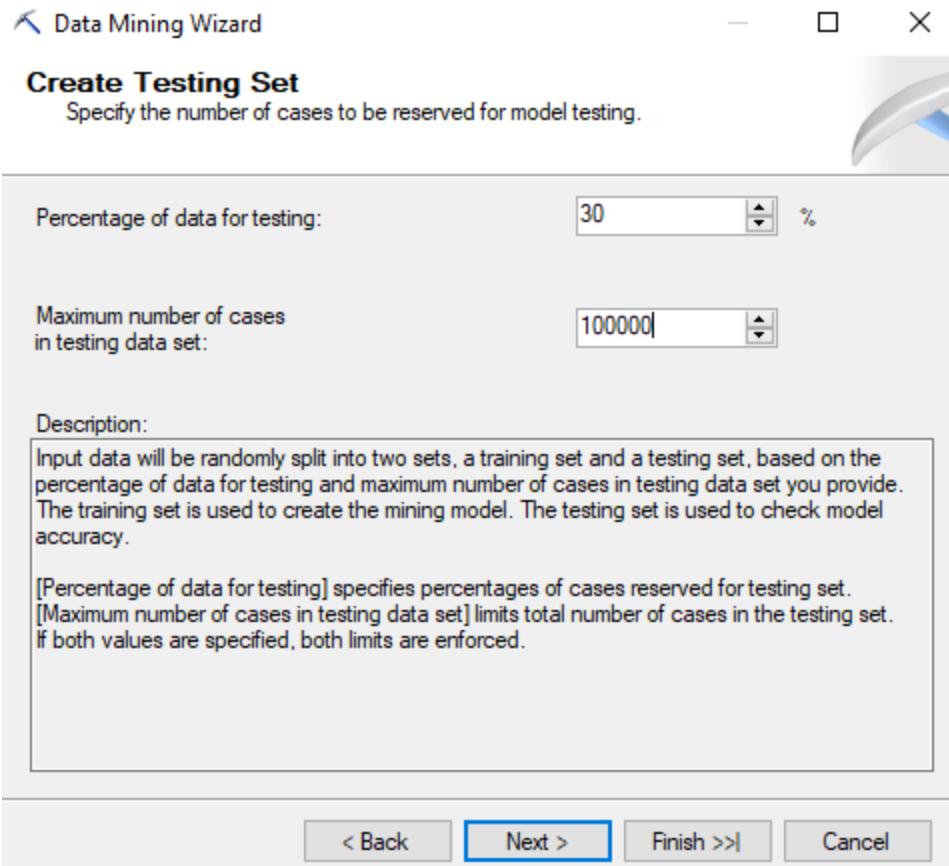
- Chọn thuộc tính quyết định để phân lớp dữ liệu bằng cách click vào Predictable, như hình bên dưới ta sẽ chọn thuộc tính quyết định phân lớp là Road surface conditions, sau đó nhấn vào suggest để được gợi ý chọn lựa các thuộc tính kèm theo.



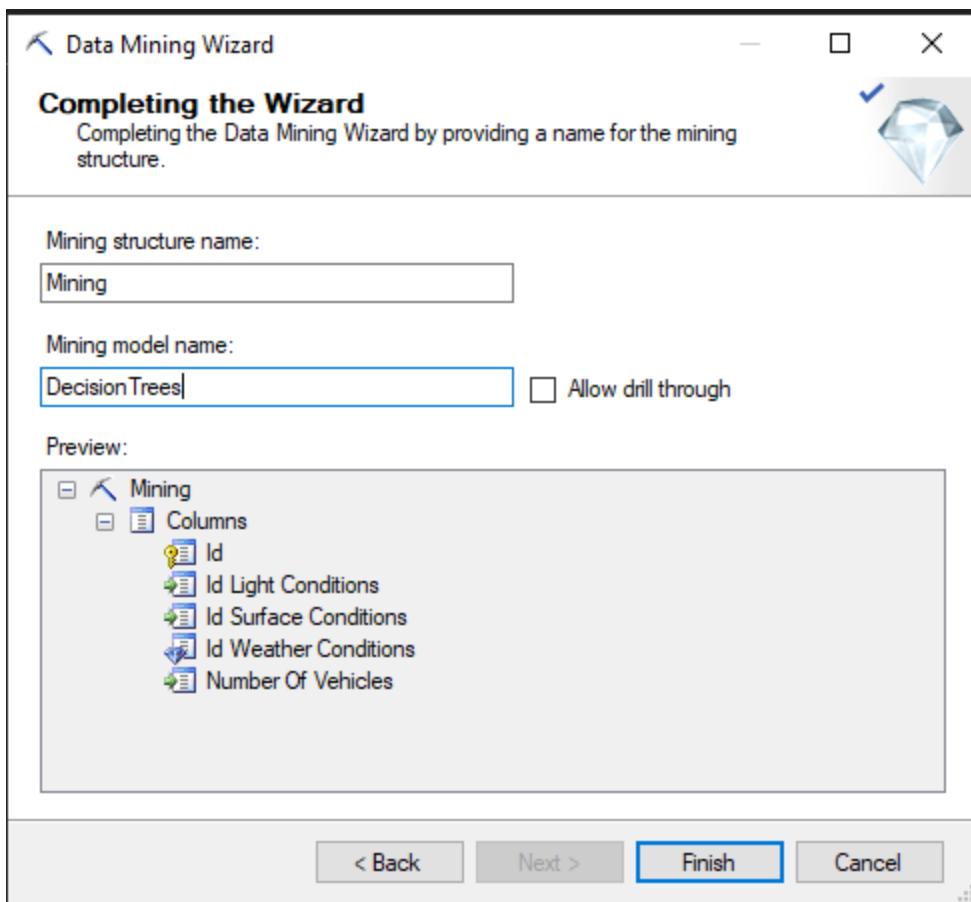
- Nhấn chọn detect để chuyển kiểu dữ liệu các cột cần thiết sang discrete (dữ liệu chắc chắn rời rạc)



