

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM
KHOA CÔNG NGHỆ THÔNG TIN



HCMUTE
MÔN HỌC: KHO DỮ LIỆU
(DATA MINING)

BÁO CÁO CUỐI KÌ

ĐỀ TÀI: PHÂN TÍCH VÀ XÂY DỰNG KHO DỮ LIỆU TỪ DỮ LIỆU
“DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS”

GVHD: NGUYỄN VĂN THÀNH

SVTH:	MSSV
Nguyễn Thanh Bình	20133025
Nguyễn Nhật Triều	20133102
Đoàn Quốc Trung	20133104

Lớp thứ 5_Tiết 1-4

Tp.Hồ Chí Minh, tháng 5 năm 2023

NHẬN XÉT CỦA GV:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

ĐIỂM

GV ký tên

MỤC LỤC

1.	Tổng quan về dataset	1
1.1.	Lý do chọn dữ liệu	1
1.2.	Nguồn dữ liệu sử dụng.....	1
1.2.1.	Giới thiệu về nơi cấp dữ liệu	1
1.2.2.	Cách tải dữ liệu để thực hiện kho dữ liệu	2
1.3.	Mô tả chi tiết dữ liệu.....	2
1.3.1.	Thông số của tập dữ liệu.....	2
1.3.2.	Dữ liệu sau khi trích xuất.....	3
1.3.3.	Mô tả chi tiết các thuộc tính trong tập dữ liệu	3
1.4.	Giới thiệu về công nghệ sử dụng	7
1.4.1.	Visual Studio 2022	7
1.4.2.	SQL Server 2022	7
1.4.3.	Power BI.....	7
1.4.4.	Visual Studio Code.....	8
1.4.5.	Python.....	8
1.4.6.	SQL (Ngôn ngữ truy vấn có cấu trúc)	9
2.	Thiết kế xây dựng kho dữ liệu	10
2.1.	Tiền xử lý dữ liệu.....	10
2.2.	Thiết kế các Dim và Fact	11
2.2.1.	Các bảng Dim	11
2.2.1.1.	Cấu trúc DimCategory	11
2.2.1.2.	Cấu trúc bảng DimCustomer.....	12
2.2.1.3.	Cấu trúc bảng DimDate.....	12
2.2.1.4.	Cấu trúc bảng DimDepartment	12
2.2.1.5.	Cấu trúc bảng DimLocation	12
2.2.1.6.	Cấu trúc bảng DimProduct.....	13
2.2.2.	Các bảng Fact	13
2.2.2.1.	Cấu trúc bảng FactSales	13
2.2.2.2.	Cấu trúc bảng FactDelivery	14
2.3.	Lược đồ Constellation.....	15
3.	Tích hợp dữ liệu vào kho (SSIS)	15
3.1.	Quá trình tạo mới project SSIS	15
3.2.	Tích hợp dữ liệu từ file csv vào Stage	16
3.3.	Tích hợp dữ liệu từ Stage và Warehouse	27
4.	Phân tích dữ liệu (SSAS).....	31
4.1.	Quá trình xây dựng mô hình	31
4.2.	Quá trình xây dựng khối - Cube	35
5.	Xây dựng Dashboard với PowerBI.....	44

5.1.	Trang chủ.....	44
5.1.1.	Dashboard Sales Report.....	44
5.1.1.1.	Biểu đồ thể hiện tỷ lệ của Quantity Order theo Customer Country.....	45
5.1.1.2.	Biểu đồ thể hiện tổng doanh thu cửa hàng qua các năm	46
5.1.1.3.	Biểu đồ thể hiện doanh thu theo từng loại khách hàng	46
5.1.1.4.	Biểu đồ thể hiện doanh thu theo từng phương thức thanh toán.....	47
5.1.1.5.	Biểu đồ thể hiện doanh thu các khách hàng mua hàng nhiều nhất.....	48
5.1.1.6.	Biểu đồ thể hiện số lượng đơn đặt hàng theo trạng thái đơn hàng	48
5.1.2.	Dashboard Profit Report	49
5.1.2.1.	Bảng xếp hạng lợi nhuận của các sản phẩm được bán	49
5.1.2.2.	Biểu đồ thể hiện tổng lợi nhuận theo từng doanh mục sản phẩm.....	50
5.1.2.3.	Biểu đồ theo dõi lợi nhuận theo tháng	51
5.1.3.	Dashboard Delivery Report	52
5.1.3.1.	Biểu đồ thể hiện số ngày trung bình giao hàng theo phương thức vận chuyển.....	52
5.1.3.2.	Biểu đồ thể hiện tỷ lệ tổng số đơn hàng theo trạng thái vận chuyển.....	53
5.1.3.3.	Biểu đồ thể hiện số lượng đơn đặt hàng theo loại khách hàng và phương thức vận chuyển.....	54
5.1.3.4.	Biểu đồ thể hiện số lượng theo phương thức vận chuyển	54
5.1.3.5.	Biểu đồ thể hiện số lượng đơn đặt hàng theo doanh mục sản phẩm và trạng thái đơn hàng.....	56
5.1.3.6.	Biểu đồ thể hiện số lượng đơn hàng theo trạng thái giao hàng và phương thức vận chuyển.....	56
5.1.3.7.	Biểu đồ thể hiện số lượng đặt hàng đến từng quốc gia	57
6.	Kết luận	58
6.1.	Kết quả đạt được	58
6.2.	Hạn chế	58
6.3.	Bảng phân công nhiệm vụ	58
7.	Tài liệu tham khảo	60

1. Tổng quan về dataset

1.1. Lý do chọn dữ liệu

Bán lẻ là một lĩnh vực kinh doanh rất quan trọng trong nền kinh tế của một quốc gia. Trong lĩnh vực này, các cửa hàng bán lẻ luôn đóng vai trò quan trọng trong việc cung cấp hàng hóa và dịch vụ cho khách hàng. Hiện nay, việc bán lẻ đang ngày một phát triển lớn mạnh, điều đó được thể hiện rõ khi ngày một nhiều cửa hàng bán lẻ khai trương.

Tuy nhiên, không phải việc khai trương một cửa hàng bán lẻ thì sẽ được nhiều người quan tâm và mua mà phải có sự hiểu biết sâu sắc về hành vi mua sắm của khách hàng và cách hoạt động của thị trường bán lẻ.

Trong bối cảnh đó, việc xây dựng và phân tích kho dữ liệu bán hàng của một cửa hàng bán lẻ sẽ là một công cụ hữu ích để giúp các nhà quản lý cửa hàng hiểu rõ hơn về hoạt động của cửa hàng, từ đó đưa ra những quyết định kinh doanh thông minh hơn.

Chính vì vậy, nhóm quyết định chọn tập dữ liệu “DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS” làm dữ liệu để xây dựng kho dữ liệu báo cáo đồ án cuối kỳ môn này.

1.2. Nguồn dữ liệu sử dụng

1.2.1. Giới thiệu về nơi cấp dữ liệu

Kaggle là một nền tảng trực tuyến phổ biến dành cho những người đam mê khoa học dữ liệu và máy học. Nó cung cấp một cộng đồng đa dạng gồm các chuyên gia dữ liệu với nhiều tài nguyên và công cụ để khám phá, phân tích và cộng tác trên các bộ dữ liệu và dự án máy học. Kaggle cung cấp một kho dữ liệu khổng lồ từ nhiều miền khác nhau, cho phép người dùng truy cập và làm việc với dữ liệu trong thế giới thực. Nó cũng tổ chức các cuộc thi nơi các nhà khoa học dữ liệu có thể thể hiện kỹ năng của họ và cạnh tranh để giành giải thưởng. Ngoài

ra, Kaggle cung cấp một nền tảng để người dùng chia sẻ các phân tích dữ liệu và mô hình học máy của họ thông qua sổ ghi chép, thúc đẩy chia sẻ kiến thức và cộng tác trong cộng đồng. Với cộng đồng sôi động, bộ sưu tập tập dữ liệu phong phú và các tính năng tương tác, Kaggle là một nền tảng dành cho những người đam mê dữ liệu cũng như các chuyên gia.

1.2.2. Cách tải dữ liệu để thực hiện kho dữ liệu

Link tải dataset: <https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

Dữ liệu sau khi tải từ trang kaggle về gồm có 3 tệp csv lần lượt sau:
DataCoSupplyChainDataset.csv, DescriptionDataCoSupplyChain.csv,
tokenized_access_logs.csv

Nhưng ở đây để thực hiện đề tài, nhóm chỉ lấy dữ liệu bên trong tệp DataCoSupplyChainDataset.csv

1.3. Mô tả chi tiết dữ liệu

Tập dữ liệu "DataCo Smart Supply Chain for Big Data Analysis" là một bộ dữ liệu bao gồm các thông tin liên quan đến đơn đặt hàng của một chuỗi bán lẻ lớn, được tải lên trang Kaggle. Dữ liệu có thể được sử dụng để phân tích doanh số của chuỗi bán lẻ, quản lý các đơn hàng được vận chuyển đến các khu vực và đưa ra chiến lược kinh doanh về trong tương lai.

1.3.1. Thông số của tập dữ liệu

Tập DataCoSupplyChainDataset.csv gồm 51 cột và hơn 180.000 bản ghi, trong đó mỗi bản ghi trong tập dữ liệu chứa thông tin về một đơn hàng cụ thể, bao gồm định danh đơn hàng, định danh khách hàng, ngày đặt hàng, mã thành phố, phương thức thanh toán, tổng giá trị đơn hàng, số lượng sản phẩm, mã sản phẩm, danh mục sản phẩm,...

1.3.2. Dữ liệu sau khi trích xuất

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		Type	Days for ship	Days for shipment	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk	Category Id	Category Name	Customer City	Customer Country	Customer Email	Customer Fname	Customer Lname	Customer Address	Customer City	Customer Country	Customer Email	Customer Fname	Customer Lname
2	0	DEBIT	3	4	91.25	314.64	Advance s	0	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Cally	20755	Holloway	XXXXXXXXXX	Consumer PR	5365	Nobl		
3	1	TRANSFER	5	4	-249.09	311.36	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Irene	19492	Luna	XXXXXXXXXX	Consumer PR	2679	Rusti		
4	2	CASH	4	4	-247.78	309.72	Shipping c	0	73	Sporting C San Jose	EE. UU.	XXXXXXXXXX	Gillian	19491	Maldonad	XXXXXXXXXX	Consumer CA	8510	Roun	95	
5	3	DEBIT	3	4	22.86	304.81	Advance s	0	73	Sporting C Los Angeli	EE. UU.	XXXXXXXXXX	Tana	19490	Tate	XXXXXXXXXX	Home Off CA	3200	Amb	90	
6	4	PAYMENT	2	4	134.21	298.25	Advance s	0	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Orli	19489	Hendricks	XXXXXXXXXX	Corporate PR	8671	Iron		
7	5	TRANSFER	6	4	18.58	294.98	Shipping c	0	73	Sporting C Tonawanc	EE. UU.	XXXXXXXXXX	Kimberly	19488	Flowers	XXXXXXXXXX	Consumer NY	2122	Hazy	14	
8	6	DEBIT	2	1	95.18	288.42	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Constance	19487	Terrell	XXXXXXXXXX	Home Off PR	1879	Gree		
9	7	TRANSFER	2	1	68.43	285.14	Late deliv	1	73	Sporting C Miami	EE. UU.	XXXXXXXXXX	Erica	19486	Stevens	XXXXXXXXXX	Corporate FL	7595	Cott	33	
10	8	CASH	3	2	133.72	278.59	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Nichole	19485	Olsen	XXXXXXXXXX	Corporate PR	2051	Dust		
11	9	CASH	2	1	132.15	275.31	Late deliv	1	73	Sporting C San Ramo	EE. UU.	XXXXXXXXXX	Oprah	19484	Delacruz	XXXXXXXXXX	Corporate CA	9139	Blue	94	
12	10	TRANSFER	6	2	130.58	272.03	Shipping c	0	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Germane	19483	Short	XXXXXXXXXX	Corporate PR	4058	Quie		
13	11	TRANSFER	5	2	45.69	268.76	Late deliv	1	73	Sporting C Freeport	EE. UU.	XXXXXXXXXX	Freya	19482	Robbins	XXXXXXXXXX	Consumer NY	3243	Shad	11	
14	12	TRANSFER	4	2	21.76	262.2	Late deliv	1	73	Sporting C Salinas	EE. UU.	XXXXXXXXXX	Cassandra	19481	Jensen	XXXXXXXXXX	Corporate CA	131	Sunny	93	
15	13	DEBIT	2	1	24.58	245.81	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Natalie	19480	McFadden	XXXXXXXXXX	Corporate PR	2531	Wish		
16	14	TRANSFER	2	1	16.39	327.75	Late deliv	1	73	Sporting C Peabody	EE. UU.	XXXXXXXXXX	Kimberley	19479	Sharpe	XXXXXXXXXX	Corporate MA	6417	Silve	1	
17	15	DEBIT	2	1	-259.58	324.47	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Sade	19478	Lancaster	XXXXXXXXXX	Corporate PR	257	Harve		
18	16	PAYMENT	5	2	-246.36	321.2	Late deliv	1	73	Sporting C Canovana	Puerto Ric	XXXXXXXXXX	Brynne	19477	Giles	XXXXXXXXXX	Corporate PR	7342	Hazy		
19	17	CASH	2	1	23.84	317.92	Late deliv	1	73	Sporting C Paramount	EE. UU.	XXXXXXXXXX	Clara	19476	Bird	XXXXXXXXXX	Corporate CA	7787	Lazy	90	
20	18	DEBIT	2	1	102.26	314.64	Late deliv	1	73	Sporting C Caguas	Puerto Ric	XXXXXXXXXX	Bo	19475	Griffin	XXXXXXXXXX	Consumer PR	5136	Rusti		
21	19	PAYMENT	0	0	87.18	311.36	Shipping c	0	73	Sporting C Mount Pr	EE. UU.	XXXXXXXXXX	Kim	19474	Simon	XXXXXXXXXX	Consumer IL	1723	Tawr	60	
22	20	TRANSFER	0	0	154.86	309.72	Shipping c	0	73	Sporting C Long Beac	EE. UU.	XXXXXXXXXX	Kellie	19473	Farmer	XXXXXXXXXX	Corporate CA	7111	Silen	90	

1.3.3. Mô tả chi tiết các thuộc tính trong tập dữ liệu

Tập dữ liệu gồm các thuộc tính sau:

Tên thuộc tính	Mô tả
Type	Loại giao dịch
Days for ship(real)	Số ngày vận chuyển thực tế
Days for shipment (scheduled)	Số ngày dự kiến giao hàng
Benefit per order	Thu nhập trên mỗi đơn hàng đặt hàng
Sales per customer	Tổng doanh số bán hàng trên mỗi khách hàng
Delivery Status	Trạng thái giao hàng: Giao hàng nhanh, Giao hàng trễ, Hủy giao hàng, Giao hàng đúng hạn
Late_delivery_risk	Biến phân loại cho biết liệu việc gửi hàng có chậm trễ hay không (1: chậm trễ, 0: không chậm trễ)
Category Id	Mã danh mục sản phẩm
Category Name	Tên của danh mục sản phẩm
Customer City	Thành phố nơi khách hàng thực hiện mua hàng
Customer Country	Quốc gia nơi khách hàng thực hiện mua hàng
Customer Email	Email của khách hàng
Customer Fname	Tên của khách hàng

Customer Id	Mã khách hàng
Customer Lname	Họ của khách hàng
Customer Password	Mật khẩu khách hàng được che giấu
Customer Segment	Loại khách hàng: Người tiêu dùng, doanh nghiệp, văn phòng, gia đình
Customer State	Tiểu bang mà khách hàng mua hàng
Customer Street	Đường phố mà khách hàng mua hàng
Customer Zipcode	Mã bưu chính của khách hàng
Department Id	Mã phòng ban của cửa hàng
Latitude	Kinh độ tương ứng với vị trí của cửa hàng
Longitude	Kinh độ tương ứng với vị trí của cửa hàng
Market	Thị trường mà đơn hàng được giao đến: Châu Phi, Châu Âu, LATAM, Châu Á Thái Bình Dương, USCA
Order City	Thành phố giao đến của đơn hàng
Order Country	Quốc gia giao đến của đơn hàng
Order Customer Id	Mã đặt hàng của khách hàng
Order date (DateOrders)	Ngày đặt hàng
Order Id	Mã đơn hàng
Order Item Cardprod Id	Mã sản phẩm được tạo ra thông qua đầu đọc RFID
Order Item Discount	Giá trị chiết khấu cho đơn hàng
Order Item Discount Rate	Tỷ lệ chiết khấu cho sản phẩm trong một đơn hàng
Order Item Id	Mã sản phẩm trong một đơn hàng
Order Item Product Price	Giá của sản phẩm trước khi được giảm giá
Order Item Profit Ratio	Tỷ lệ lợi nhuận của sản phẩm trong một đơn hàng
Order Item Quantity	Số lượng sản phẩm trong một đơn hàng
Sales	Giá trị doanh số bán hàng
Order Item Total	Tổng số tiền trên mỗi đơn hàng

Order Profit Per Order	Lợi nhuận trên mỗi đơn hàng
Order Region	Vùng trên thế giới mà đơn hàng được giao đến: Đông Nam Á, Nam Á, Đại Dương, Đông Á, Tây Á, Tây của Mỹ, Trung tâm Mỹ, Tây Phi, Trung Phi, Bắc Phi, Tây Âu, Bắc Âu, Caribe, Nam Mỹ, Đông Phi, Nam Âu, Đông của Mỹ, Canada, Nam Phi, Trung Á, Châu Âu, Trung Mỹ, Đông Âu, Nam của Mỹ
Order State	Tiểu bang của vùng nơi đơn hàng được giao
Order Status	Trạng thái đơn hàng: Đã hoàn thành, đang chờ, đã đóng, đang chờ thanh toán, đã hủy, đang xử lý, nghi ngờ gian lận, đang giữ, đánh giá thanh toán
Product Card Id	Mã sản phẩm
Product Category Id	Mã danh mục sản phẩm
Product Description	Mô tả sản phẩm
Product Image	Ảnh sản phẩm
Product Name	Tên sản phẩm
Product Price	Giá sản phẩm
Product Status	Trạng thái hàng tồn kho: Nếu là 1 thì không có sẵn, 0 thì sản phẩm có sẵn
Shipping date (DateOrders)	Ngày và giờ chính xác của việc vận chuyển
Shipping Mode	Các chế độ vận chuyển: Lớp tiêu chuẩn, lớp đầu tiên, lớp thứ hai, cùng ngày

Product Price là giá bán đề xuất của sản phẩm đó trên thị trường, đây là giá trị được đưa ra cho khách hàng trước khi họ thực hiện bất kỳ hoạt động mua hàng nào. Đây là giá trị thường được đặt bởi nhà sản xuất hoặc nhà bán lẻ dựa trên chi phí sản xuất, lợi nhuận mong muốn và các yếu tố thị trường.

Order Item Product Price là giá bán của sản phẩm đó tại thời điểm khách hàng đặt mua sản phẩm. Đây là giá trị thực tế mà khách hàng phải trả cho sản phẩm đó, bao gồm cả các giảm giá, khuyến mại hoặc phí vận chuyển. Trong một đơn hàng, mỗi mặt hàng có thể có một giá bán khác nhau tùy thuộc vào các yếu tố khác nhau, chẳng hạn như số lượng mặt hàng, chiết khấu hay khuyến mại được áp dụng.

Order Item Profit Ratio là chỉ số đo lường lợi nhuận của một sản phẩm trong một đơn hàng. Nó được tính bằng cách chia lợi nhuận thu về từ sản phẩm đó trong đơn hàng cho giá bán của sản phẩm đó trong đơn hàng. Nói cách khác, Order Item Profit Ratio cho chúng ta biết bao nhiêu phần trăm của giá bán của một sản phẩm được cộng vào lợi nhuận sau khi trừ đi tất cả các chi phí liên quan đến sản phẩm đó trong đơn hàng.

Công thức tính Order Item Profit Ratio như sau:

$$\text{Order Item Profit Ratio} = (\text{Giá bán} - \text{Tổng chi phí liên quan đến sản phẩm}) / \text{Giá bán} * 100\%$$

Trong đó:

- Giá bán là giá bán của sản phẩm trong đơn hàng.
- Tổng chi phí liên quan đến sản phẩm bao gồm các chi phí như chi phí sản xuất, chi phí vận chuyển, chi phí quản lý và các chi phí khác liên quan đến sản phẩm đó trong đơn hàng.

Ví dụ, nếu một sản phẩm được bán trong một đơn hàng với giá bán là 100 đô la, và tổng chi phí liên quan đến sản phẩm đó trong đơn hàng là 50 đô la, thì Order Item Profit Ratio sẽ là:

$$\text{Order Item Profit Ratio} = (100 - 50) / 100 * 100\% = 50\%$$

Điều này có nghĩa là, 50% giá trị của sản phẩm đó được tính vào lợi nhuận của đơn hàng sau khi trừ đi tất cả các chi phí liên quan đến sản phẩm đó. Càng

cao tỷ lệ lợi nhuận này thì sản phẩm đó càng mang lại lợi nhuận cao cho doanh nghiệp trong đơn hàng đó.

