

KHO VÀ KHAI PHÁ DỮ LIỆU

PHÂN TÍCH DỮ LIỆU DOANH THU BÁN HÀNG CỦA CÁC CỬA HÀNG TẠI MỸ

GVHD: PHAN ĐÌNH VĂN

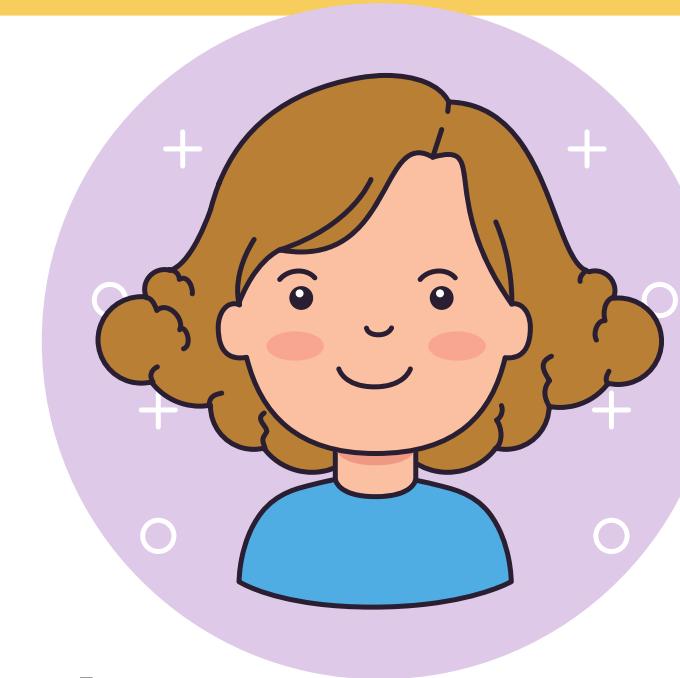


ETL

THÀNH VIÊN NHÓM 3



LÊ THỊ KIỀU LINH



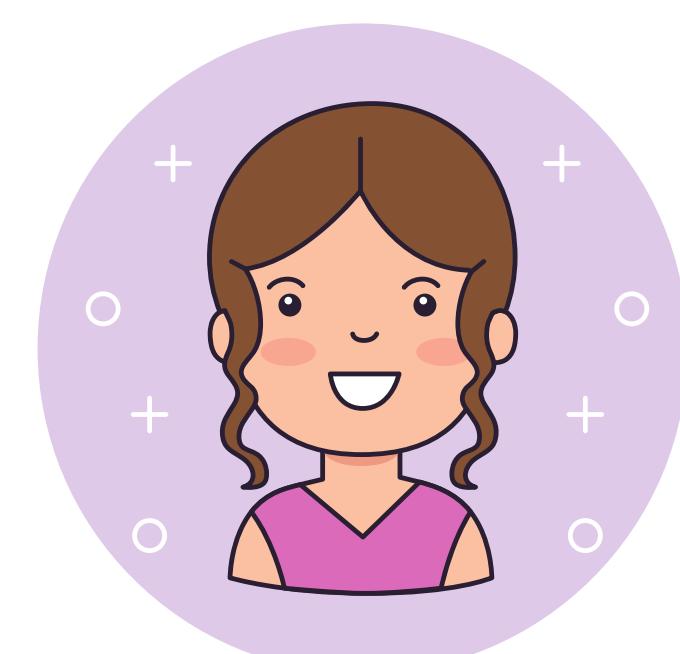
TRẦN THỊ ANH THƯ



PHẠM THỊ VÂN



TẠ THỊ TRANG



NGUYỄN THỊ BẢO TRÂM

PHẦN TRĂM ĐÓNG GÓP

LÊ THỊ KIỀU LINH

ETL ProETL

20%

TRẦN THỊ ANH THƯ

OLAP Project + Pivot Table

20%

PHẠM THỊ VÂN

MDX

20%

TẠ THỊ TRANG

Dashboard

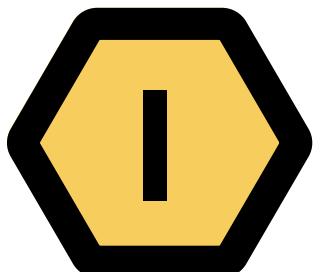
20%

NGUYỄN THỊ BẢO TRÂM

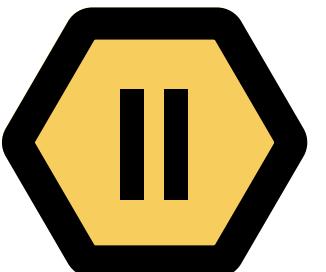
Dashboard

20%

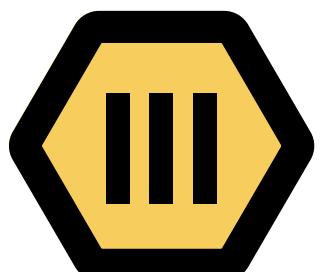
NỘI DUNG BÁO CÁO



GIẢI THÍCH DATA



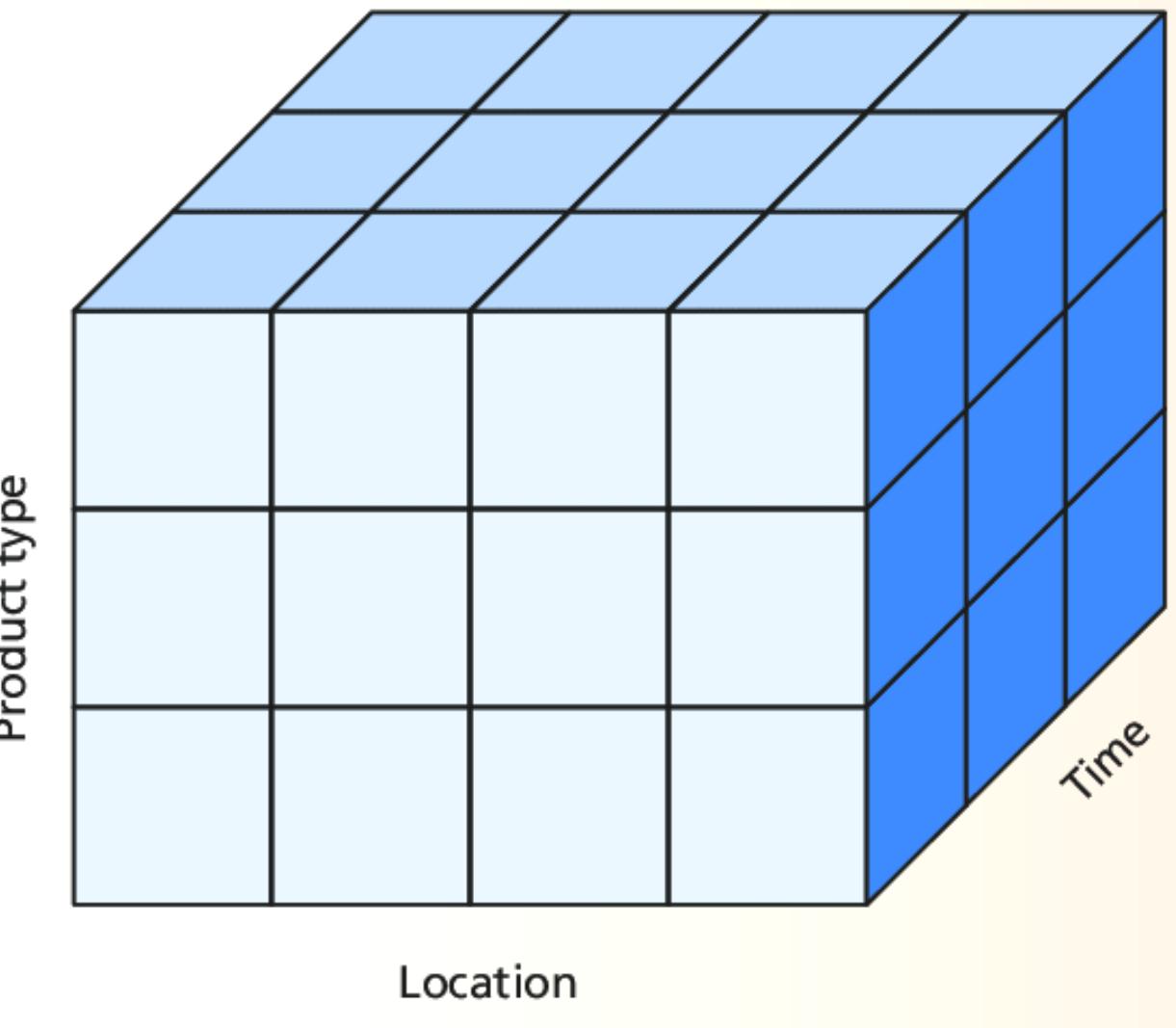
ETL PROJECT



**OLAP PROJECT - PIVOT
TABLE - MDX**



DASHBOARD - KẾT LUẬN



| | Market | Market Size | Profit | Margin | Sales | COGS | Total Expe | Marketing | Inventor |
|-----------------|--------|--------------|--------|--------|-------|------|------------|-----------|----------|
| 203 Connecticut | East | Small Market | 107 | 176 | 292 | 116 | 69 | 38 | 96 |
| 203 Connecticut | East | Small Market | 75 | 135 | 225 | 120 | 60 | 29 | 114 |
| 203 Connecticut | East | Small Market | 122 | 195 | 325 | 123 | 73 | 42 | 113 |
| 203 Connecticut | East | Small Market | 105 | 174 | 289 | 125 | 69 | 37 | 116 |
| 203 Connecticut | East | Small Market | 104 | 135 | 223 | 90 | 56 | 29 | 114 |
| 203 Connecticut | East | Small Market | 104 | 135 | 223 | 90 | 56 | 29 | 113 |
| 203 Connecticut | East | Small Market | 135 | 155 | 275 | 103 | 64 | 33 | 113 |
| 203 Connecticut | East | | | | | 125 | 73 | 41 | 111 |
| 203 Connecticut | East | | | | | 130 | 73 | 42 | 113 |
| 203 Connecticut | East | | | | | 20 | 16 | 5 | 80 |
| 203 Connecticut | East | | | | | 36 | 21 | 10 | 80 |
| 203 Connecticut | East | | | | | 19 | 15 | 5 | 81 |
| 203 Connecticut | East | Small Market | 27 | 36 | 64 | 24 | 18 | 6 | 80 |
| 203 Connecticut | East | Small Market | 49 | 54 | 96 | 36 | 21 | 10 | 80 |
| 203 Connecticut | East | Small Market | -2 | 75 | 128 | 53 | 77 | 48 | 59 |
| 203 Connecticut | East | Small Market | 1 | 84 | 144 | 60 | 83 | 54 | 60 |
| 203 Connecticut | East | Small Market | 2 | 86 | 147 | 61 | 84 | 55 | 61 |
| 203 Connecticut | East | Small Market | 11 | 105 | 201 | 75 | 103 | 68 | 52 |
| 203 Connecticut | East | Small Market | 4 | 90 | 165 | 65 | 87 | 51 | 50 |
| 203 Connecticut | East | Small Market | 0 | 87 | 160 | 63 | 87 | 52 | 51 |
| 203 Connecticut | East | Small Market | 0 | 87 | 160 | 63 | 87 | 52 | 51 |

I. GIẢI THÍCH DỮ LIỆU

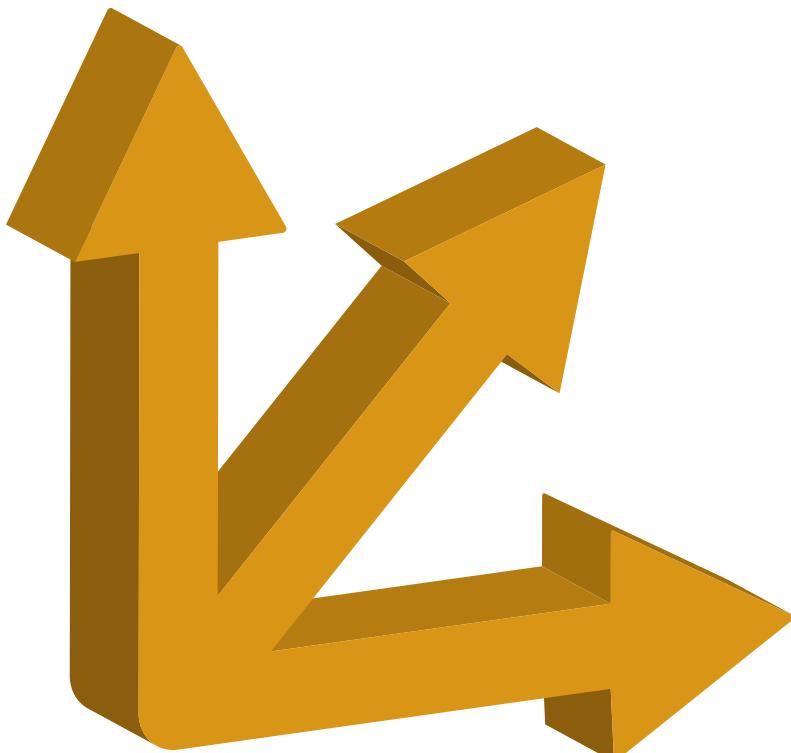
**DATA CHỨA THÔNG TIN VỀ DOANH SỐ BÁN HÀNG (ĐƠN VỊ TÍNH:
ĐÔ LA) TRÊN CÁC CỬA HÀNG Ở MỸ TỪ NĂM 2010 ĐẾN 2011. DỮ LIỆU
ĐƯỢC TRÍCH XUẤT TỪ NGÀY 1 THÁNG 1 NĂM 2010 ĐẾN NGÀY 1
THÁNG 12 NĂM 2011.**