- Chọn phân chia giữa tập test tập train và số trường hợp test



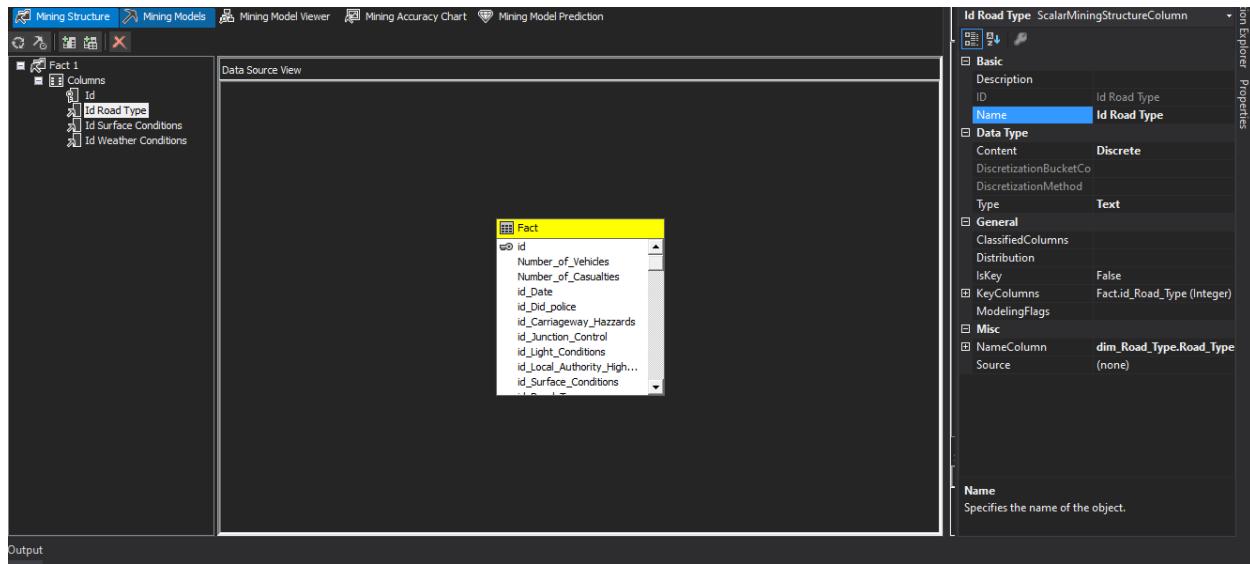
- Đặt tên cho model



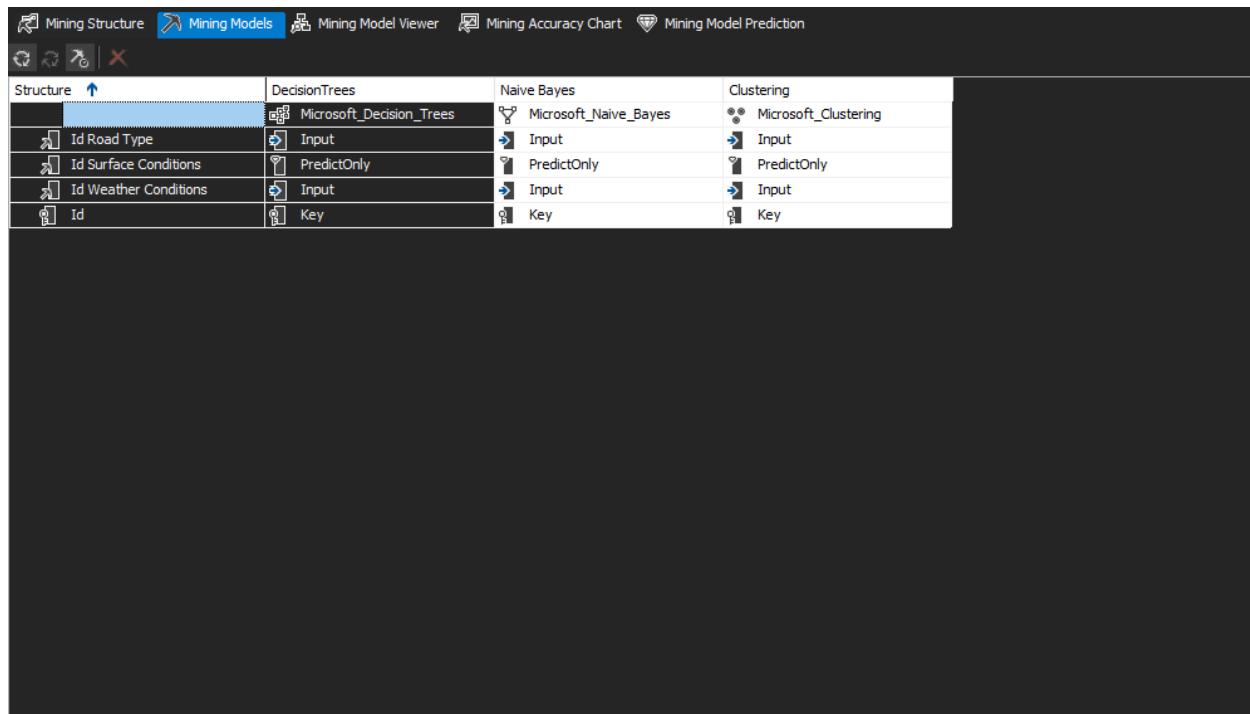
- Giao diện sau khi tạo hoàn thành

Fact 1 MiningStructure	
<input checked="" type="checkbox"/> Advanced	ErrorConfiguration (default) Language
<input checked="" type="checkbox"/> Basic	Description ID      Fact 1 Name    Fact 1