1.4. Giới thiệu về công nghệ sử dụng

1.4.1. Visual Studio 2022

Visual Studio 2022 là môi trường phát triển tích hợp (IDE) hàng đầu được phát triển bởi Microsoft. Nó cung cấp một nền tảng mạnh mẽ cho nhà phát triển để tạo ra các ứng dụng đa dạng. Với những tính năng mạnh mẽ, giao diện thân thiện và hỗ trợ ngôn ngữ đa dạng, Visual Studio 2022 giúp việc lập trình, gỡ lỗi và kiểm thử trở nên hiệu quả trên nhiều nền tảng khác nhau. Nó cung cấp các công cụ tăng cường năng suất, khả năng gỡ lỗi tiên tiến và tích hợp mượt mà với các framework phổ biến, là sự lựa chọn ưu tiên cho các chuyên gia phát triển phần mềm.

1.4.2. SQL Server 2022

SQL Server 2022 là một hệ quản trị cơ sở dữ liệu quan hệ (RDBMS) mạnh mẽ và mở rộng được phát triển bởi Microsoft. Nó cung cấp cho tổ chức một nền tảng an toàn và hiệu quả để lưu trữ, quản lý và phân tích các khối lượng dữ liệu lớn. Với những tính năng tiên tiến như xử lý trong bộ nhớ, bảo mật nâng cao và tối ưu hóa truy vấn thông minh, SQL Server 2022 cho phép doanh nghiệp trích xuất thông tin quý giá từ dữ liệu và đưa ra quyết định dựa trên dữ liệu. Nó cũng cung cấp tích hợp mượt mà với các sản phẩm và dịch vụ đám mây của Microsoft, là sự lựa chọn đáng tin cậy cho các giải pháp cơ sở dữ liệu cấp doanh nghiệp.

1.4.3. Power BI

Power BI là một công cụ phân tích dữ liệu và trực quan hóa dữ liệu mạnh mẽ được phát triển bởi Microsoft. Nó cho phép người dùng kết nối với các nguồn dữ liệu khác nhau, biến đổi và làm sạch dữ liệu, và tạo ra các báo cáo và bảng điều khiển tương tác. Với giao diện thân thiện và khả năng kéo và thả, Power BI cho

phép người dùng phân tích dữ liệu, khám phá thông tin quan trọng và đưa ra quyết định dựa trên dữ liệu. Nó cung cấp một loạt các biểu đồ, khả năng phân tích tiên tiến và khả năng chia sẻ và cộng tác trên các báo cáo. Power BI được sử dụng rộng rãi trong các tổ chức để thu được những thông tin quý giá từ dữ liệu và thúc đẩy các chiến lược dựa trên dữ liệu.

1.4.4. Visual Studio Code

Visual Studio Code, thường được gọi là VS Code, là một trình soạn thảo mã nguồn mạnh mẽ và nhẹ nhàng được phát triển bởi Microsoft. Nó được sử dụng rộng rãi bởi các nhà phát triển trên nhiều nền tảng và ngôn ngữ lập trình. Với giao diện tùy chỉnh, tích hợp Git sẵn có và thư viện mở rộng đa dạng, VS Code cung cấp một trải nghiệm lập trình mượt mà và hiệu quả. Nó cung cấp tính năng như IntelliSense để hoàn thành mã, khả năng gỡ lỗi và một hệ sinh thái phong phú của các tiện ích mở rộng, cho phép nhà phát triển tùy chỉnh môi trường lập trình theo nhu cầu cụ thể của họ. Dù là làm việc trên dự án nhỏ hay ứng dụng quy mô lớn, Visual Studio Code là một lựa chọn phổ biến để lập trình hiệu quả và thú vị.

1.4.5. Python

Python là một ngôn ngữ lập trình thông dịch, linh hoạt và dễ học được phát triển bởi Guido van Rossum. Với cú pháp đơn giản và dễ đọc, Python là ngôn ngữ lập trình phổ biến trong lĩnh vực phân tích dữ liệu, trí tuệ nhân tạo và phát triển ứng dụng web. Python cung cấp một thư viện phong phú và mạnh mẽ, bao gồm NumPy, Pandas và TensorFlow, giúp xử lý dữ liệu, tính toán khoa học và xây dựng mô hình máy học. Với cộng đồng lớn và sự hỗ trợ tận tâm từ cộng đồng lập trình viên, Python trở thành công cụ lý tưởng cho việc phát triển ứng dụng, tạo ra các công cụ phân tích dữ liệu và xây dựng các hệ thống thông minh.

1.4.6. SQL (Ngôn ngữ truy vấn có cấu trúc)

SQL (Structured Query Language) là một ngôn ngữ truy vấn được thiết kế để quản lý và tương tác với cơ sở dữ liệu quan hệ. Với SQL, người dùng có thể truy xuất, thêm, sửa đổi và xóa dữ liệu từ cơ sở dữ liệu. SQL cung cấp các câu lệnh mạnh mẽ như SELECT, INSERT, UPDATE và DELETE để thực hiện các hoạt động trên dữ liệu. Ngoài ra, SQL cũng hỗ trợ các khối lệnh điều kiện, các phép toán logic và các hàm tính toán để thao tác và truy xuất dữ liệu một cách linh hoạt. SQL được sử dụng rộng rãi trong việc xây dựng và quản lý các hệ thống quản lý cơ sở dữ liệu (DBMS) và là công cụ không thể thiếu trong lĩnh vực quản lý dữ liệu và phân tích dữ liệu.

2. Thiết kế xây dựng kho dữ liệu

2.1. Tiền xử lý dữ liệu

Load dữ liệu từ file csv bằng python để xử lý

```
import pandas as pd
import numpy as np

data = pd.read_csv('DataCoSupplyChainDataset.csv', encoding=('ISO-8859-1'))
data
```

	Type	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk	Category Id	Category Name	Customer City	...	Order Zipcode	Product Card Id	Product Category Id	Product Description
0	DEBIT	3	4	91.250000	314.640015	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73	NaN
1	TRANSFER	5	4	-249.089996	311.359985	Late delivery	1	73	Sporting Goods	Caguas	...	NaN	1360	73	NaN
2	CASH	4	4	-247.779999	309.720001	Shipping on time	0	73	Sporting Goods	San Jose	...	NaN	1360	73	NaN
3	DEBIT	3	4	22.860001	304.809998	Advance shipping	0	73	Sporting Goods	Los Angeles	...	NaN	1360	73	NaN
4	PAYMENT	2	4	134.210007	298.250000	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73	NaN
...
180514	CASH	4	4	40.000000	399.980011	Shipping on time	0	45	Fishing	Brooklyn	...	NaN	1004	45	NaN
180515	DEBIT	3	2	-613.770019	395.980011	Late delivery	1	45	Fishing	Bakersfield	...	NaN	1004	45	NaN

Chuyển dữ liệu 2 cột order date (DateOrders) và cột shipping date (DateOrders) từ kiểu Datetime sang kiểu Date

```
Xử lý kiểu dữ liệu cột order date (DateOrders) và cột shipping date (DateOrders)

data['order date (DateOrders)'] = pd.to_datetime(data['order date (DateOrders)'])
data['shipping date (DateOrders)'] = pd.to_datetime(data['shipping date (DateOrders)'])

n = len(data)
for i in range(n):
    order_date = data.loc[i, 'order date (DateOrders)'].date()
    data.loc[i, 'order date (DateOrders)'] = order_date

    shipping_date = data.loc[i, 'shipping date (DateOrders)'].date()
    data.loc[i, 'shipping date (DateOrders)'] = shipping_date
```

Tạo thêm 1 file Location.csv để thuận tiện cho việc tạo các bảng Dim

Tạo file Location.csv

```

n = len(data)
lst = []
for i in range(n):
    unique_val = data.loc[i, ['Market', 'Order City', 'Order Country', 'Order Region', 'Order State']]
    unique_val_list = unique_val.tolist()
    if len(lst) == 0:
        lst.append(unique_val_list)
    else:
        if unique_val_list not in lst:
            lst.append(unique_val_list)
lst.head()

```

[] Python

```

df = pd.DataFrame(lst, columns = ['Market', 'Order City', 'Order Country', 'Order Region', 'Order State'])
df['LocationID'] = range(1, len(df) + 1)

```

[51] Python

```

path='D:\\SV3_2\\'
csv_data = df.to_csv(path+'Location.csv')

```

[52] Python

Các dòng đầu của file csv vừa mới tạo

```

data_location = pd.read_csv("Location.csv", encoding=('ISO-8859-1'))
data_location.head()

```

Python

	Market	Order City	Order Country	Order Region	Order State	LocationID
0	Pacific Asia	Bekasi	Indonesia	Southeast Asia	Java Occidental	1
1	Pacific Asia	Bikaner	India	South Asia	Rajasth	2
2	Pacific Asia	Townsville	Australia	Oceania	Queensland	3
3	Pacific Asia	Toowoomba	Australia	Oceania	Queensland	4
4	Pacific Asia	Guangzhou	China	Eastern Asia	Guangdong	5

2.2. Thiết kế các Dim và Fact

LAPTOP-MB5F72F2\...dbo.DimCategory

	Column Name	Data Type	Allow Nulls
PK	CategoryKey	int	<input type="checkbox"/>
	CategoryID	float	<input checked="" type="checkbox"/>
	CategoryName	nvarchar(255)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

2.2.1. Các bảng Dim

2.2.1.1. Cấu trúc DimCategory

2.2.1.2. Cấu trúc bảng DimCustomer

	Column Name	Data Type	Allow Nulls
▶	CustomerID	float	<input checked="" type="checkbox"/>
	CustomerCity	nvarchar(255)	<input checked="" type="checkbox"/>
	CustomerCountry	nvarchar(255)	<input checked="" type="checkbox"/>
	CustomerState	nvarchar(255)	<input checked="" type="checkbox"/>
	CustomerStreet	nvarchar(255)	<input checked="" type="checkbox"/>
	CustomerZipcode	float	<input checked="" type="checkbox"/>
	CustomerSegement	nvarchar(255)	<input checked="" type="checkbox"/>
🔑	CustomerKey	int	<input type="checkbox"/>
	FullName	nvarchar(255)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

2.2.1.3. Cấu trúc bảng DimDate

LAPTOP-MB5F72F2...DW - dbo.DimDate			
	Column Name	Data Type	Allow Nulls
▶	Year	int	<input checked="" type="checkbox"/>
	Quarter	int	<input checked="" type="checkbox"/>
	Month	int	<input checked="" type="checkbox"/>
	Day	int	<input checked="" type="checkbox"/>
🔑	DateKey	int	<input type="checkbox"/>
			<input type="checkbox"/>

2.2.1.4. Cấu trúc bảng DimDepartment

	Column Name	Data Type	Allow Nulls
▶	DepartmentID	float	<input checked="" type="checkbox"/>
	DepartmentName	nvarchar(255)	<input checked="" type="checkbox"/>
	Latitude	float	<input checked="" type="checkbox"/>
	Longitude	float	<input checked="" type="checkbox"/>
🔑	DepartmentKey	int	<input type="checkbox"/>
			<input type="checkbox"/>

2.2.1.5. Cấu trúc bảng DimLocation

	Column Name	Data Type	Allow Nulls
►	LocationID	float	<input checked="" type="checkbox"/>
	City	nvarchar(255)	<input checked="" type="checkbox"/>
	State	nvarchar(255)	<input checked="" type="checkbox"/>
	Country	nvarchar(255)	<input checked="" type="checkbox"/>
	Region	nvarchar(255)	<input checked="" type="checkbox"/>
	Market	nvarchar(255)	<input checked="" type="checkbox"/>
🔑	LocationKey	int	<input type="checkbox"/>

2.2.1.6. Cấu trúc bảng DimProduct

	Column Name	Data Type	Allow Nulls
►	ProductName	nvarchar(255)	<input checked="" type="checkbox"/>
	ProductPrice	float	<input checked="" type="checkbox"/>
	ProductStatus	float	<input checked="" type="checkbox"/>
🔑	ProductKey	int	<input type="checkbox"/>
	CategoryKey	int	<input checked="" type="checkbox"/>
	ProductID	float	<input checked="" type="checkbox"/>

2.2.2. Các bảng Fact

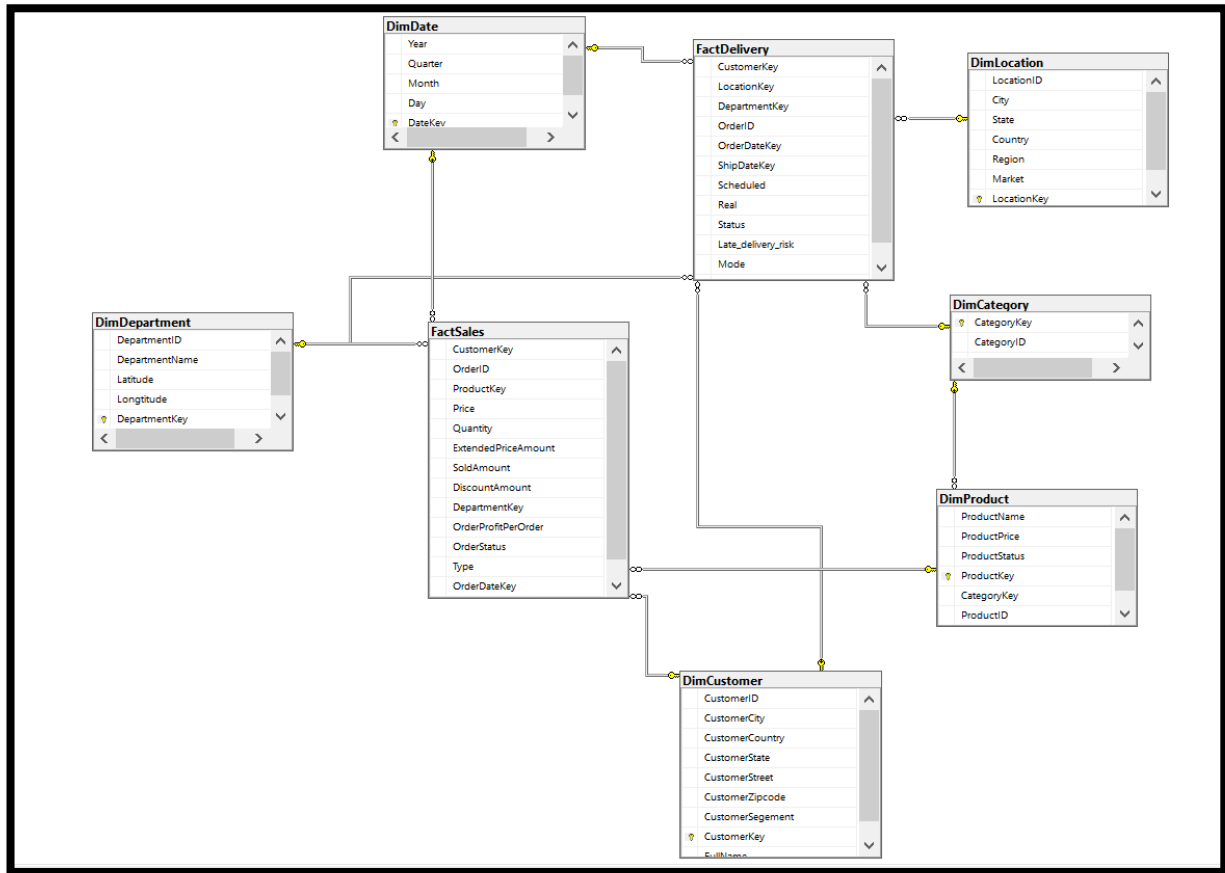
2.2.2.1. Cấu trúc bảng FactSales

	Column Name	Data Type	Allow Nulls
►	CustomerKey	int	<input checked="" type="checkbox"/>
	OrderID	float	<input checked="" type="checkbox"/>
	ProductKey	int	<input checked="" type="checkbox"/>
	Price	float	<input checked="" type="checkbox"/>
	Quantity	float	<input checked="" type="checkbox"/>
	ExtendedPriceAmount	float	<input checked="" type="checkbox"/>
	SoldAmount	float	<input checked="" type="checkbox"/>
	DiscountAmount	float	<input checked="" type="checkbox"/>
	DepartmentKey	int	<input checked="" type="checkbox"/>
	OrderProfitPerOrder	float	<input checked="" type="checkbox"/>
	OrderStatus	nvarchar(255)	<input checked="" type="checkbox"/>
	Type	nvarchar(255)	<input checked="" type="checkbox"/>
	OrderDateKey	int	<input checked="" type="checkbox"/>
	ShipDateKey	int	<input checked="" type="checkbox"/>

2.2.2.2. Cấu trúc bảng *FactDelivery*

	Column Name	Data Type	Allow Nulls
▶	CustomerKey	int	<input checked="" type="checkbox"/>
	LocationKey	int	<input checked="" type="checkbox"/>
	DepartmentKey	int	<input checked="" type="checkbox"/>
	OrderID	float	<input checked="" type="checkbox"/>
	OrderDateKey	int	<input checked="" type="checkbox"/>
	ShipDateKey	int	<input checked="" type="checkbox"/>
	Scheduled	float	<input checked="" type="checkbox"/>
	Real	float	<input checked="" type="checkbox"/>
	Status	nvarchar(255)	<input checked="" type="checkbox"/>
	Late_delivery_risk	float	<input checked="" type="checkbox"/>
	Mode	nvarchar(255)	<input checked="" type="checkbox"/>
	CategoryKey	int	<input checked="" type="checkbox"/>

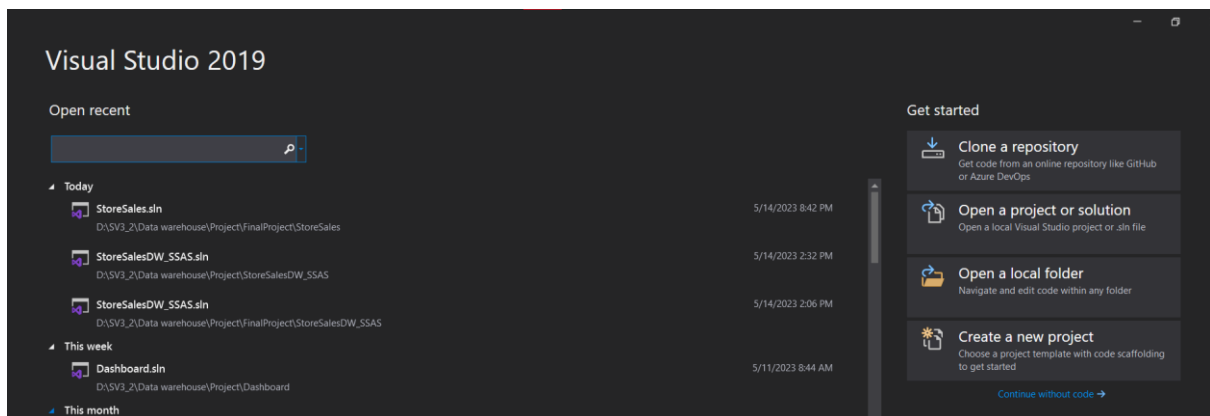
2.3. Lược đồ Constellation



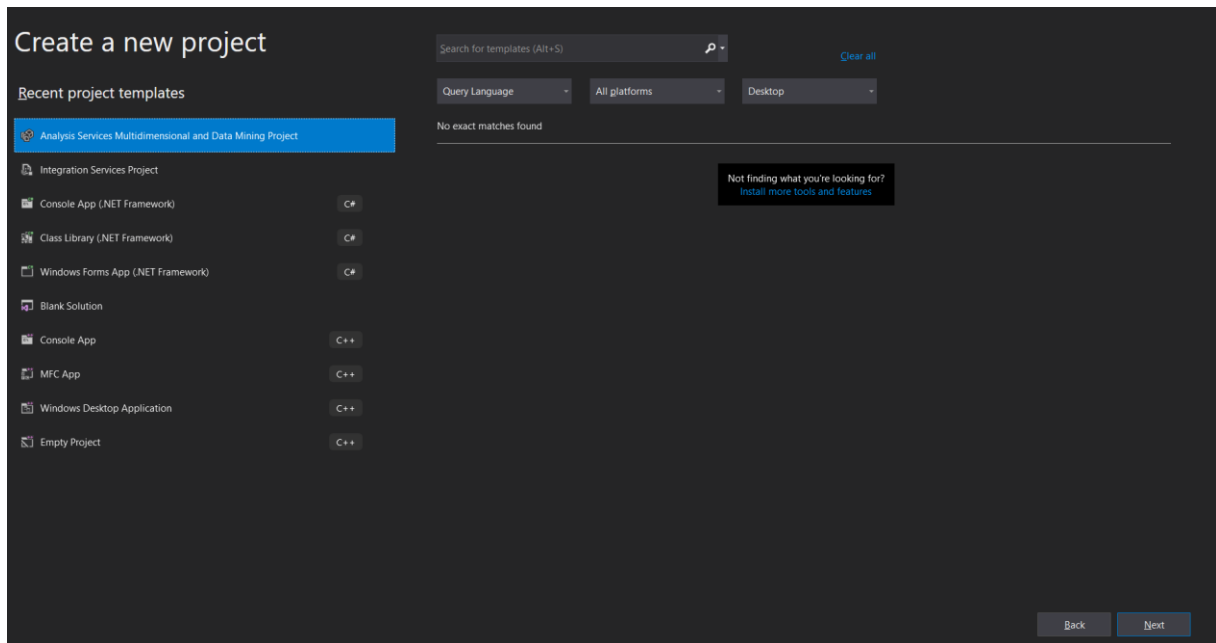
3. Tích hợp dữ liệu vào kho (SSIS)