DIMENSION

Market

Product

Date



MEASURES

Profit

Margin

Sales

COGS

Total Expenses

Marketing

Inventory

Budget Profit

Budget COGS

Budget Margin

Budget Sales



| Tên dữ liệu | Kiểu dữ liệu | Ý nghĩa |
|---------------|--------------|---|
| Area Code | int | Mã cửa hàng |
| State | varchar(50) | Tiểu bang |
| Market | varchar(50) | Khu vực |
| Market size | varchar(50) | Quy mô thị trường |
| Date | datetime | Ngày tháng |
| Profit | money | Lợi nhuận |
| Margin | money | $\text{Lãi suất} = \text{Lợi nhuận} + \text{Tổng chi phí HOẶC Doanh thu} - \text{COGS}$ |
| Sales | money | Doanh thu bán hàng |
| COGS | money | Gía vốn hàng bán |
| Total Expense | money | Tổng chi phí có được sản phẩm để bán hàng |

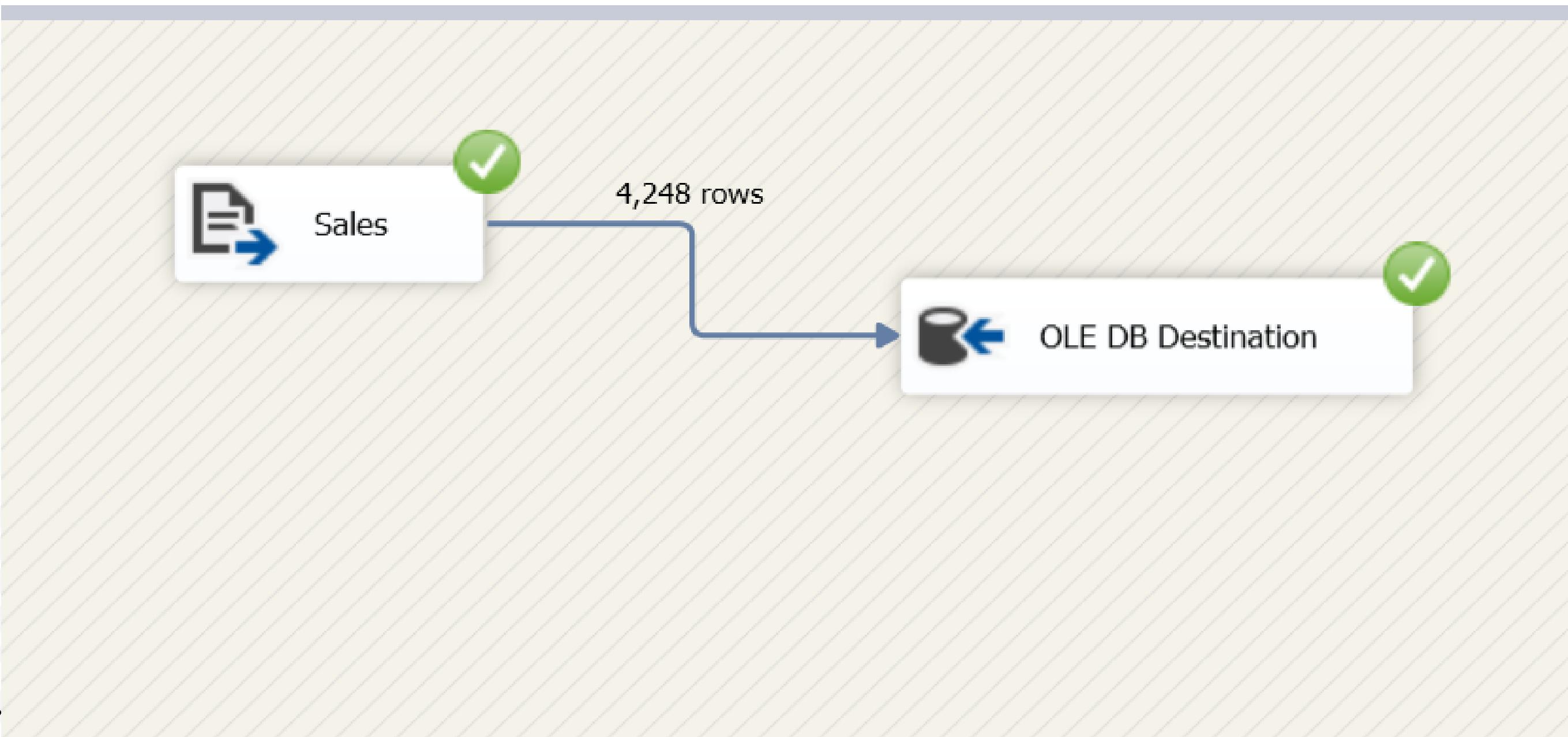
| Tên dữ liệu | Kiểu dữ liệu | Ý nghĩa |
|---------------|--------------|---|
| Marketing | money | Chi phí quảng cáo/tiếp thị |
| Inventory | int | Hàng tồn kho của sản phẩm tại thời điểm giảm giá |
| Budget Profit | money | Lợi nhuận dự kiến |
| Budget COGS | money | Gía vốn hàng bán dự kiến |
| Budget Margin | money | Biên ngân sách = Lợi nhuận dự kiến + Tổng chi phí dự kiến HOẶC Doanh thu dự kiến - Giá vốn dự kiến |
| Budget Sales | money | Doanh số bán hàng dự kiến |
| ProductID | int | Mã sản phẩm |
| Product | varchar(50) | Sản phẩm |
| Product Type | varchar(50) | Danh mục sản phẩm |
| Type | varchar(50) | Loại sản phẩm |