<input checked="" type="checkbox"/> Misc	CacheMode      KeepTrainingCases Collation HoldoutMaxCases    100000 HoldoutMaxPercent   30 HoldoutSeed        0 Source            Accident DW (Data source v)
Name Specifies the name of the object.	

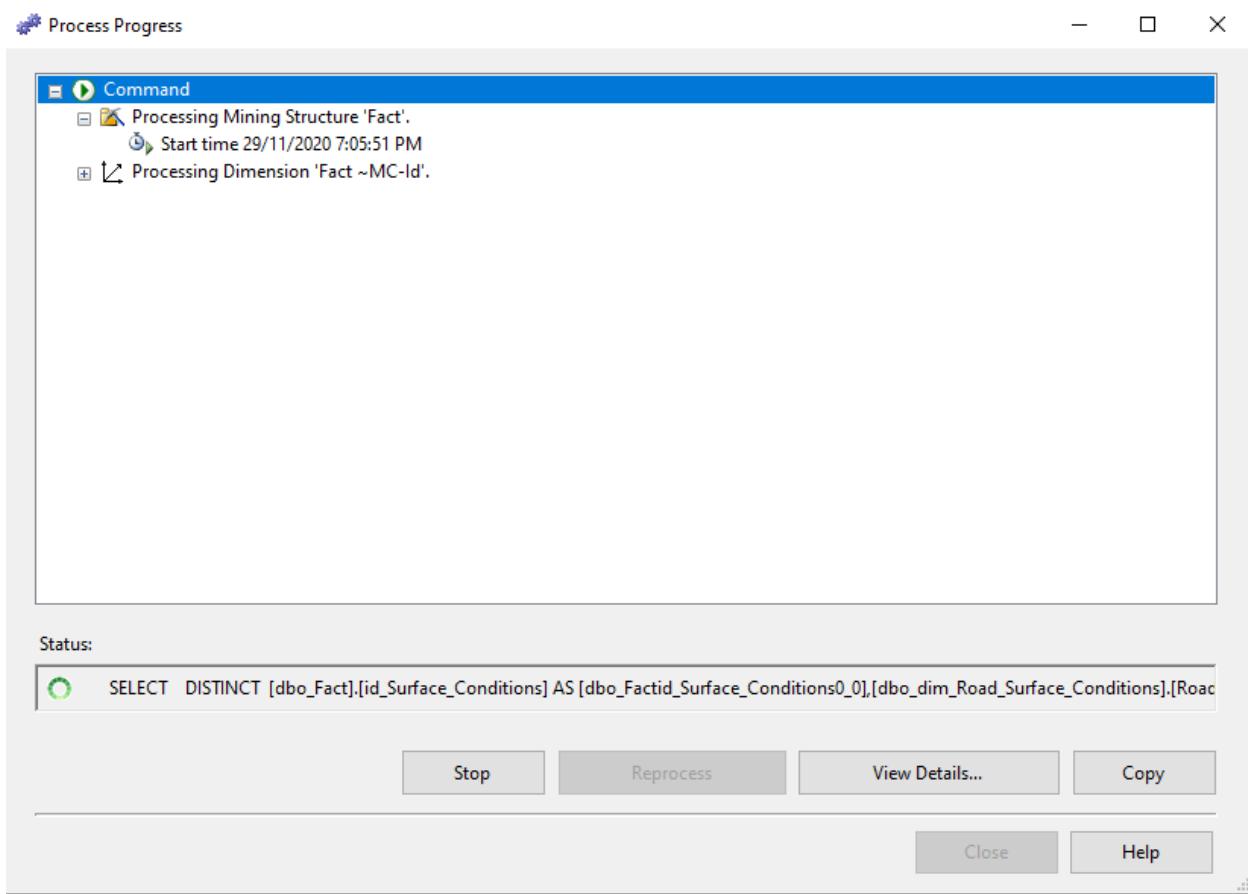
- Dữ liệu trong bảng Fact là các ID kiểu số, vì vậy ta cần đổi Type và NameColumn cho các thuộc tính.