3.1. Quá trình tạo mới project SSIS

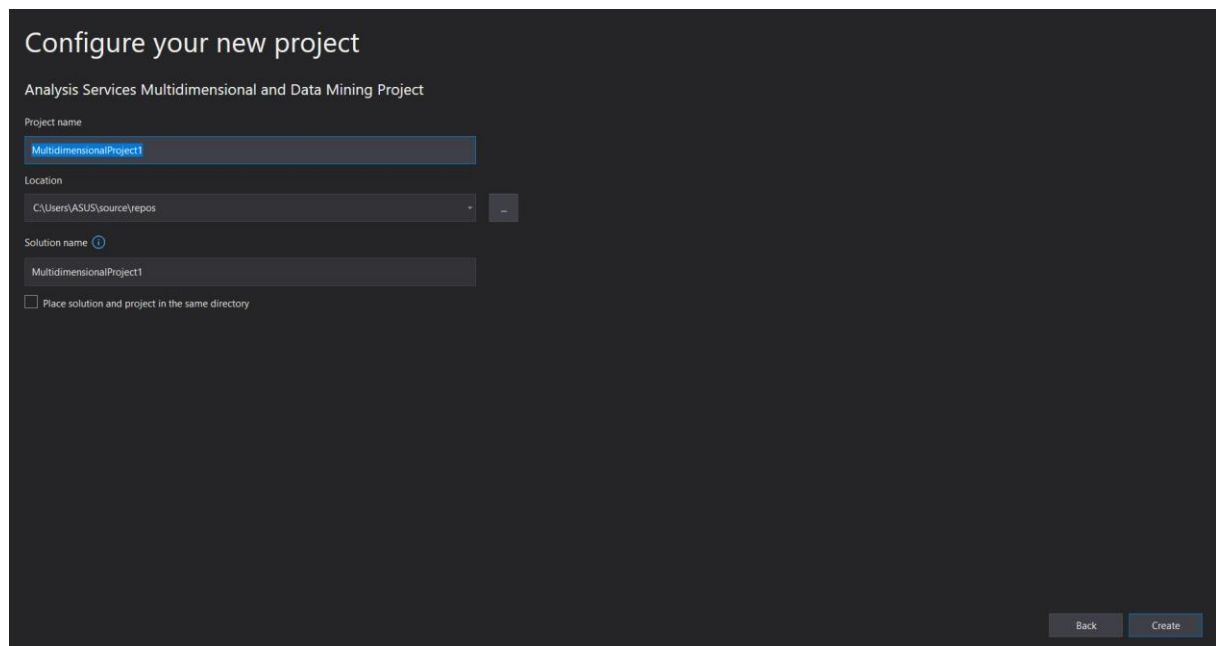
Khởi động Microsoft Visual Studio -> chọn Create a new project



Chọn Analysis Services Multidimensional and Data Mining Project. Sau đó bấm Next

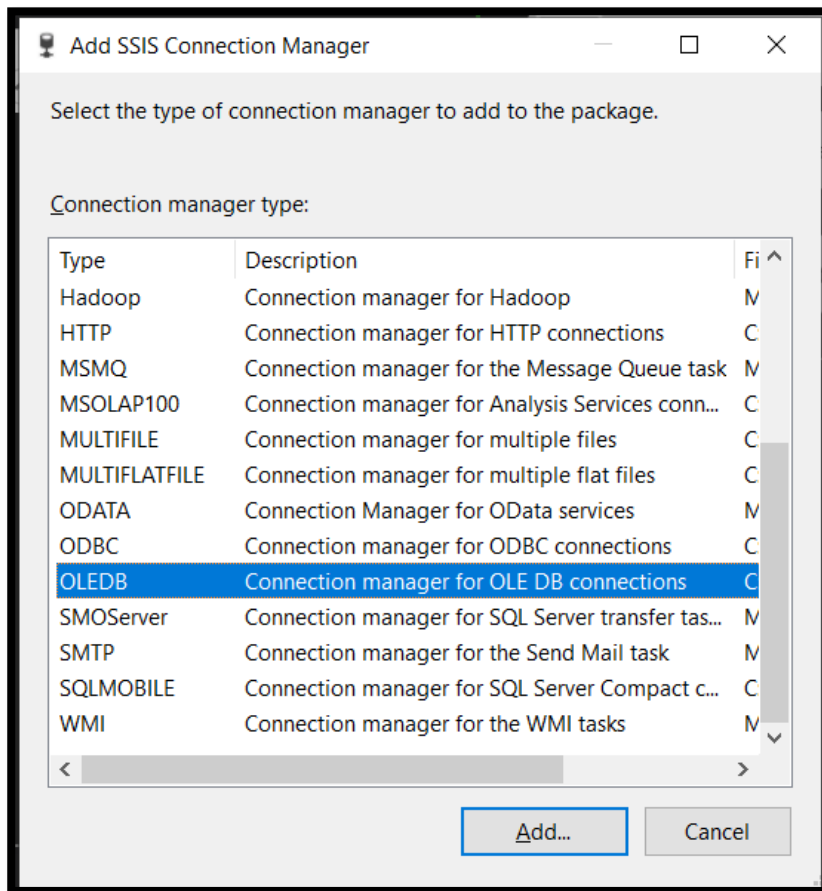


Nhập tên Project vào ô Project name sau đó bấm Create

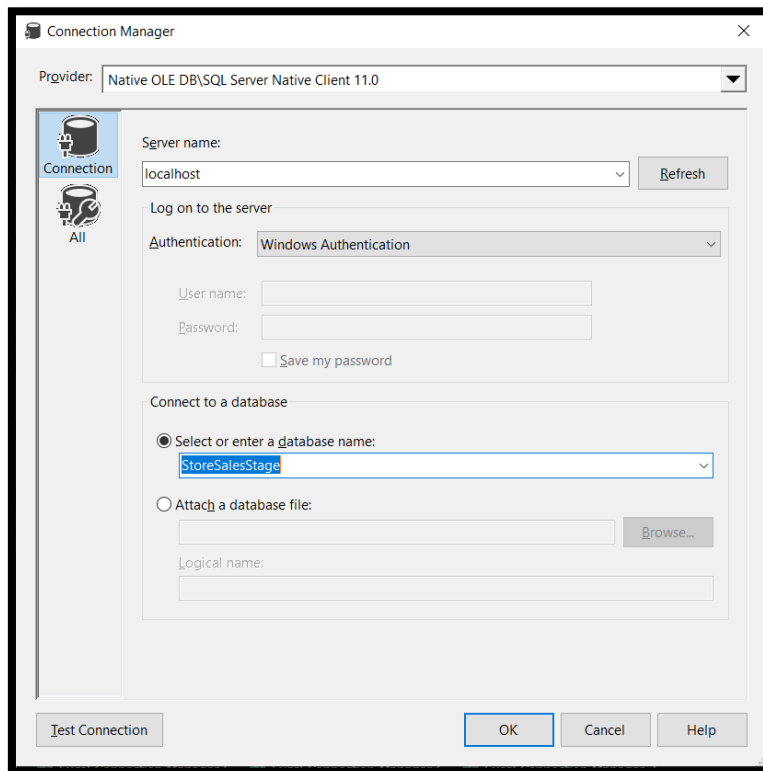


3.2. Tích hợp dữ liệu từ file csv vào Stage

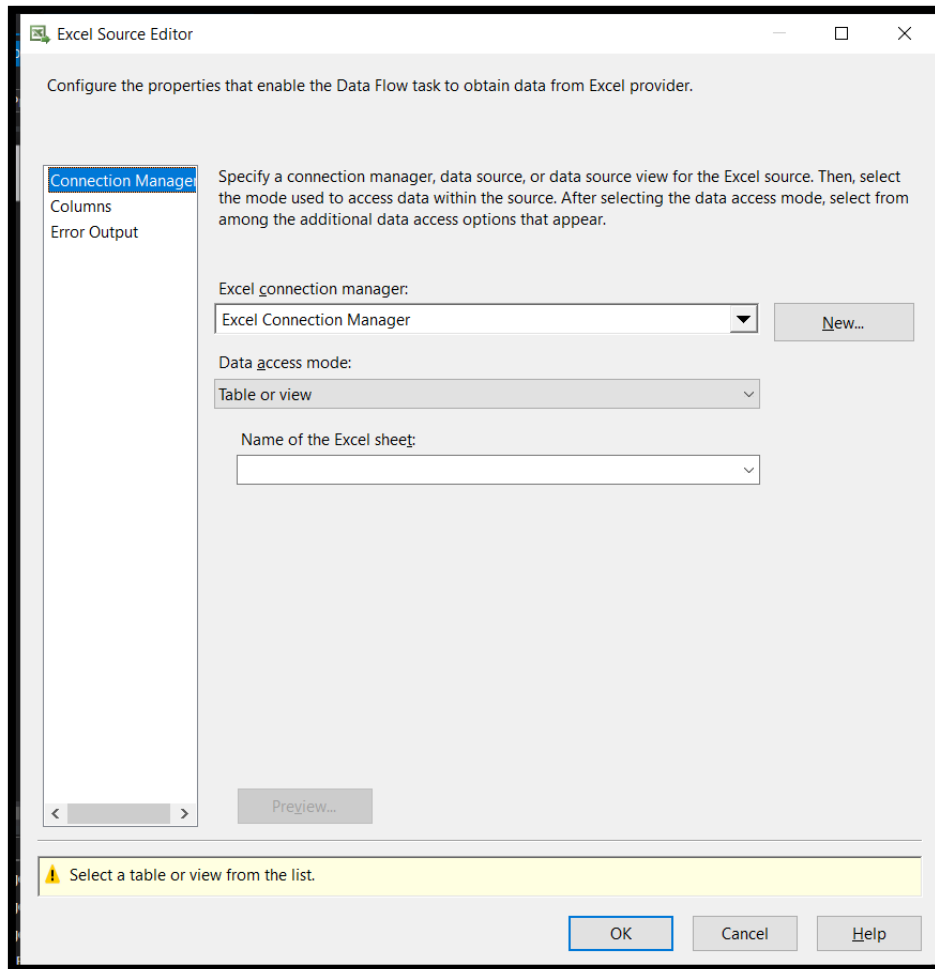
Kích chuột phải vào Connection Managers, sau đó chọn New Connection Manager, chọn OLEDB



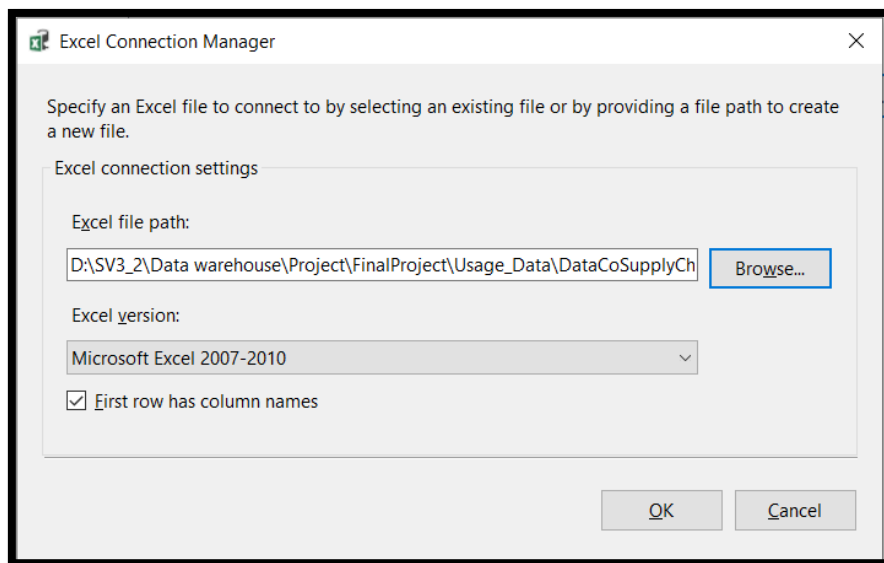
Chọn Add->New. Chọn Provider: Native OLE DB\SQL Server Native Client 11.0. Sau đó nhập tên Server name ở và chọn database muốn connect đến để lưu Stage ở hộp thoại Connec to a database -> nhấn OK



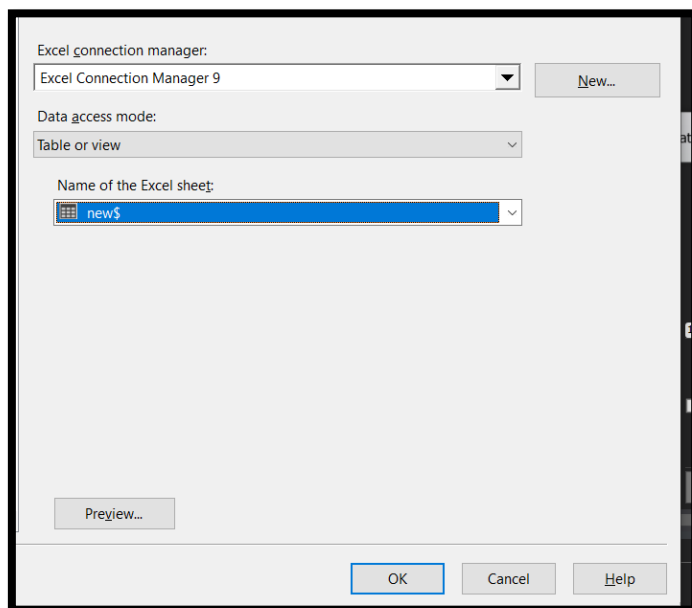
Nhấn chuột phải vào SSIS Packages->New SSIS Package. Một cửa sổ Package mới hiện ra. Kéo thả Data Flow Task ở hộp thoại SSIS Toolbox vào cửa sổ mới tạo và Rename cho nó. Nhấn đúp vào Data Flow Task, Kéo thả Excel Source ở hộp thoại SSIS Toolbox vào đó. Nháy đúp vào Excel Source, hộp thoại sau xuất hiện.



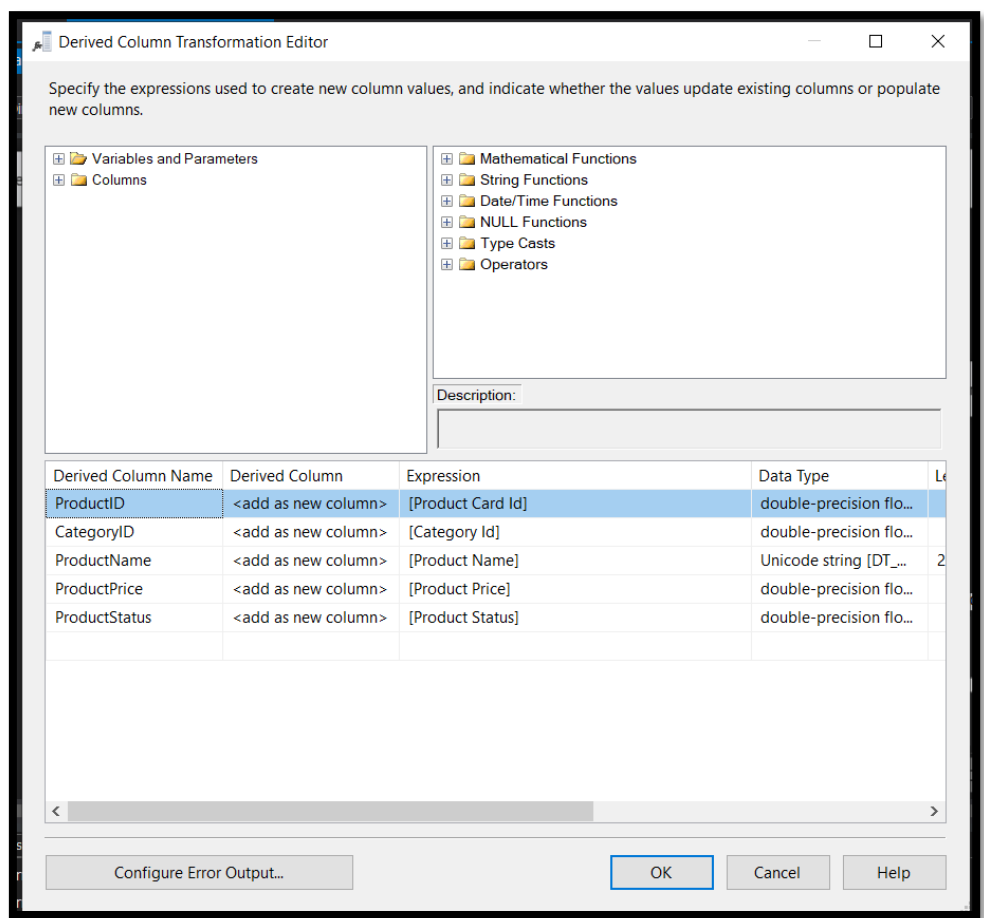
Chọn New ở hộp thoại connection manager và trở tới nơi chứa file excel ở Excel file path và chọn version excel cho phù hợp-> Nhấn Ok



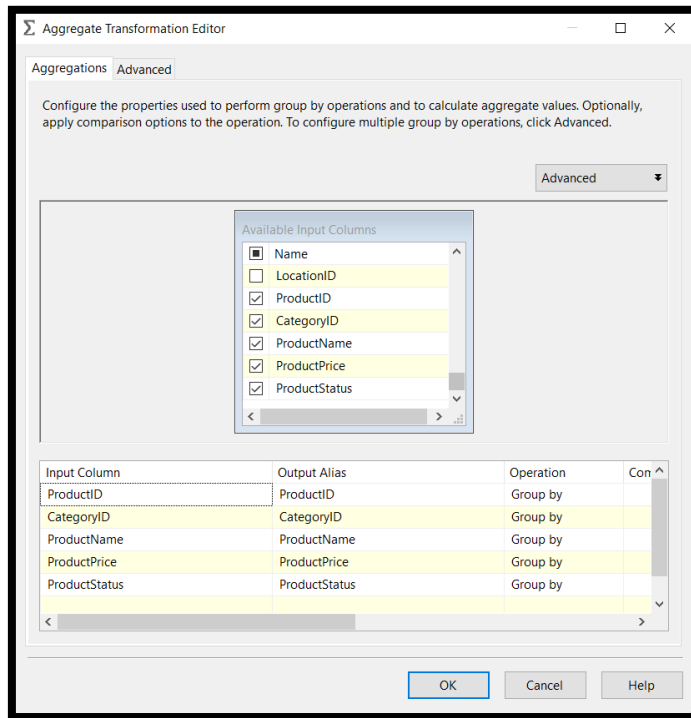
Chọn Sheet ở hộp thoại Name of the Excel sheet -> Nhấn Ok



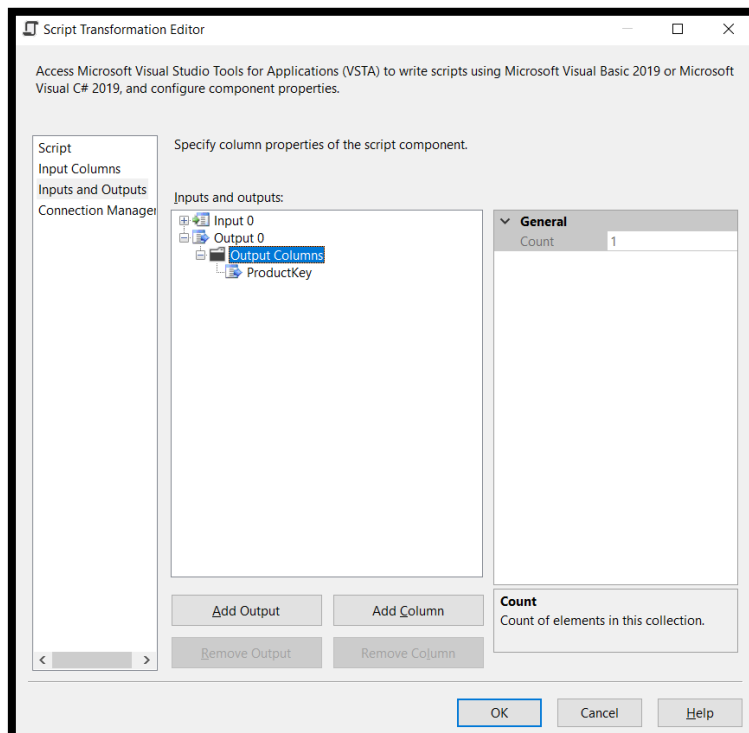
Sau đó Kéo hộp thoại Derived Column vào và điều chỉnh như hình dưới cho DimProduct