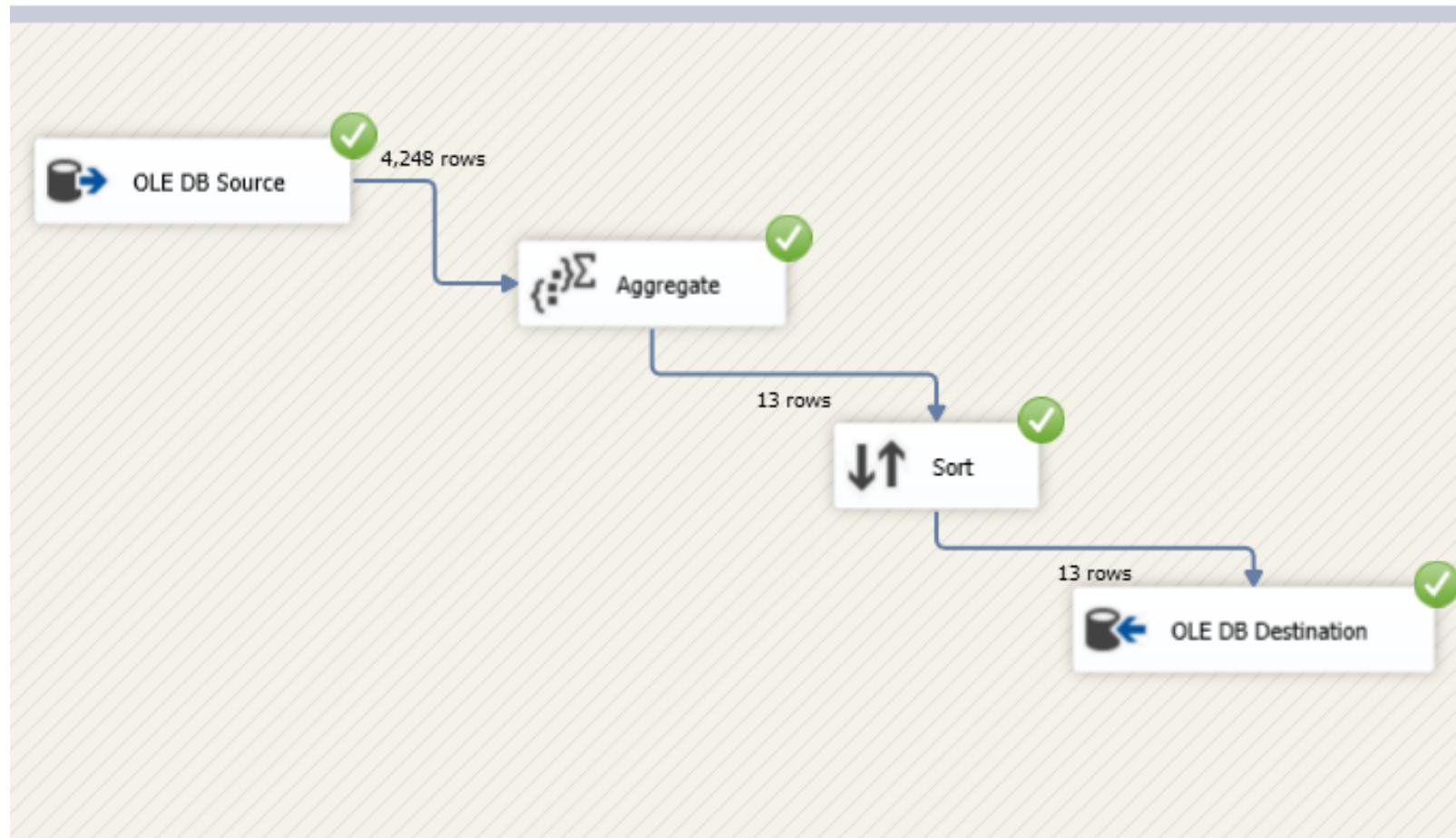
II. ETL PROJECT

1. SALES SOURCE

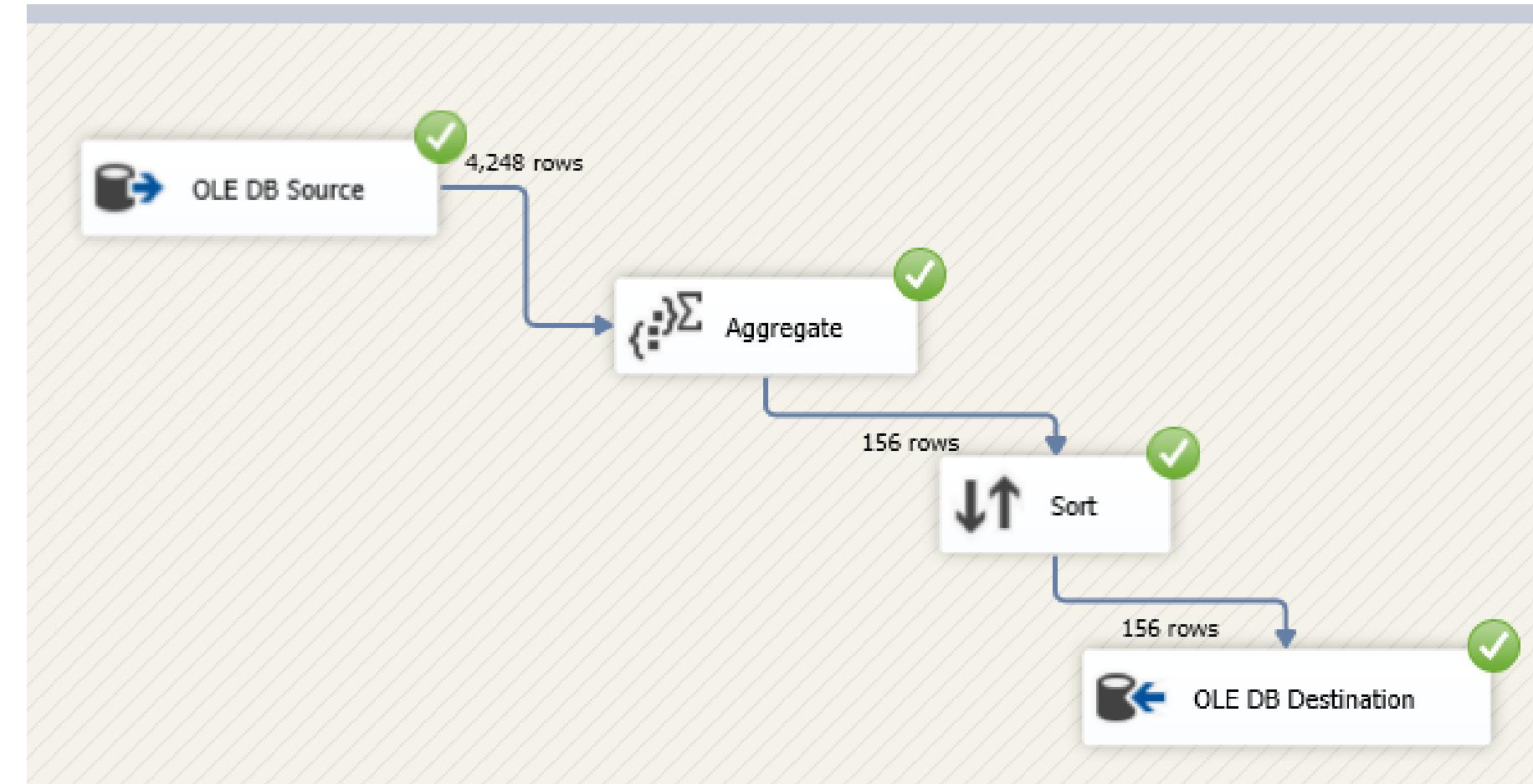
Data Flow Task:  Sales



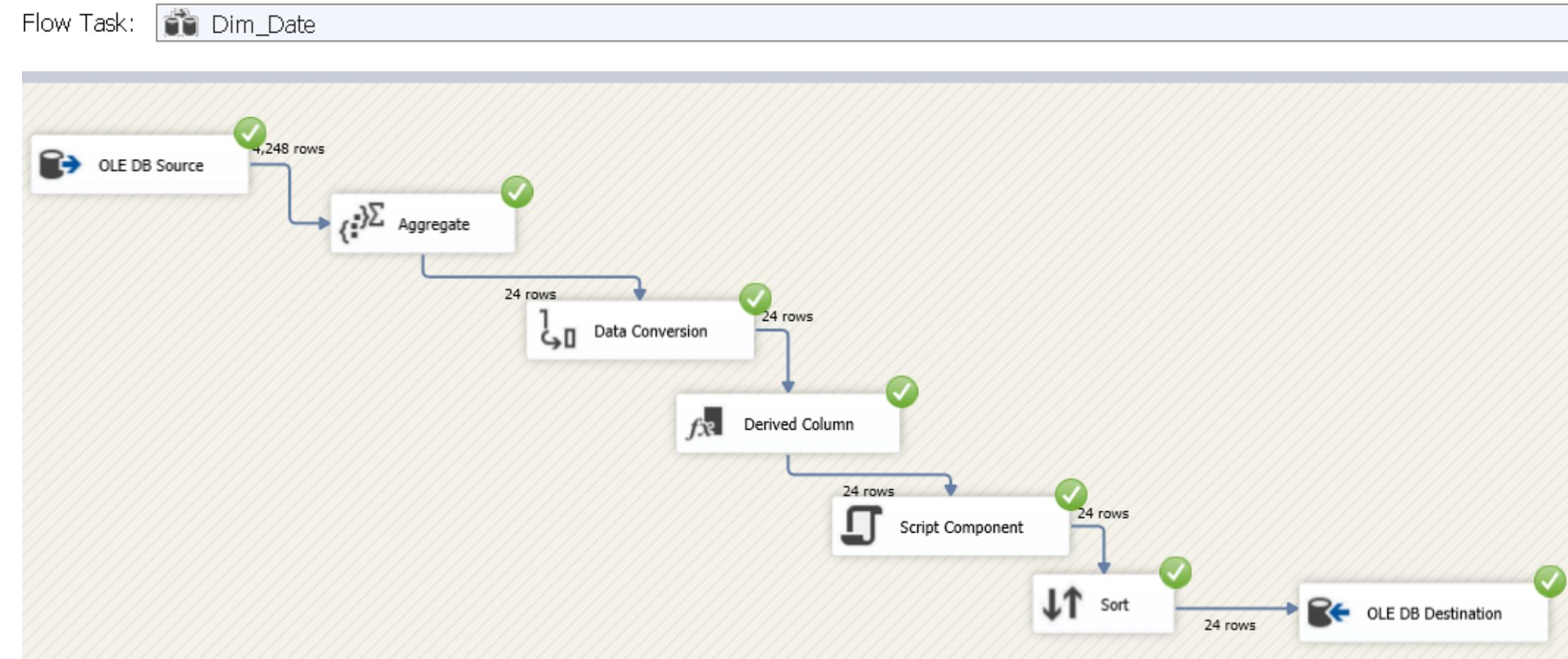
Flow Task: Dim_Product



v Task: Dim_Market

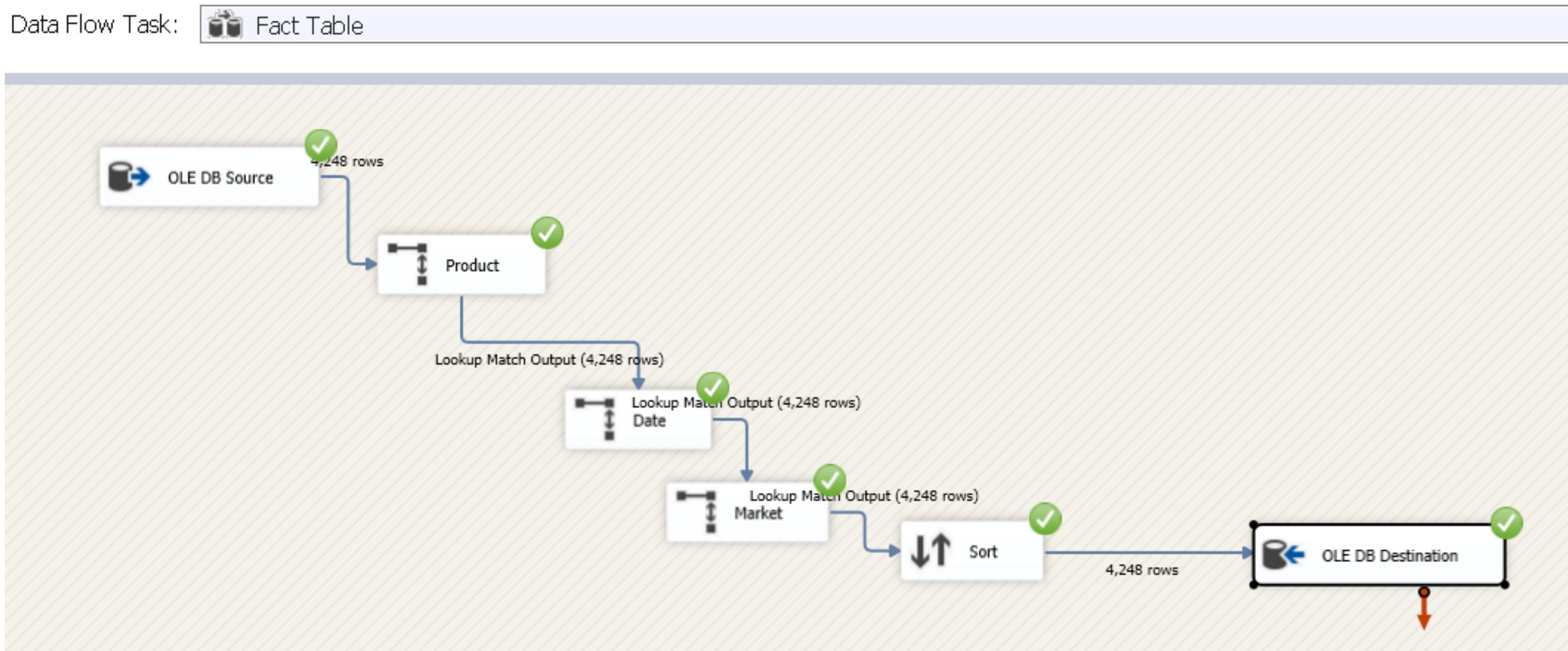


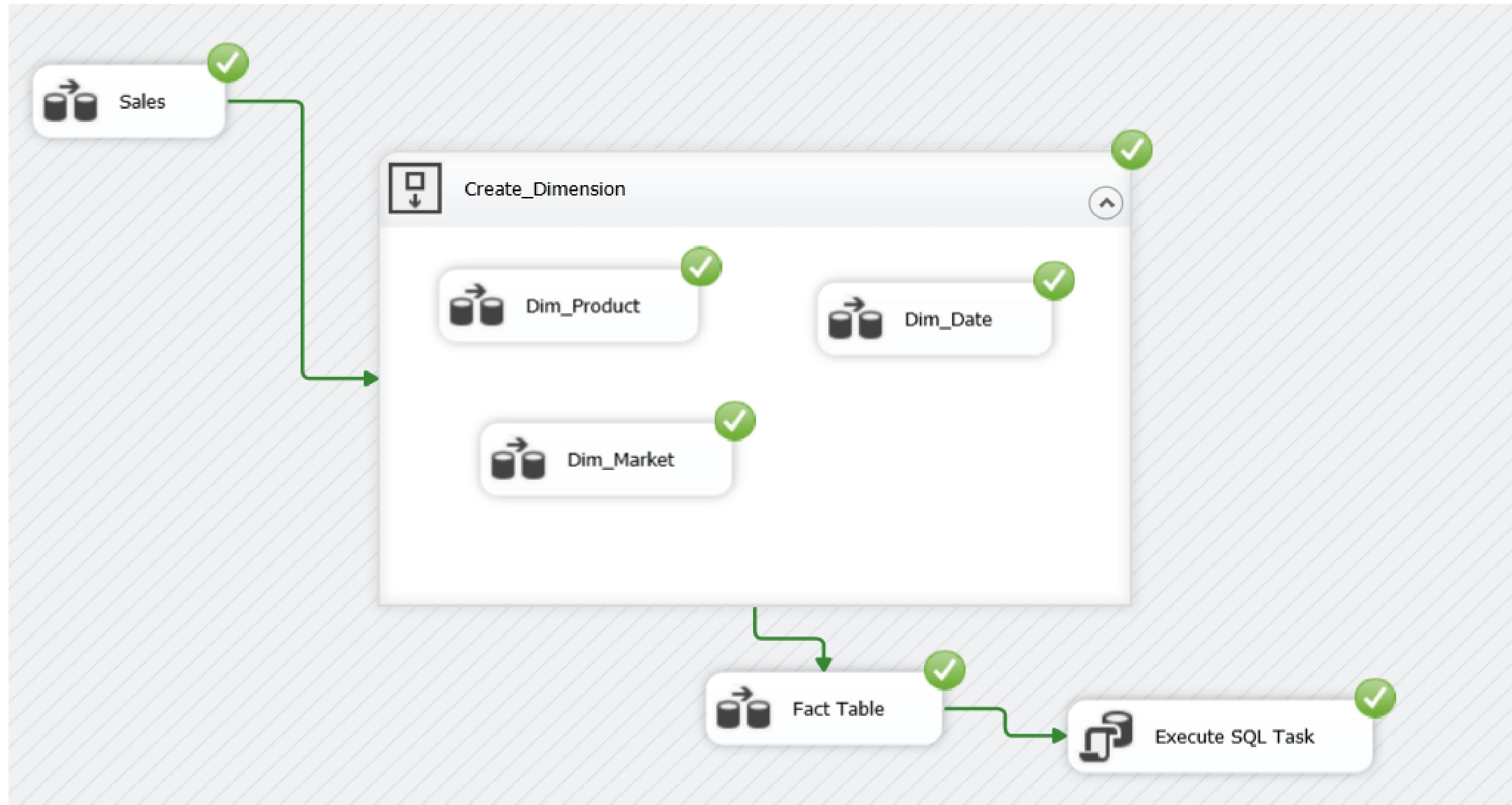
Flow Task: Dim_Date



2. DIMENSION

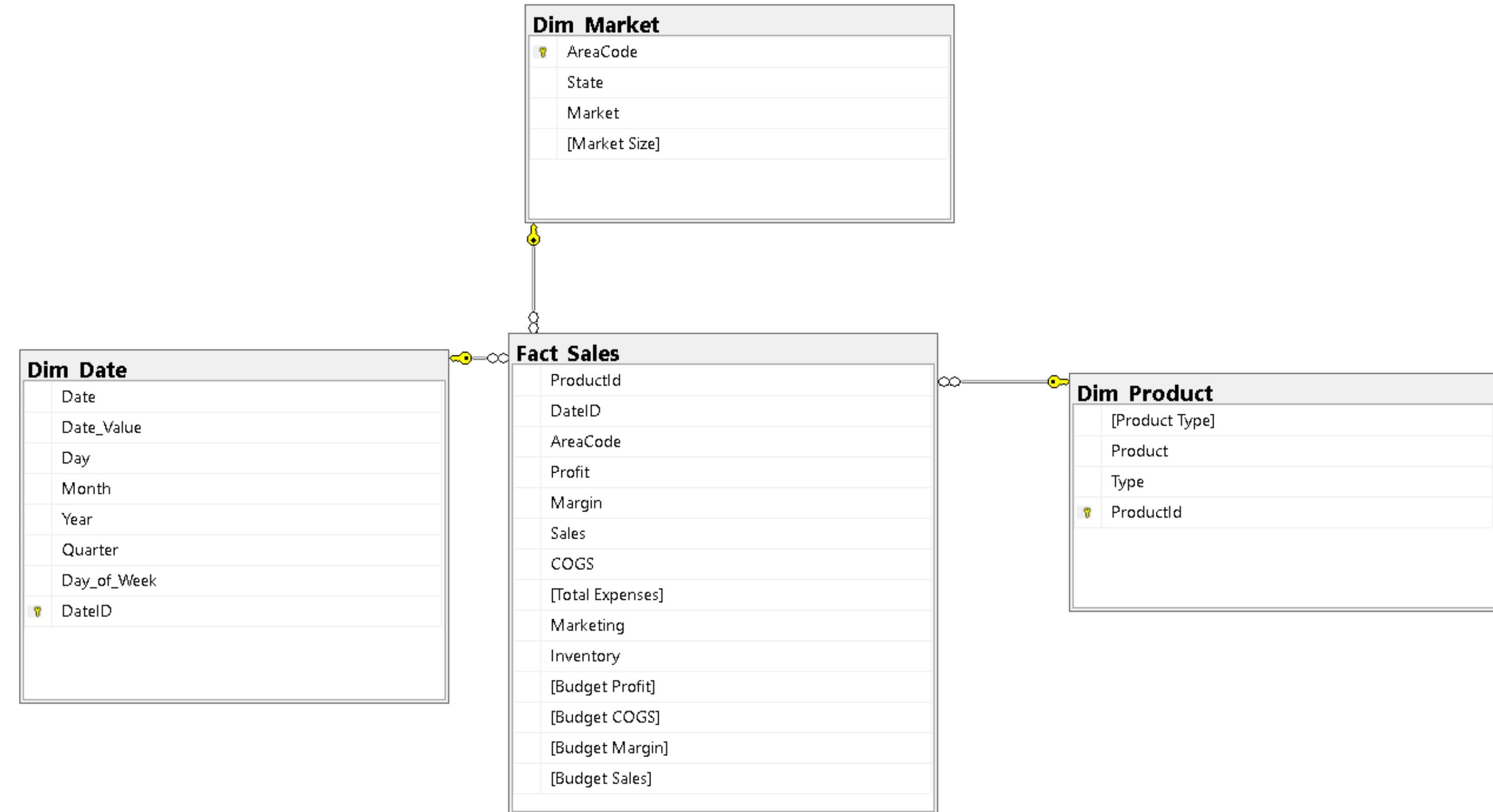
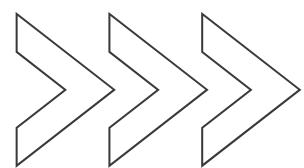
3.FACT TABLE





