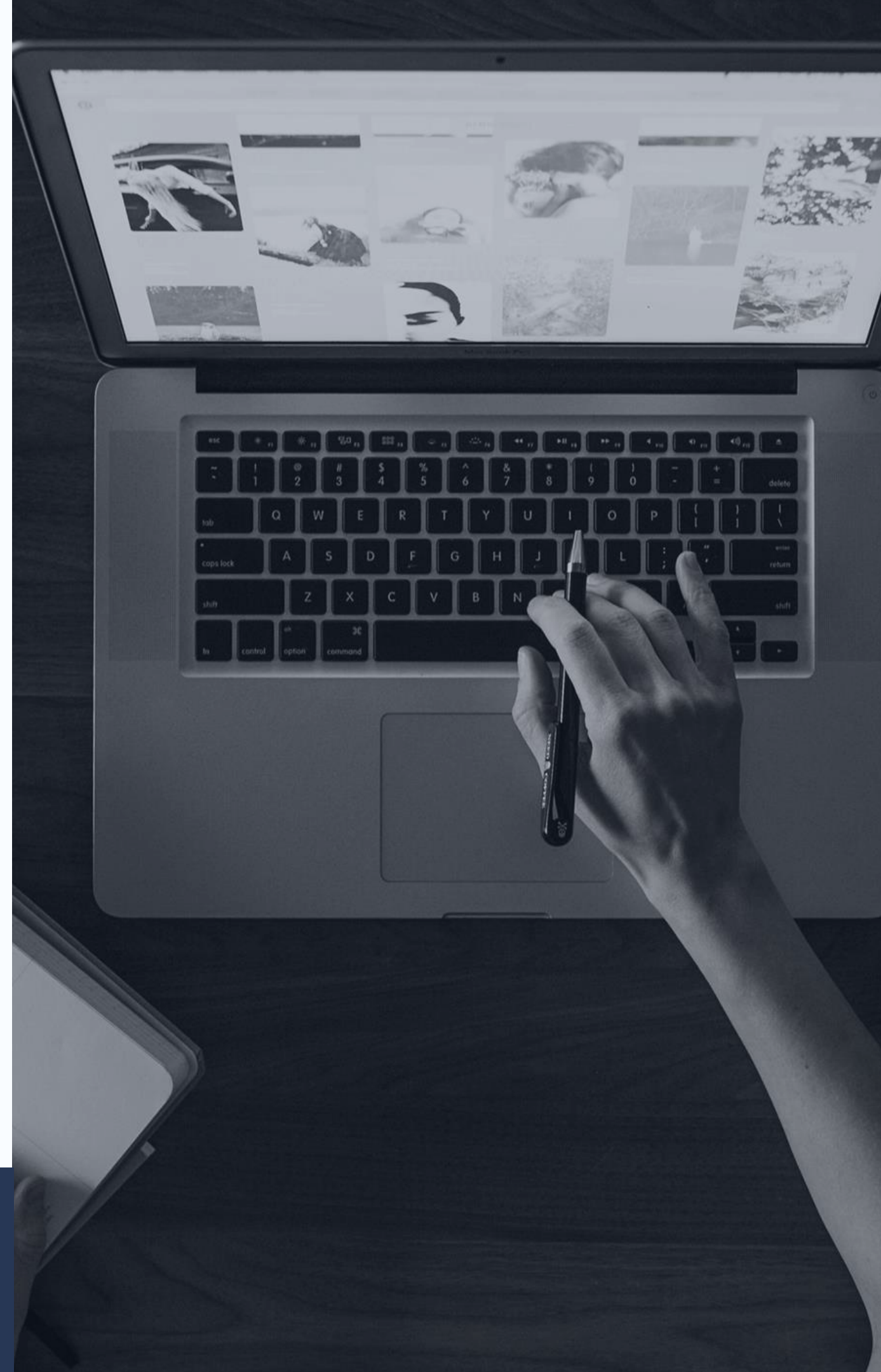


**PHÂN TÍCH CHUỖI CUNG ỨNG**

# DATA WAREHOUSE


*GVHD: Th.S Nguyễn Văn Thành*




**MEMBER**



**Lê Phước Yên**  
20133119



**Phan Thanh Tín**  
20133095



**Vũ Trung Kiên**  
20133060



**Nguyễn Đức Linh**  
20133007

# NỘI DUNG BÁO CÁO

---

- **Vấn đề phân tích**
- **Tập dữ liệu**
- **Phân tích và thiết kế kho dữ liệu**
- **Trực quan hoá để giải quyết các câu hỏi**
- **Tổng kết**

# VẤN ĐỀ PHÂN TÍCH

## CÁC CÂU HỎI ĐẶT RA

---

Những sản phẩm nào bán chạy nhất?  
Những khách hàng nào mang lại doanh thu cao nhất ?

Loại hình thanh toán nào có doanh thu cao nhất ?

Khu vực nào có lượng mua hàng cao ?

Doanh thu, lợi nhuận theo thời gian ?

Những sản phẩm mang lại doanh thu cao nhất ?



---

TẬP DỮ LIỆU (DATASET)

---

# DATAACO SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS

---

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

Tập dữ liệu là tập hợp các hoá đơn của khách hàng và khá là chi tiết.

180520 dòng dữ và 53 thuộc tính.



# TIỀN SỬ LÝ DỮ LIỆU

## SỬ DỤNG PYTHON ĐỂ XỬ LÝ

---

### Nhận xét:

- Nhận thấy trong tập dữ liệu của nhóm có rất nhiều thuộc tính, các giá trị trùng lặp, và bị thiếu. Do đó nhóm đã đi đến quyết định tạo ra các Sub Excel để giảm thiểu việc thao tác dễ dàng hơn. Ở mỗi sub, nhóm đã phân tích và thực hiện các phương pháp để tạo ra tập dữ liệu phù hợp

### Tạo Sub\_ship












```
1 Ship = dataf[['Days for shipping (real)', 'Days for shipment (scheduled)', 'Delivery Status', 'Shipping Mode']]
2 unique_Ship = Ship[['Days for shipping (real)', 'Days for shipment (scheduled)', 'Delivery Status', 'Shipping Mode']].drop_duplicates()
3 unique_Ship.shape
```

(26, 4)

```
1 values = np.random.permutation(np.arange(100000, 1000000))
2 random_values = np.random.choice(values, size=26, replace=False)
3 unique_Ship['Ship Id'] = random_values
4 dataf = pd.merge(dataf, unique_Ship, on=['Days for shipping (real)', 'Days for shipment (scheduled)', 'Delivery Status', 'Shipping Mode'], how='left')
5 unique_Ship.to_csv('Sub_Ship.csv', index=False)
```

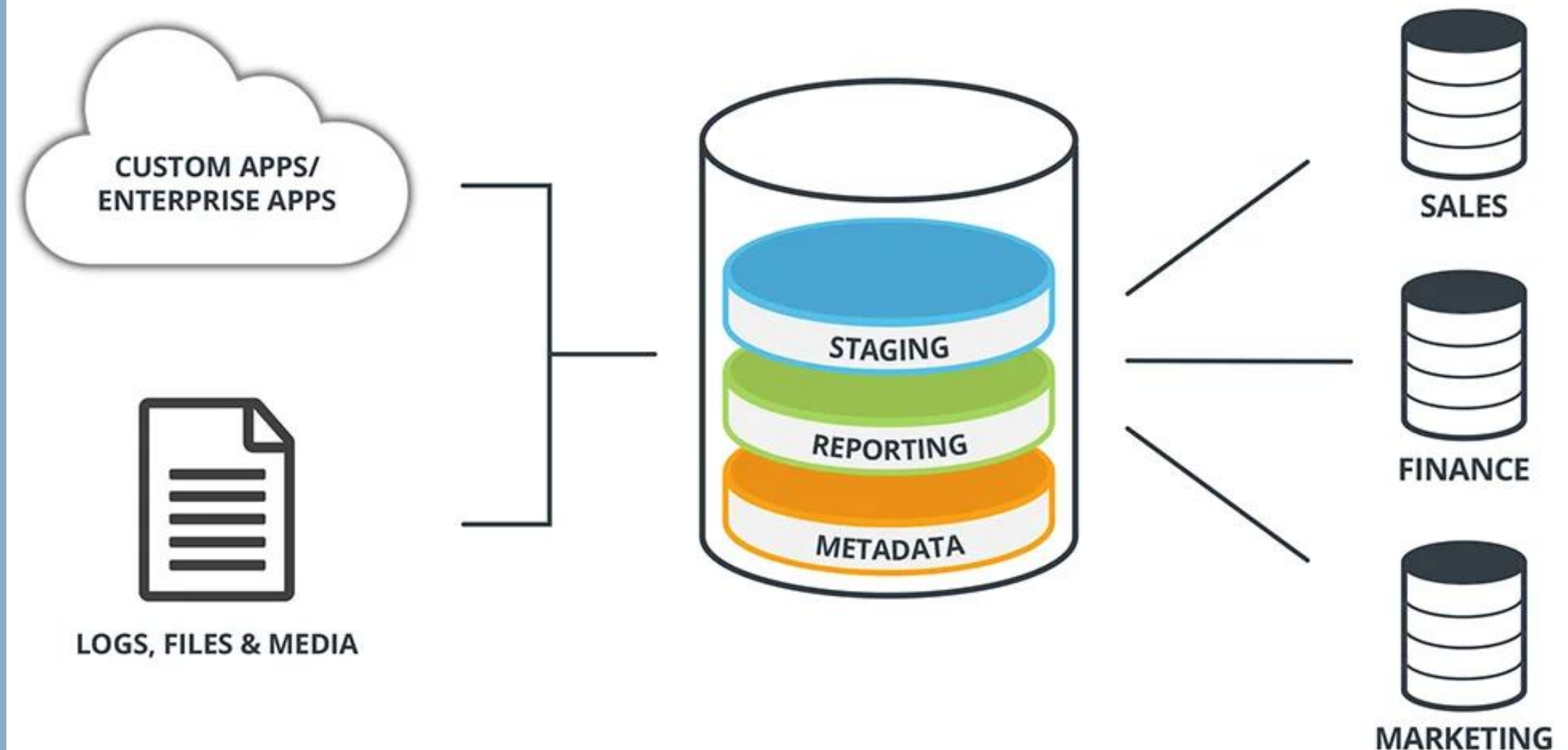
# KẾT QUẢ ĐẠT ĐƯỢC

Sau khi áp dụng tương tự cho các thuộc tính khác để sinh ra các Sub\_Excel.