- Chọn mục Mining model và thêm hai mô hình khác là Naïve Bayes và Gom cụm

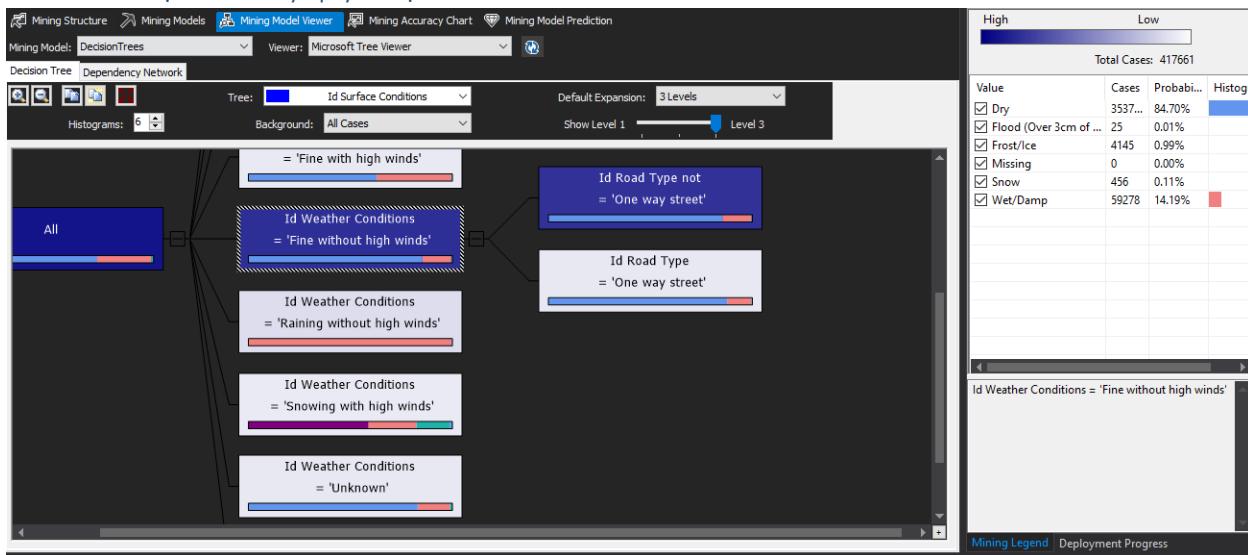


- Thực hiện quá trình khai phá dữ liệu

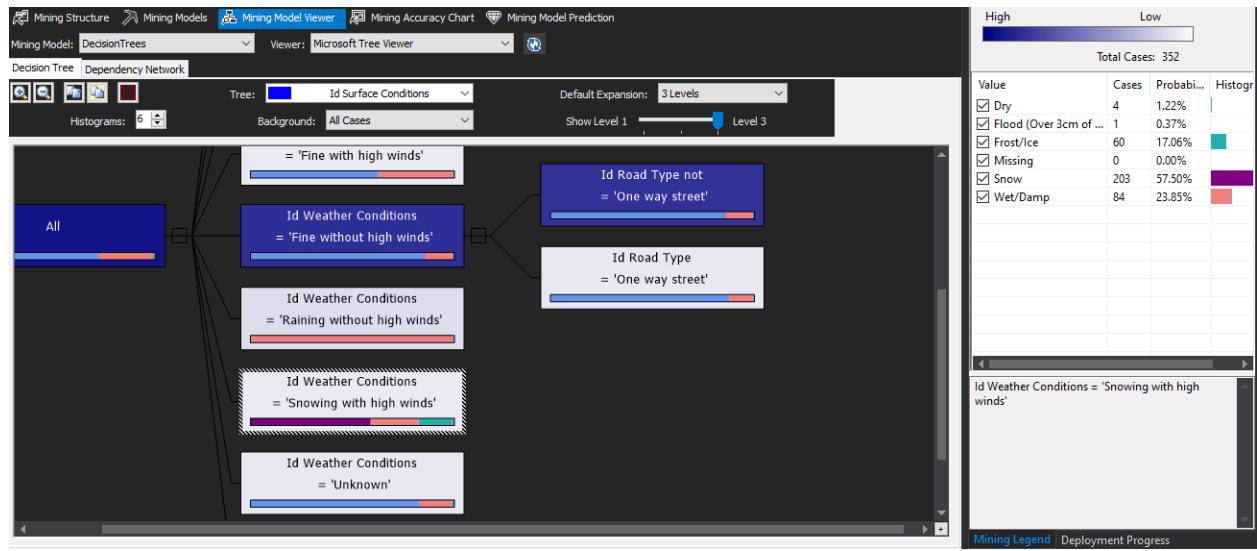


## 2. Tạo báo cáo với SSRS Project

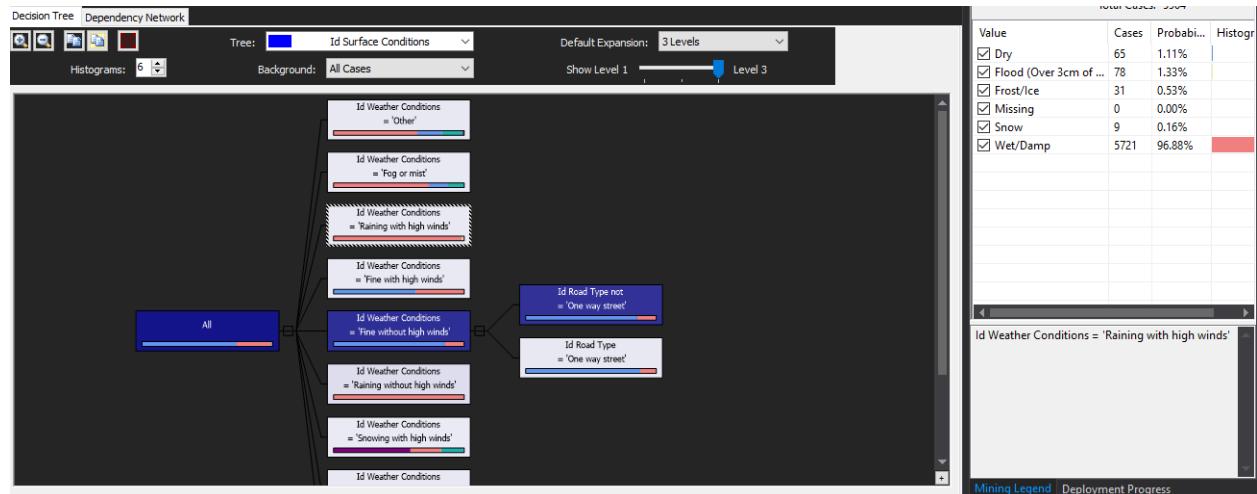
### 2.1. Thuật toán cây quyết định



- Kết luận: Có đến 84,70% số nạn nhân trong các vụ tai nạn với