4. MÔ HÌNH STAR SCHEMA

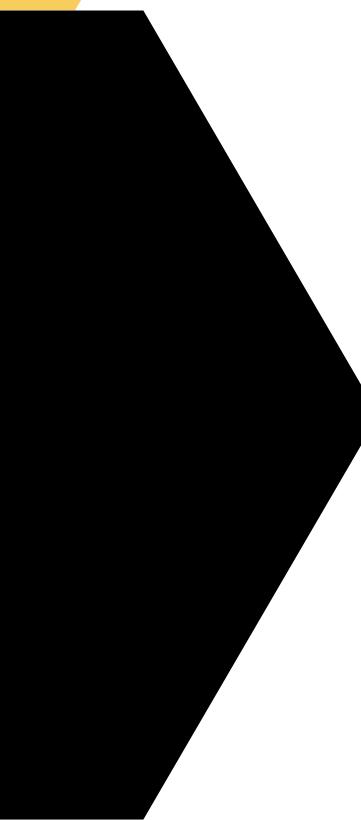
5. MÔ HÌNH DIAGRAM



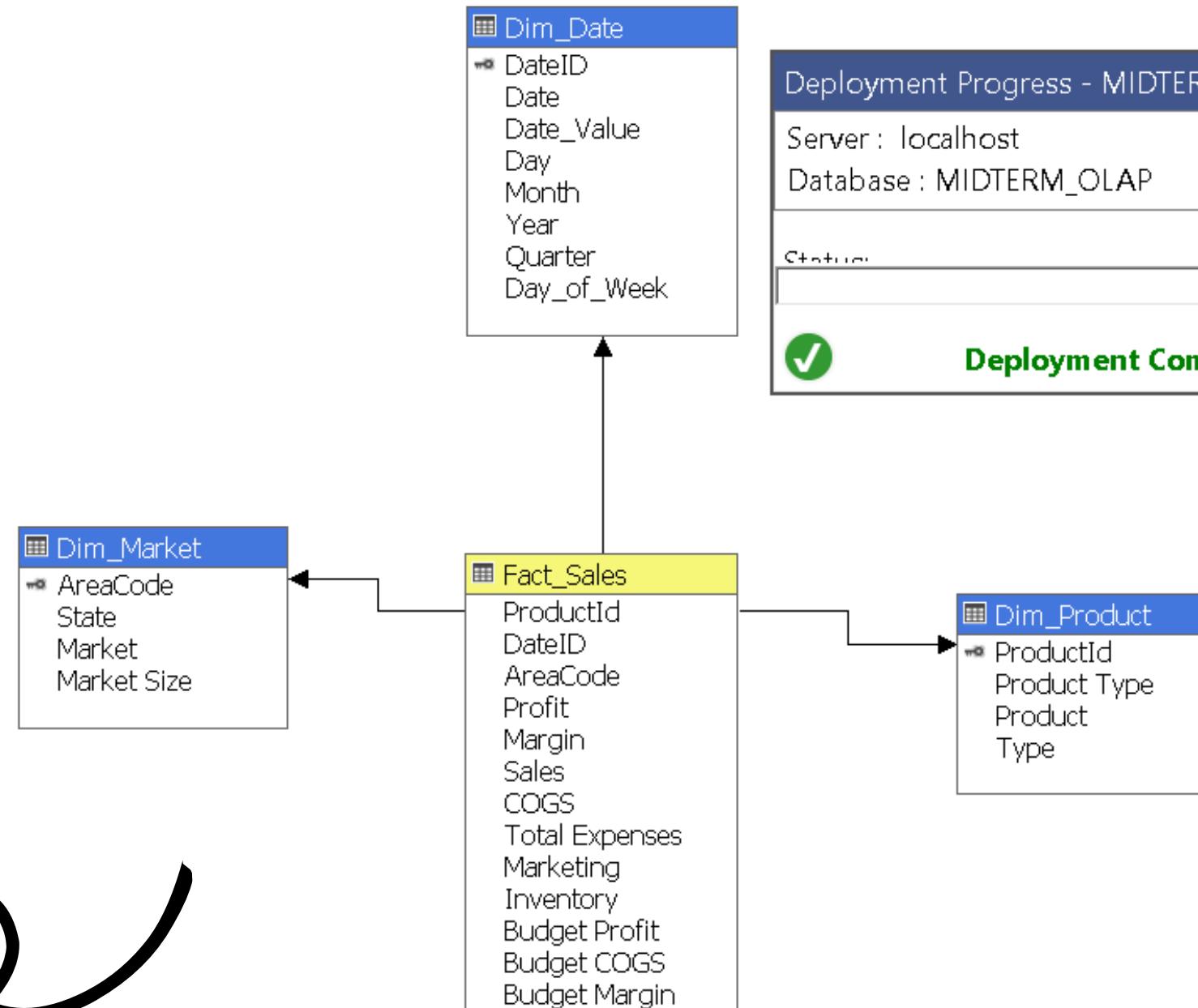


III. OLAP PROJECT

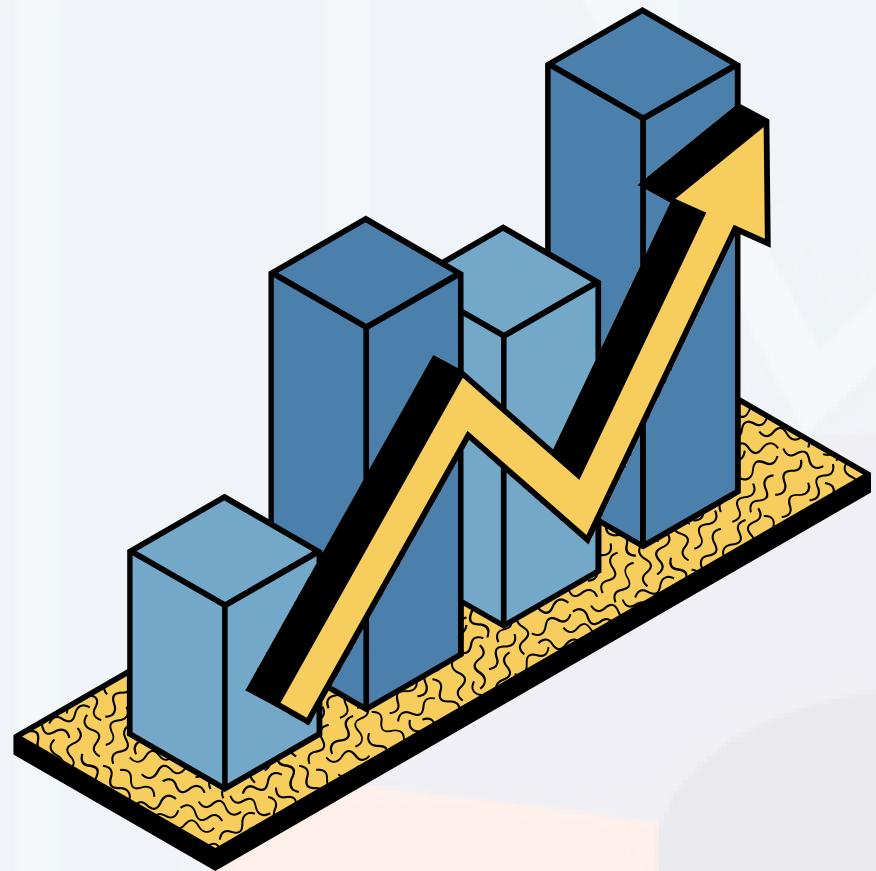
1. OLAP



OLAP Cube



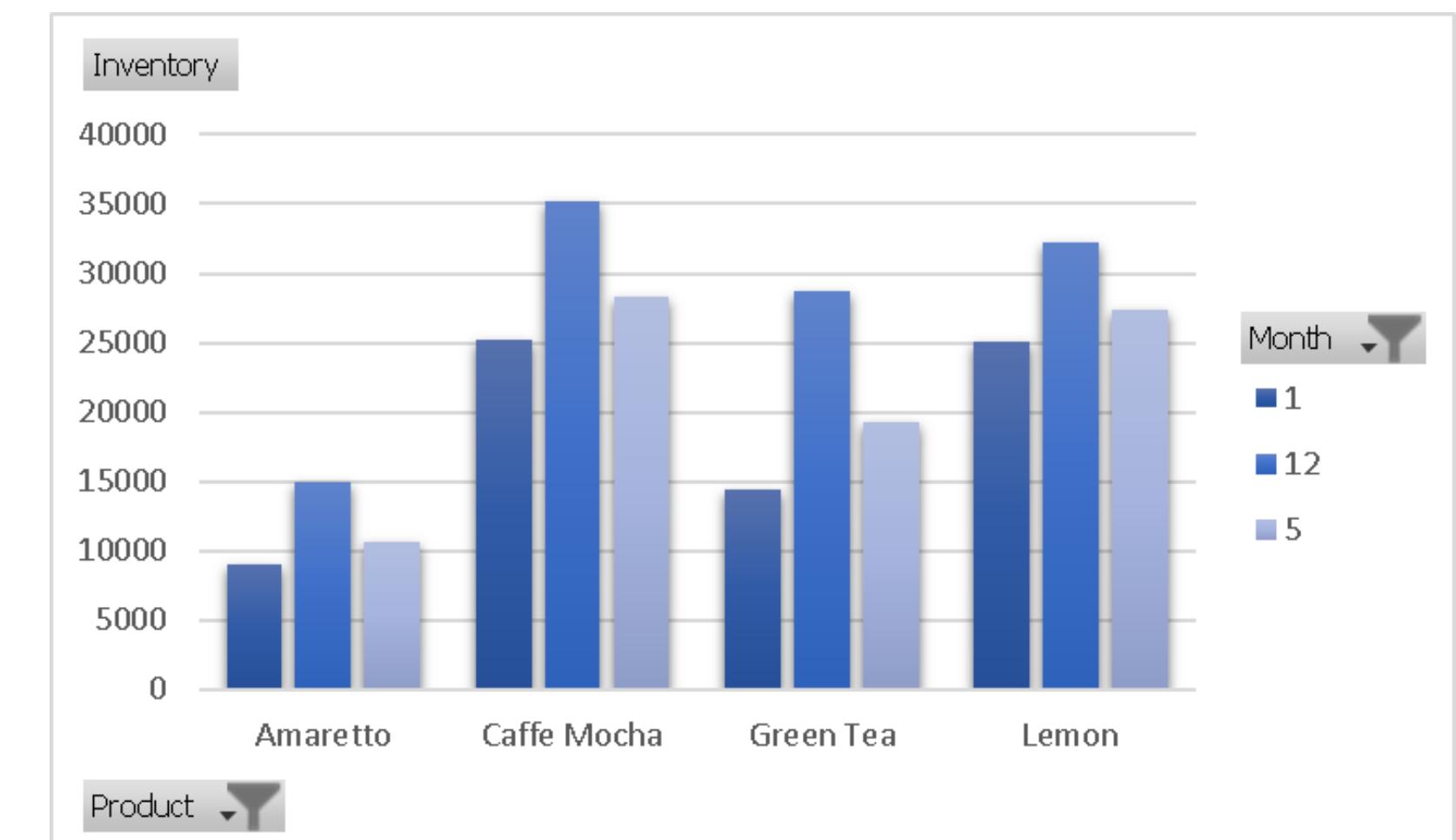
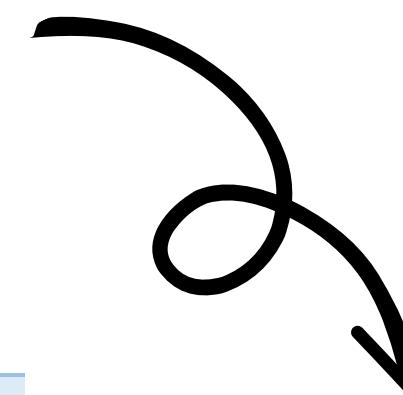
OLAP model