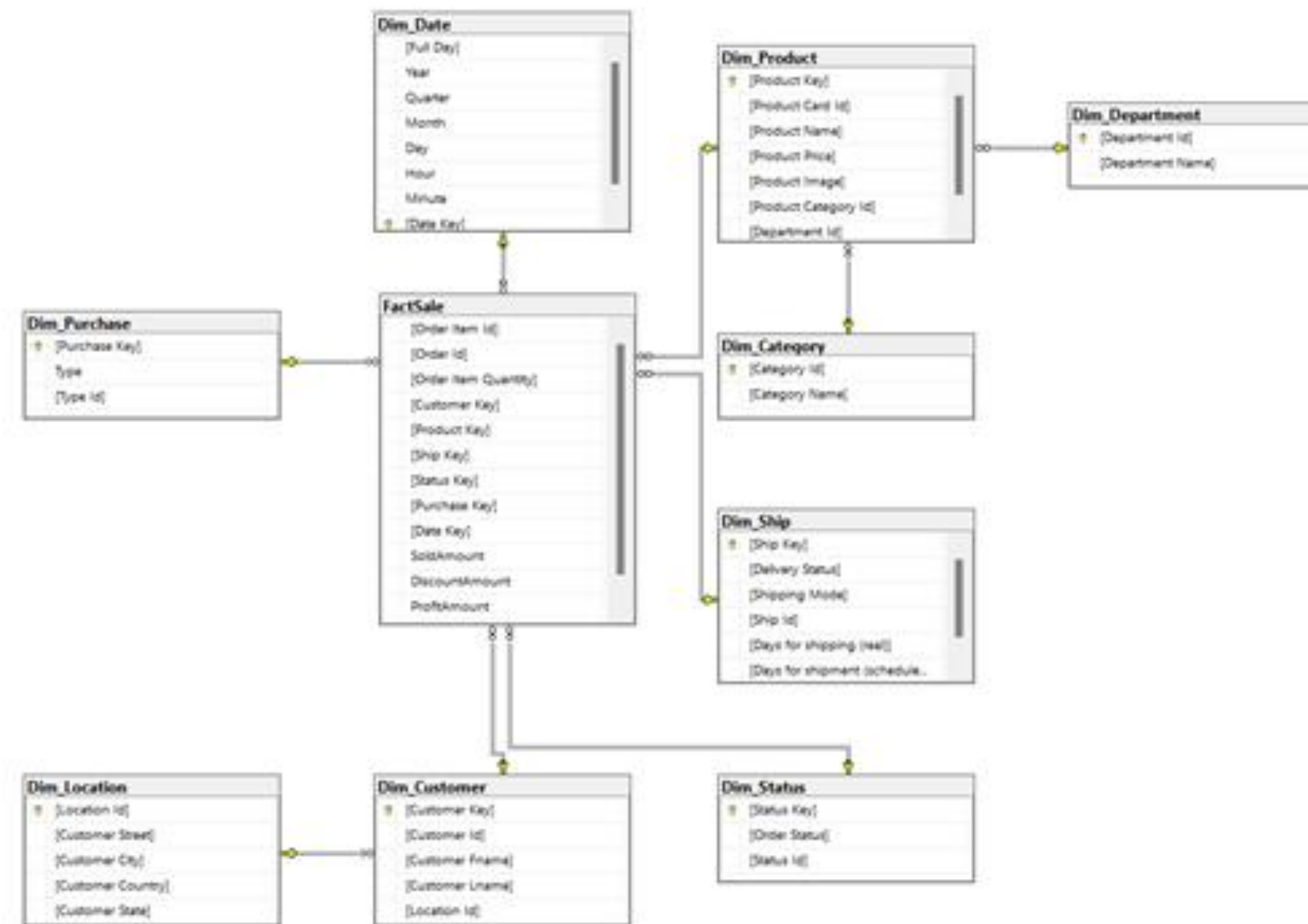
 DataChainCoSupply.csv	05/12/2019 6:54 SA	Microsoft Excel C...	93.663 KB
 Sub_Category.csv	12/05/2023 10:17 CH	Microsoft Excel C...	1 KB
 Sub_Customer.csv	12/05/2023 10:17 CH	Microsoft Excel C...	541 KB
 Sub_Department.csv	12/05/2023 10:17 CH	Microsoft Excel C...	1 KB
 Sub_Location.csv	12/05/2023 10:17 CH	Microsoft Excel C...	587 KB
 Sub_Order.csv	12/05/2023 10:17 CH	Microsoft Excel C...	15.437 KB
 Sub_Product.csv	12/05/2023 10:17 CH	Microsoft Excel C...	15 KB
 Sub_Purchase.csv	12/05/2023 10:17 CH	Microsoft Excel C...	1 KB
 Sub_Ship.csv	12/05/2023 10:17 CH	Microsoft Excel C...	2 KB
 Sub_Status.csv	12/05/2023 10:17 CH	Microsoft Excel C...	1 KB
 XuLiDaTa.ipynb	12/05/2023 10:56 CH	Jupyter Source File	27 KB



# PHÂN TÍCH VÀ THIẾT KẾ DATA WAREHOUSE

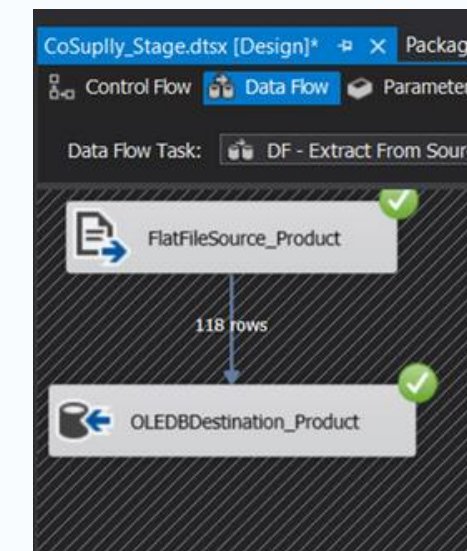
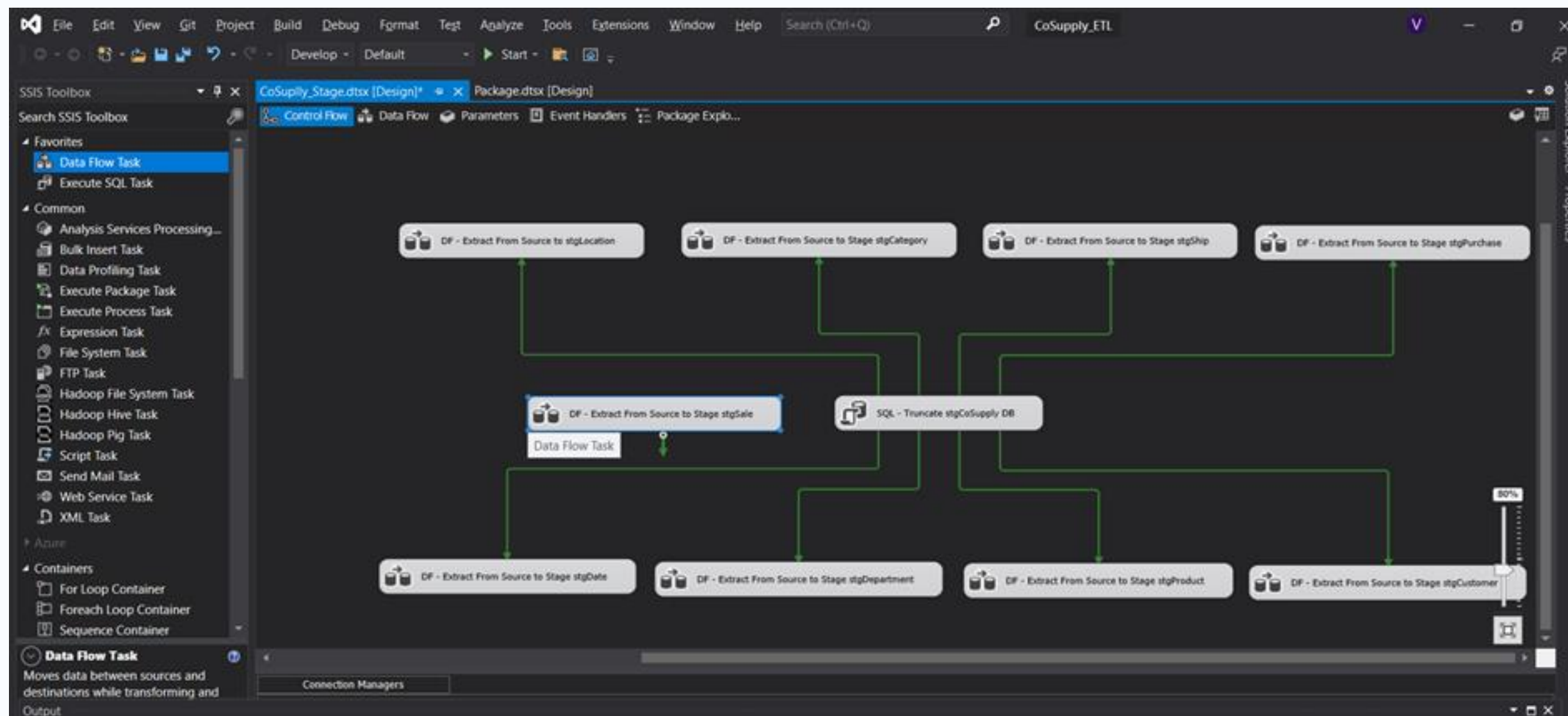


# MÔ HÌNH STAR SCHEMA





# QUÁ TRÌNH SSIS



ata into a relational database using an OLE DB provider.

DB connection manager, a data source, or a data source view, and select the data using the SQL command access mode, specify the SQL command either by typing using Query Builder. For fast-load data access, set the table update options.

ion manager:

H09.CoSupply\_Stage

de:

table or the view:

s or views could be loaded

Create Table

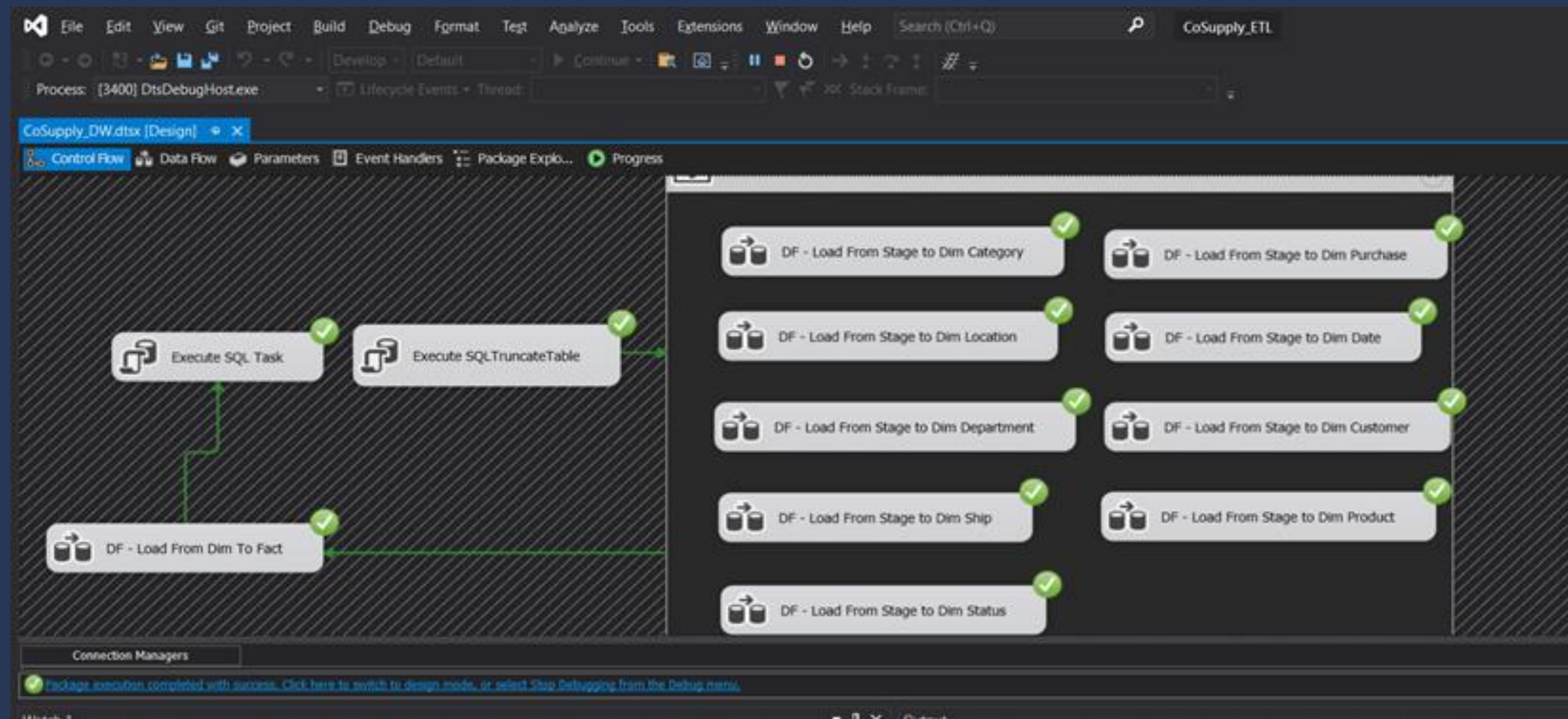
```
CREATE TABLE [Stg_Product] (  
    [Product Card Id] int,  
    [Product Name] varchar(50),  
    [Product Price] real,  
    [Product Image] varchar(255),  
    [Product Category Id] int,  
    [Department Id] int  
)
```