điều kiện thời tiết “Fine without high wind” và điều kiện mặt đường “Dry”



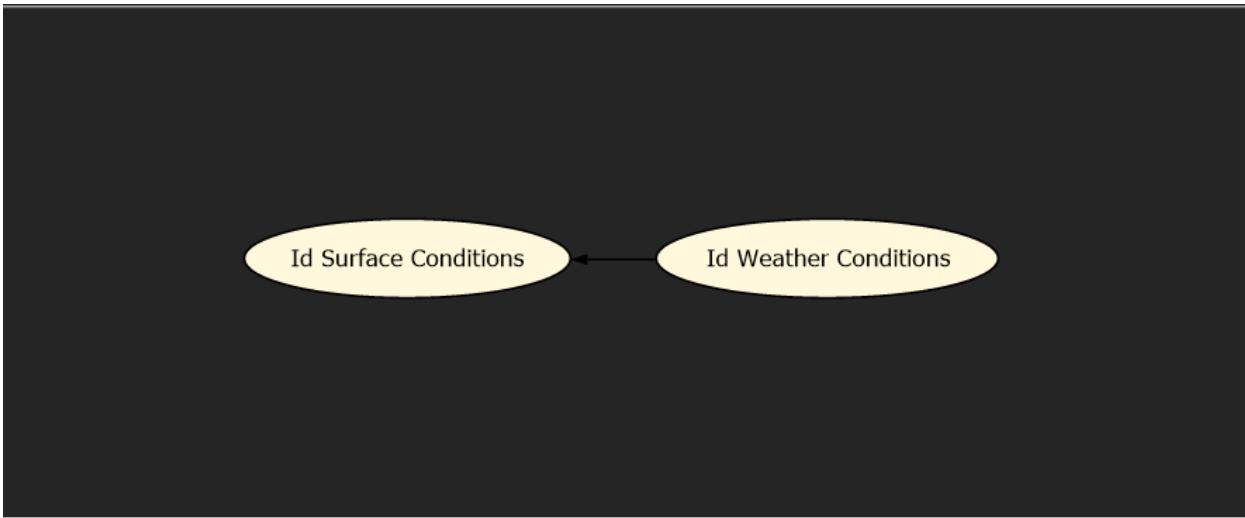
- Kết luận: Nếu thời tiết “Snow with highwind” hoặc “Snow without highwin” thì sẽ có khoảng hơn 50% xác xuất mặt đường bị ngập tuyết.



- Kết luận: Nếu thời tiết là “Raining with highwin” hoặc “Raining without highwind” thì gần như mặt đường sẽ bị trơn trượt.