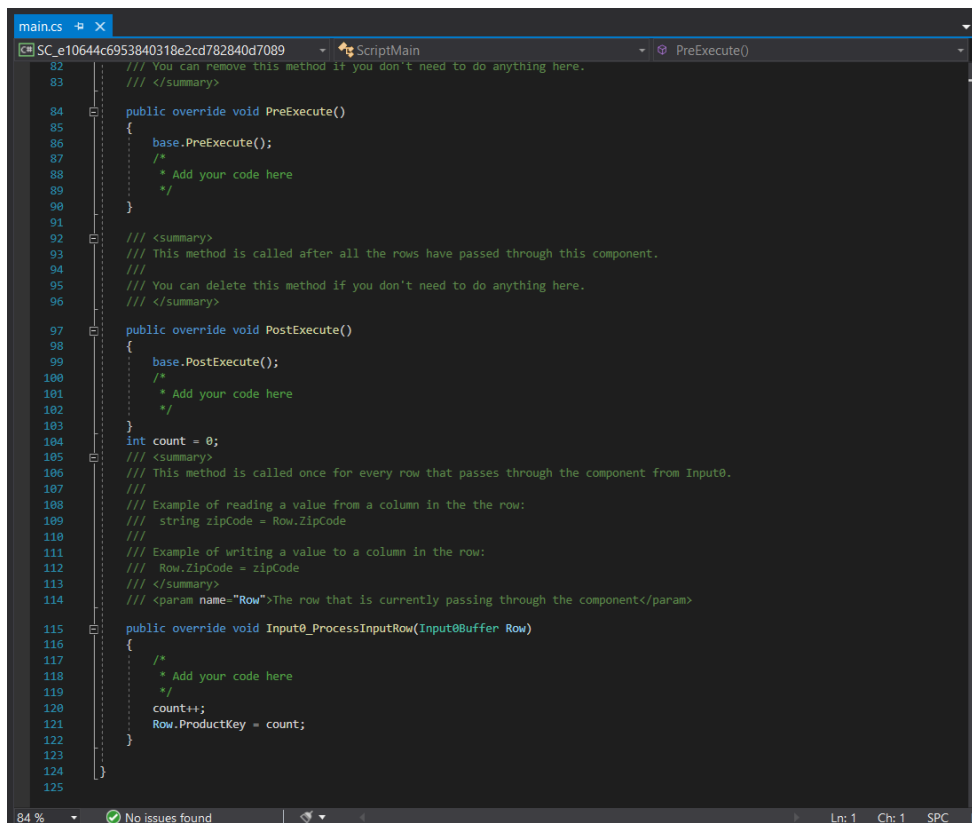
Sau đó kéo hộp thoại Aggregate vào và điều chỉnh như hình dưới cho DimProduct



Sau đó kép hộp thoại Script Component và điều chỉnh như hình dưới cho DimProduct

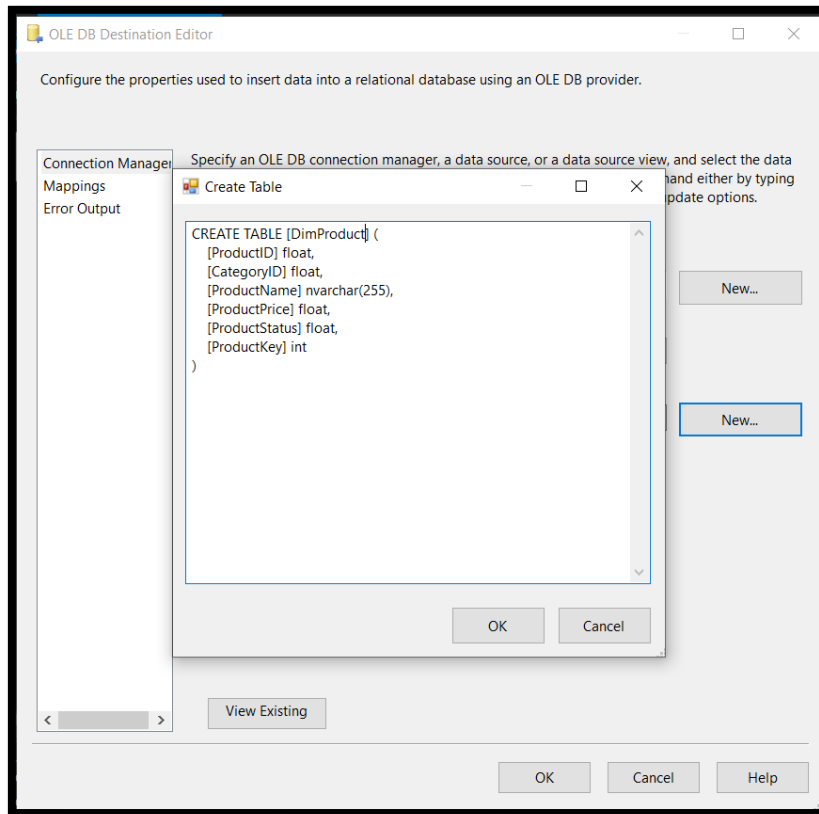


Nhấn vào Script và chọn Edit Script và sửa file main như hình dưới

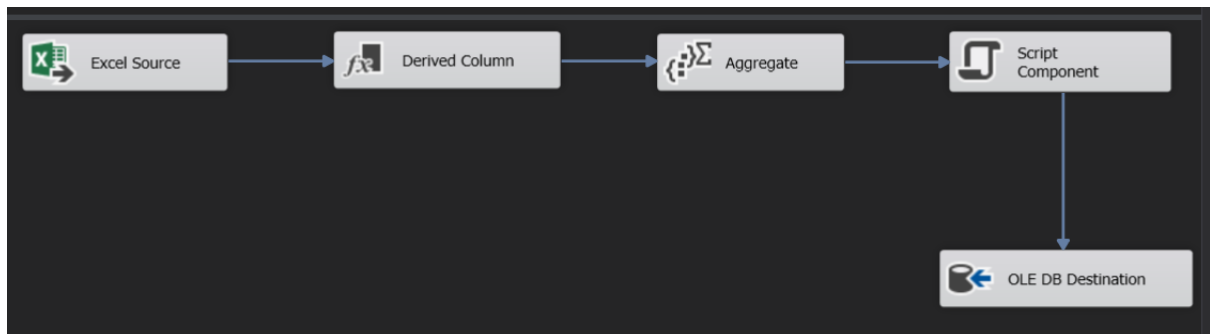


```
main.cs
SC_e10644c6953840318e2cd782840d7089 ScriptMain PreExecute()
82 /// You can remove this method if you don't need to do anything here.
83 /// </summary>
84 public override void PreExecute()
85 {
86     base.PreExecute();
87     /*
88     * Add your code here
89     */
90 }
91
92 /// <summary>
93 /// This method is called after all the rows have passed through this component.
94 ///
95 /// You can delete this method if you don't need to do anything here.
96 /// </summary>
97 public override void PostExecute()
98 {
99     base.PostExecute();
100    /*
101    * Add your code here
102    */
103 }
104 int count = 0;
105 /// <summary>
106 /// This method is called once for every row that passes through the component from Input0.
107 ///
108 /// Example of reading a value from a column in the row:
109 /// string zipCode = Row.ZipCode
110 ///
111 /// Example of writing a value to a column in the row:
112 /// Row.ZipCode = zipCode
113 /// </summary>
114 /// <param name="Row">The row that is currently passing through the component</param>
115 public override void Input0_ProcessInputRow(Input0Buffer Row)
116 {
117     /*
118     * Add your code here
119     */
120     count++;
121     Row.ProductKey = count;
122 }
123
124
125
84 % No issues found Ln: 1 Ch: 1 SPC
```

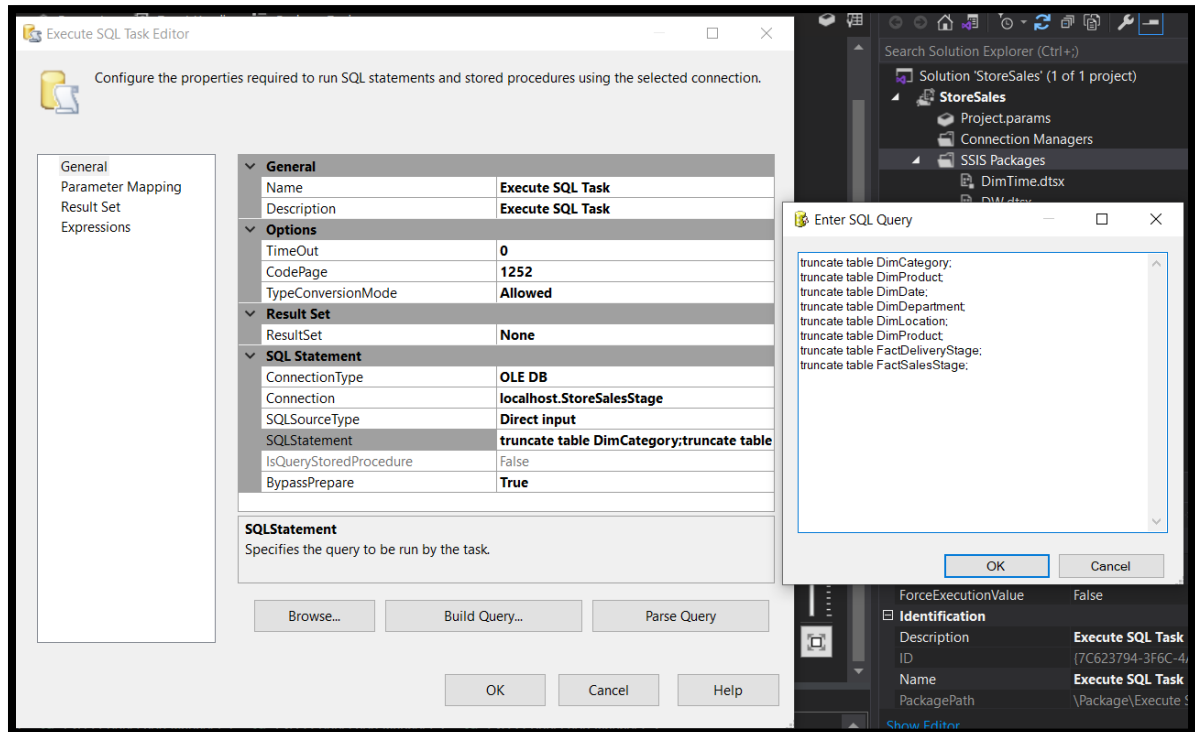
Sau đó Kéo OLE DB Destination Editor vào. Chọn connection ở OLE DB Connection manager, sau đó bấm New, điền vào như sau:



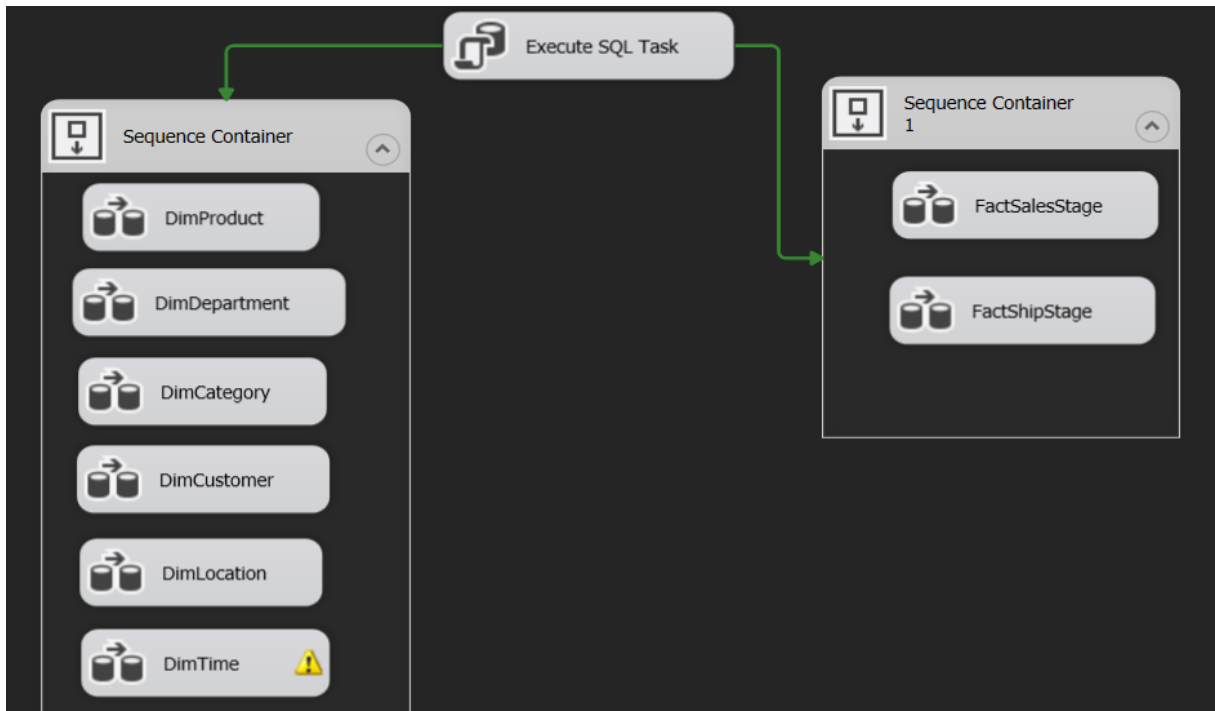
Sau đó nhấn Ok. Ta được DataFlow



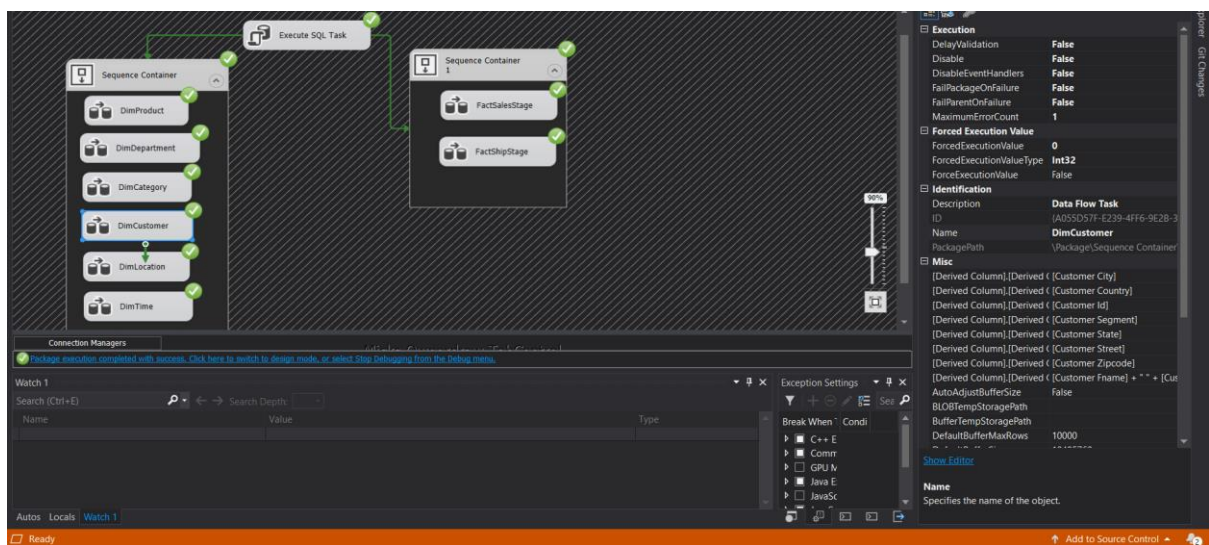
Các Dim và Fact khác ta thực hiện tương tự. Sau đó, Ta tạo Execute SQL Task như sau:



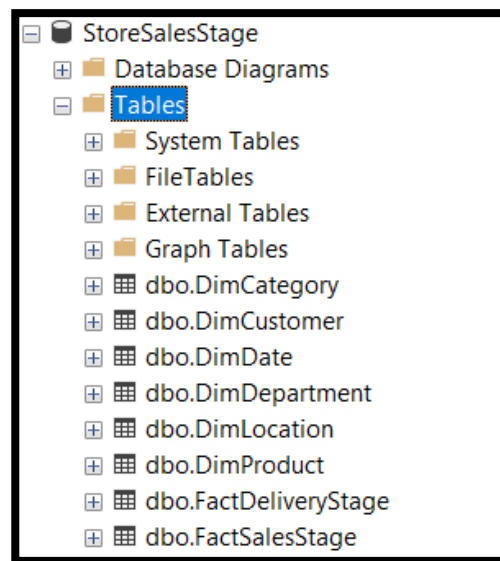
Ta sẽ có được quá trình Stage từ file excel vào Stage như hình bên dưới