OK Cancel



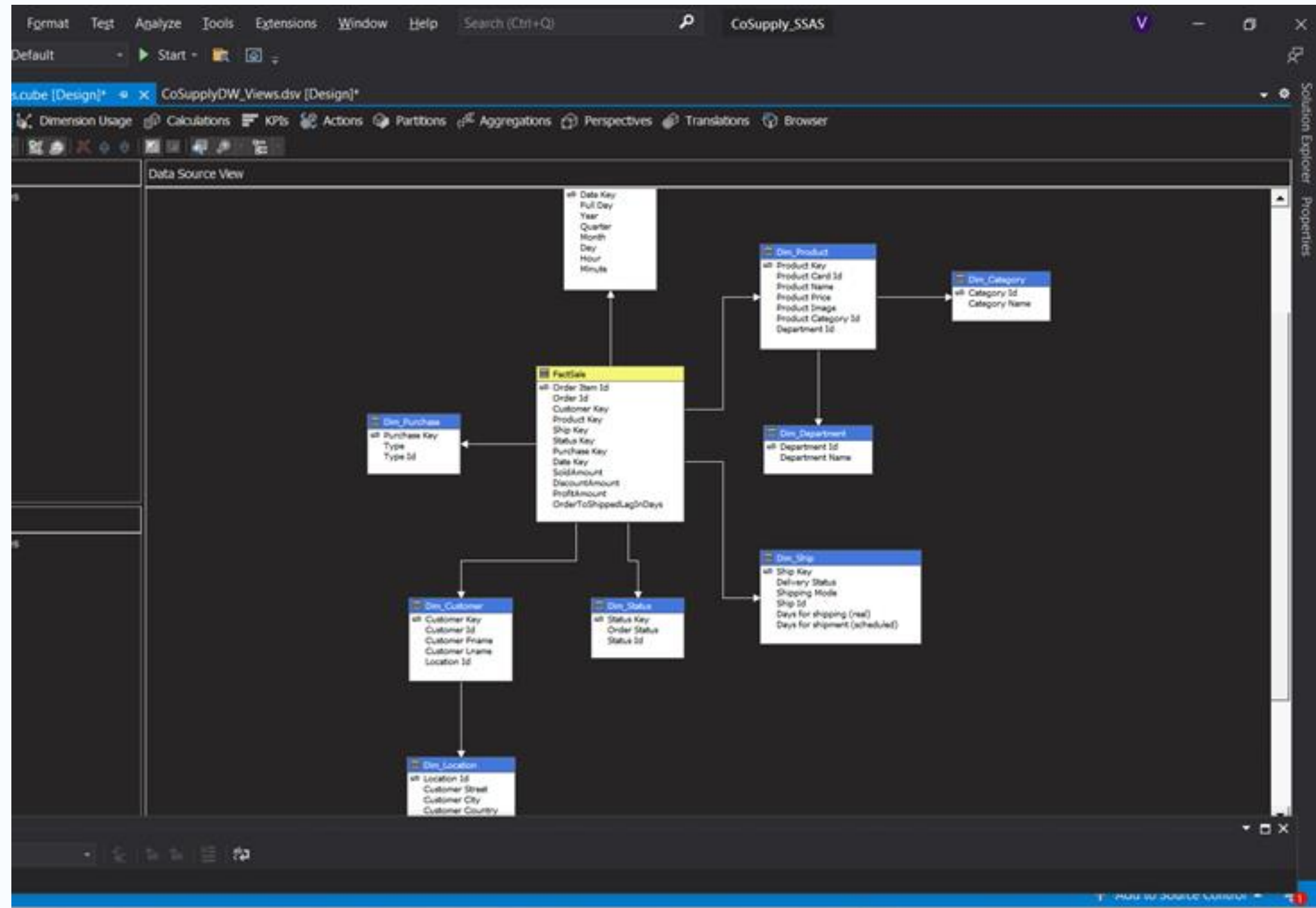
# QUÁ TRÌNH SSIS

ĐỔ DỮ LIỆU TỪ SUPPLY\_STG  
VÀO SUPPLY\_DW

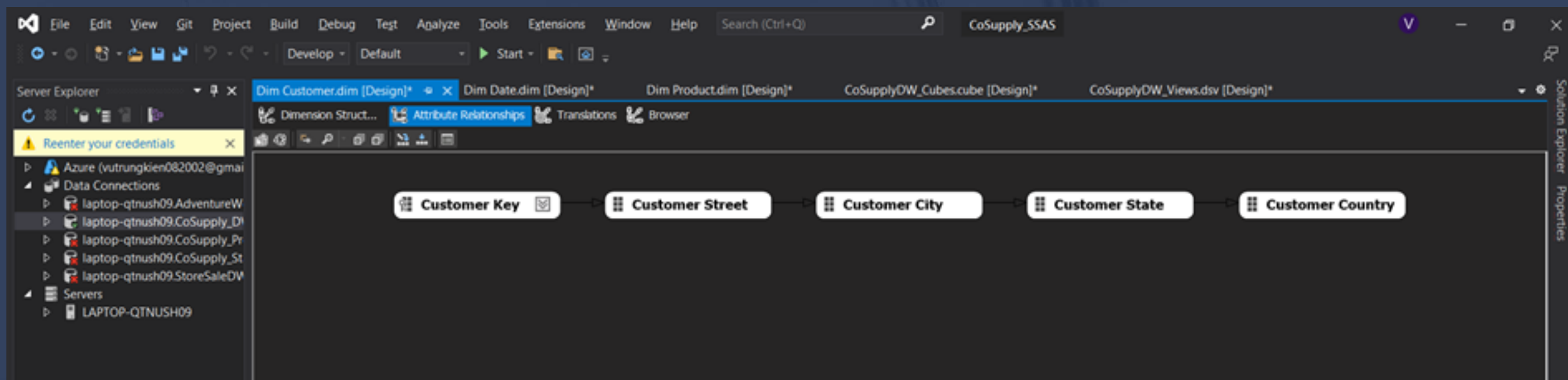
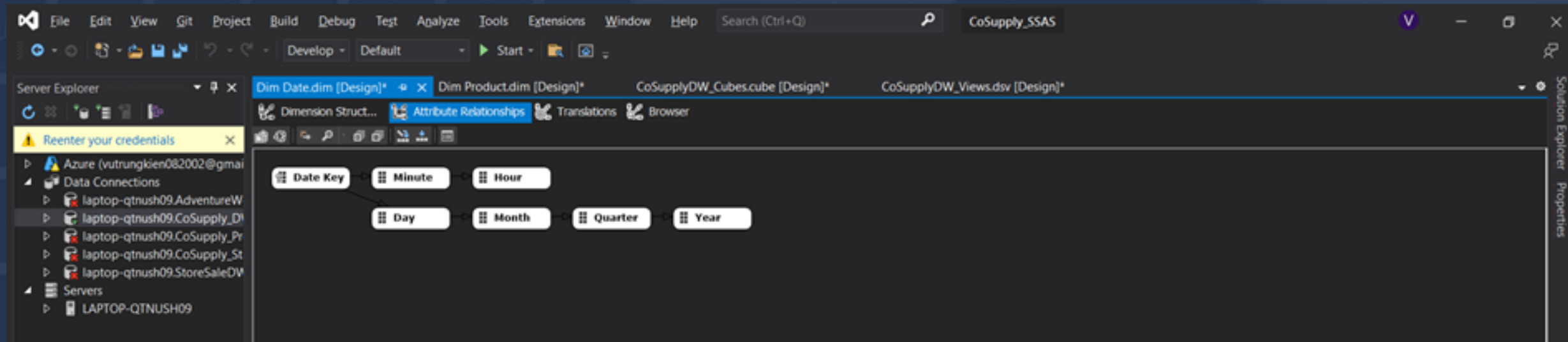


# QUÁ TRÌNH SSAS

## TẠO DATA CUBE



# PHÂN CẤP CHO DIM DATE VÀ DIM LOCATION





# PHÂN TÍCH TRẢ LỜI CÁC CÂU HỎI

**Tổng số doanh thu trong từng tiểu bang của mỗi nước năm 2017**



The screenshot shows a BI tool interface with a menu bar (Build, Debug, Test, Analyze, Tools, Extensions, Window, Help) and a search bar. The main workspace displays a cube design for 'CoSupplyDW\_Cubes.cube' with dimensions 'Dim Customer.dim', 'CoSupplyDW\_Cubes.cube', and 'Dim Date.dim'. A table on the right shows the results of a query, filtered by 'Year' = '2017'. The table has columns for 'Customer Country', 'Customer State', and 'Sold Amount'.

Customer Country	Customer State	Sold Amount
EE. UU.	MN	48178.63
EE. UU.	MO	93525.89
EE. UU.	MT	5926.66
EE. UU.	NC	121436
EE. UU.	ND	14225.9
EE. UU.	NJ	230359.2
EE. UU.	NM	62336.56
EE. UU.	NV	107521.1
EE. UU.	NY	755217.1
EE. UU.	OH	256332.8
EE. UU.	OK	17226.75
EE. UU.	OR	113622.7
EE. UU.	PA	242399.5
EE. UU.	RI	16722.3
EE. UU.	SC	44697.97
EE. UU.	TN	90841.23
EE. UU.	TX	571680.9
EE. UU.	UT	70908.14
EE. UU.	VA	122494.1
EE. UU.	WA	64112.45
EE. UU.	WI	62025.44
EE. UU.	WV	14843.84
Puerto Rico	PR	4595629



# PHÂN TÍCH TRẢ LỜI CÁC CÂU HỎI

**Thống kê doanh thu theo các phương thức giao hàng trong năm 2016**



The screenshot displays the CoSupply SSAS application interface. The top menu bar includes options like Build, Debug, Test, Analyze, Tools, Extensions, Window, and Help. The main workspace shows a cube design for 'CoSupplyDW\_Cubes.cube' with various tabs like Cube Struct..., Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and a Browser button. A table on the right shows the results of a query, filtered by the year 2016.