2. PIVOT TABLE VÀ ĐỒ THỊ

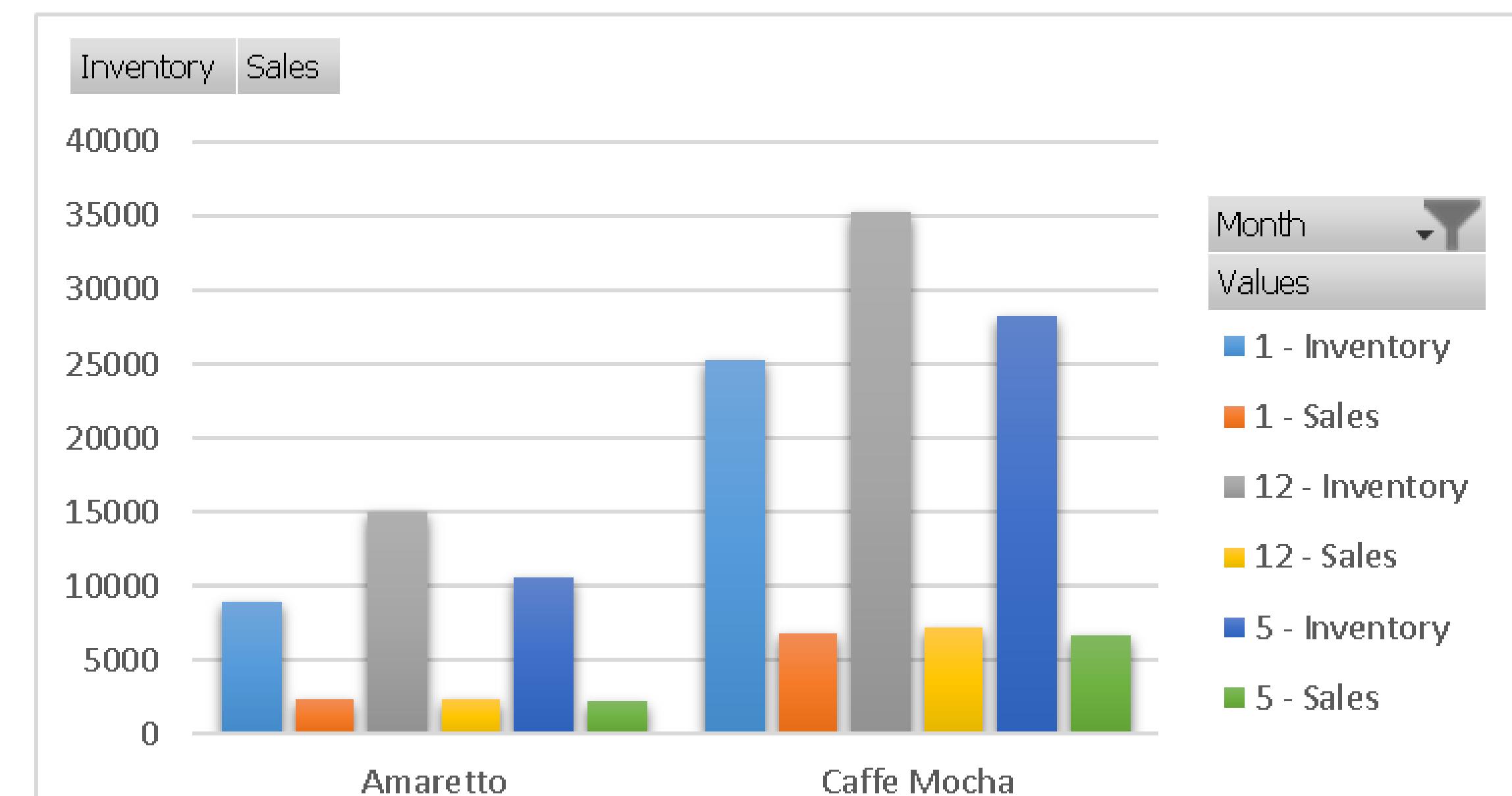
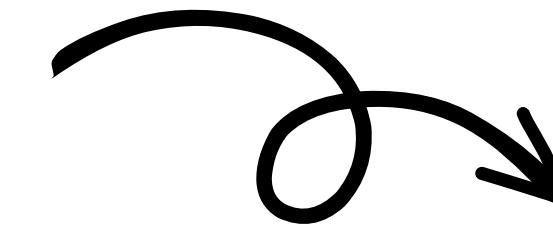
a.Tổng số hàng tồn kho của: Amaretto, CafeMocha, Green Tea, Lemon trong tháng 1, tháng 12 và tháng 5

| Inventory | Column Labels | | | |
|--------------------|---------------|---------------|--------------|---------------|
| Row Labels | 1 | 12 | 5 | Grand Total |
| Amaretto | 8990 | 15014 | 10570 | 34574 |
| Caffe Mocha | 25250 | 35200 | 28280 | 88730 |
| Green Tea | 14462 | 28762 | 19270 | 62494 |
| Lemon | 25044 | 32166 | 27346 | 84556 |
| Grand Total | 73746 | 111142 | 85466 | 270354 |



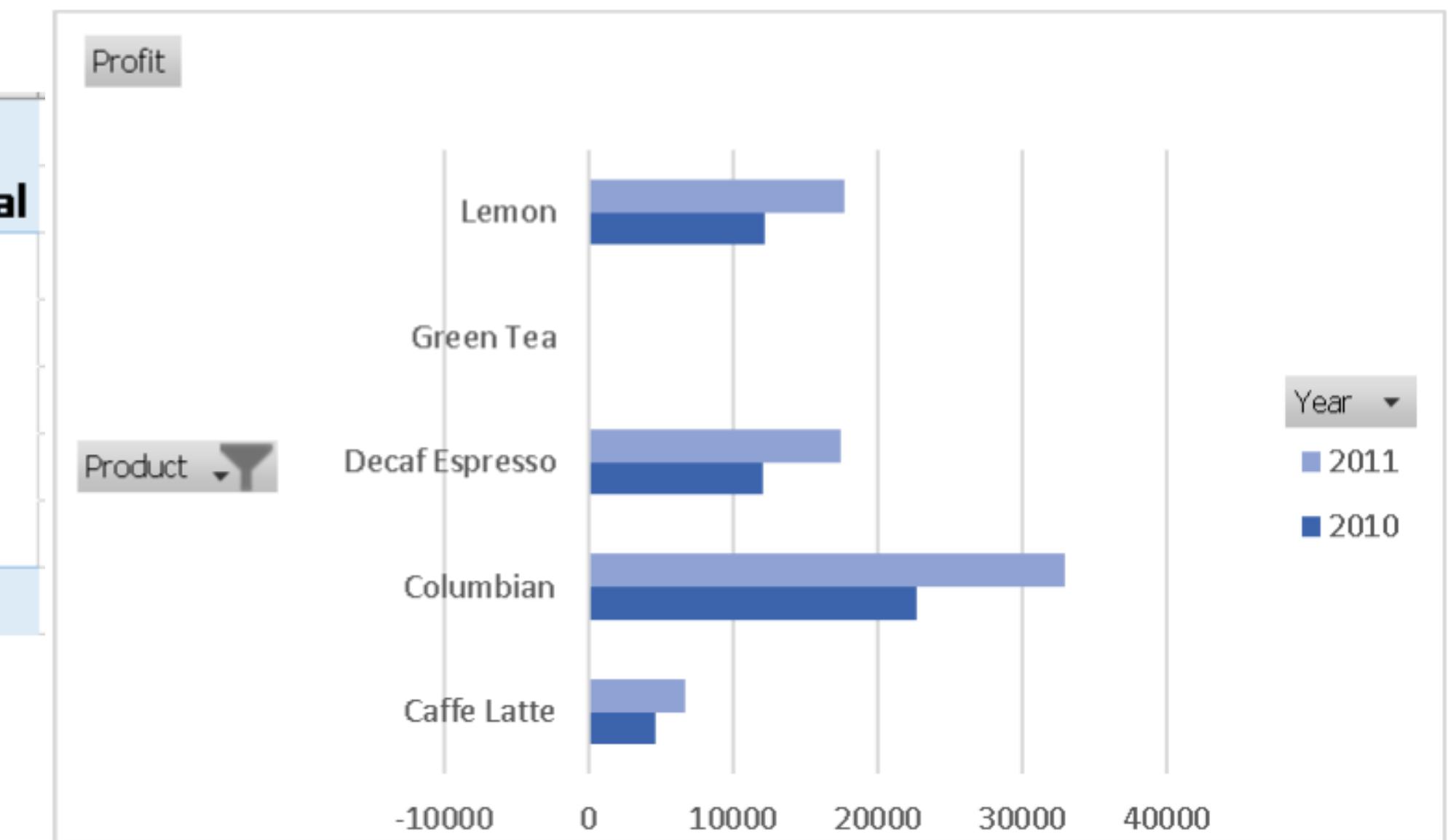
a.Tổng số hàng tồn kho, doanh thu của: Amaretto, CafeMocha, Green Tea, Lemon trong tháng 1, tháng 12 và tháng 5

| | | Column Labels | | | | | |
|--------------------|--------------|---------------|--------------|-------------|-----------------|-------------|--------------|
| | | 1 | 12 | 5 | Total Inventory | Total Sales | |
| Row Labels | Inventory | Sales | Inventory | Sales | Inventory | Sales | |
| Amaretto | 8990 | 2292 | 15014 | 2262 | 10570 | 2122 | 34574 |
| Caffe Mocha | 25250 | 6824 | 35200 | 7201 | 28280 | 6673 | 88730 |
| Grand Total | 34240 | 9116 | 50214 | 9463 | 38850 | 8795 | 27374 |



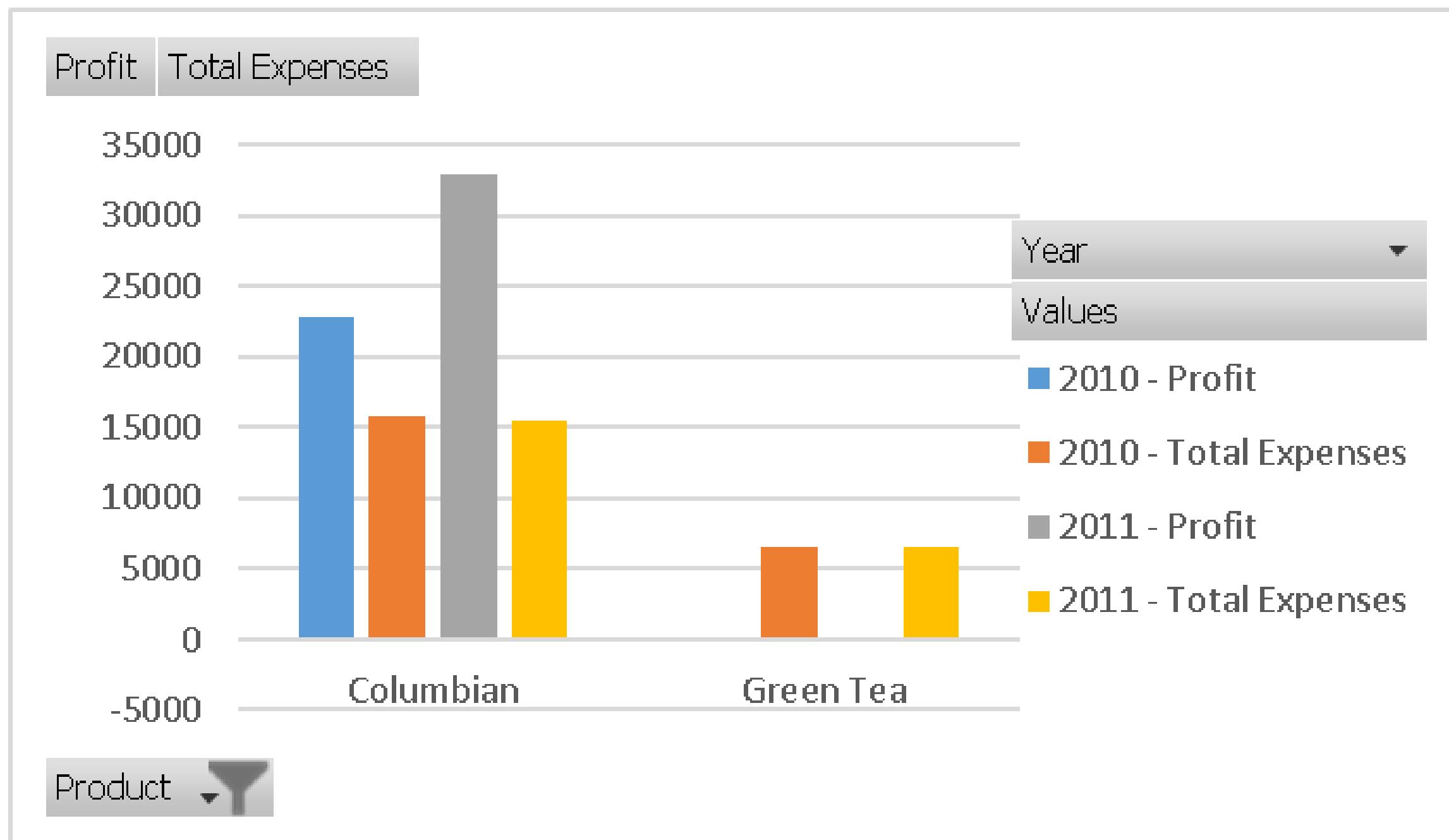
b.Tổng lợi nhuận của các sản phẩm trong năm 2010 và 2011

| Profit | Column Labels | | |
|--------------------|---------------|--------------|---------------|
| Row Labels | 2010 | 2011 | Grand Total |
| Caffe Latte | 4636 | 6739 | 11375 |
| Columbian | 22777 | 33027 | 55804 |
| Decaf Espresso | 12025 | 17477 | 29502 |
| Green Tea | -90 | -141 | -231 |
| Lemon | 12195 | 17674 | 29869 |
| Grand Total | 51543 | 74776 | 126319 |