Nhấn F5 để thực thi

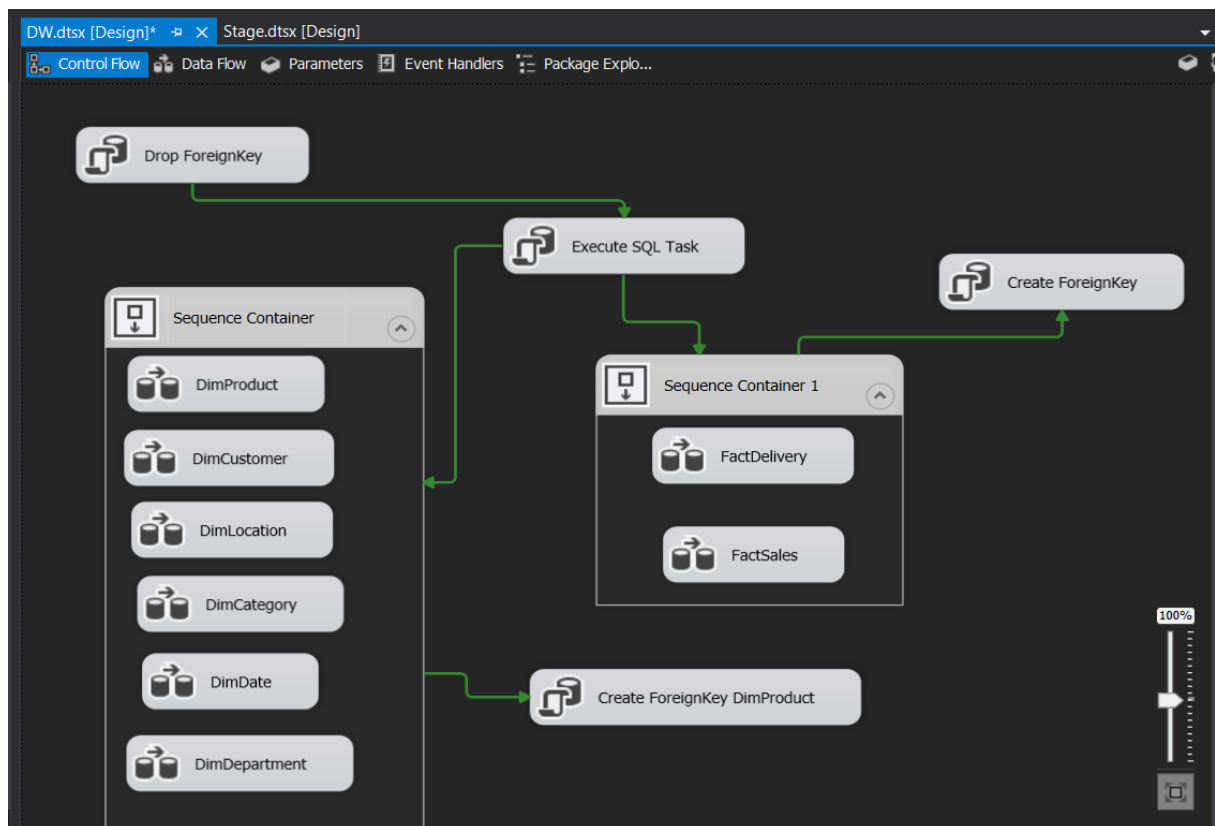


Sau khi thực thi ta được các bảng trong CSDL Stage

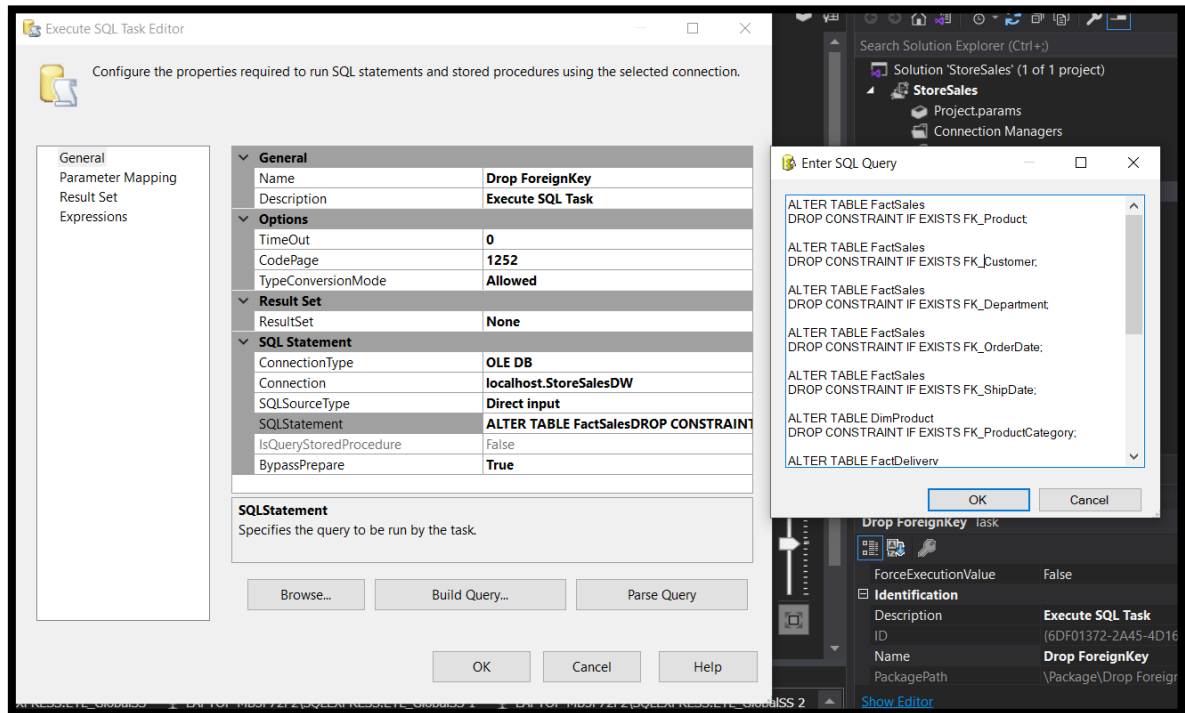


3.3. Tích hợp dữ liệu từ Stage và Warehouse

DataFlow cho cả quá trình

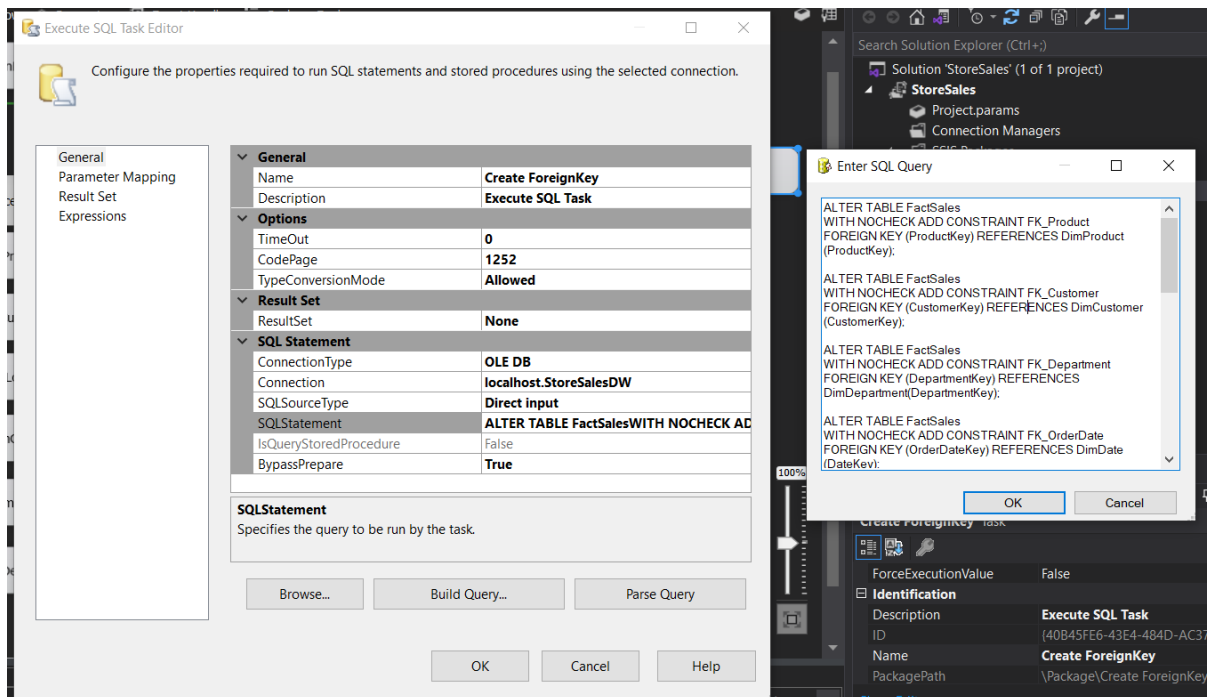


Drop ForeignKey dùng để xóa tất cả các ForeignKey có trước đó

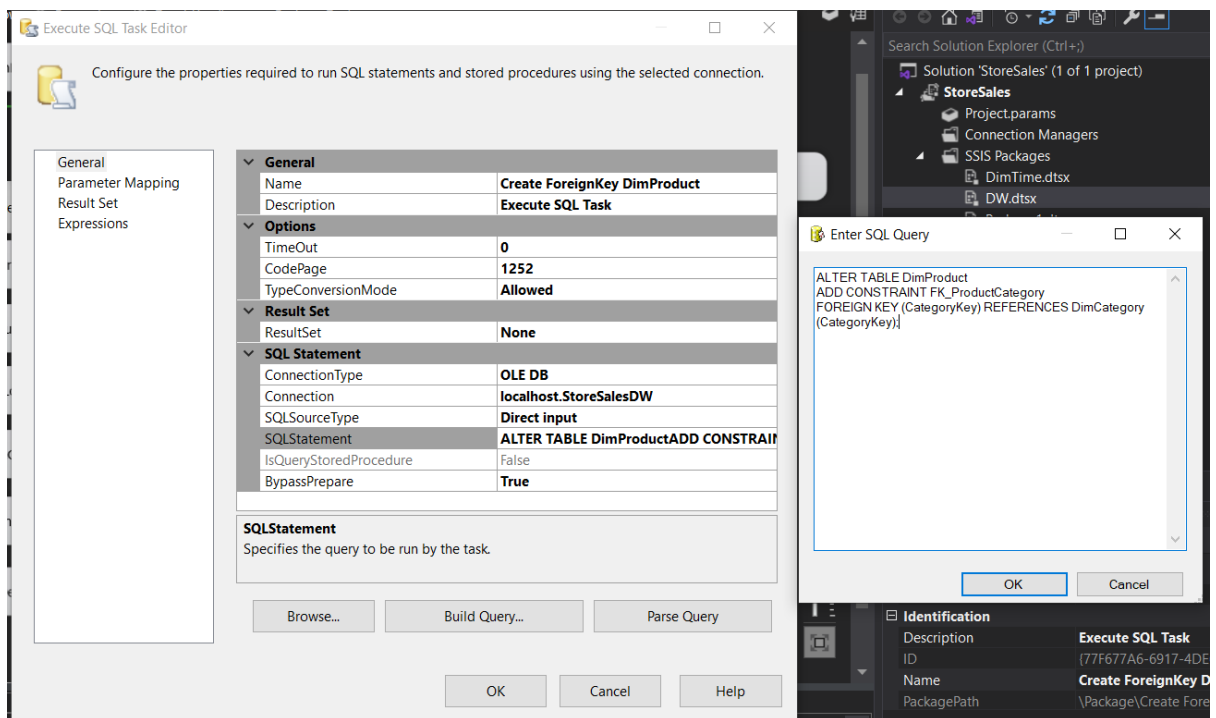


Các DimProduct, DimCustomer, DimLocation, DimCategory, DimDate, DimDepartment, FactDelivery, FactSales dùng để load dữ liệu từ Stage vào Warehouse

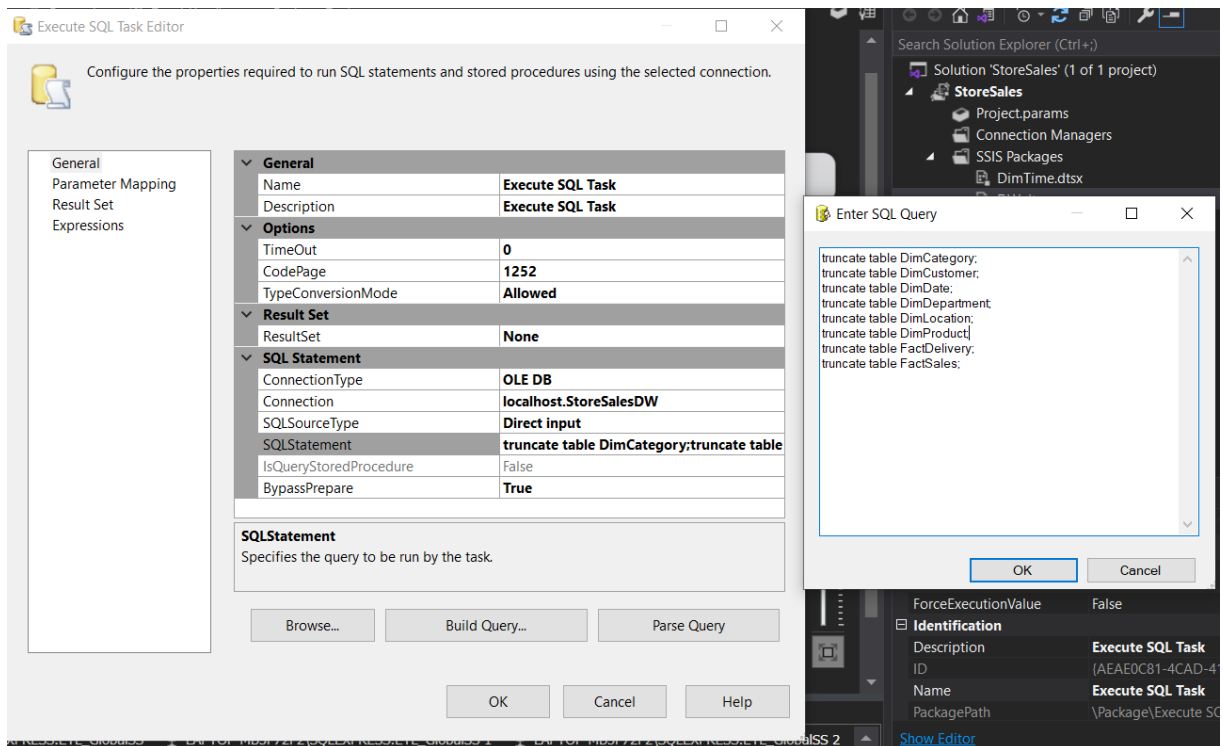
Create ForeignKey dùng để tạo khóa ngoại giữa các bảng Dim và Fact trong kho



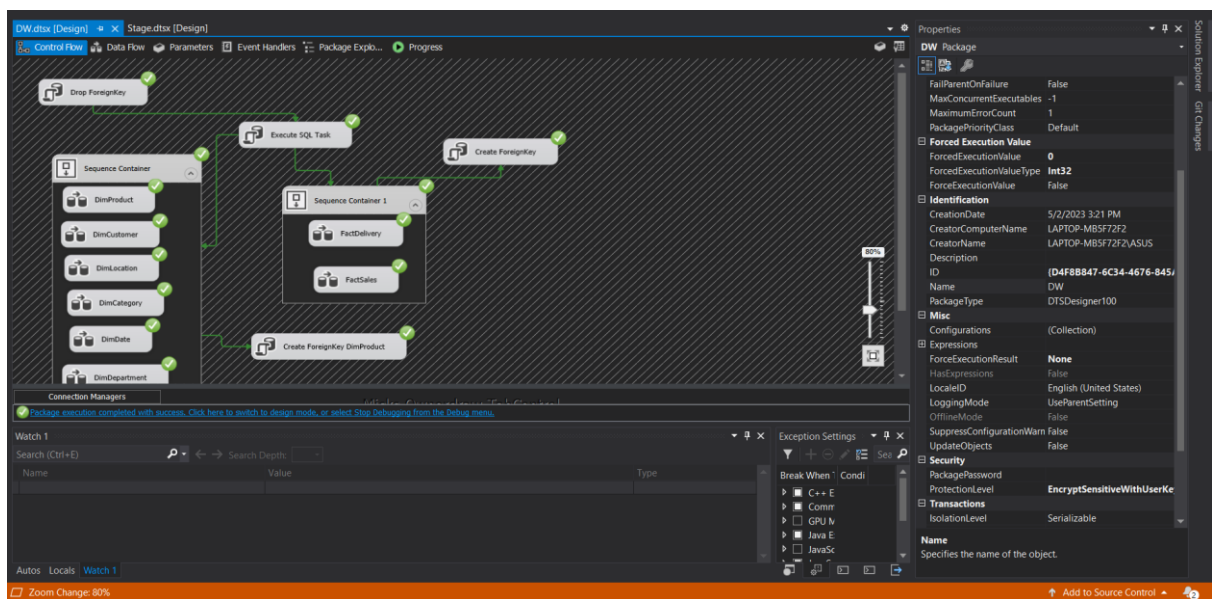
Create ForeignKey DimProduct dùng để tạo khóa ngoại giữa DimProduct và DimCategory trong kho



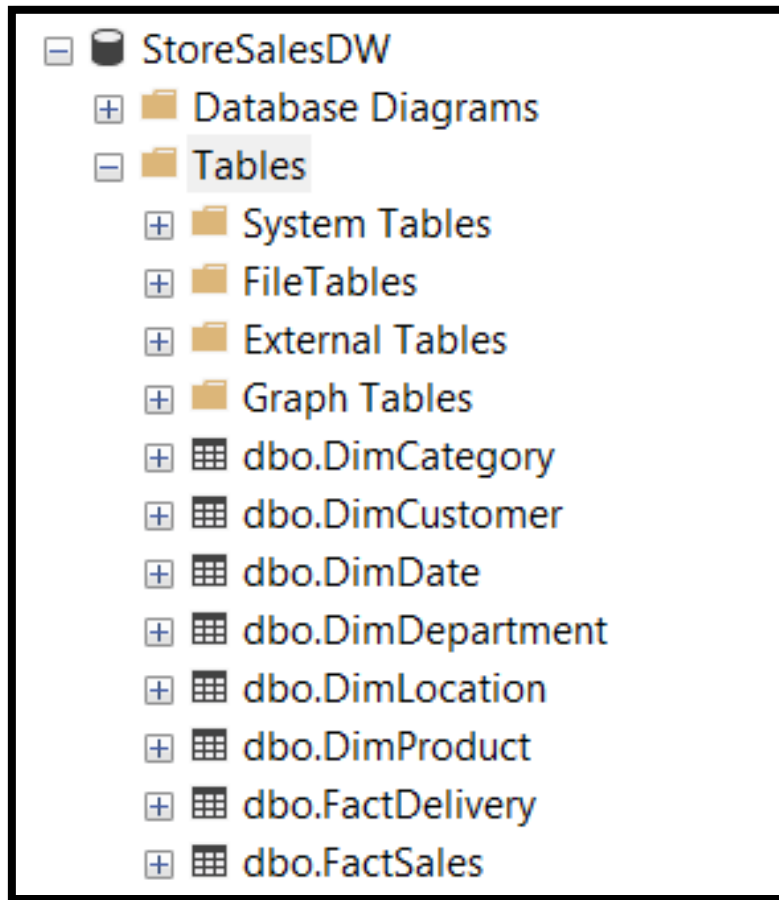
Execute SQL Task dùng để xóa các table nếu đã có trong kho trước đó



Nhấn F5 để thực thi



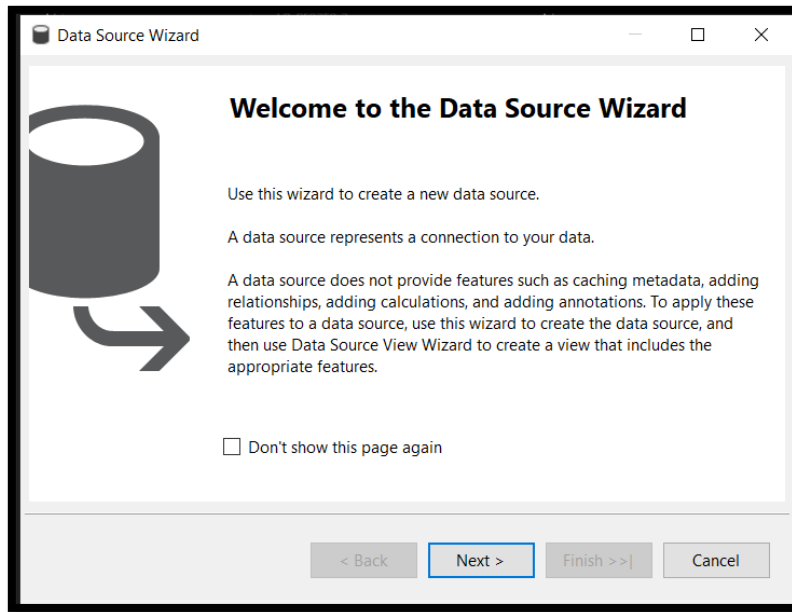
Sau khi thực thi ta được kho dữ liệu trong Database



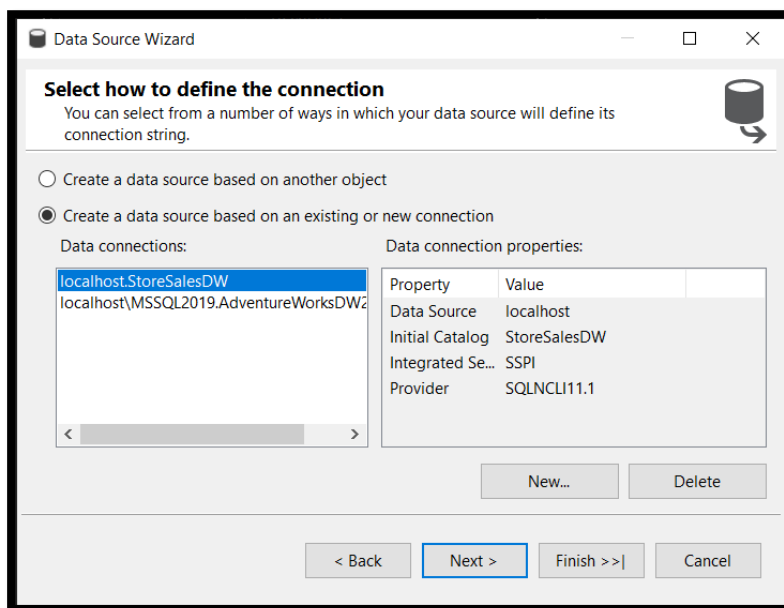
4. Phân tích dữ liệu (SSAS)

4.1. Quá trình xây dựng mô hình

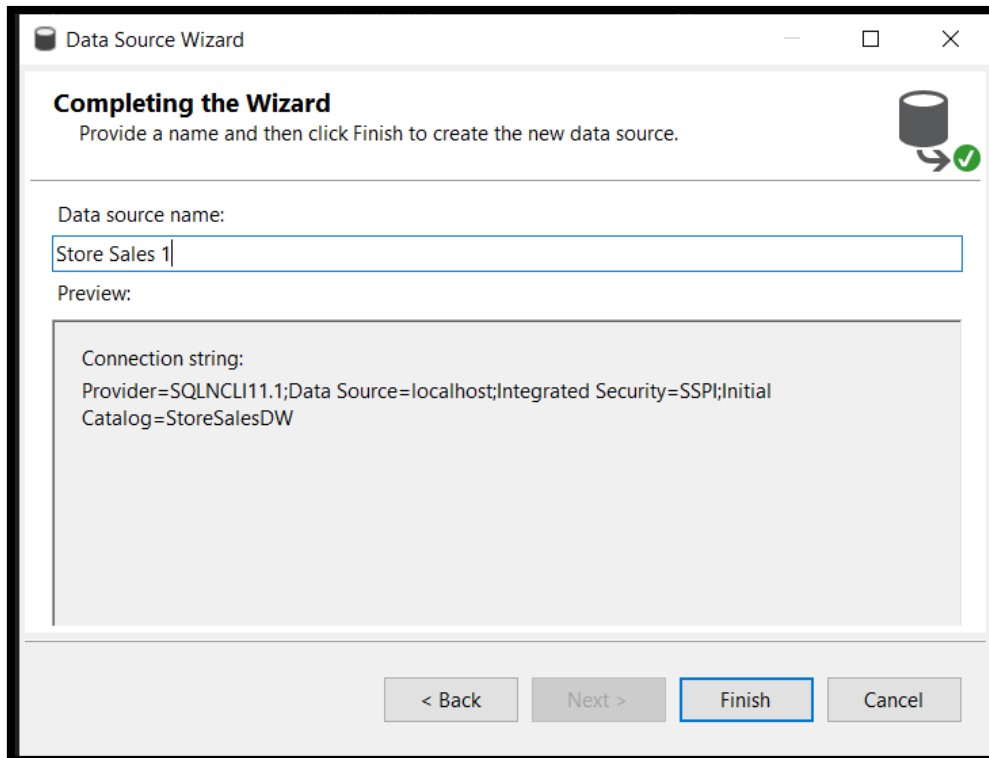
Chọn New DataSource View...



Chọn Next. Nhấn chọn DataSource đã tạo ở trước rồi chọn Next



Sau đó chọn Use the service account → Finish



The screenshot shows the 'Data Source Wizard' window at the 'Completing the Wizard' step. The window title is 'Data Source Wizard'. The main heading is 'Completing the Wizard' with a subtitle 'Provide a name and then click Finish to create the new data source.' and a database icon with a green checkmark. Below this, there is a text input field for 'Data source name:' containing 'Store Sales 1'. Underneath is a 'Preview:' section with a light gray background containing the following text: 'Connection string: Provider=SQLNCLI11.1;Data Source=localhost;Integrated Security=SSPI;Initial Catalog=StoreSalesDW'. At the bottom, there are four buttons: '< Back' (disabled), 'Next >' (disabled), 'Finish' (active/highlighted), and 'Cancel' (disabled).

Data Source Wizard

Completing the Wizard
Provide a name and then click Finish to create the new data source.