Dimension	Hierarchy	Operator	Filter Expression
Dim Date	Year	Equal	{ 2016 }
<Select dimension>			

Delivery Status	Fact Sale Count
Advance shipp...	14199
Late delivery	34446
Shipping canc...	2624
Shipping on tl...	11281

# TỔNG DOANH THU THEO TRẠNG THÁI GIAO HÀNG VÀ PHƯƠNG THỨC THANH TOÁN

The screenshot shows the CoSupply\_SSAS cube design tool. The left pane displays the cube structure with the following hierarchy:

- CoSupplyDW\_Cubes
  - Measures
    - Fact Sale
      - Discount Amount
      - Fact Sale Count
      - Maximum Sold Amount
      - Order Id
      - Order To Shipped Lag In Days
      - Profit Amount
      - Sold Amount
  - KPIs
    - Dim Customer
    - Dim Date
      - Date Key
      - Day
      - Month
      - Quarter
      - Year
      - Date\_Hierarchy

The right pane shows the 'Fact Sale' measure selected, with a table of values:

Delivery Status	Fact Sale Count
Advance shipp...	14199
Late delivery	34446
Shipping canc...	2624
Shipping on t...	11281

The screenshot shows the CoSupply\_SSAS cube design tool. The left pane displays the cube structure with the following hierarchy:

- CoSupplyDW\_Cubes
  - Measures
    - Fact Sale
      - Discount Amount
      - Fact Sale Count
      - Maximum Sold Amount
      - Order Id
      - Order To Shipped Lag In Days
      - Profit Amount
      - Sold Amount
  - KPIs
    - Dim Customer
    - Dim Date
    - Dim Product
    - Dim Purchase
      - Purchase Key
      - Type
      - Type Id
    - Dim Ship

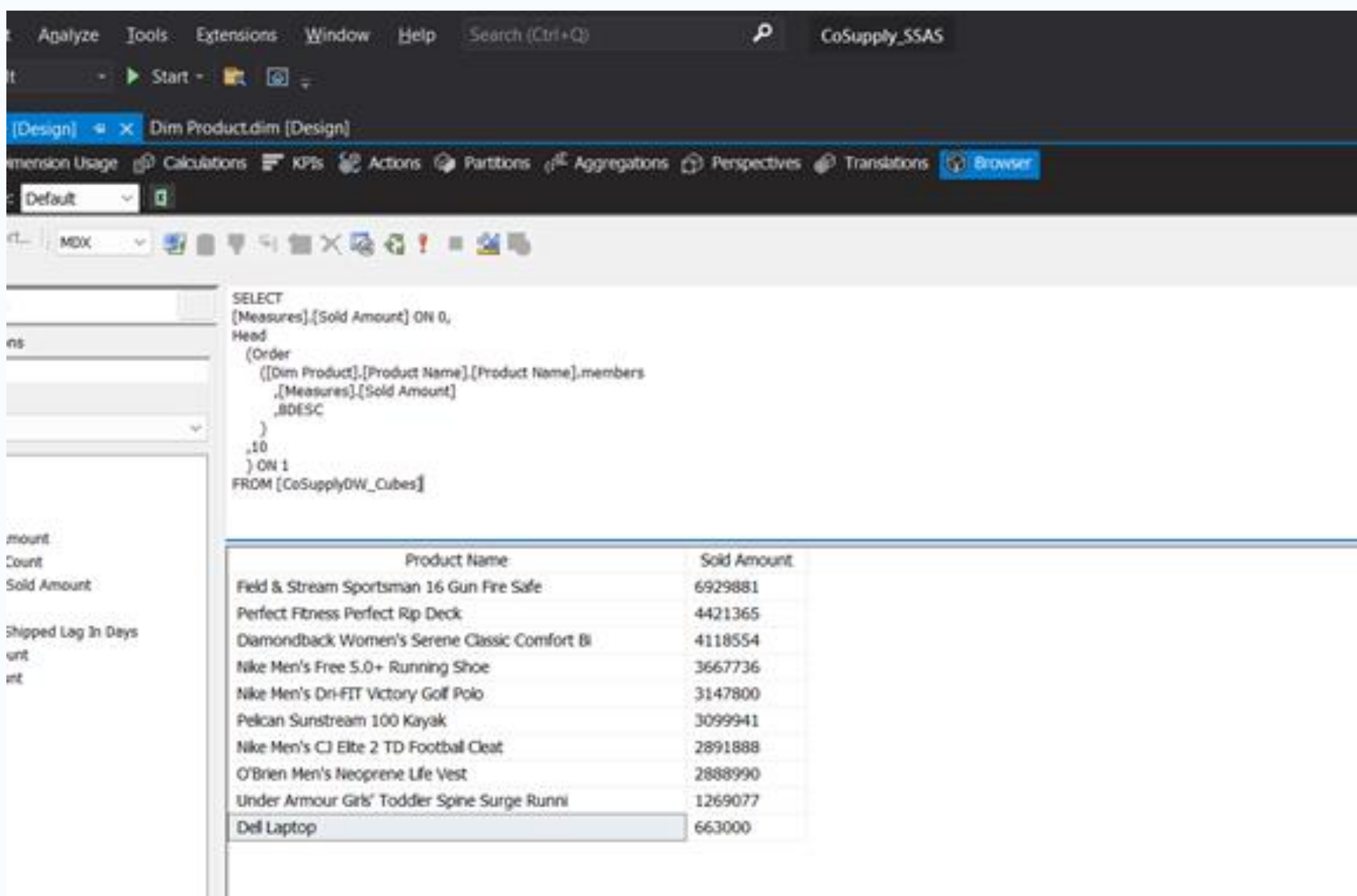
The right pane shows the 'Fact Sale' measure selected, with a table of values:

Type	Fact Sale Count
CASH	6633
DEBIT	24604
PAYMENT	14314
TRANSFER	16999



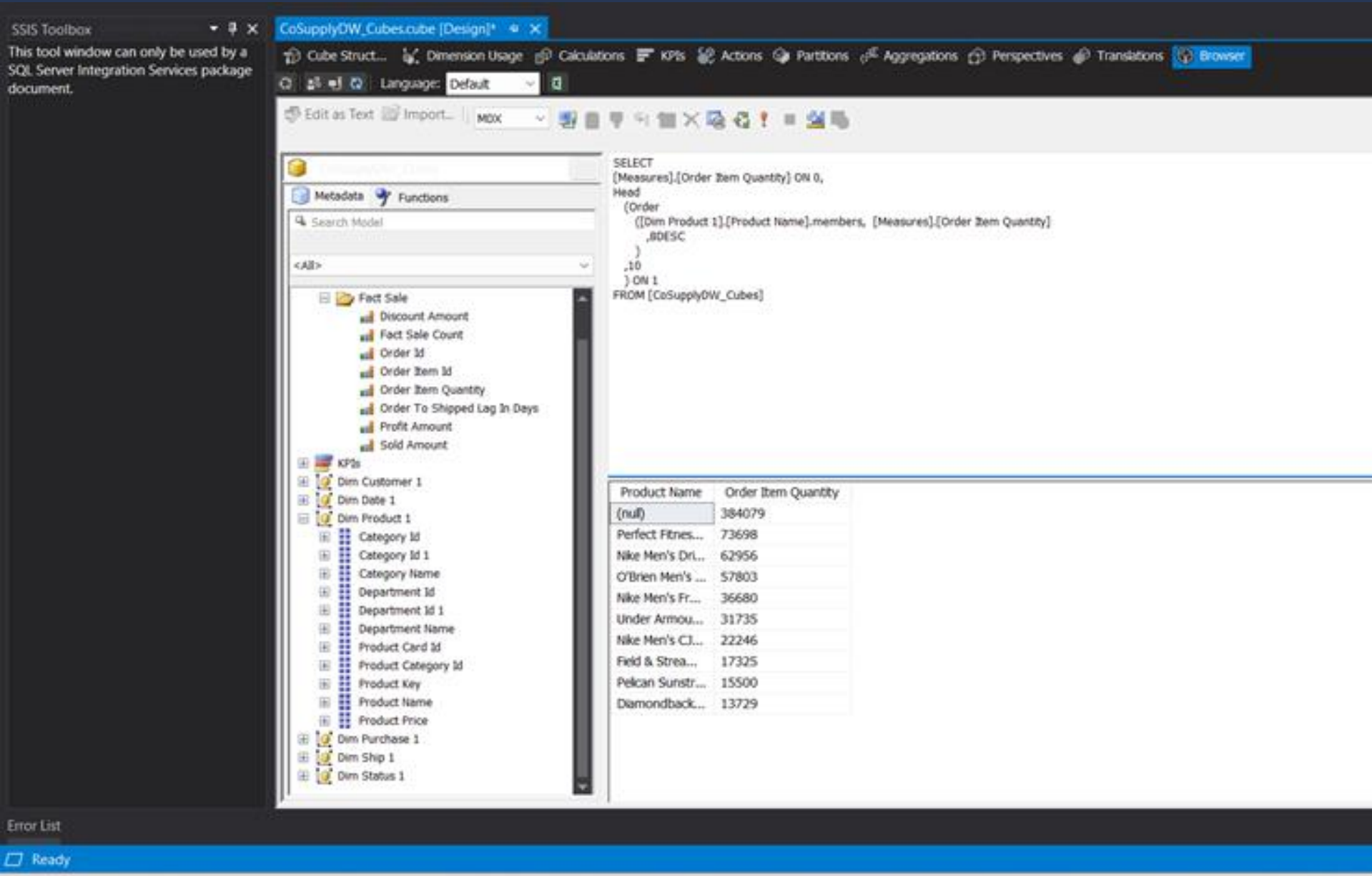
# THỐNG KÊ THEO SẢN PHẨM

## Top 10 sản phẩm có doanh thu cao nhất



SELECT  
[Measures].[Sold Amount] ON 0,  
Head  
(Order  
([Dim Product].[Product Name].[Product Name].members  
,[Measures].[Sold Amount]  
),BDESC  
)  
,10  
) ON 1  
FROM [CoSupplyDW\_Cubes]

Product Name	Sold Amount
Field & Stream Sportsman 16 Gun Fire Safe	6929881
Perfect Fitness Perfect Rip Deck	4421365
Diamondback Women's Serene Classic Comfort Bi	4118554
Nike Men's Free 5.0+ Running Shoe	3667736
Nike Men's Dri-FIT Victory Golf Polo	3147800
Pelican Sunstream 100 Kayak	3099941
Nike Men's CJ Elite 2 TD Football Cleat	2891888
O'Brien Men's Neoprene Life Vest	2888990
Under Armour Girls' Toddler Spine Surge Runni	1269077
Dell Laptop	663000



SELECT  
[Measures].[Order Item Quantity] ON 0,  
Head  
(Order  
([Dim Product 1].[Product Name].members, [Measures].[Order Item Quantity]  
,BDESC  
)  
,10  
) ON 1  
FROM [CoSupplyDW\_Cubes]

Product Name	Order Item Quantity
(null)	384079
Perfect Fitness...	73698
Nike Men's Dri...	62956
O'Brien Men's ...	57803
Nike Men's Fr...	36680
Under Armou...	31735
Nike Men's CJ...	22246
Field & Strea...	17325
Pelican Sunstr...	15500
Diamondback...	13729

## Top 10 sản phẩm bán chạy nhất

# DOANH THU VÀ LỢI NHUẬN THEO THỜI GIAN

CoSupplyDW\_Cubes

Metadata

Search Model

<All>

CoSupplyDW\_Cubes

- Measures
  - Fact Sale
    - Discount Amount
    - Fact Sale Count
    - Order Id
    - Order To Shipped Lag In Days
    - Profit Amount
    - Sold Amount
- KPIs
- Dim Customer
- Dim Date
  - Date Key
  - Day
  - Month
  - Quarter
  - Year
  - Date\_Hierarchy
- Dim Product

Dimension	Hierarchy	Operator
Dim Date	Year	Equal
<Select dimension>		

Month	Sold Amount
1	1029710
10	1073991
11	626905.4
12	503906.4
2	992546.3
3	1048019
4	1038334
5	1105488
6	1032101
7	1104393
8	1109356
9	1143795