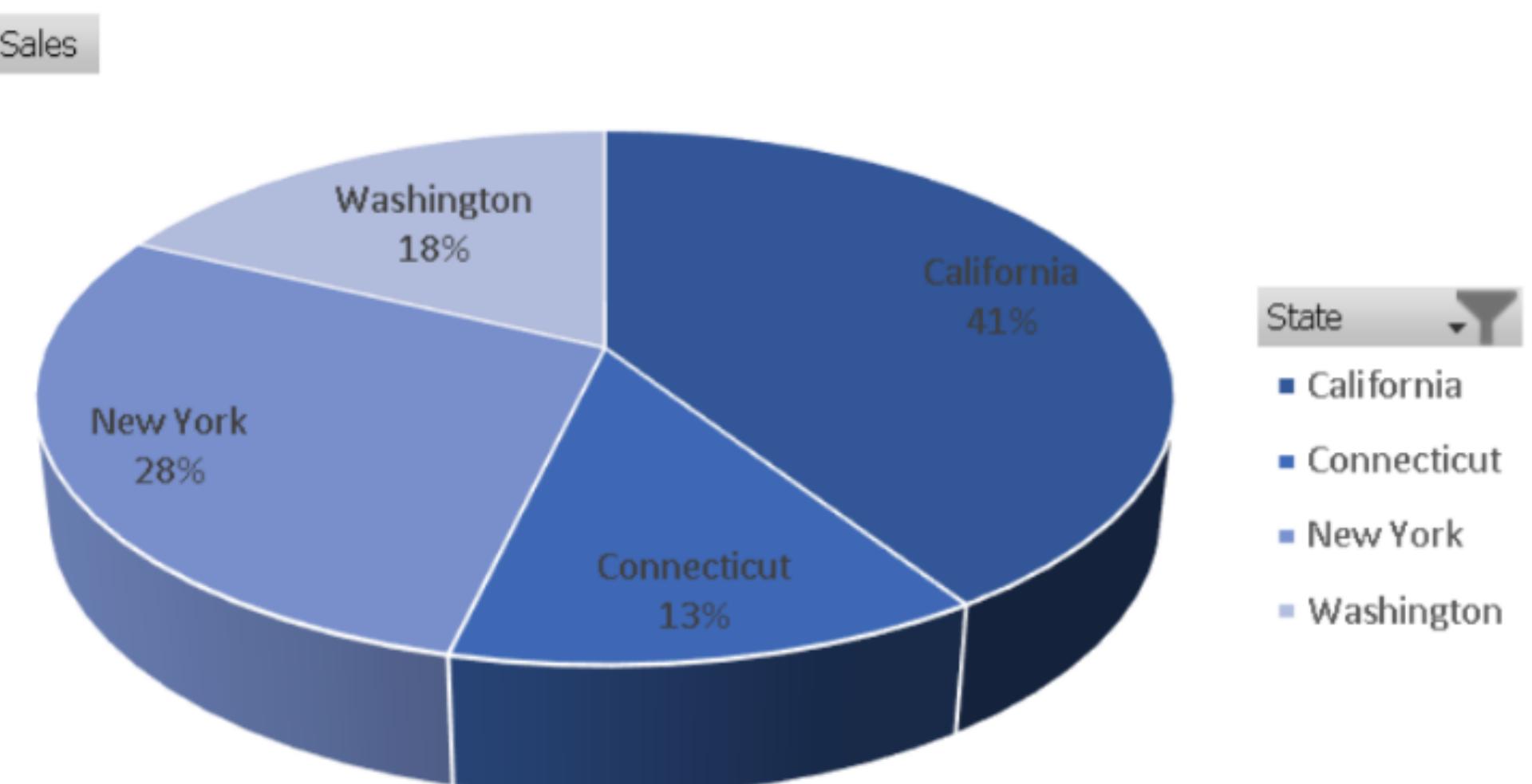
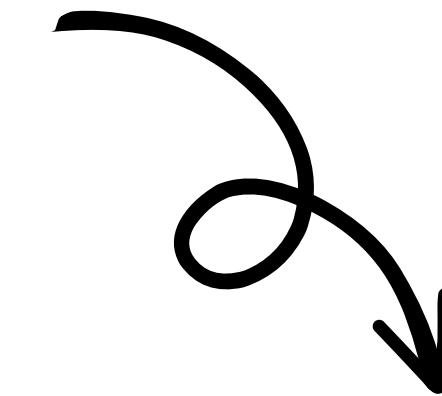
**b.Tổng lợi nhuận,
doanh thu của các
sản phẩm trong
năm 2010 và 2011**

| Row Labels | Column Labels | | 2010 | | 2011 | | Total Profit | Total Expenses |
|--------------------|---------------|----------------|--------------|----------------|--------------|--------------|--------------|----------------|
| | Profit | Total Expenses | Profit | Total Expenses | | | | |
| Columbian | 22777 | 15849 | 33027 | 15524 | 55804 | 31373 | | |
| Green Tea | -90 | 6616 | -141 | 6491 | -231 | 13107 | | |
| Grand Total | 22687 | 22465 | 32886 | 22015 | 55573 | 44480 | | |

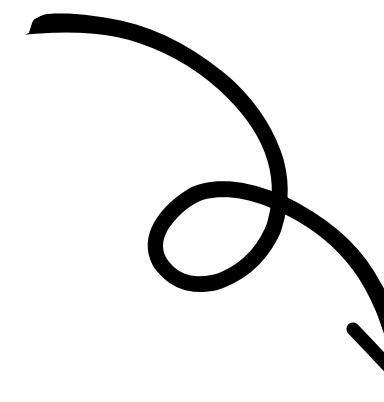


c.Tổng Doanh thu theo quý của tại các cửa hàng trong khu vực

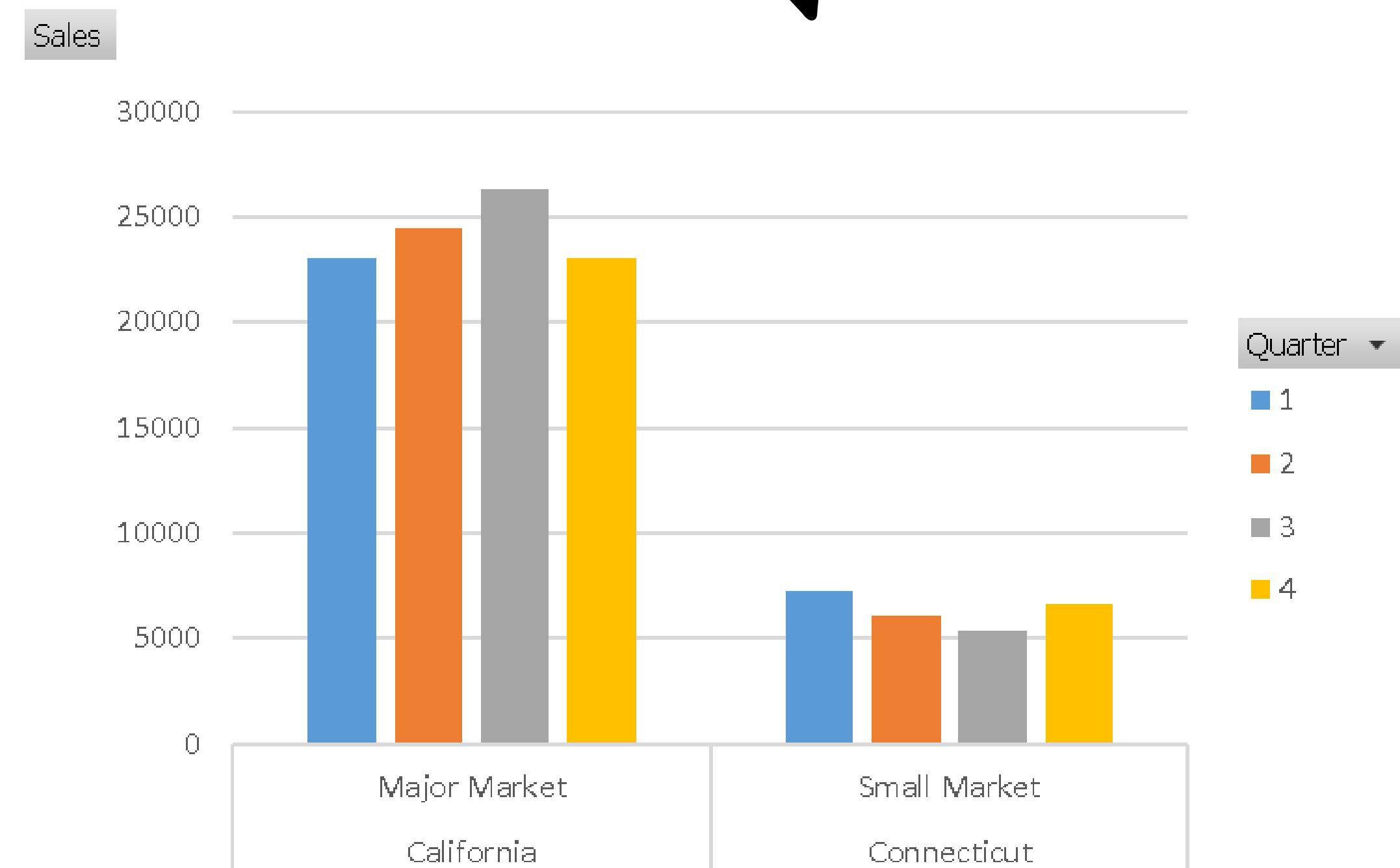
| Sales | Column Labels | | | | | |
|--------------------|---------------|--------------|--------------|--------------|---------------|--|
| Row Labels | 1 | 2 | 3 | 4 | Grand Total | |
| California | 23029 | 24503 | 26328 | 23032 | 96892 | |
| Connecticut | 7235 | 6123 | 5401 | 6670 | 25429 | |
| New York | 16046 | 18299 | 18776 | 17731 | 70852 | |
| Washington | 10069 | 9462 | 9655 | 9744 | 38930 | |
| Grand Total | 56379 | 58387 | 60160 | 57177 | 232103 | |



| Sales | Column Labels ▾ | | | | |
|--------------|-----------------|-------|-------|-------|-------------|
| Row Labels | 1 | 2 | 3 | 4 | Grand Total |
| California | 23029 | 24503 | 26328 | 23032 | 96892 |
| Major Market | 23029 | 24503 | 26328 | 23032 | 96892 |
| Connecticut | 7235 | 6123 | 5401 | 6670 | 25429 |
| Small Market | 7235 | 6123 | 5401 | 6670 | 25429 |
| Grand Total | 30264 | 30626 | 31729 | 29702 | 122321 |



c.Tổng Doanh
thu theo quý
của tại các
cửa hàng
trong khu vực



3.MDX

```
//1. Total Profit by Product Type  
select [Measures].[Profit] on columns,  
       NONEMPTY ([Dim Product].[Product Type].[Product Type]) on rows  
from [Midterm OLAP View];
```



| Messages | | Results |
|------------|--|---------|
| | | Profit |
| Coffee | | 74683 |
| Espresso | | 68620 |
| Herbal Tea | | 63254 |
| Tea | | 52986 |

```
//2. Total Sales by years  
select [Measures].[Sales] on columns,  
       NONEMPTY ([Dim Date].[Year].[Year]) on rows  
from [Midterm OLAP View];
```



| Messages | | Results |
|----------|--|---------|
| | | Sales |
| 2010 | | 401159 |
| 2011 | | 418652 |

3.MDX

//3. Total COGS by Product and Month

```
select NONEMPTY ([Dim Date].[Month].[Month]) ON COLUMNS,  
       NONEMPTY ([Dim Product].[Product].[Product]) ON ROWS  
from [Midterm_Olap_View]  
where [Measures].[COGS];
```



| | 1 | 10 | 11 | 12 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Amaretto | 990 | 1078 | 1052 | 1094 | 1008 | 988 | 1044 | 1066 | 1102 | 1098 | 1118 | 1094 |
| Caffe Latte | 1314 | 1190 | 1186 | 1254 | 1266 | 1302 | 1310 | 1352 | 1310 | 1372 | 1236 | 1202 |
| Caffe Mocha | 2928 | 3240 | 3158 | 3208 | 3050 | 3034 | 3040 | 3054 | 3222 | 3322 | 3566 | 3168 |
| Columbian | 3776 | 3834 | 3786 | 4016 | 3772 | 3814 | 3898 | 4034 | 4340 | 4500 | 4560 | 4066 |
| Chamomile | 2512 | 2706 | 2620 | 2614 | 2562 | 2536 | 2544 | 2538 | 2644 | 2706 | 2822 | 2582 |
| Darjeeling | 2412 | 2552 | 2334 | 2446 | 2450 | 2468 | 2464 | 2576 | 2698 | 2742 | 2826 | 2566 |
| Decaf Espresso | 2750 | 2712 | 2596 | 2652 | 2750 | 2786 | 2768 | 2834 | 2864 | 2926 | 2930 | 2750 |
| Decaf Irish Cream | 2288 | 2272 | 2234 | 2442 | 2332 | 2372 | 2508 | 2550 | 2684 | 2872 | 2670 | 2344 |
| Earl Grey | 2240 | 2324 | 2230 | 2438 | 2266 | 2286 | 2354 | 2450 | 2486 | 2516 | 2500 | 2464 |
| Green Tea | 1514 | 1592 | 1574 | 1680 | 1492 | 1508 | 1578 | 1604 | 1602 | 1636 | 1658 | 1640 |
| Lemon | 3192 | 3352 | 3300 | 3552 | 3242 | 3274 | 3376 | 3478 | 3616 | 3770 | 3842 | 3500 |
| Mint | 1550 | 1582 | 1488 | 1572 | 1568 | 1584 | 1606 | 1690 | 1806 | 1882 | 1936 | 1666 |
| Regular Espresso | 854 | 850 | 852 | 846 | 856 | 868 | 860 | 886 | 888 | 902 | 880 | 856 |