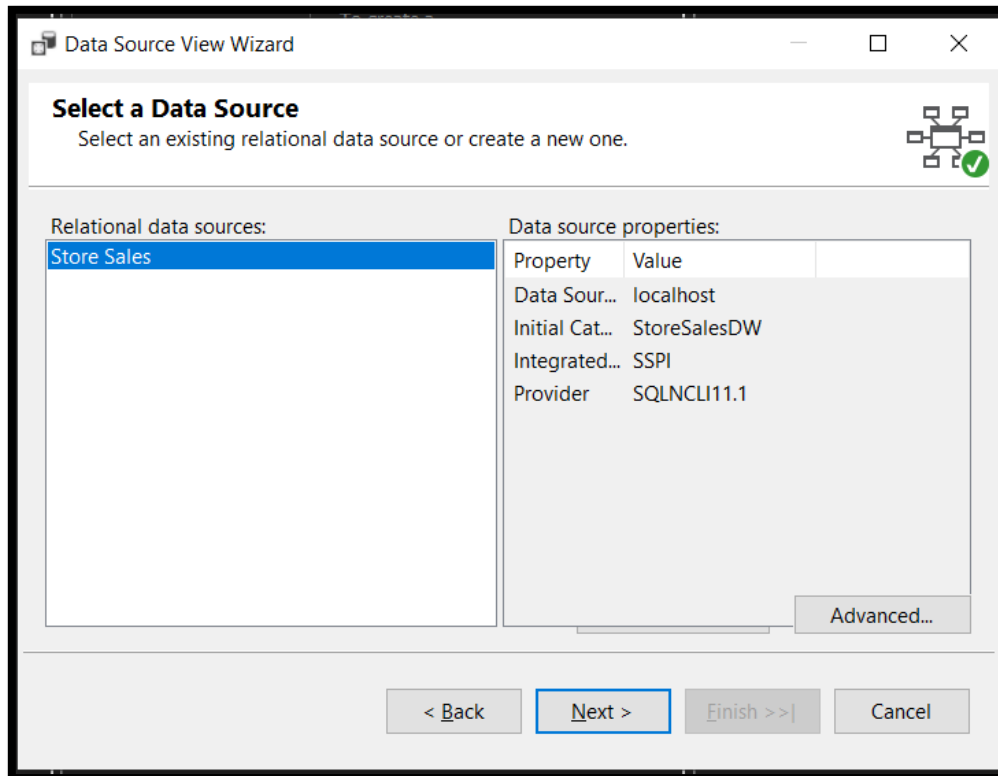
Data source name:
Store Sales 1

Preview:

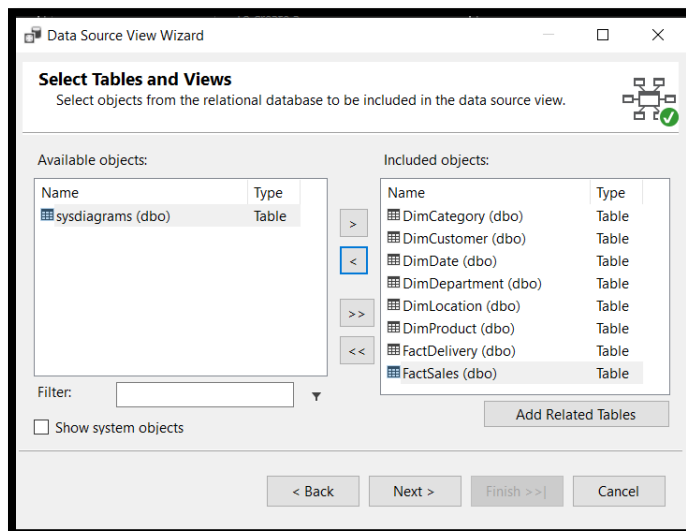
Connection string:
Provider=SQLNCLI11.1;Data Source=localhost;Integrated Security=SSPI;Initial Catalog=StoreSalesDW

< Back Next > **Finish** Cancel

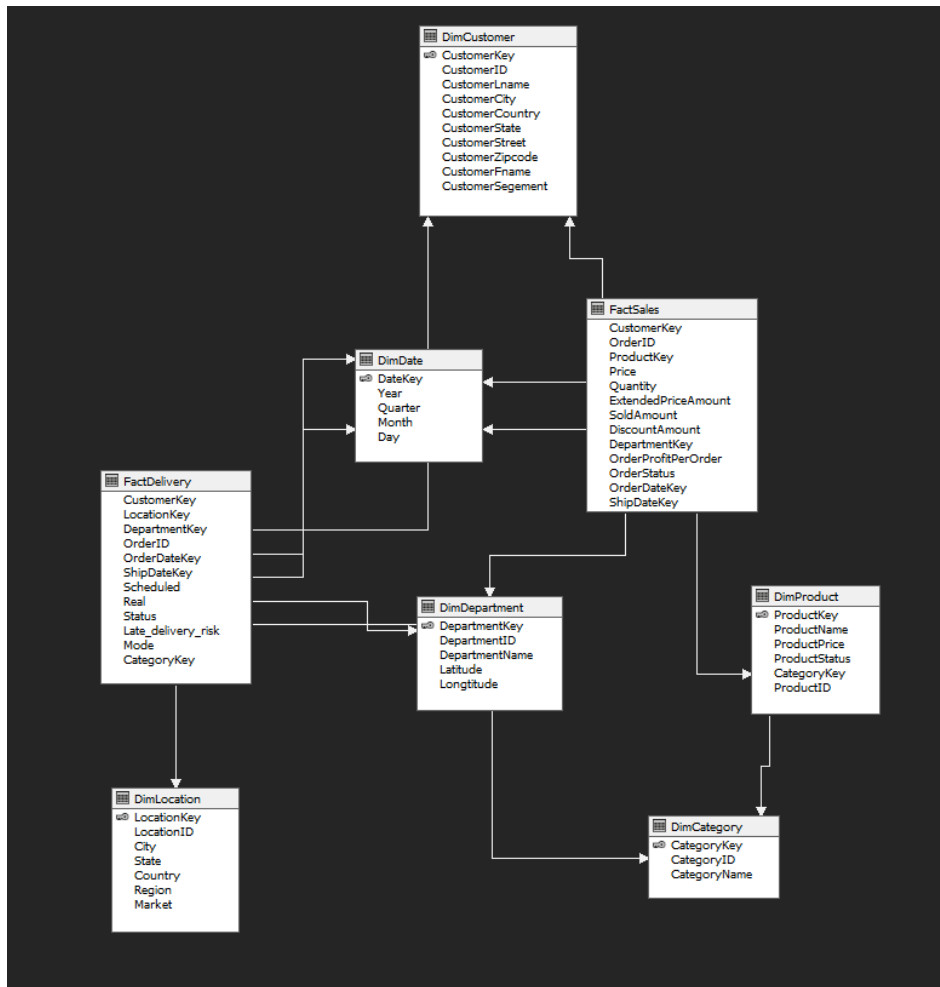
Nhấn chuột phải và Data Source View → Next



Chọn Next → nhấn dấu mũi tên >>



Chọn Next -> Finish

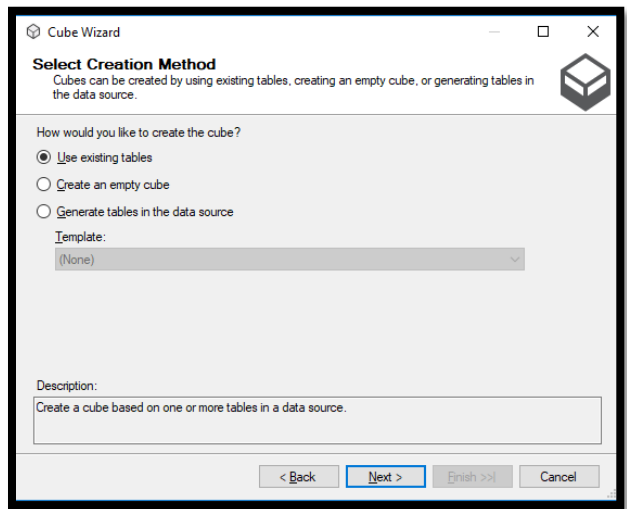


DataSource Views sẽ hiện như hình trên

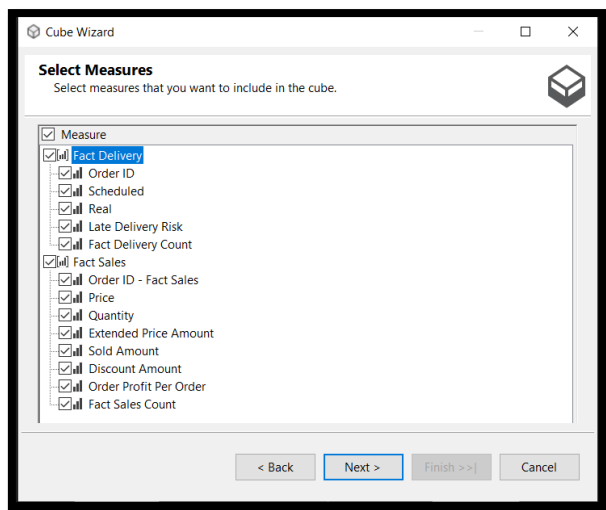
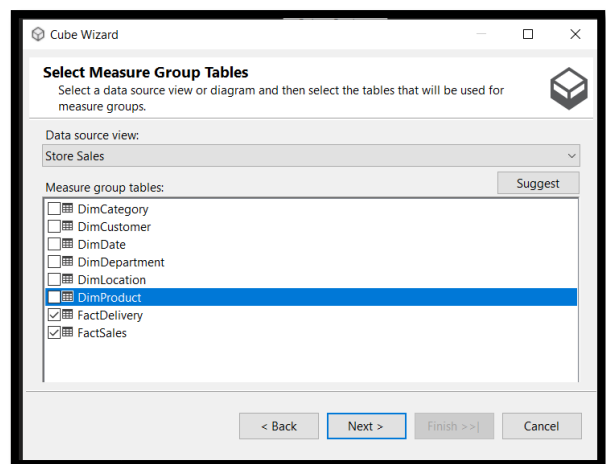
4.2. Quá trình xây dựng khối - Cube

Nhấn chuột phải vào Cubes chọn New Cube... dialog hiện ra và nhấn Next

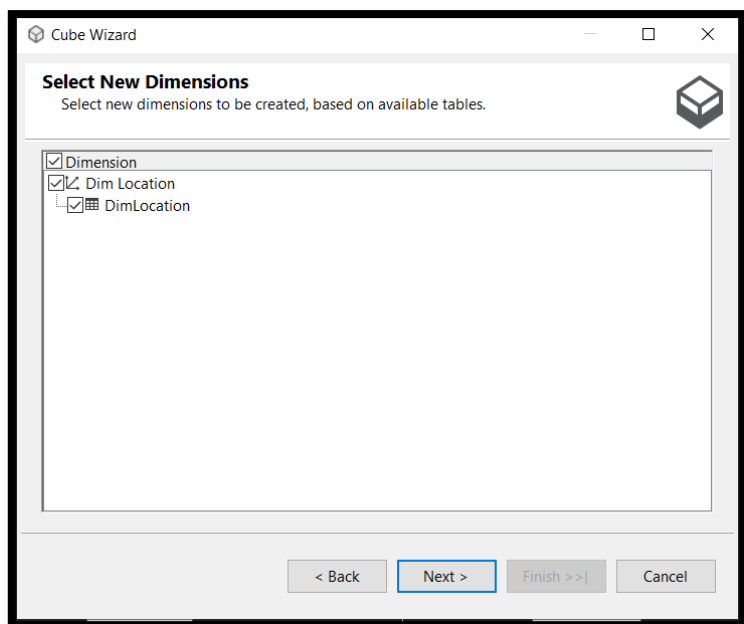
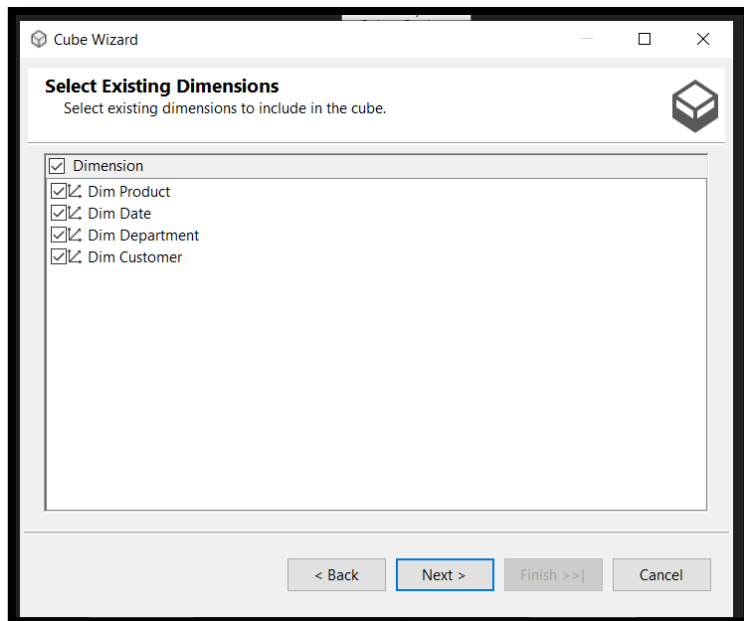
Chọn sử dụng bảng có sẵn (Use existing table) và nhấn Next:



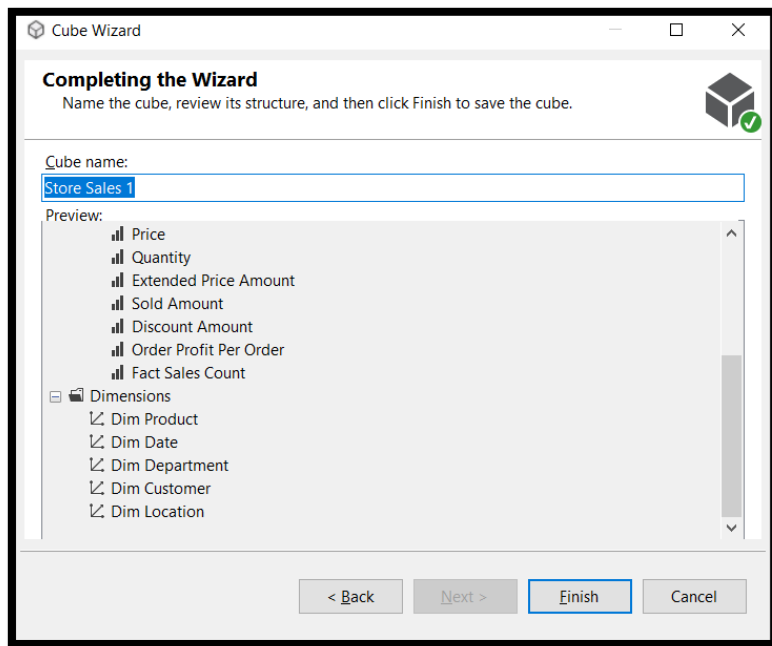
Chọn Measure Group Table và nhấn Next:



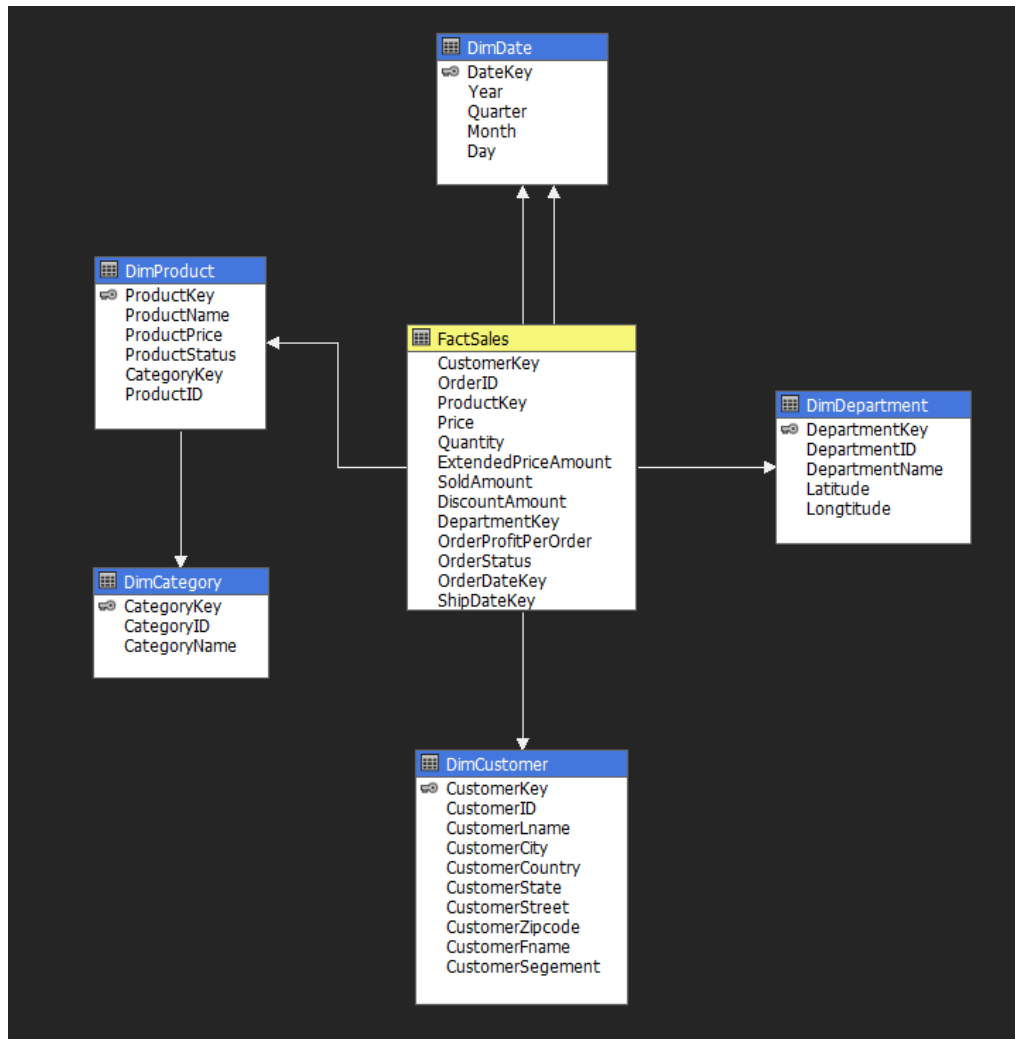
Chọn Demision:



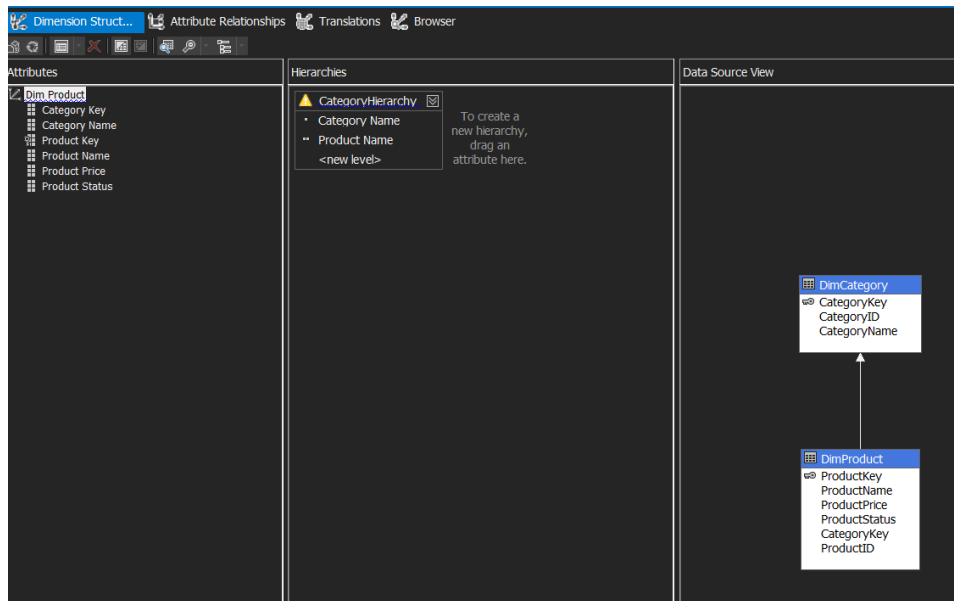
Nhấn Next . Sau đó nhập tên Cube và nhấn Finish để hoàn thành