CoSupplyDW\_Cubes

Metadata

Search Model

Measure Group:

<All>

CoSupplyDW\_Cubes

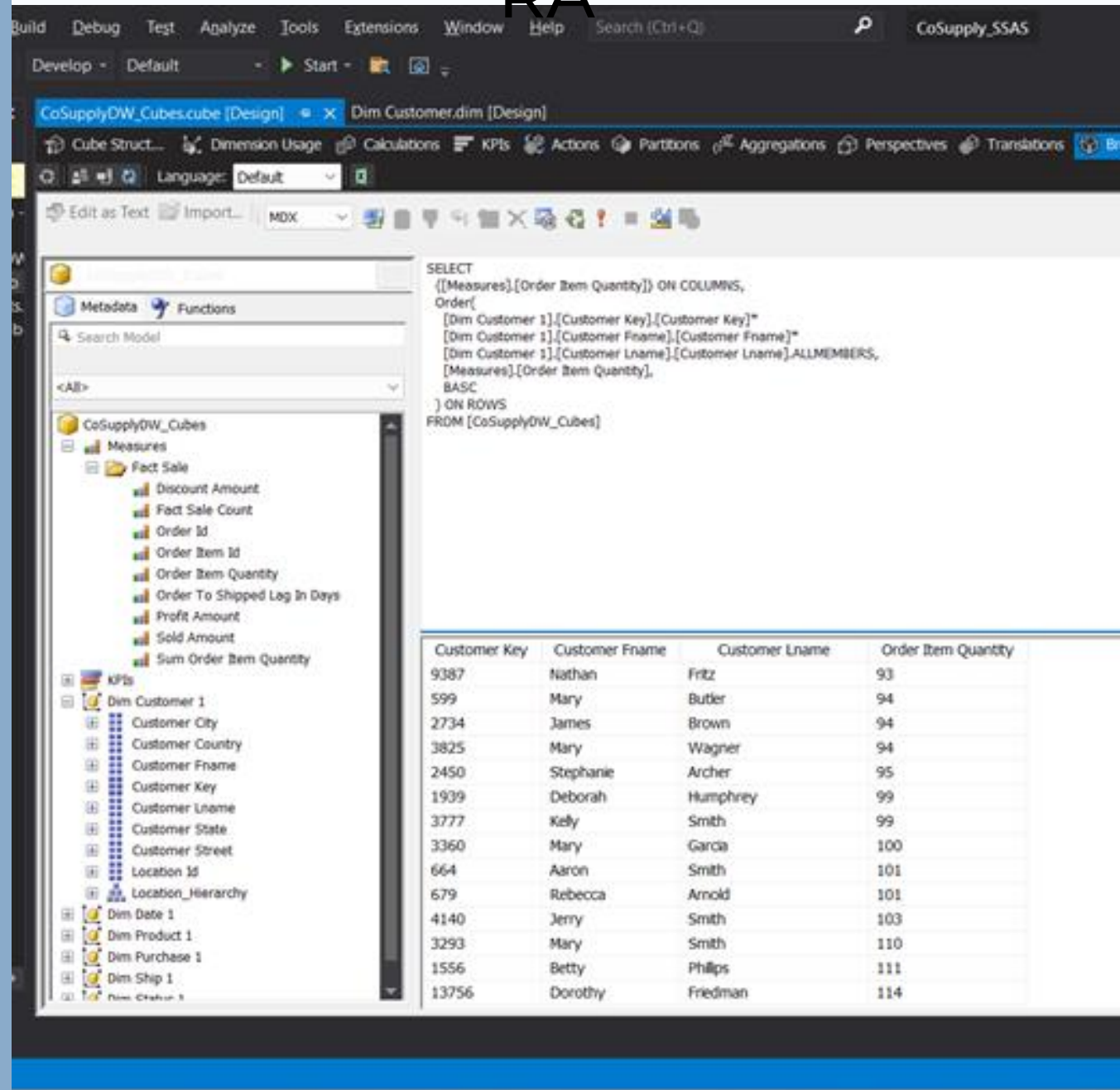
- Measures
  - Fact Sale
    - Discount Amount
    - Fact Sale Count
    - Order Id
    - Order Item Id
    - Order Item Quantity
    - Order To Shipped Lag In Days
    - Profit Amount
    - Sold Amount
    - Sum Order Item Quantity
- KPIs
- Dim Customer 1
- Dim Date 1
  - Date Key
  - Day
  - Month
  - Quarter

Dimension	Hierarchy	Operator
Dim Date 1	Year	Equal
<Select dimension>		

Month	Profit Amount
1	128364.1
10	126307.7
11	74239.34
12	73053.17
2	128489.9
3	132650.7
4	129572.4
5	128855.1
6	122523.8
7	125494.8
8	146588.6
9	137235.3

# THỐNG KÊ THEO KHÁCH HÀNG

## DOANH SỐ BÁN RA



SELECT

{[Measures].[Order Item Quantity]} ON COLUMNS,

Order{

[Dim Customer 1].[Customer Key].[Customer Key]\*

[Dim Customer 1].[Customer Fname].[Customer Fname]\*

[Dim Customer 1].[Customer Lname].[Customer Lname].ALLMEMBERS,

[Measures].[Order Item Quantity],

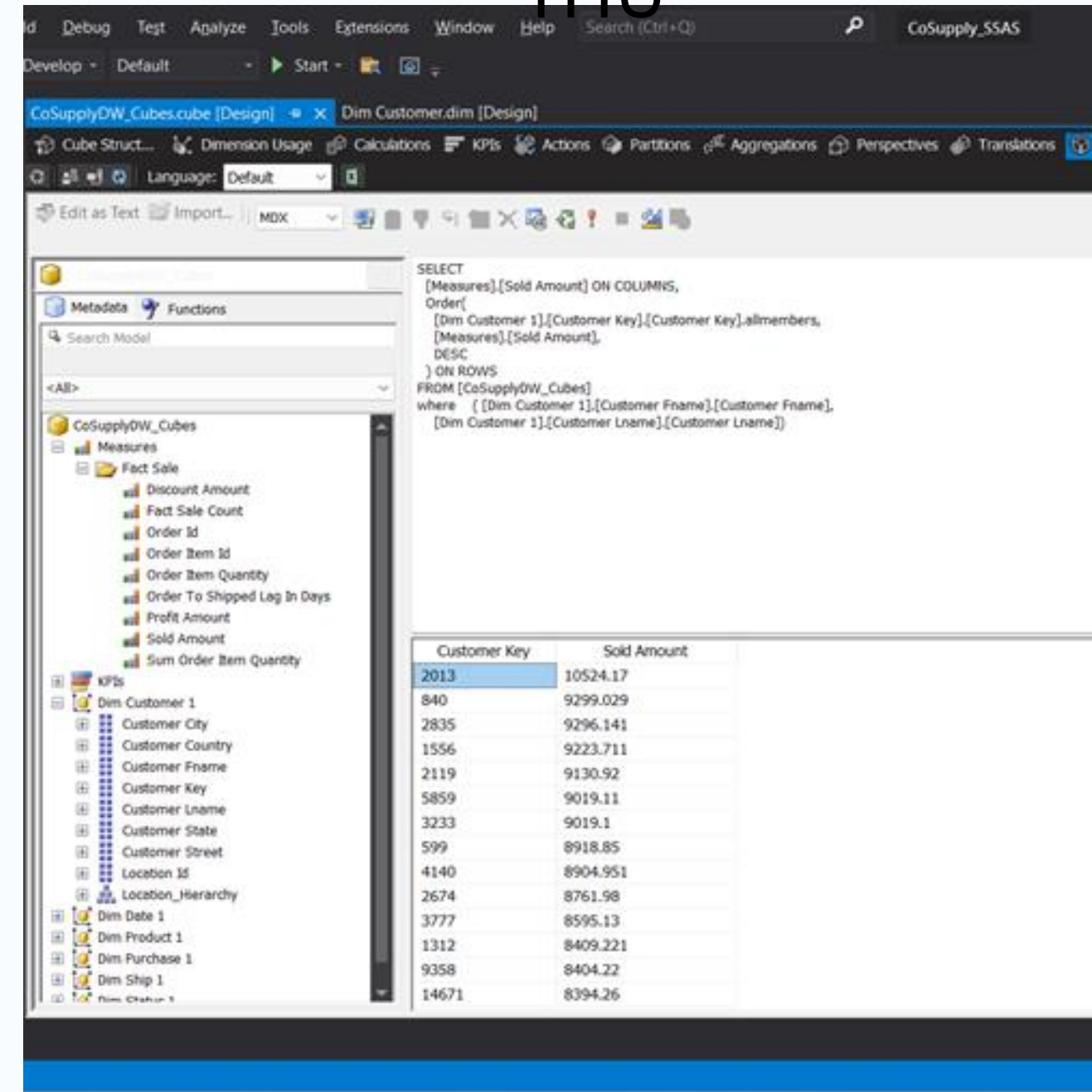
BASC

} ON ROWS

FROM [CoSupplyDW\_Cubes]

Customer Key	Customer Fname	Customer Lname	Order Item Quantity
9387	Nathan	Fritz	93
599	Mary	Butler	94
2734	James	Brown	94
3825	Mary	Wagner	94
2450	Stephanie	Archer	95
1939	Deborah	Humphrey	99
3777	Kelly	Smith	99
3360	Mary	Garcia	100
664	Aaron	Smith	101
679	Rebecca	Arnold	101
4140	Jerry	Smith	103
3293	Mary	Smith	110
1556	Betty	Phillips	111
13756	Dorothy	Friedman	114

## DOANH THU



SELECT

[Measures].[Sold Amount] ON COLUMNS,

Order{

[Dim Customer 1].[Customer Key].[Customer Key].allmembers,

[Measures].[Sold Amount],

DESC

} ON ROWS

FROM [CoSupplyDW\_Cubes]

