**TRƯỜNG ĐẠI HỌC KINH TẾ - ĐẠI HỌC ĐÀ NẴNG**

**KHOA THỐNG KÊ – TIN HỌC**

**–––––––––––––––––––––––––––––––**



**THỰC HÀNH THIẾT KẾ KHO VÀ PHÂN TÍCH DỮ LIỆU KINH DOANH**

**PHÂN TÍCH DỮ LIỆU BÁN THỨC UỐNG CỦA CÁC CỬA HÀNG TẠI MỸ 2010-2011**

**GVHD :** Đặng Quốc Viện

**Lớp :** 46K21.2

**Nhóm 8 :** Trần Quốc Hải (Nhóm trưởng)

Dương Thị Kim Cúc

Nguyễn Thị Diễm Quỳnh

Phạm Thị Vân

Nguyễn Thị Thúy Hằng

Hoàng Thị Ngọc Oanh

MỤC LỤC

[**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU DOANH THU BÁN HÀNG CỦA CÁC CỬA HÀNG TẠI MỸ NĂM 2010 VÀ 2011** 3](#_Toc153305790)

[**I.** **GIỚI THIỆU DỮ LIỆU:** 3](#_Toc153305791)

[**II.** **ERD:** 4](#_Toc153305792)

[**III.** **ETL (Phân tích, chuyển đổi và đưa dữ liệu vào database):** 5](#_Toc153305793)

[***1.******Mô hình ETL tổng hợp:*** *5*](#_Toc153305794)

[***2.*** ***Thực hiện quy trình ETL tạo các dimension:*** *6*](#_Toc153305795)

[**IV.** **OLAP:** 12](#_Toc153305796)

[**V.** **MDX:** 15](#_Toc153305797)

[***1.******Top 3 sản phẩm có lợi nhuận nhiều nhất:*** *15*](#_Toc153305798)

[***2.******Top 3 khu vực có doanh thu bán hàng cao nhất:*** *15*](#_Toc153305799)

[***3.******Tổng doanh thu của các tiểu bang theo quý:*** *15*](#_Toc153305800)

[**VI.** **PIVOT TABLE:** 16](#_Toc153305801)

[***1.******Lợi nhuận của từng sản phẩm theo năm:*** *16*](#_Toc153305802)

[***2.******Ngân sách giá vốn cho từng loại sản phẩm:*** *17*](#_Toc153305803)

[***3.******Doanh thu theo thị trường:*** *19*](#_Toc153305804)

[**VII.** **DASHBOARD:** 20](#_Toc153305805)

[**VIII.** **REPORTS:** 22](#_Toc153305806)

# **ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU DOANH THU BÁN HÀNG CỦA CÁC CỬA HÀNG TẠI MỸ NĂM 2010 VÀ 2011**

# **GIỚI THIỆU DỮ LIỆU:**

* **Nội dung**: Tập dữ liệu này chứa thông tin về doanh số bán thức uống (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.
* **Mục tiêu**: Xây dựng hình ảnh trực quan để trả lời các câu hỏi như:

**+** Tiểu bang nào/ Cửa hàng nào ở Mỹ mang lại doanh số bán hàng cao nhất?

**+** Tổng chi phí/ giá vốn/ lợi nhuận/… trong khoảng thời gian 2010-2011 dao động trong khoảng bao nhiêu?

**+** Hàng tồn kho qua các tháng trong giai đoạn 2010-2011 chênh lệch như thế nào? Khoảng thời gian nào có lượng hàng tồn kho cao nhất và thấp nhất?

**+** Ngân sách lợi nhuận/ giá vốn/ doanh thu/… là bao nhiêu?

**+** ……

* **Nguồn dữ liệu**: Kaggle
* **Kích thước dữ liệu**: 20 cột & 4248 dòng
* **Mô tả thuộc tính**:

|  |  |  |
| --- | --- | --- |
| **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 bán |
| Profit | money | Lợi nhuận |
| Margin | money | Ký quỹ = Lợi nhuận + Tổng chi phí  **HOẶC** Doanh thu - COGS |
| Sales | money | Doanh thu bán hàng |
| COGS | money | Giá vốn hàng bán |
| Total Expenses | money | Tổng chi phí có được sản phẩm để bán hàng |
| 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 | Giá 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 |
| Product Type | varchar (50) | Danh mục sản phẩm |
| ProductID | int | Mã sản phẩm |
| Product | varchar(50) | Sản phẩm |
| Type | varchar (50) | Loại sản phẩm |

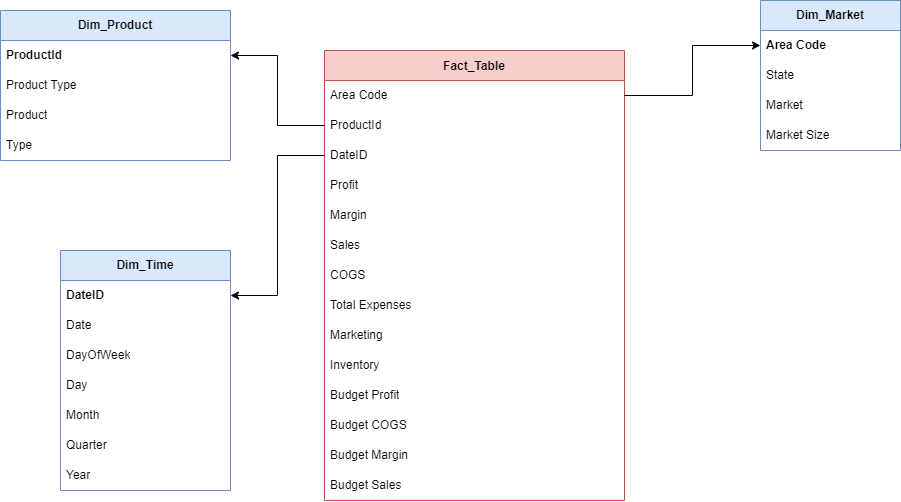
* Có 4 khu vực là : Đông, Tây, Nam , Trung tâm
* Market size: 2 loại là small market: thị trường nhỏ và major market là thị trường lớn