## 2.2. Thuật toán Naïve bayes

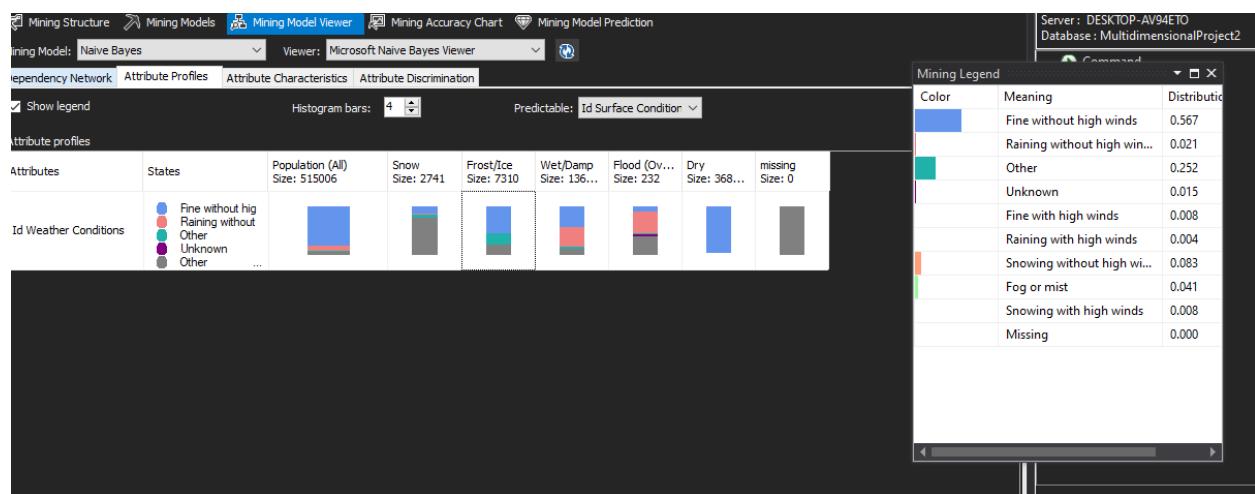
- Kết quả



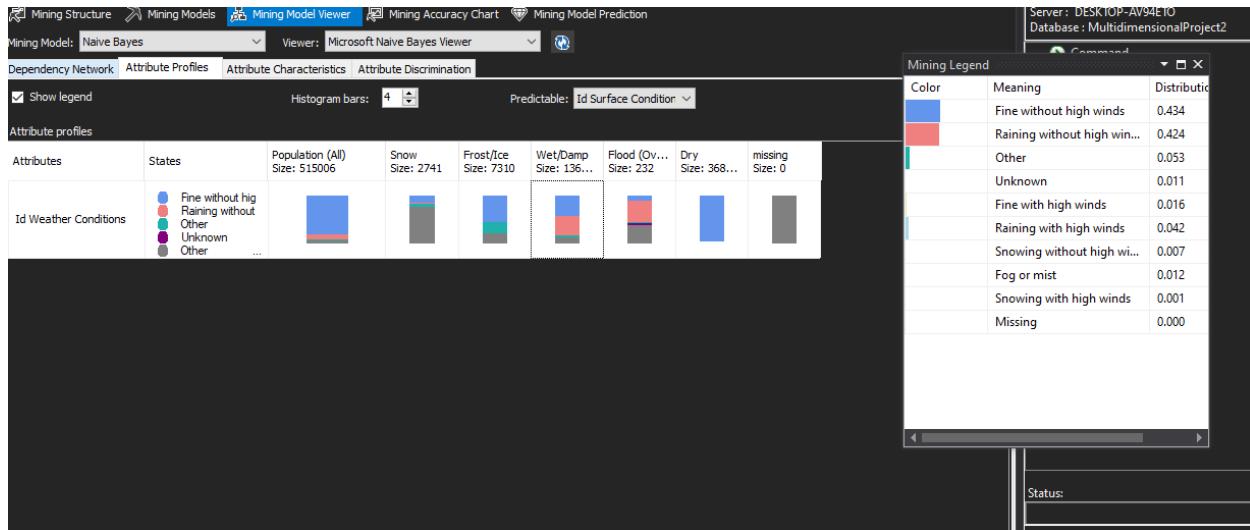
Ta thấy như trên thì điều kiện thời tiết ảnh hưởng trực tiếp đến điều kiện mặt đường



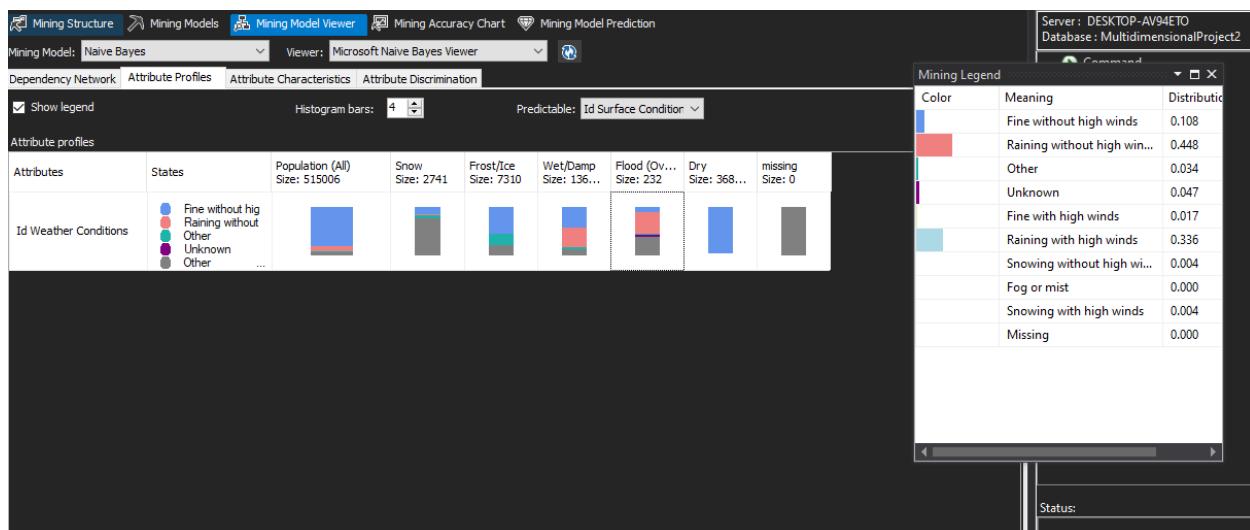
- Với điều kiện mặt đường là snow thì hầu như kết quả test thời tiết sẽ là Snow without high winds