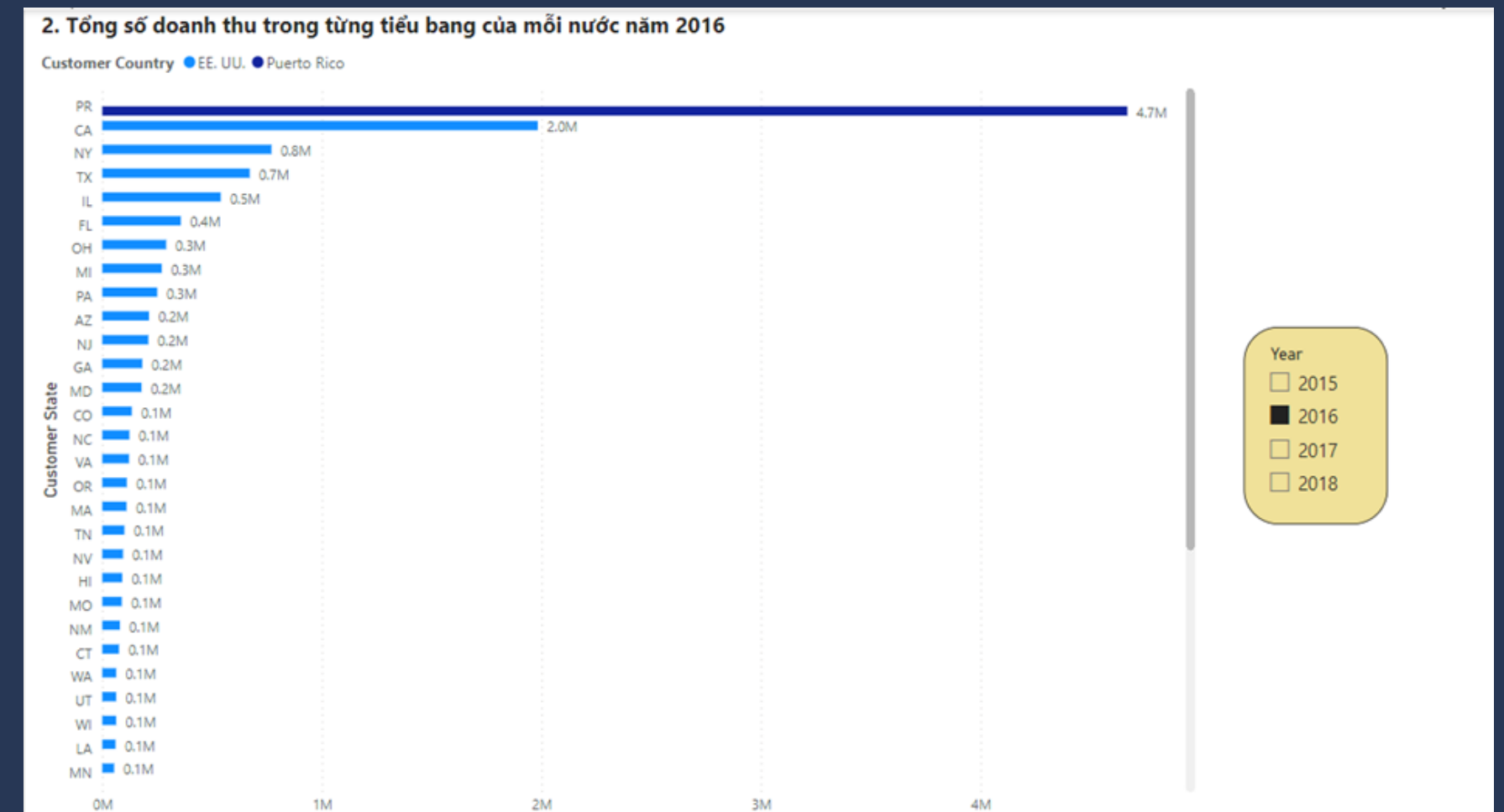
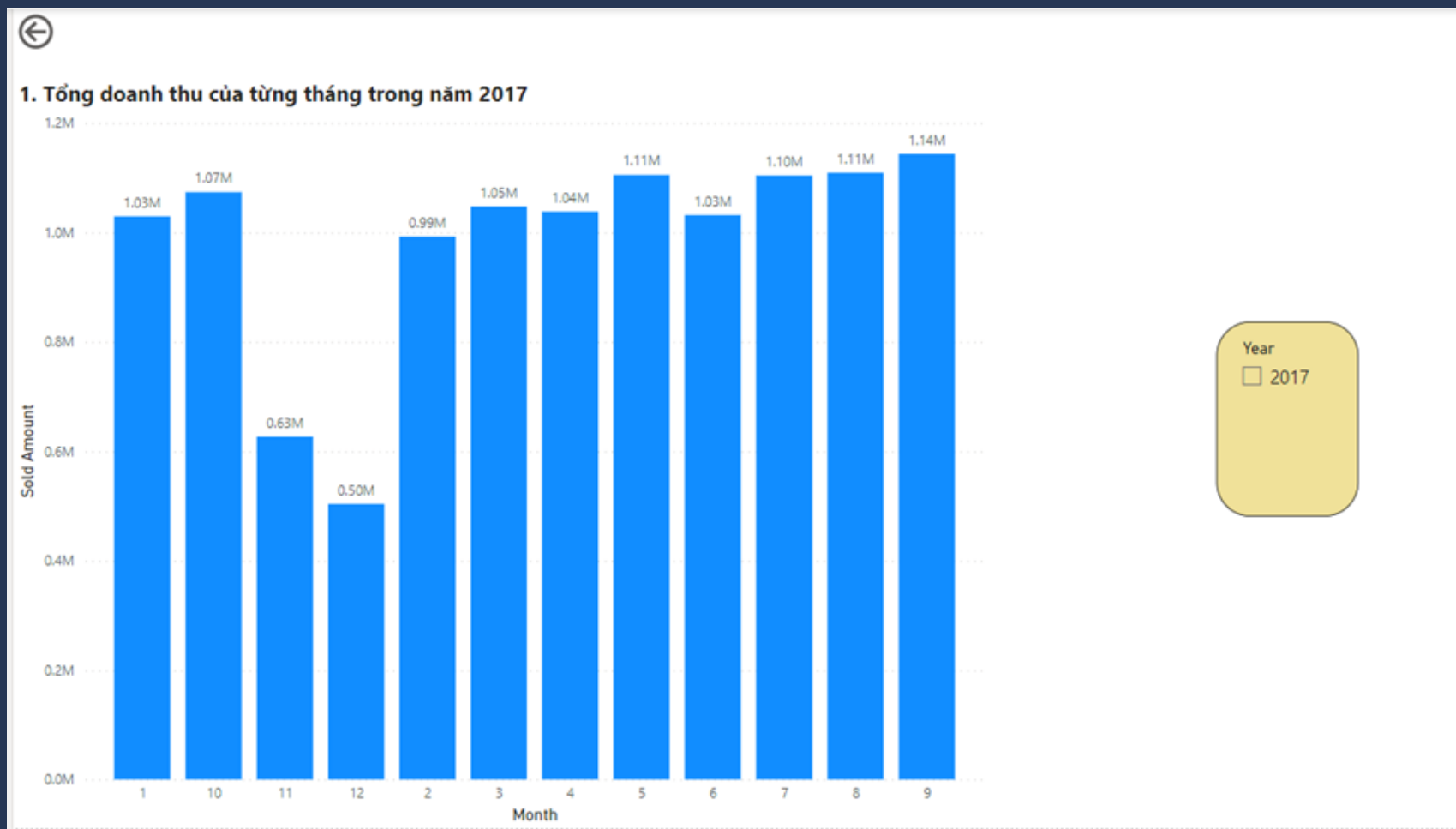
where ([Dim Customer 1].[Customer Fname].[Customer Fname],

[Dim Customer 1].[Customer Lname].[Customer Lname])

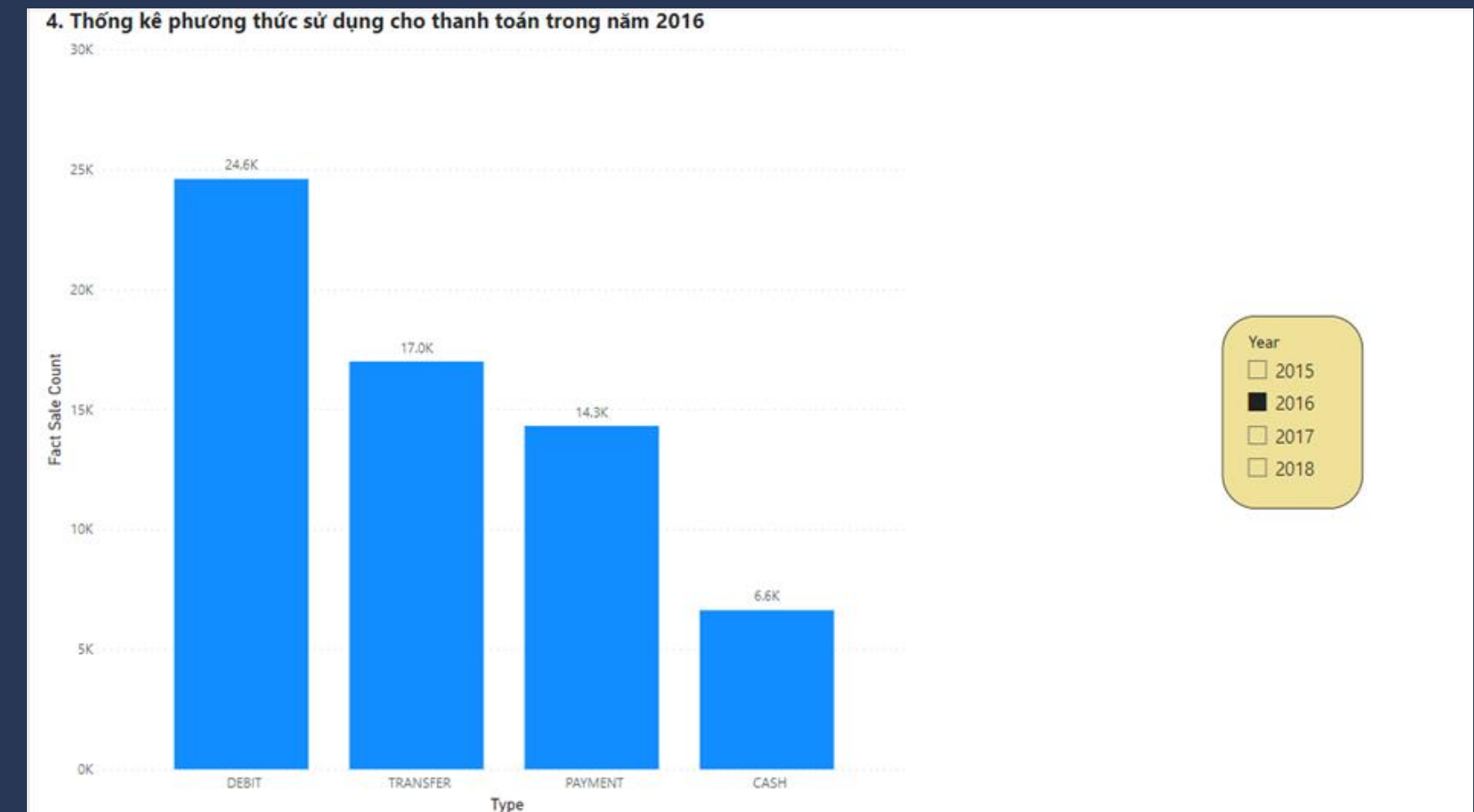
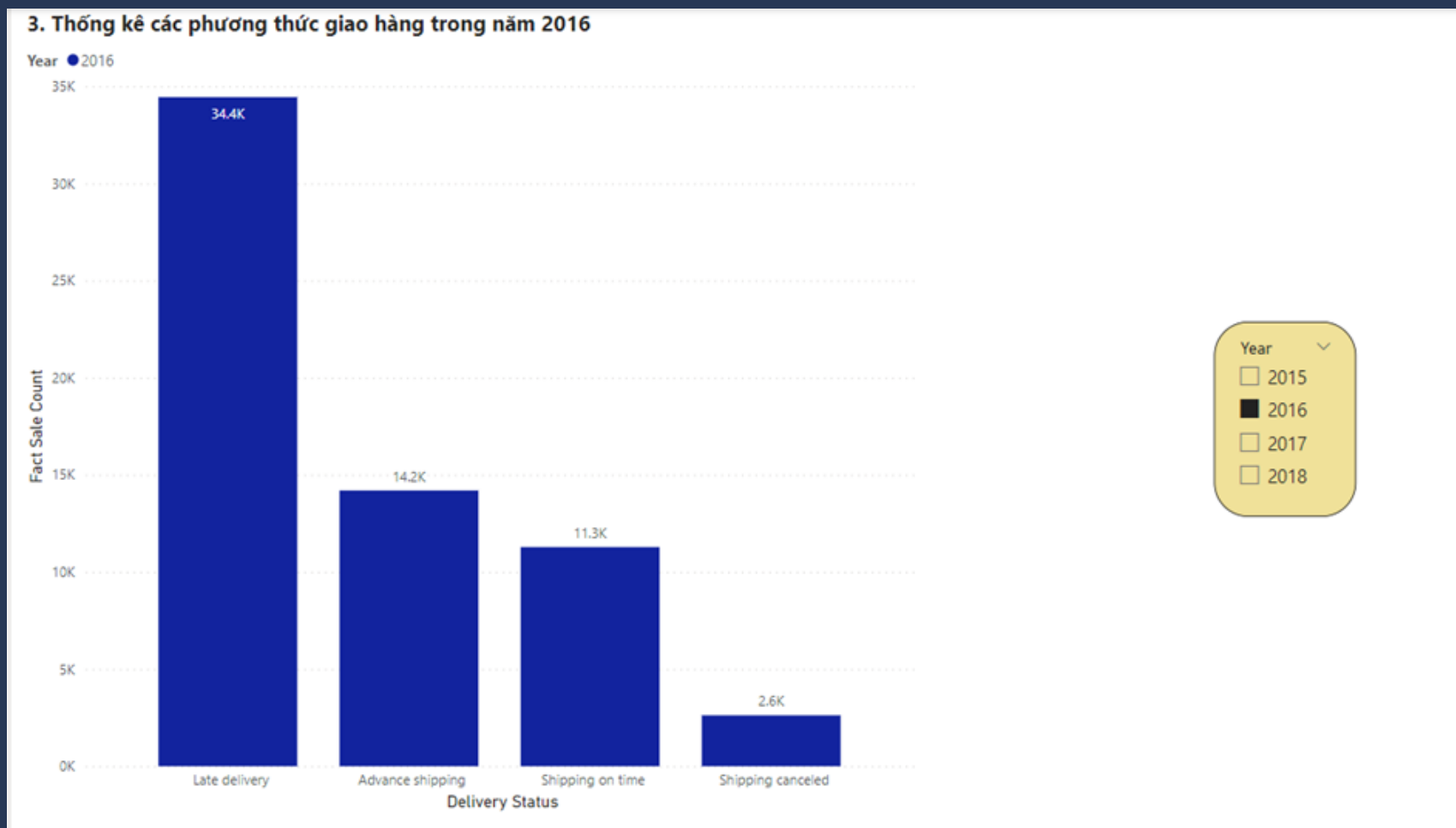
Customer Key	Sold Amount
2013	10524.17
840	9299.029
2835	9296.141
1556	9223.711
2119	9130.92
5859	9019.11
3233	9019.1
599	8918.85
4140	8904.951
2674	8761.98
3777	8595.13
1312	8409.221
9358	8404.22
14671	8394.26