3.MDX

//4. Total Sales by State and Quarter

```
select NONEMPTY ([Dim Date].[Quarter].[Quarter]) on columns,  
       NONEMPTY ([Dim Market].[State].[State]) on rows  
from [Midterm OLAP View]  
where [Measures].[Sales];
```



| | Messades | Results | 1 | 2 | 3 | 4 |
|---------------|----------|---------|-------|-------|-------|---|
| California | | 23029 | 24503 | 26328 | 23032 | |
| Colorado | | 12377 | 11901 | 11789 | 12112 | |
| Connecticut | | 7235 | 6123 | 5401 | 6670 | |
| Florida | | 8606 | 9601 | 10411 | 8825 | |
| Illinois | | 16428 | 17989 | 18654 | 16812 | |
| Iowa | | 12864 | 13826 | 14701 | 13359 | |
| Louisiana | | 6053 | 5426 | 5489 | 6193 | |
| Massachusetts | | 7626 | 7432 | 7568 | 7339 | |
| Missouri | | 6382 | 6197 | 5996 | 6072 | |
| Nevada | | 13722 | 15298 | 16317 | 14822 | |
| New Hampshire | | 3441 | 3781 | 4122 | 3543 | |
| New Mexico | | 3626 | 4042 | 4419 | 3805 | |
| New York | | 16046 | 18299 | 18776 | 17731 | |
| Ohio | | 8903 | 8502 | 8325 | 8787 | |
| Oklahoma | | 6170 | 6682 | 7313 | 7298 | |
| Oregon | | 10541 | 9866 | 9831 | 10661 | |
| Texas | | 9384 | 9246 | 9688 | 9092 | |
| Utah | | 8770 | 8690 | 8765 | 9159 | |
| Washington | | 10069 | 9462 | 9655 | 9744 | |
| Wisconsin | | 8483 | 8207 | 8540 | 7839 | |

3.MDX

//5. Total Inventory and Margin by Product

```
select {[Measures].[Inventory],[Margin]} on columns,  
      NONEMPTY ([Dim Product].[Product].[Product]) on rows  
from [Midterm_OLAP_View]
```



| | Inventory | Margin |
|-------------------|-----------|--------|
| Amaretto | 142276 | 12950 |
| Caffe Latte | 55294 | 19824 |
| Caffe Mocha | 362996 | 45084 |
| Columbian | 339624 | 77252 |
| Chamomile | 273656 | 42552 |
| Darjeeling | 283328 | 41064 |
| Decaf Espresso | 308124 | 43162 |
| Decaf Irish Cream | 322054 | 31370 |
| Earl Grey | 218066 | 36786 |
| Green Tea | 259308 | 12444 |
| Lemon | 344898 | 52418 |
| Mint | 210414 | 15030 |
| Regular Espresso | 63334 | 13102 |

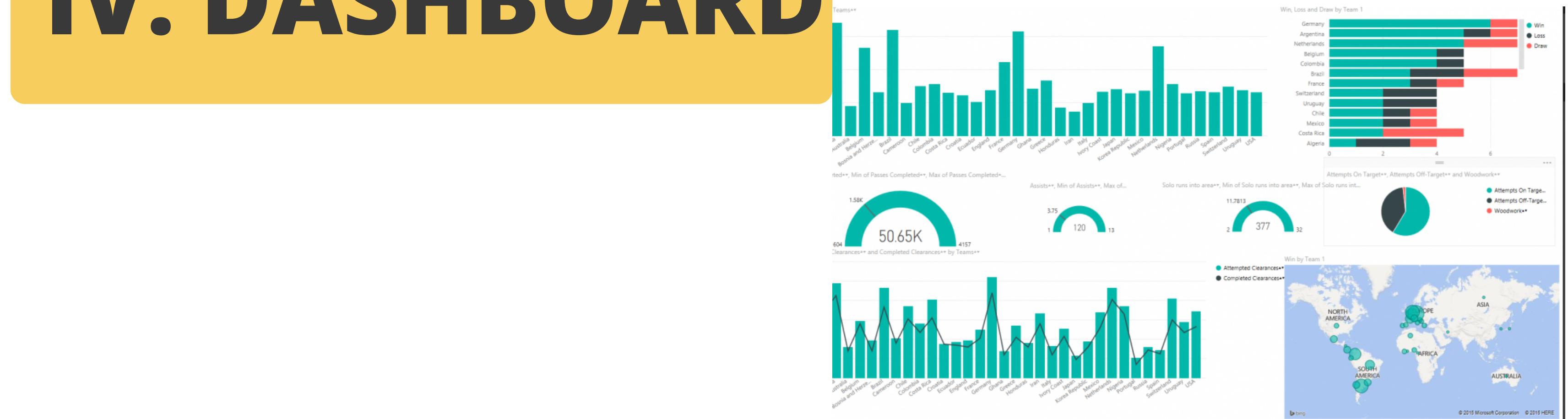
3.MDX

```
//6. Total Sales by years of New York
select {[Measures].[Sales],[Marketing]} on columns,
NONEMPTY ([Dim Date].[Year].[Year]) on rows
from [Midterm_Olap_View]
where [Dim Market].[State].[New York];
```



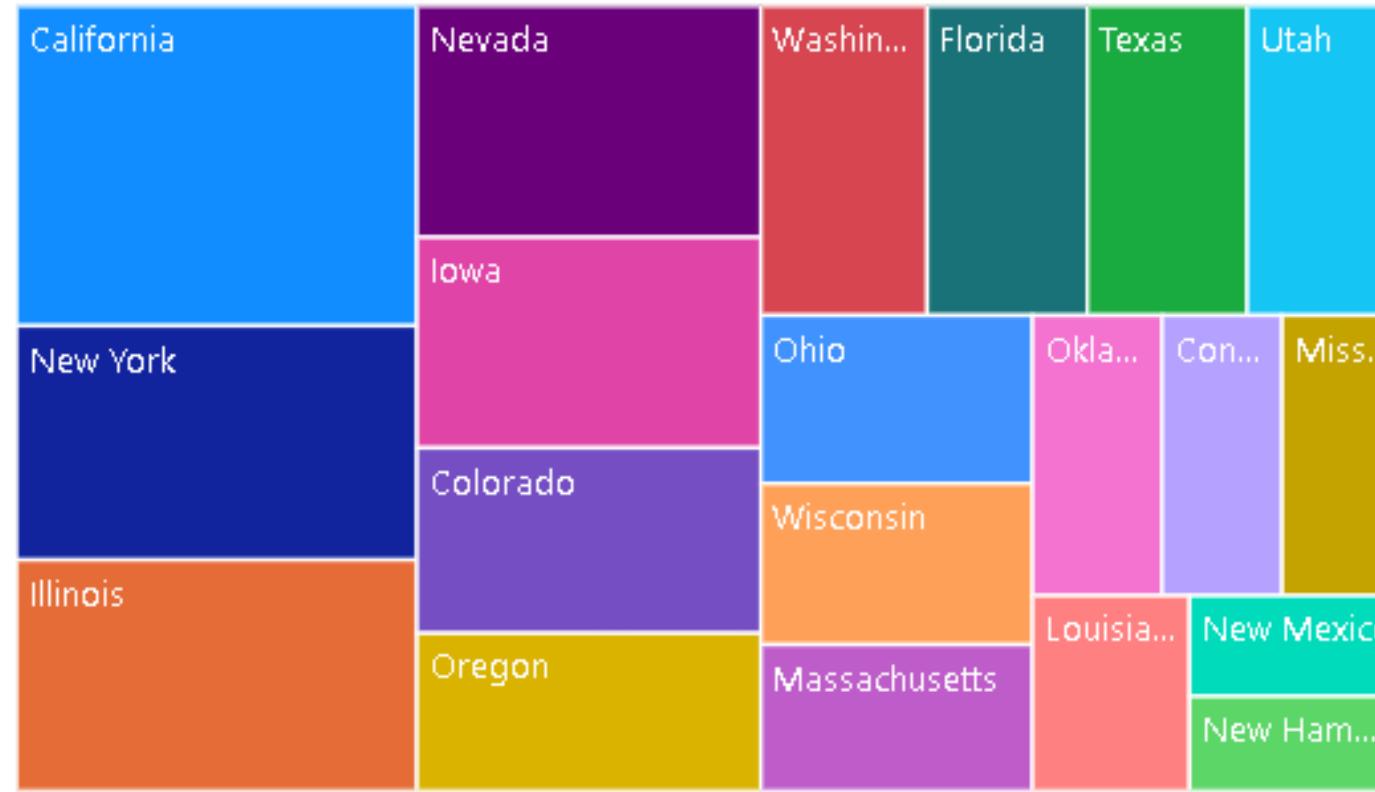
| | Messades | Results |
|------|----------|-----------|
| | Sales | Marketing |
| 2010 | 34698 | 6522 |
| 2011 | 36154 | 6522 |