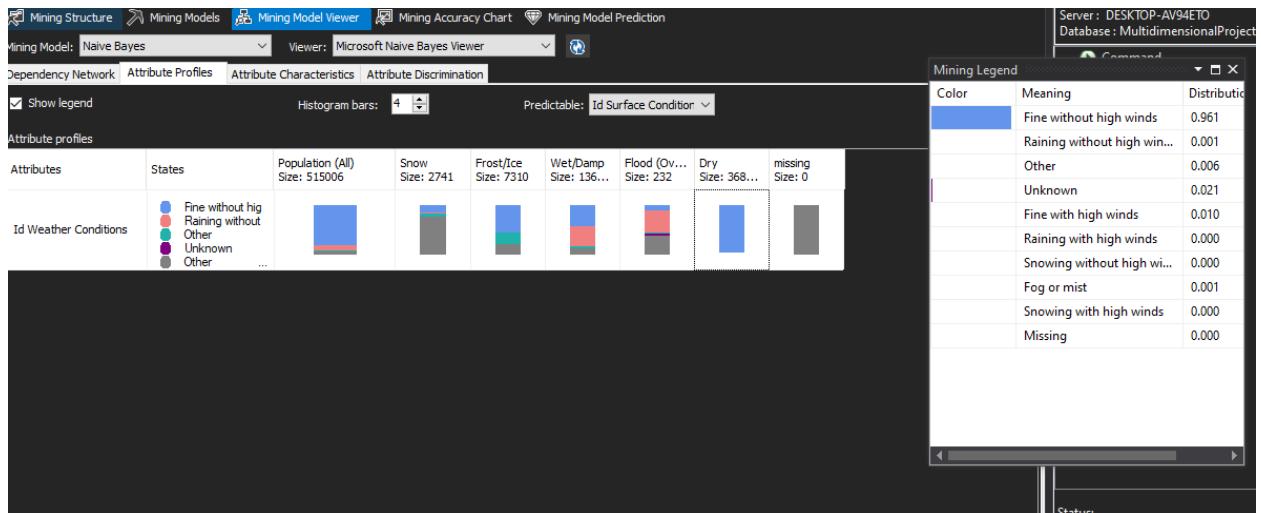
- Với điều kiện mặt đường là Frost/Ice kết quả của thời tiết test được phần lớn là Fine without high winds



- Với điều kiện mặt đường là wet/damp thì kết quả của thời tiết test được phần lớn là Fine without high winds và Raining without high winds



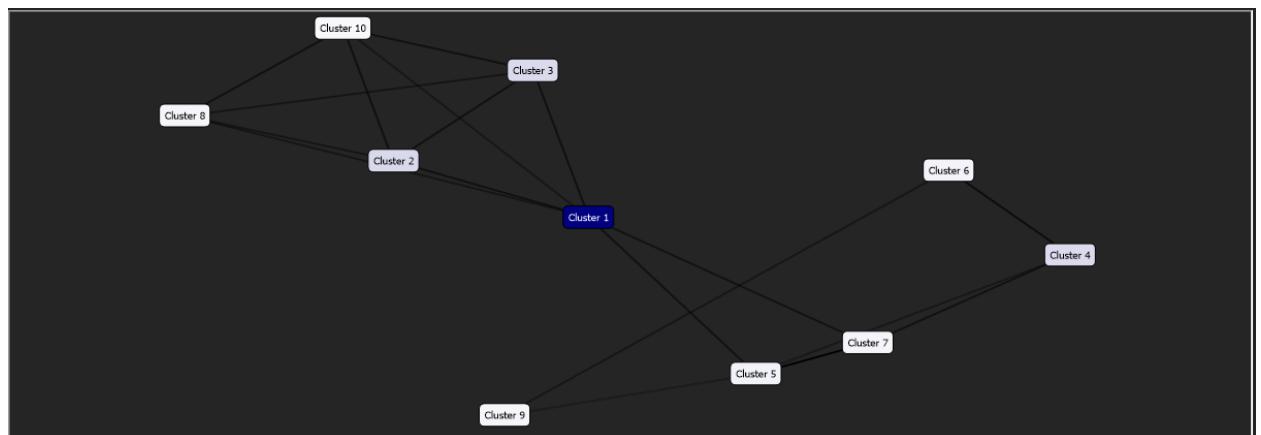
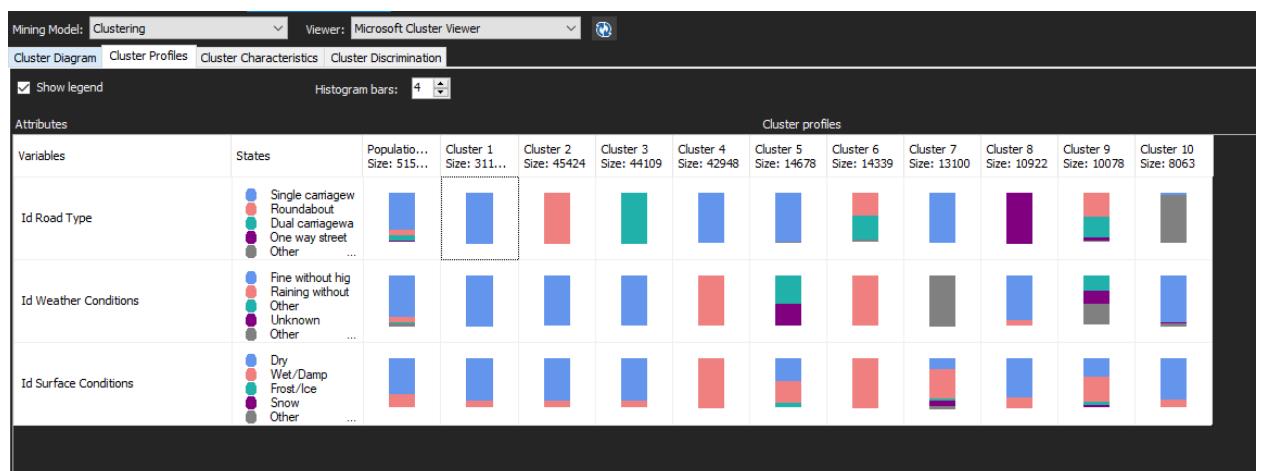
- Với điều kiện mặt đường là wet/damp thì kết quả của thời tiết test được phần lớn liên quan đến Raining