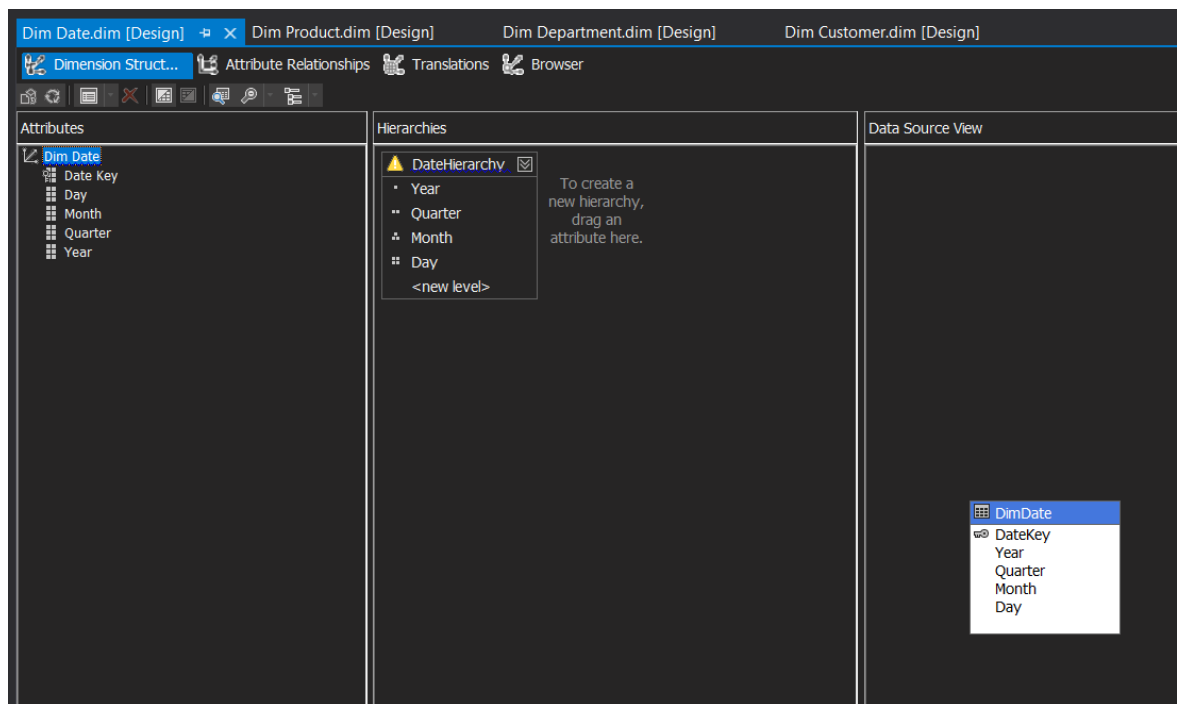
Cubes đã tạo



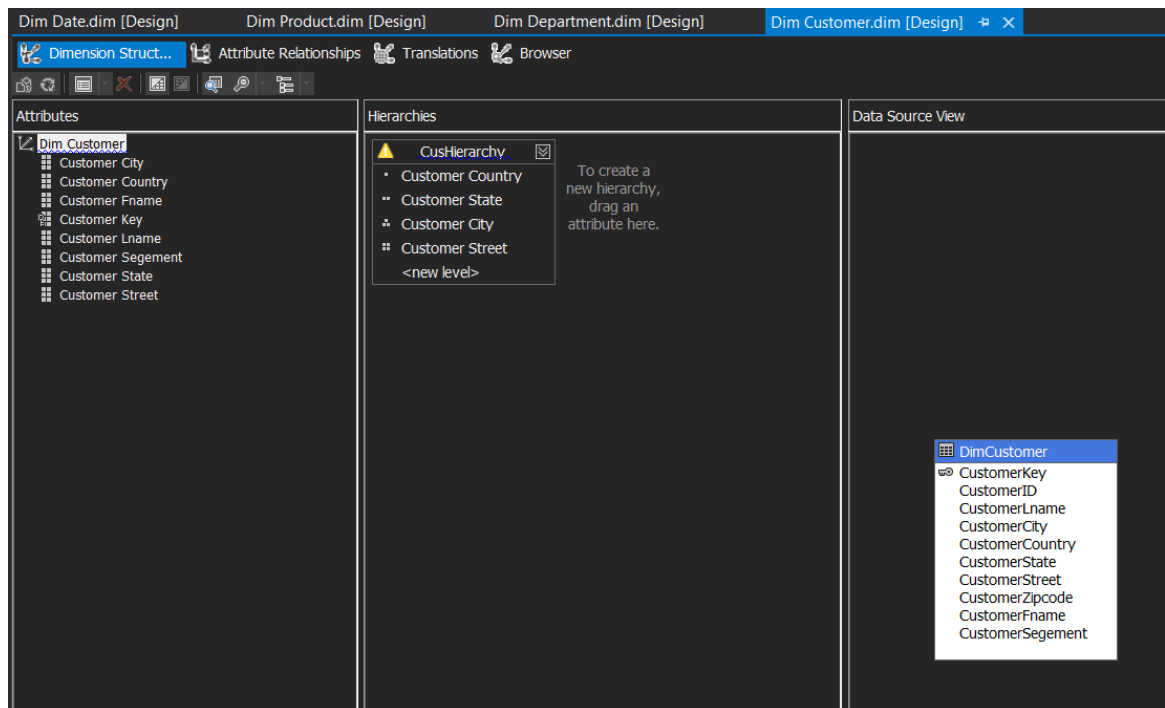
Nháy đúp chuột vào Dim Product.dim trong dialog Dimensions và tạo Hierarchies



Nháy đúp chuột vào Dim Date.dim trong dialog Dimensions và tạo Hierarchies



Nháy đúp chuột vào Dim Customer.dim trong dialog Dimensions và tạo Hierarchies



Lần lượt bấm vào từng Dim vừa tạo Hierarchies và nhấn nút  bên góc trái màn hình

Process Dimension - Dim Product

Object list:

Object Name	Type	Process Options	Settings
<input checked="" type="checkbox"/> Dim Product	Dimension	Process Update	

Remove

Impact Analysis...

Batch Settings Summary

Processing order:

Parallel

Transaction mode:

(Default)

Dimension errors:

(Default)

Dimension key error log path :

(Default)

Process affected objects:

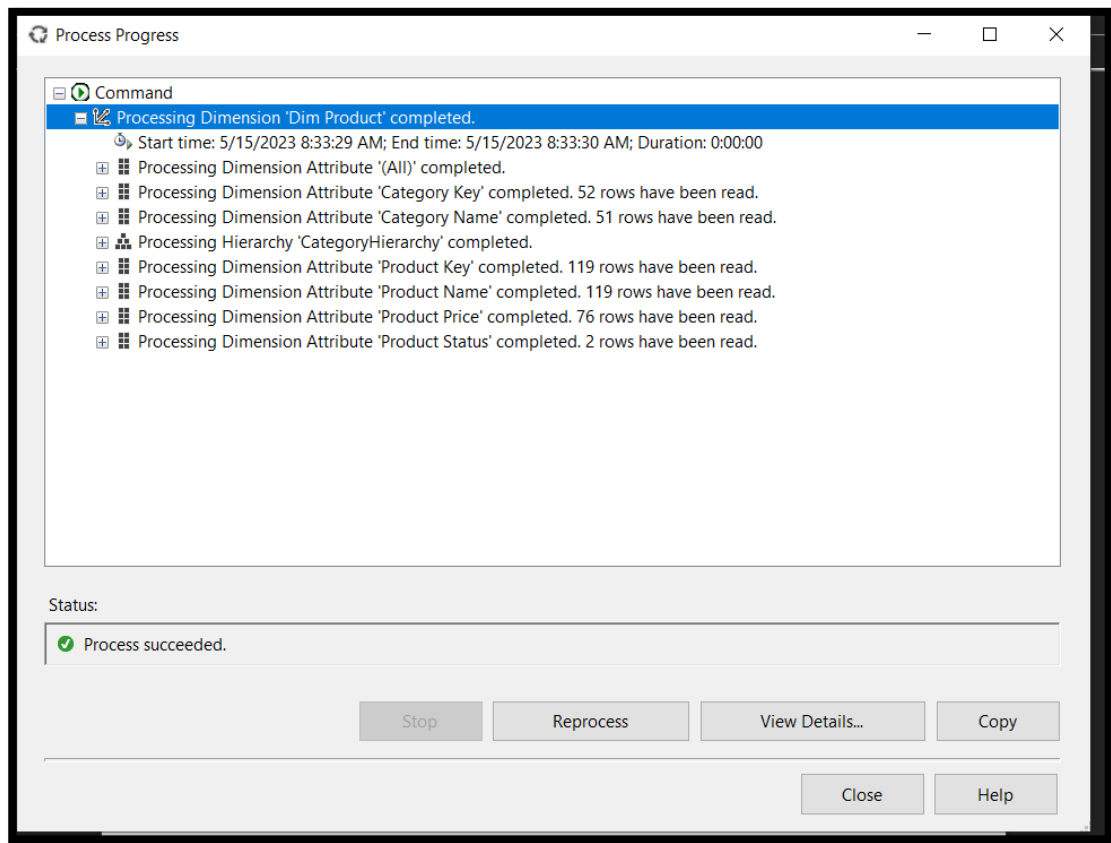
Do not process

Change Settings...

Run...

Close

Nhấn Run. Sau đó nhấn chuột phải vào tên Project và nhấn Deploy



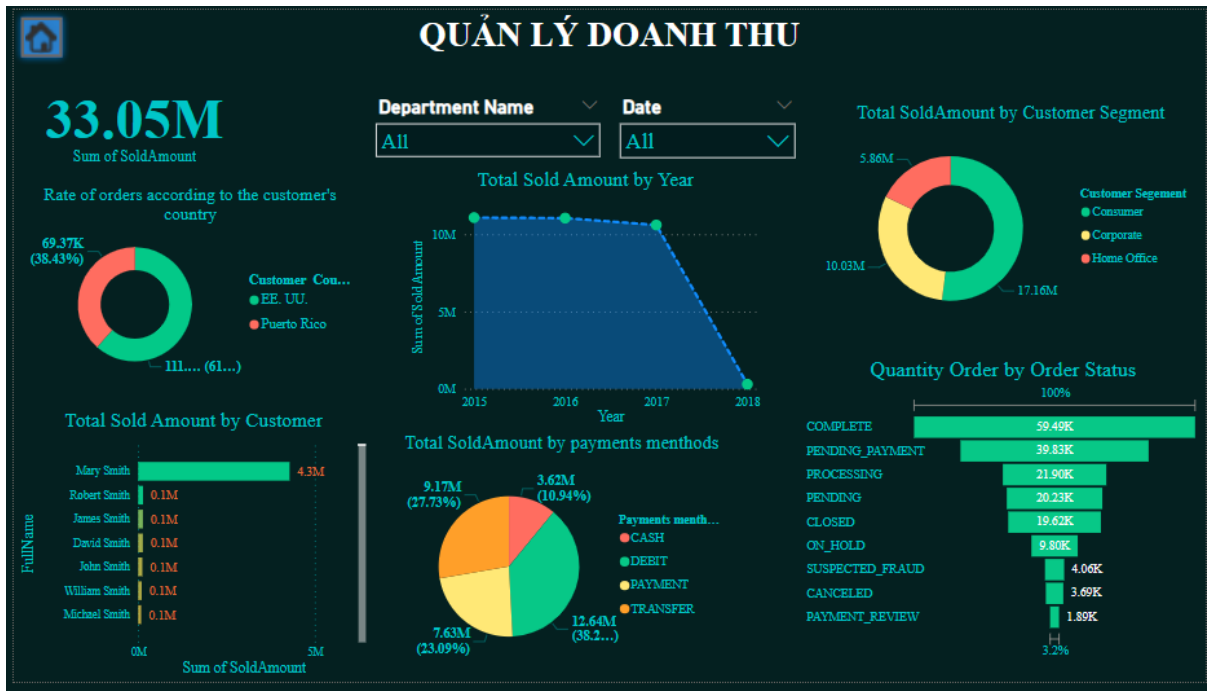
5. Xây dựng Dashboard với PowerBI

5.1. Trang chủ



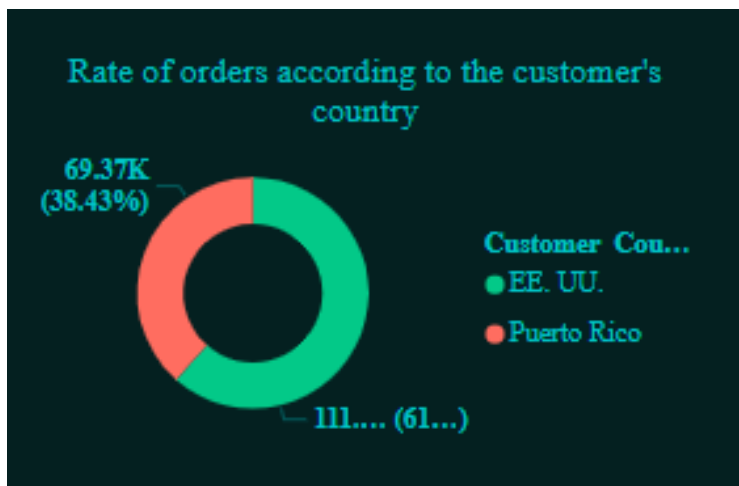
Thông tin trang chủ bao gồm các số liệu về vị trí giao, các danh mục sản phẩm, số lượng sản phẩm, số lượng order, số lượng sản phẩm đã đặt hàng, các cửa hàng bán các sản phẩm. Trang chủ còn bao gồm 3 dashboard: Sales Report, Profit Report và Delivery Report.

5.1.1. Dashboard Sales Report



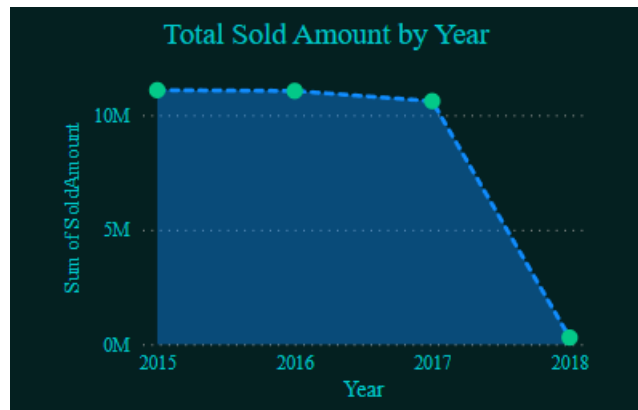
Dashboards Sales Report: gồm doanh số bán hàng và các biểu đồ liên quan đến doanh số và có thể theo dõi theo từng cửa hàng và từng mốc thời gian cụ thể

5.1.1.1. Biểu đồ thể hiện tỷ lệ của Quantity Order theo Customer Country



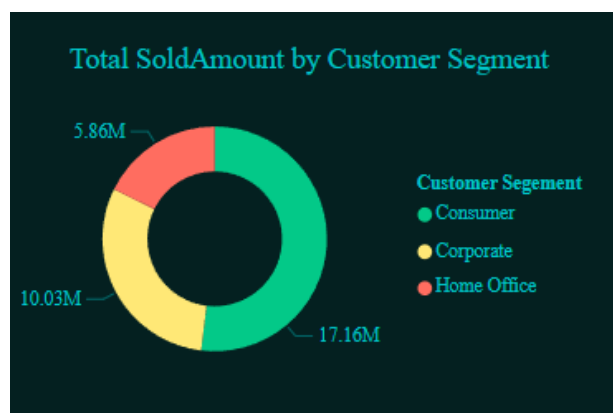
Nhận xét: Ta thấy có 69373 số đơn hàng chiếm 38.43% số đơn hàng được đặt hàng từ Puerto Rico. Và 61.57% số đơn hàng được đặt từ EE. UU. Điều này có thể cho thấy rằng thị trường tập trung chủ yếu vào khách hàng ở EE.UU. Sắp tới có thể mở rộng thị trường sang các quốc gia để phát triển doanh thu.

5.1.1.2. Biểu đồ thể hiện tổng doanh thu cửa hàng qua các năm



Nhận xét: Ta thấy hầu hết từ 2015 – 2017 thì doanh thu năm ở mức ổn định hơn 10M không có biến đổi nhiều. Nếu trong những năm tiếp theo không có sự phát triển thì cần phải xem xét lại chất lượng sản phẩm và cách tiếp cận với khách hàng

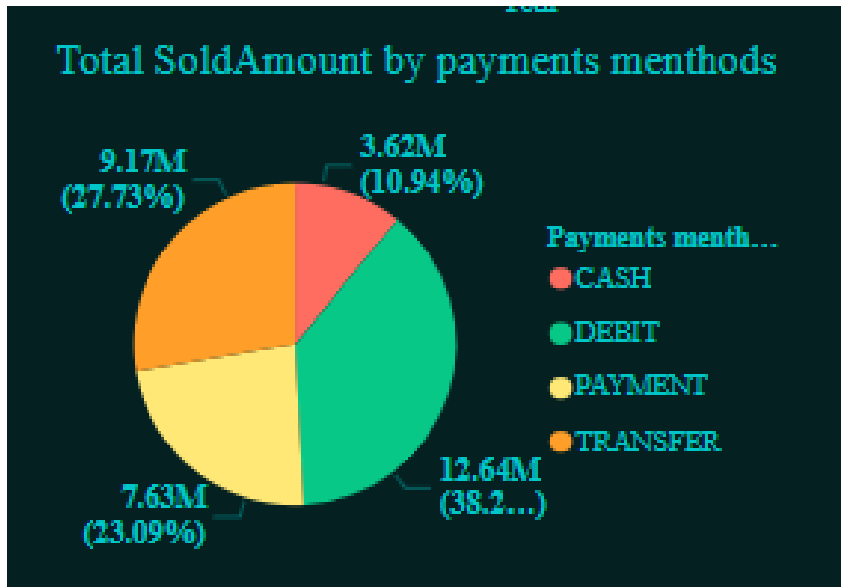
5.1.1.3. Biểu đồ thể hiện doanh thu theo từng loại khách hàng



Nhận xét: Ta thấy khách hàng thông thường chiếm doanh thu lớn nhất với 17.16M chiếm 51,93% tiếp sau đó là công ty với 10,03M chiếm 30,35%. Và phần còn lại là các văn phòng. Điều này cho thấy rằng doanh nghiệp của bạn có thể đang tập trung chủ yếu vào khách hàng cá nhân và các công ty. Để tăng doanh thu và thì cần phải đa dạng hóa khách hàng khác và cần tập trung vào việc cải thiện dịch vụ

và sản phẩm để giữ chân được các khách hàng hiện tại và thu hút được khách hàng mới.

5.1.1.4. Biểu đồ thể hiện doanh thu theo từng phương thức thanh toán



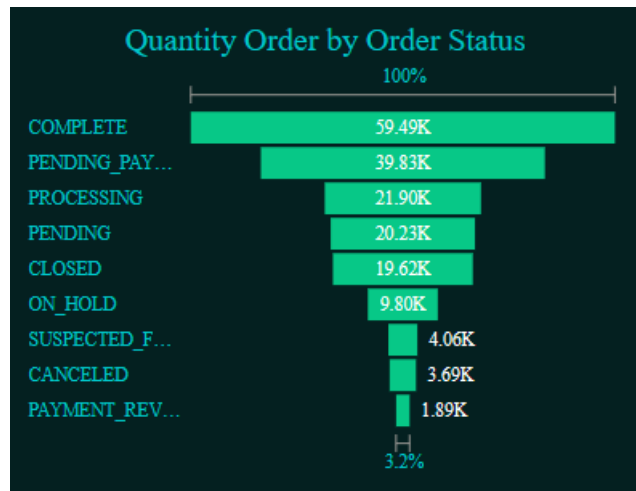
Nhận xét: Ta thấy khách hàng thanh toán qua ví trả sau chiếm doanh thu nhiều nhất với doanh thu 12.64M chiếm 38,24%. Điều này cho thấy rằng phương thức thanh toán này đang trở thành một lựa chọn phổ biến trong việc thanh toán cho sản phẩm hoặc dịch vụ. Tiếp theo là phương thức chuyển khoản chiếm 27,73% tiếp theo là trả qua thẻ với 7.63M chiếm 23,09% và phần còn lại là tiền mặt chiếm 10,94%.

5.1.1.5. Biểu đồ thể hiện doanh thu các khách hàng mua hàng nhiều nhất



Nhận xét: Ta thấy khách hàng Mary Smith mua hàng nhiều nhất chiếm 4.3M doanh thu của Công ty. Hầu hết các khách hàng hiện tại không có sự chênh lệch nhiều. Điều này cho thấy cần có thay đổi chiến lược thu hút với các khách hàng hiện tại và các khách hàng tương lai.

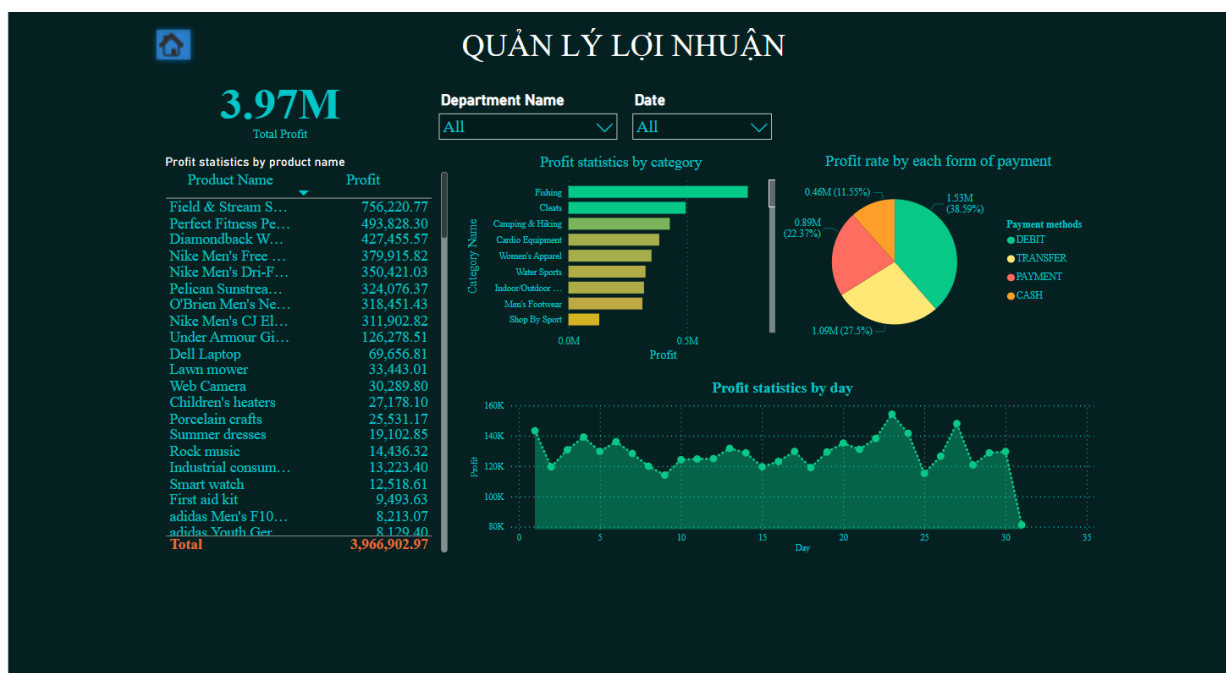
5.1.1.6. Biểu đồ thể hiện số lượng đơn đặt hàng theo trạng thái đơn hàng



Nhận xét: Hầu hết số đơn hàng đã được hoàn thành với 59.49K số đơn hàng. Tiếp theo chờ thanh toán chiếm 39.83K số đơn hàng. Trong khi đó trạng thái đơn hàng hủy chiếm 3.69K số đơn hàng và tiếp theo là số đơn hàng hủy và số đơn hàng tài khoản chưa xác minh. Điều này cho thấy rằng tỷ lệ hủy đơn và các trường hợp

nhì ngờ gian lận là khá thấp trong doanh nghiệp của bạn. Tuy nhiên, để tăng hiệu quả kinh doanh doanh nghiệp nên tập trung vào việc giảm thiểu số lượng đơn hàng bị hủy và đảm bảo rằng quá trình thanh toán được hoàn tất nhanh chóng để khách hàng có trải nghiệm mua hàng tốt hơn.