# TRỰC QUAN HOÁ BẰNG POWER BI



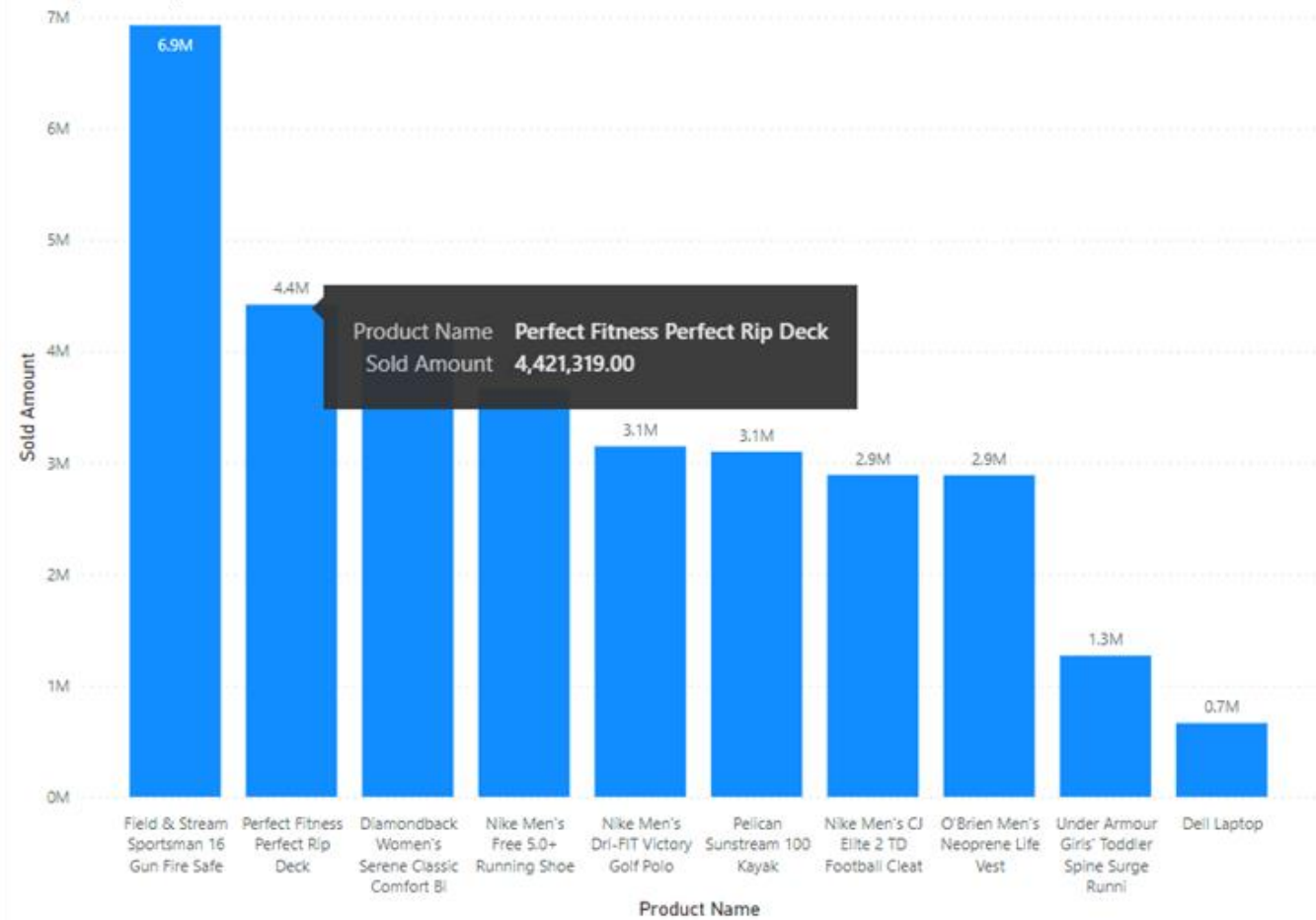
# TRỰC QUAN HOÁ BẰNG POWER BI



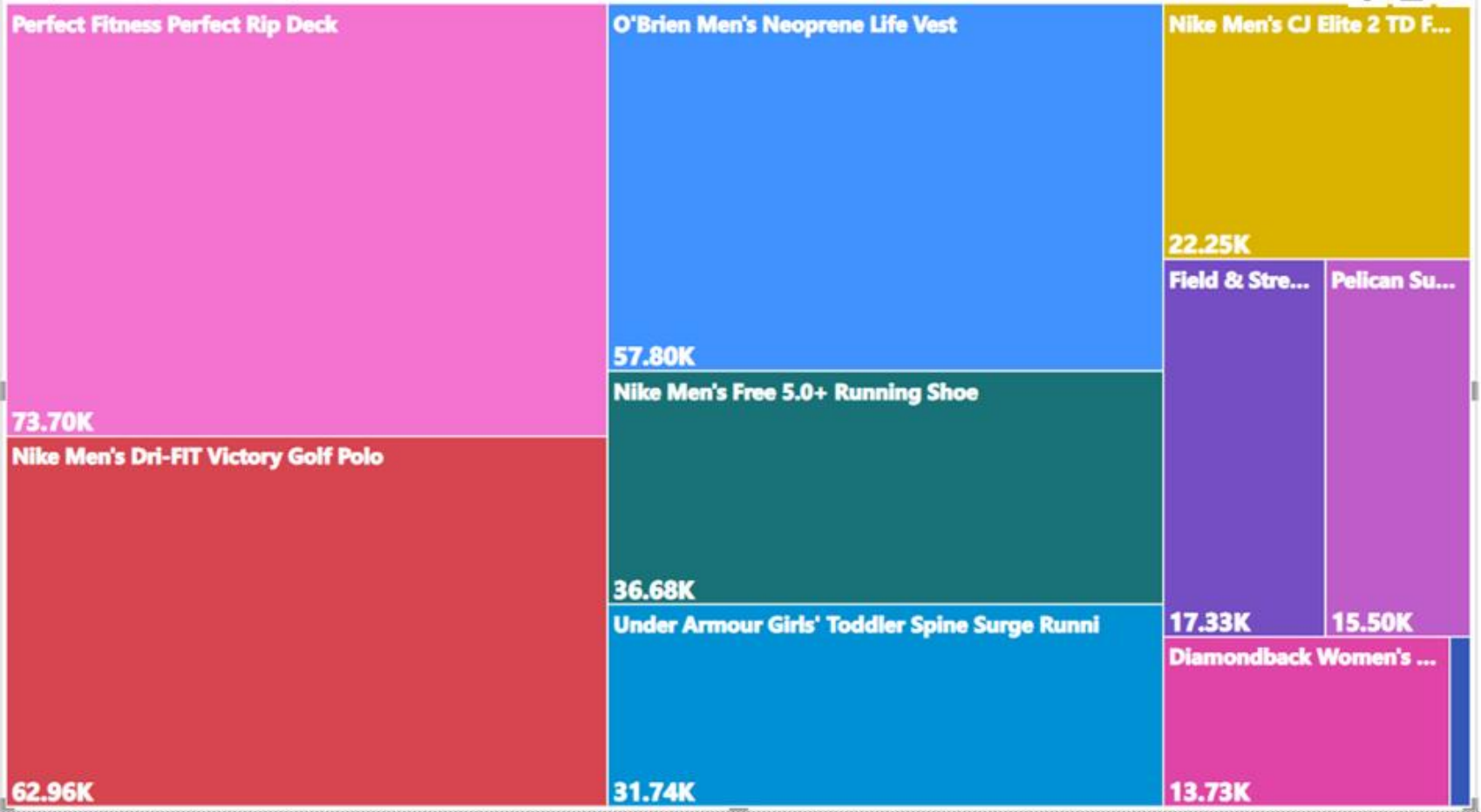


# TRỰC QUAN HOÁ BẰNG POWER BI

5. Top 10 sản phẩm có doanh thu cao nhất

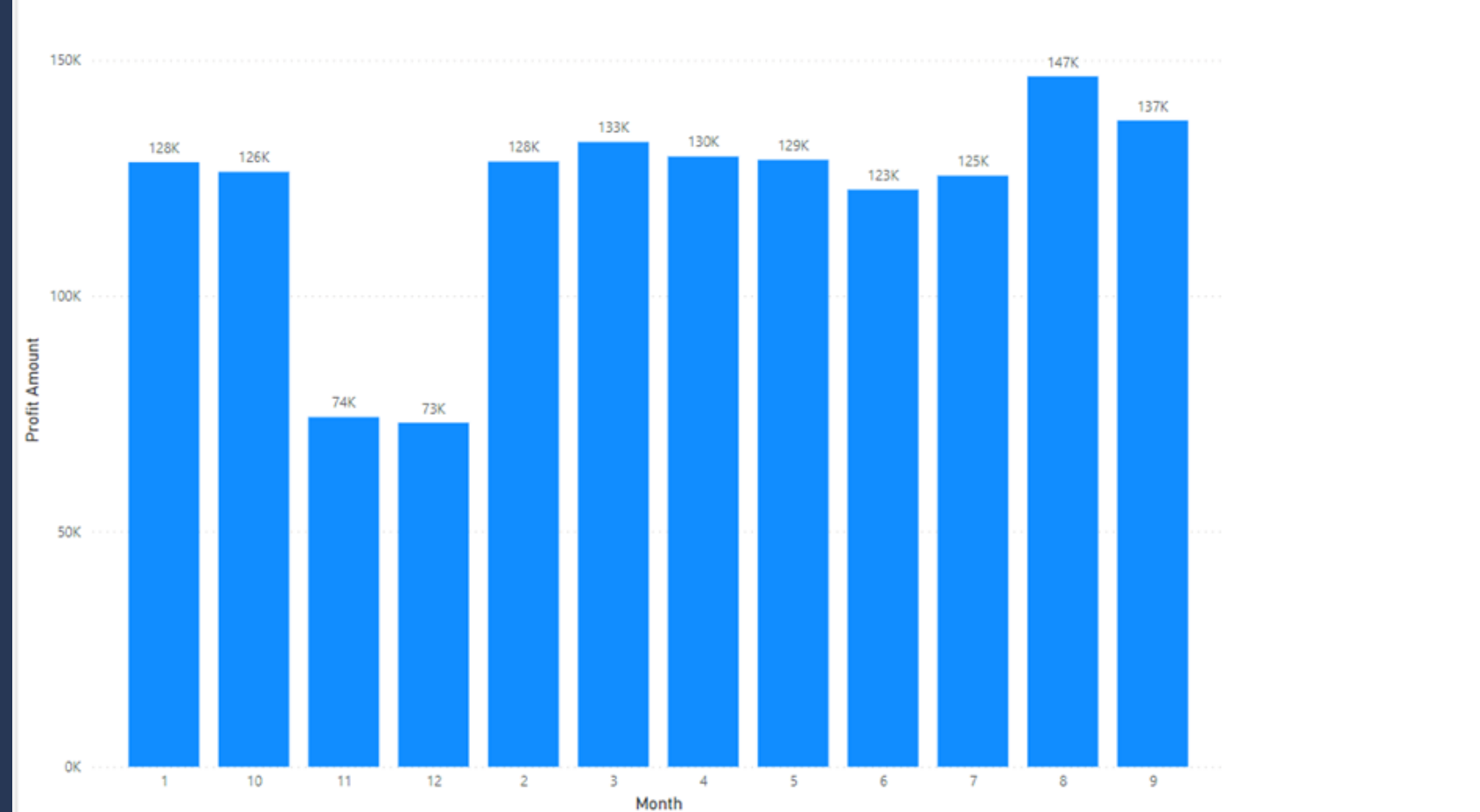


6. Top 10 sản phẩm bán chạy nhất

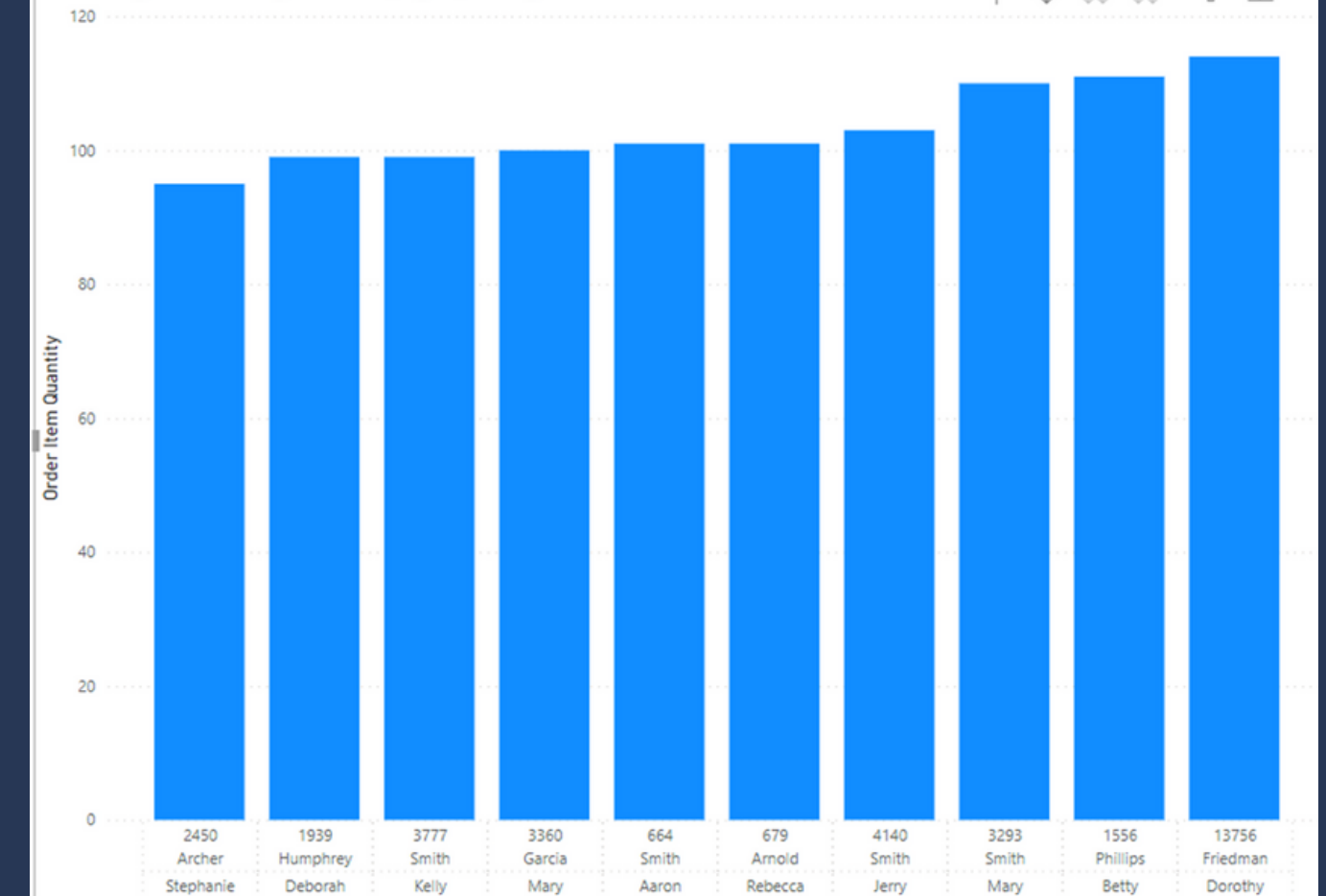


# TRỰC QUAN HOÁ BẰNG POWER BI

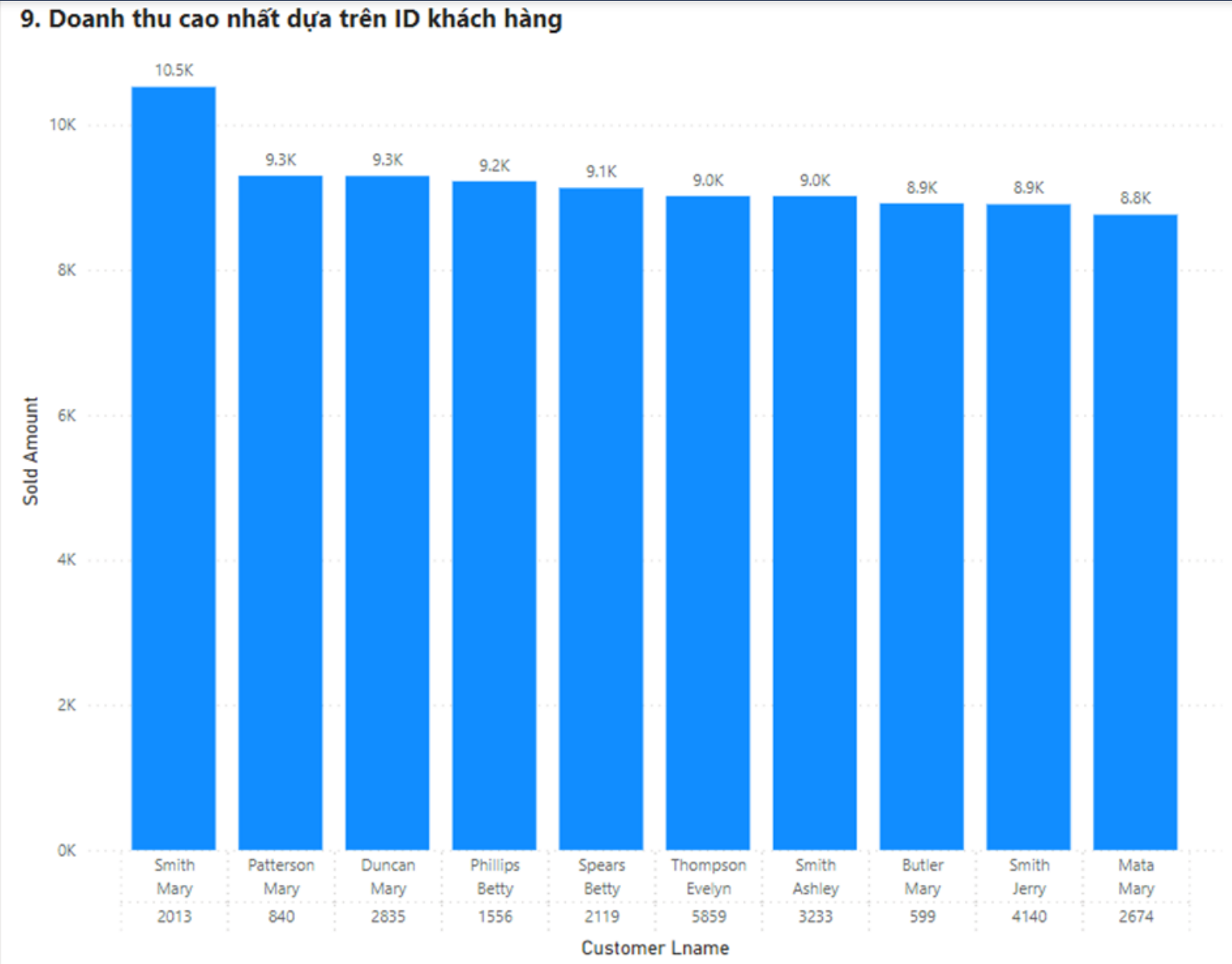
7. Tổng lợi nhuận theo từng tháng trong năm 2017



8. Những khách hàng có số lượng đặt hàng cao nhất



# TRỰC QUAN HOÁ BẰNG POWER BI



---

# TỔNG KẾT

## Kết quả đạt được:

- Có kiến thức về xây dựng kho dữ liệu.
- Xây dựng được 1 kho dữ liệu cơ bản.
- Khai thác thông tin từ kho dữ liệu.
- Trực quan hoá cơ bản.

## Hạn chế:

- Còn thiếu kinh nghiệm trong việc phân tích để xây dựng kho.
- Kho dữ liệu chưa được tối ưu.
- Các quá trình đẩy dữ liệu còn khá sơ sài.
- Trực quan hoá chưa được tỉ mỉ.

# PHƯƠNG HƯỚNG PHÁT TRIỂN



Sử dụng các thuật toán  
sử dụng SSAS áp dụng  
vào kho dữ liệu để  
phân tích chuyên sâu  
về nghiệp vụ doanh  
nghiệp



Thực hiện các truy vấn  
MDX phức tạp và hữu  
ích hơn đối với doanh  
nghiệp



Nếu đủ kiến thức có thể  
xây dựng được  
Business Driven



Từ các kết luận rút ra  
có thể lập các phương  
án marketing cho  
doanh nghiệp

**THANKS FOR WATCHING**