- Với điều kiện mặt đường là wet/damp thì kết quả của thời tiết test được phân lớn là Fine without high winds

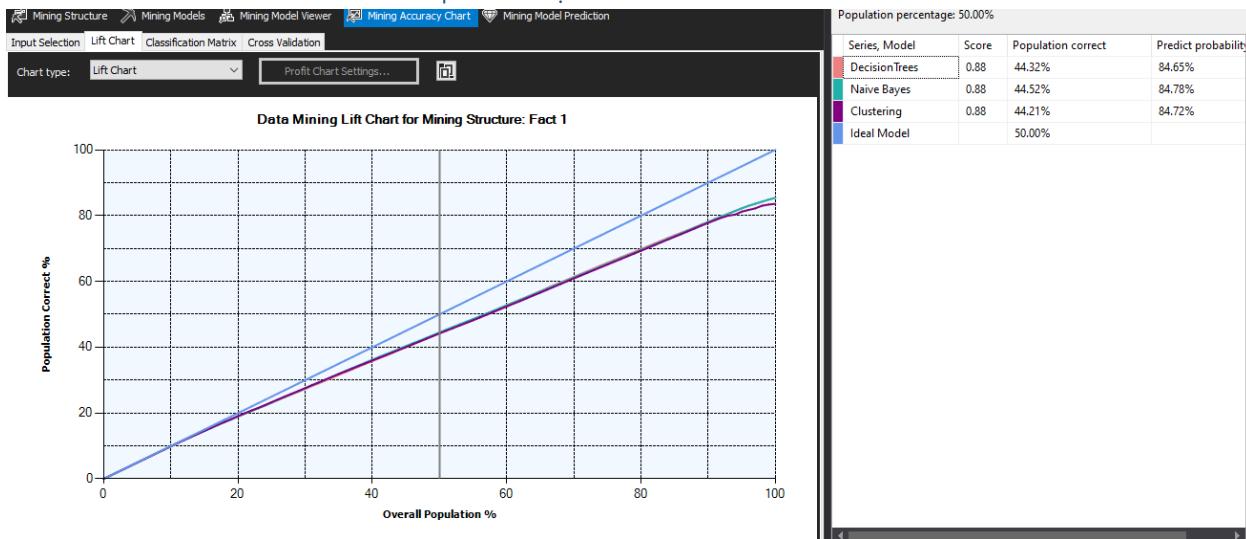
### 2.3. Thuật toán gom cụm

- Kết quả quá trình gom cụm



Ta thấy được 2 cụm cộng đồng mạng là cụm Cluster 2,3,10,8 (cụm 1) và cụm Cluster 4,5,6,7,9 (cụm 2). Ở cụm cộng đồng thứ nhất, ta thấy được điểm tương đồng của các cluster trong cụm là có Id Weather conditions là Fine without high winds và Id Surface Conditions là Dry. Còn ở cụm cộng đồng thứ hai các cluster có điểm tương đồng là Cluster 7 với 5, 6 với 4, 9. Từ sự phân chia này ta có thể tập trung vào những điểm tương đồng lớn từng cụm cộng đồng mạng để giải quyết các vấn đề ấy. Từ đó giảm thiểu được tai nạn đường bộ. Cluster 1 là trung tâm của mạng. Nghĩa là Cluster 1 có đường đi ngắn nhất đến tất cả các node trong mạng (Betweenness centrality). Từ đó nếu ta giải quyết được vấn đề của Cluster 1 ta sẽ giải quyết được phần lớn vấn đề của cả cộng đồng mạng.

#### 2.4. So sánh các mô hình khai phá dữ liệu



Từ biểu đồ trên ta thấy được độ chính xác của ba mô hình gần như bằng nhau ở mức 84%