5.1.2. Dashboard Profit Report



Dashboard Profit Report: gồm lợi nhuận đơn hàng và các biểu đồ liên quan đến lợi nhuận và có thể theo dõi theo từng cửa hàng và từng tháng năm cụ thể.

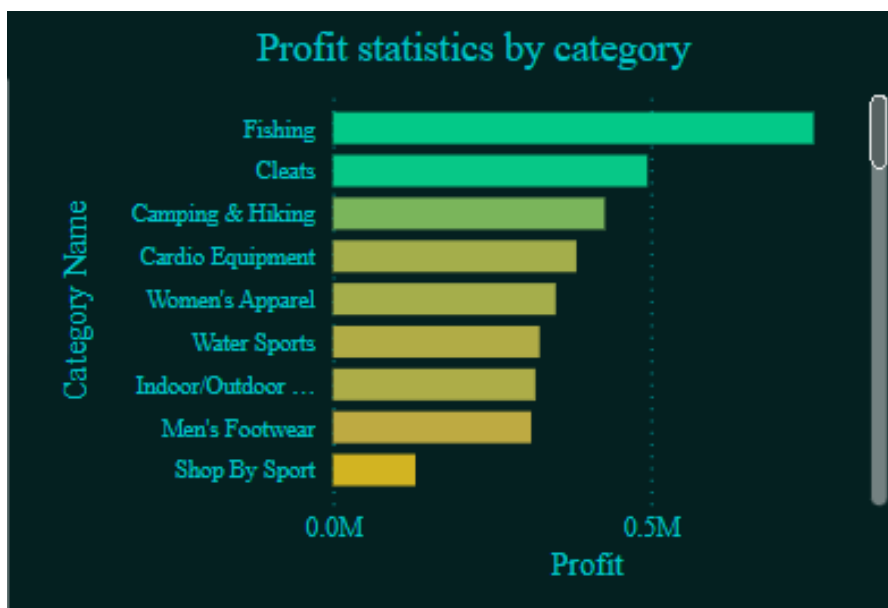
5.1.2.1. Bảng xếp hạng lợi nhuận của các sản phẩm được bán

Ta thấy Children's heaters có lợi nhuận cao nhất với 27178. Ta thấy cần thu hút các sản phẩm khác cho các khách hàng mới, khách hàng tiềm năng để tăng doanh thu lợi nhuận.

Profit statistics by product name

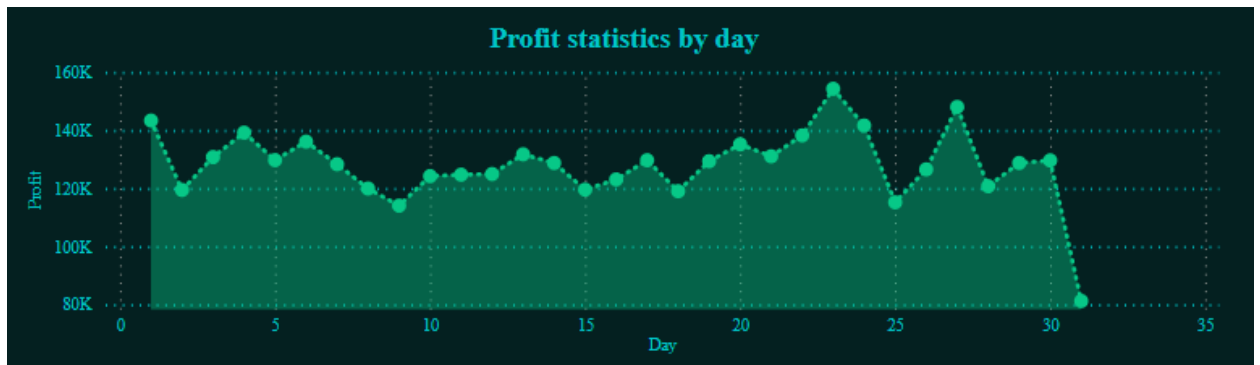
Product Name	Profit
Children's heaters	27,178.10
Porcelain crafts	25,531.17
Summer dresses	19,102.85
Rock music	14,436.32
Industrial consum...	13,223.40
Smart watch	12,518.61
First aid kit	9,493.63
adidas Men's F10...	8,213.07
adidas Youth Ger...	8,129.40
LIJA Women's E...	6,752.33
Titleist Pro V1x ...	6,671.09
DVDs	6,655.43
Titleist Pro V1 Hi...	6,093.15
Nike Men's Comf...	5,747.98
Titleist Pro V1x ...	4,592.28
Under Armour W...	4,469.77
Under Armour H...	4,067.61
Under Armour W...	4,030.38
Under Armour M...	3,745.34
Titleist Pro V1x ...	3,699.61
TYR Boys' Team ...	3,658.98
Total	3,966,902.97

5.1.2.2. Biểu đồ thể hiện tổng lợi nhuận theo từng doanh mục sản phẩm



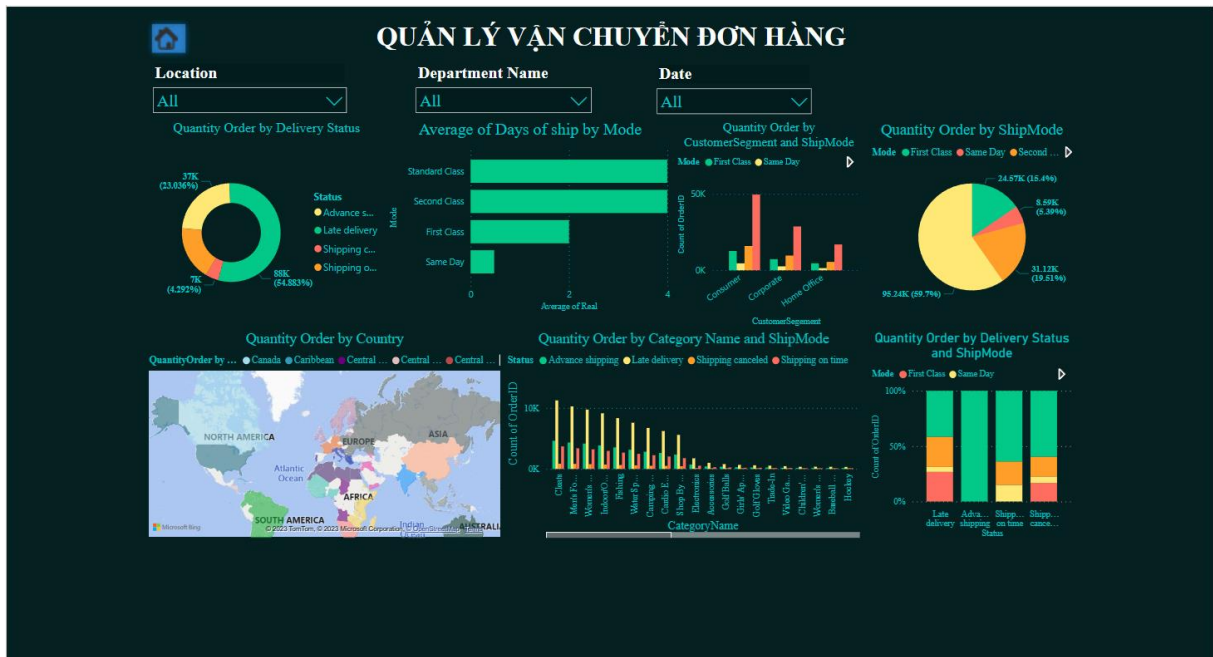
Nhận xét: Ta thấy lợi nhuận của Fishing là cao nhất với 756220.77. Tiếp theo là các mặt hàng như Cleats, Camping & Hiking,... Cần đẩy mạnh tiếp tục các doanh mục hàng với lợi nhuận cao và cần có thu hút các mặt hàng lợi nhuận đến các khách hàng khác hoặc có thể cân nhắc bỏ mặt hàng đó.

5.1.2.3. Biểu đồ theo dõi lợi nhuận theo tháng



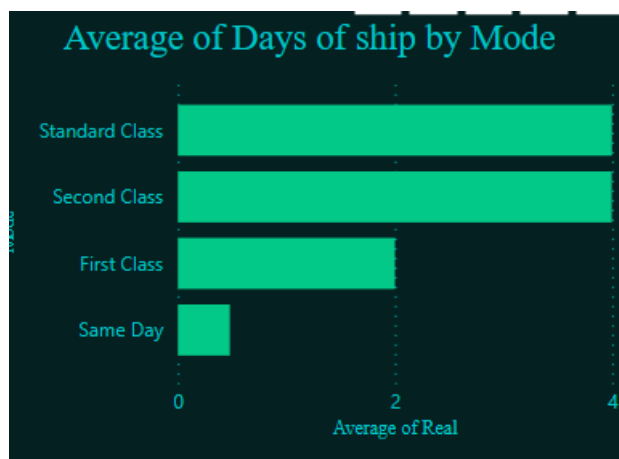
Nhận xét: Hầu hết lợi nhuận các ngày đều trên 120K. Có ngày lợi nhuận cao nhất gần 160K. Còn một vài ngày lợi nhuận dưới 120K. Cần đẩy mạnh đẩy chiến dịch, thu hút các khách hàng để tăng lợi nhuận.

5.1.3. Dashboard Delivery Report



Dashboard Delivery Report: gồm các phương thức vận chuyển đơn hàng và các biểu đồ liên quan đến ngày giao, số ngày giao và có thể theo dõi theo từng cửa hàng và từng tháng năm và vị trí cụ thể.

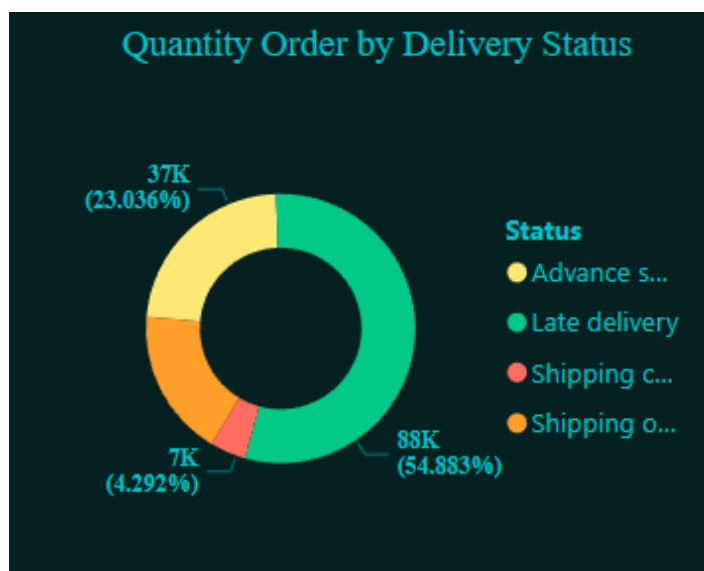
5.1.3.1. Biểu đồ thể hiện số ngày trung bình giao hàng theo phương thức vận chuyển



Nhận xét: Ta thấy phương thức vận chuyển standard class và second class trung bình có 4 ngày để hoàn thành đơn hàng. Còn First Class nhanh hơn chỉ 2 ngày và

Same Day chưa đầy 1 ngày đã giao hàng. Nếu có thể thì cần giảm thời gian second class để thu hút nhiều khách hàng hơn.

5.1.3.2. *Biểu đồ thể hiện tỷ lệ tổng số đơn hàng theo trạng thái vận chuyển*



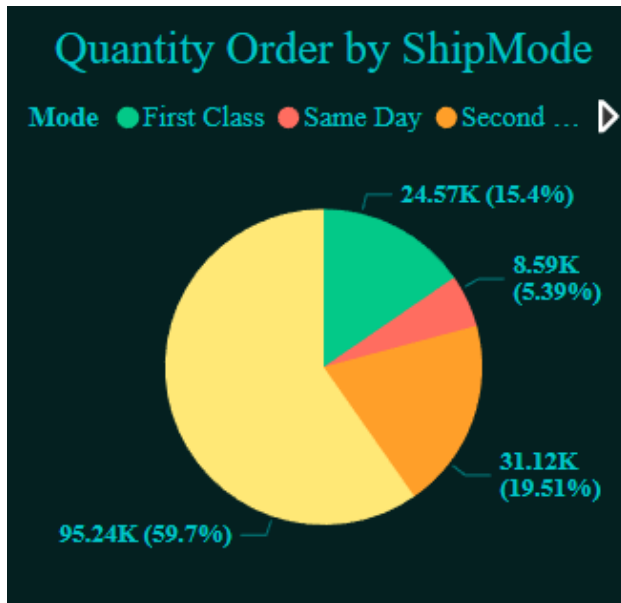
Nhận xét: Ta thấy số đơn hàng hoàn thành giao trễ rất nhiều chiếm 54.883%, tiếp theo là đơn hàng giao trước thời gian chiếm 23.036%, tiếp theo là giao hàng đúng thời gian với 28K số hóa đơn chiếm 17.79% và cuối cùng là đơn hàng hủy chiếm 4.292% trên tổng số đơn hàng. Để tăng chất lượng dịch vụ cần giảm số đơn hàng giao trễ vì làm ảnh hưởng rất nhiều đến chất lượng dịch vụ và cải thiện thêm nhiều đơn hàng giao trước thời gian.

5.1.3.3. *Biểu đồ thể hiện số lượng đơn đặt hàng theo loại khách hàng và phương thức vận chuyển*



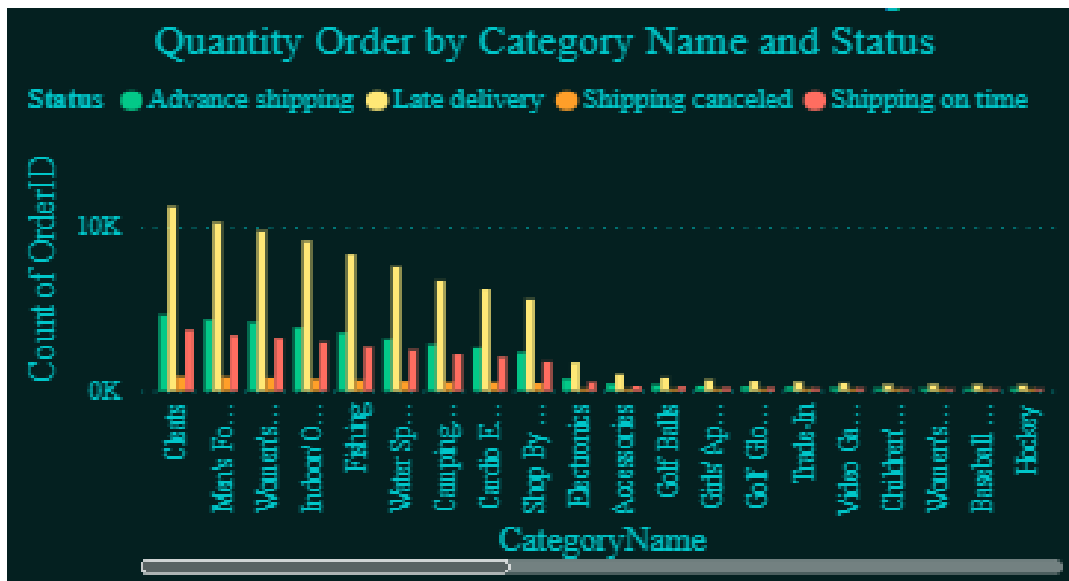
Nhận xét: Ta thấy hầu hết các loại khách hàng từ Consumer đến Home office đều đặt theo Standard Class là nhiều nhất với 49572 cho Consumer tiếp theo Corporate với 28726 và cuối cùng 16944. Tiếp theo Second Class với 15935 cho Consumer 9671 cho Corporate và 5514 cho Home Office, tiếp theo First Class, Same Day cũng giống như vậy.

5.1.3.4. *Biểu đồ thể hiện số lượng theo phương thức vận chuyển*



Nhận xét: Ta thấy số đơn hàng Standard class với 95.24K chiếm 59.7%, tiếp theo là số đơn hàng đặt theo Second class giao chiếm 19.51%, tiếp theo là First class với 24.57K số hóa đơn chiếm 15.4% và cuối cùng là Same Day chiếm 5.39% trên tổng số đơn hàng. Điều này cho thấy rằng khách hàng của doanh nghiệp của thường không có nhu cầu giao hàng nhanh và đa số chọn các phương thức vận chuyển tiêu chuẩn hoặc có thời gian giao hàng chậm hơn để giảm thiểu chi phí vận chuyển.

5.1.3.5. *Biểu đồ thể hiện số lượng đơn đặt hàng theo doanh mục sản phẩm và trạng thái đơn hàng*



Nhận xét: Hầu hết các doanh mục sản phẩm đều có số đơn hàng trễ nhiều như là Cleats với hơn 10K số đơn đặt hàng. Và sự phân bố trạng thái giao hàng của các doanh mục đều khá giống nhau từ giao hàng trễ, giao hàng trước hạn, giao hàng đúng hạn, và đơn hàng hủy. Cần giảm hạn chế thời gian giao hạn trễ để không ảnh hưởng đến trải nghiệm khách hàng và để gia tăng thêm.

5.1.3.6. *Biểu đồ thể hiện số lượng đơn hàng theo trạng thái giao hàng và phương thức vận chuyển*

Nhận xét: Ta thấy ở mục giao hàng trước hạn (Advance shipping) thì 100% đều là đặt hàng Standard Class (giao hàng tiêu chuẩn), tiếp theo ở giao hàng trễ (Late delivery) có Standard Class chiếm 41.48%, Second Class chiếm 27.29%, Same Day chiếm 4.49% và phần còn lại First class chiếm 26.775%. Tiếp theo ở giao hàng đúng hạn có chiếm tỉ lệ như sau: 63.81% cho Standard Class, 21.14%



second class và phần còn lại Same Day. Tương tự thế ở vận chuyển hủy có 59.5% Standard, 17.78% cho second, 5.78% Same Day và 16.82% cho First Class.

5.1.3.7. Biểu đồ thể hiện số lượng đặt hàng đến từng quốc gia



Nhận xét: Theo biểu đồ ta thấy các đơn hàng vận chuyển hầu hết đến các khu vực trên thế giới và khu vực West Euro có số lượng đơn hàng giao đến đó nhiều nhất đặc biệt Francia và kế tiếp khu vực theo Central America đặc biệt là Mexico. Cần phải đẩy mạnh thu hút khách hàng và có các chiến lược đến các thị trường còn ít đơn đặt hàng.

6. Kết luận

6.1. Kết quả đạt được

Nhóm đã tìm hiểu và vận dụng kiến thức về xây dựng kho dữ liệu đạt được các kết quả như sau :

- ❖ Nắm vững kiến thức và có thể vận dụng, xây dựng một kho dữ liệu hoàn chỉnh dùng để khai thác và phân tích dữ liệu.
- ❖ Trang bị kiến thức về các công cụ SSIS, SSAS.
- ❖ Xây dựng được kho dữ liệu hoàn chỉnh.
- ❖ Hiểu và sử dụng Power BI để phân tích kho dữ liệu.

6.2. Hạn chế

Do trong quá trình học, nhiều kiến thức còn chưa tiếp thu sâu, nên trong quá trình nhóm còn gặp nhiều hạn chế:

- ❖ Quá trình xây dựng kho dữ liệu chưa được tối ưu.
- ❖ Quá trình phân tích và xây dựng dashboard chưa thật sự tối ưu.

6.3. Bảng phân công nhiệm vụ

STT	MSSV	Họ tên	Nội dung	% hoàn thành
1	20133025	Nguyễn Thanh Bình	Viết báo cáo, thiết kế và phân tích dữ liệu trên dashboard trong Power BI, tìm hiểu về tập dữ liệu.	100%
2	20133102	Nguyễn Nhật Triều	Tìm hiểu về tập dữ liệu, thực hiện việc xây dựng kho dữ liệu vào kho (SSIS)	100%
3	20133104	Đoàn Quốc Trung	Tìm hiểu về tập dữ liệu, thực hiện việc phân tích dữ liệu	100%

			(SSAS) và phân tích một số biểu đồ trên dashboard	
--	--	--	--	--

7. Tài liệu tham khảo

- [1] Babuninayak. (2023, May 11). Supply Chain Report (Power BI). Kaggle. <https://www.kaggle.com/code/babuninayak820/supply-chain-report-power-bi>
- [2] Tan Duc IT School. (2021, October 21). CÁCH TẠO DASHBOARD HIỆU QUẢ TRÊN POWER BI DESKTOP | Tan Duc ITS. YouTube. https://www.youtube.com/watch?v=o_psIwSLHNM
- [3] ExcelFort. (2021b, October 2). Power BI PAGE NAVIGATION Panel | HOW TO easily create with Bookmarks & Buttons [Video]. YouTube. <https://www.youtube.com/watch?v=xCSYLrcLW00>
- [4] Các kiến thức trong môn học kho dữ liệu