IV. DASHBOARD



MARKET

Tổng cùa Sales theo State



156

Tổng số của State

4

Tổng số của Market

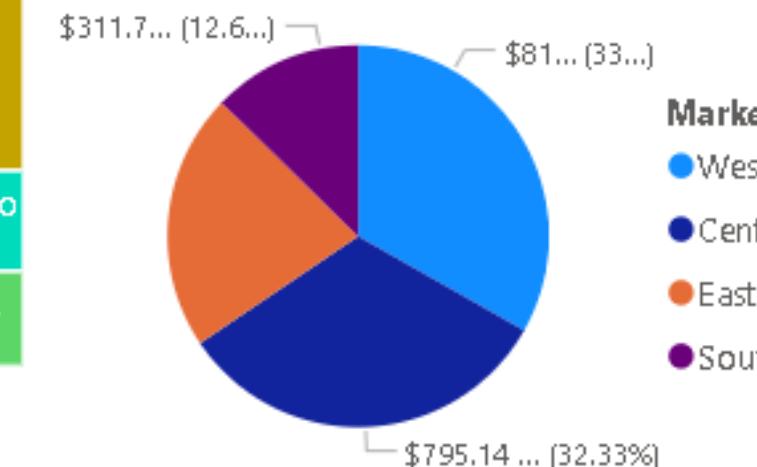
4

Tổng số của Product Type

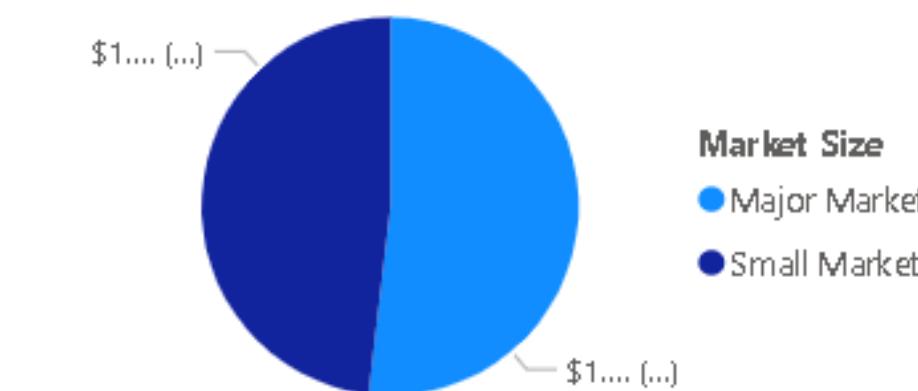
13

Tổng số của Product

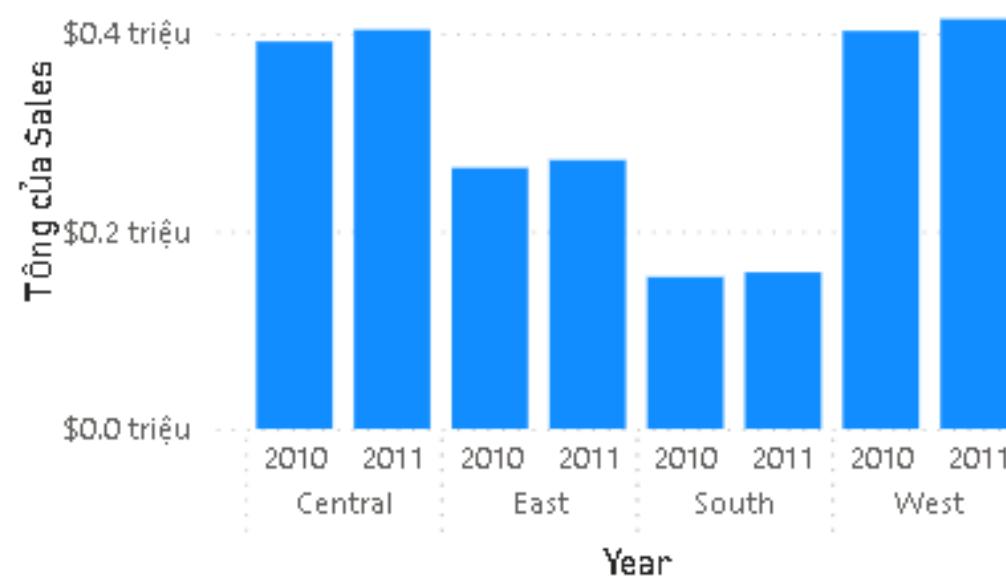
Tổng cùa Sales theo Market



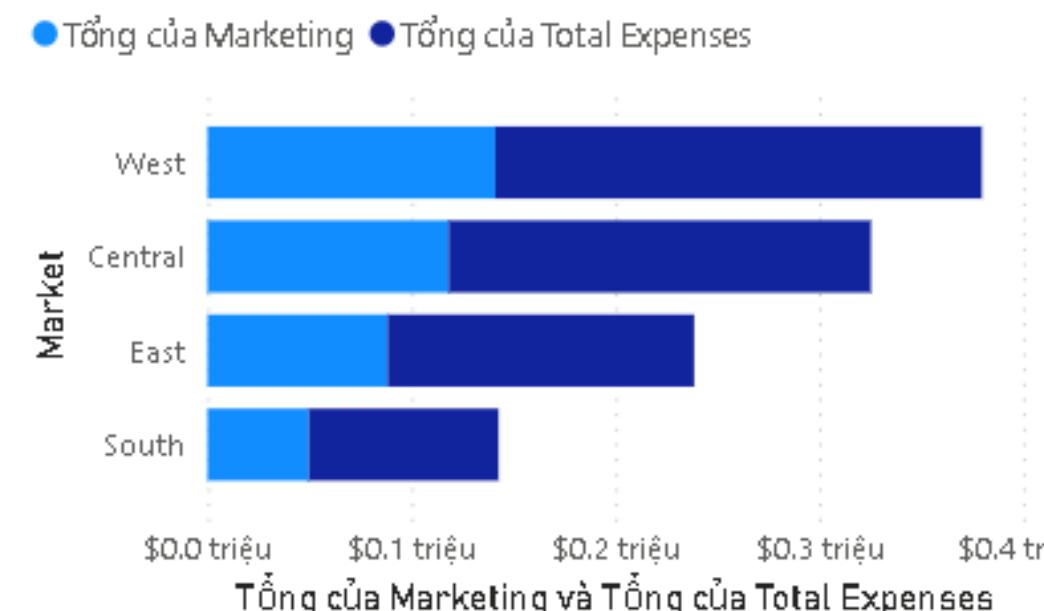
Tổng cùa Sales theo Market Size



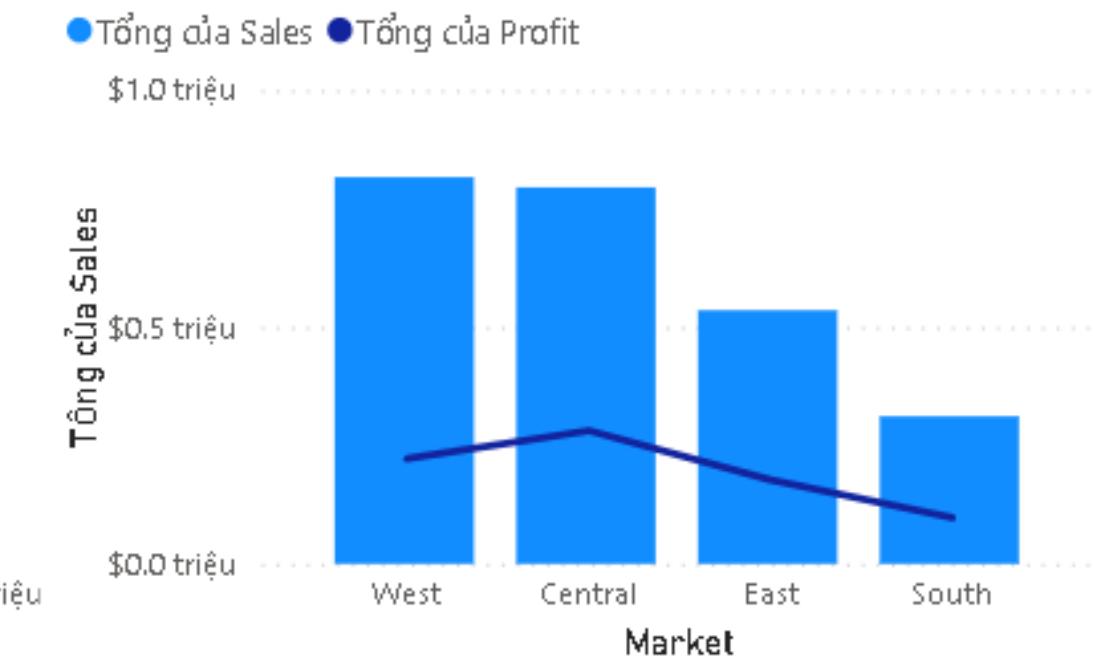
Tổng cùa Sales theo Market và Year



Tổng cùa Marketing và Tổng cùa Total Expenses theo Market



Tổng cùa Sales và Tổng cùa Profit theo Market

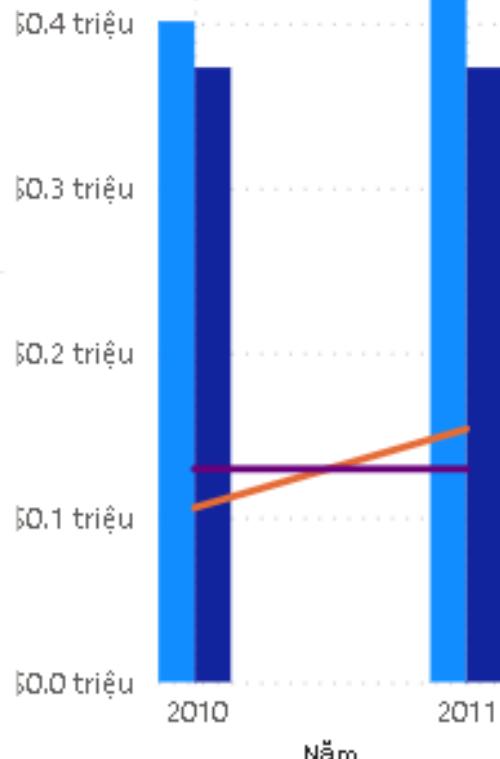


PRODUCT

PHÂN TÍCH DỮ LIỆU BÁN HÀNG CỦA CÁC CỬA HÀNG Ở US

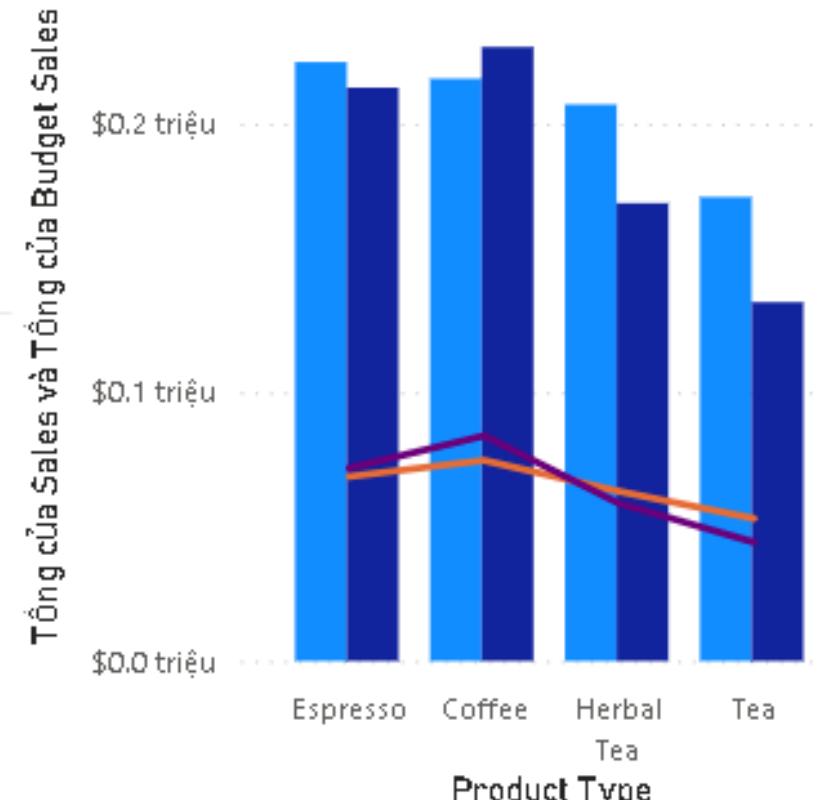
Doanh thu và doanh thu dự kiến theo năm

● Tổng của Sales ● Tổng của Budget Sales



Tổng của Sales, Tổng của Budget Sales, Tổng của Profit và Tổng của Budget Profit theo Product Type

● Tổng của Sales ● Tổng của Budget Sales ● Tổng của Profit ● Tổng của Budget Profit

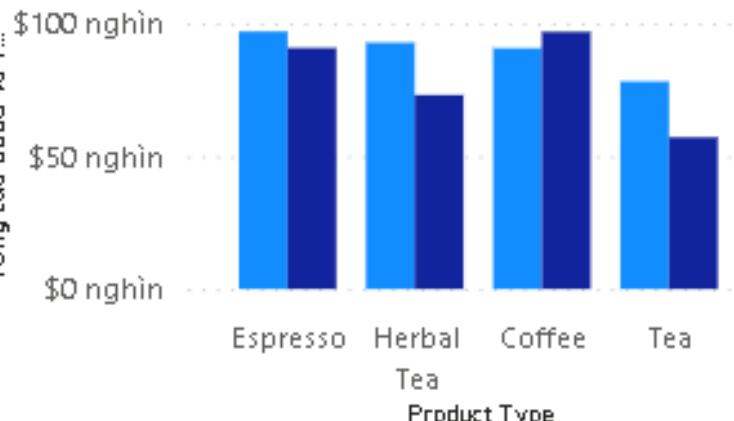


Tổng của Sales theo Product



COGS và COGS dự kiến theo loại sản phẩm

● Tổng của COGS ● Tổng của Budget COGS



156 State

4 Market

4 Product Type

13 Product

Doanh thu và doanh thu dự kiến theo loại sản phẩm

● Tổng của Sales

● Tổng của Budget Sales

● Tổng của Profit

● Tổng của Budget Profit

● Tổng của Total Expenses

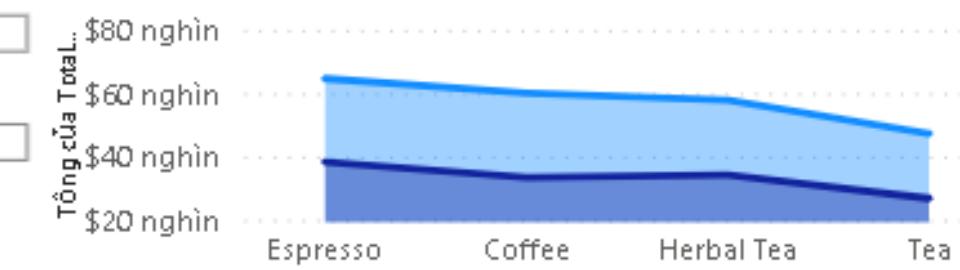
● Tổng của Marketing

● Tổng của Total Expenses

● Tổng của Marketing

Tổng của Total Expenses và Tổng của Marketing theo Product Type

● Tổng của Total Expenses ● Tổng của Marketing



California

\$96,892

Tổng của Sales

Colorado

\$48,179

Tổng của Sales

Connecticut

\$25,429

Tổng của Sales

Sales

\$418,652

Sales

\$373,080

Budget Sales

\$105,826

Profit

\$129,380

Budget Profit

\$179,336

COGS

\$158,940

Budget COGS

\$115,997

Total Expenses

\$66,237

Tổng của Marketing

KẾT LUẬN

NHẬN XÉT:

- Nhìn chung, khu vực thị trường phía Tây và khu vực Trung tâm đạt được doanh thu và lợi nhuận cao hơn các khu vực còn lại, vì được đầu các chi phí Marketing cao nên mang lại hiệu quả kinh doanh tốt
- Các dòng sản phẩm Espresso và Herbal Tea được ưa chuộng khá đồng đều của tất cả các cửa hàng ở Mỹ

GIẢI PHÁP:

- Phân bổ chi phí cho các thị trường phía Đông và phía Bắc nhiều hơn, tập trung Marketing cho các thị trường đó để đẩy mạnh hiệu quả kinh doanh.
- Dòng sản phẩm Tea vẫn còn thị hiếu khá thấp ở thị trường khu vực phía Đông và phía Bắc nên các cửa hàng cũng cần tập trung đẩy mạnh các chi phí đầu tư cũng như Marketing để mang lại hiệu quả kinh doanh tốt.



THANK YOU