# **ERD:**

Thiết kế mô hình ERD của dữ liệu bán hàng của các cửa hàng tại Mỹ năm 2010 và 2011

Gồm 3 bảng Dimension, 1 bảng Fact

* Dim\_Product (ProductId, Product Type, Product, Type)
* Dim\_Time (DateID, Date, DayOfWeek, Day, Month, Quarter, Year)
* Dim\_Market (Area Code, State, Market, Market Size)
* Fact\_Table (Area Code, ProductId, DateID, Profit, Margin, Sales, COGS, Total Expense, Marketing, Inventory, Budget Profit, Budget COGS, Budget Margin, Budget Sales)



*Hình 2.1. Mô hình ERD*

# **ETL (Phân tích, chuyển đổi và đưa dữ liệu vào database):**

## **Mô hình ETL tổng hợp:**

**A screenshot of a computer

Description automatically generated**

*Hình 3.4. Mô hình ETL tổng hợp*

## **2. Thực hiện quy trình ETL tạo các dimension:**

**2.1. Dim\_Product**

Bước 1: Thực hiện chọn các thuộc tính của Dim\_Product

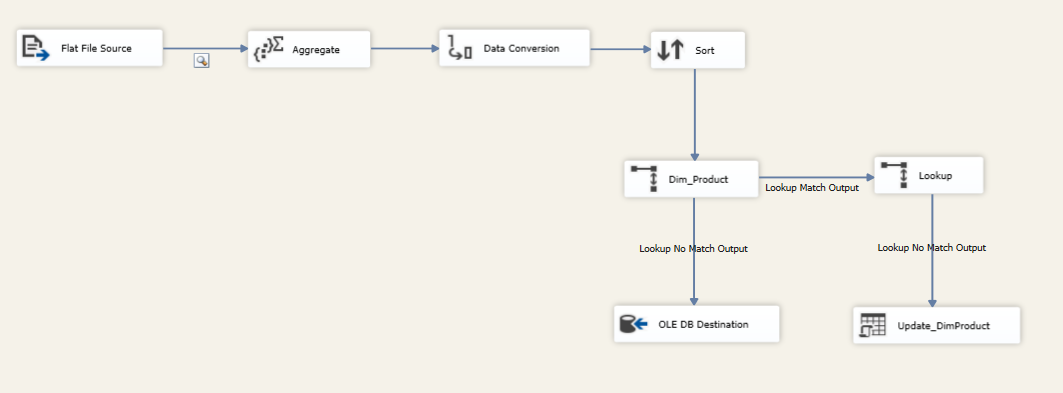
Bước 2: Thực hiện Group by các thuộc tính trên để loại bỏ những giá trị trùng lặp

Bước 3: Thực hiện chuyển đổi kiểu dữ liệu

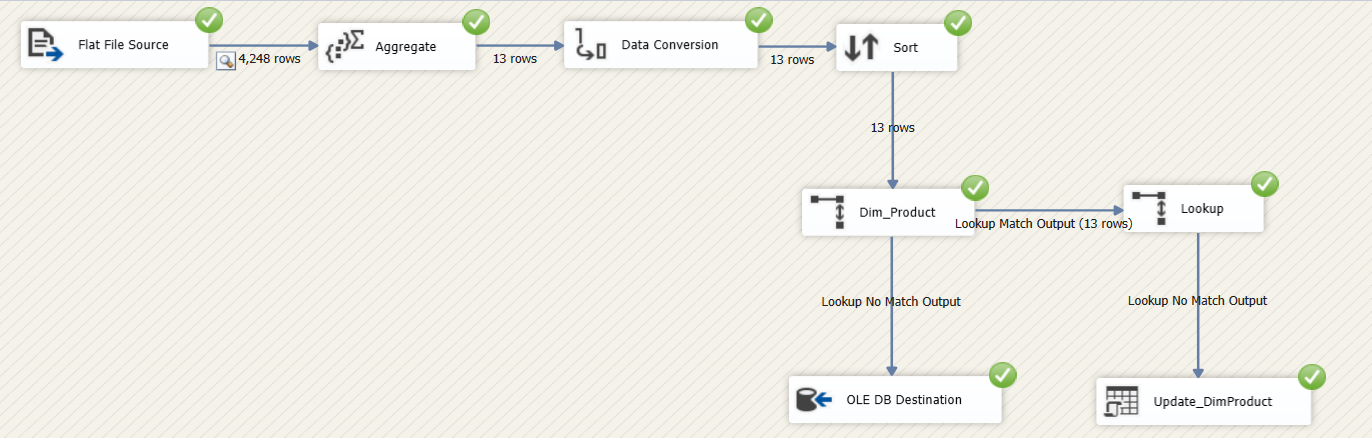
Bước 4: Sắp xếp theo ProductId

Bước 5: Kiểm tra dữ liệu có tương thích với ProductId trong cơ sở dữ liệu đích không. Nếu không thì thực hiện đổ dữ liệu vào bảng Dim\_Product.

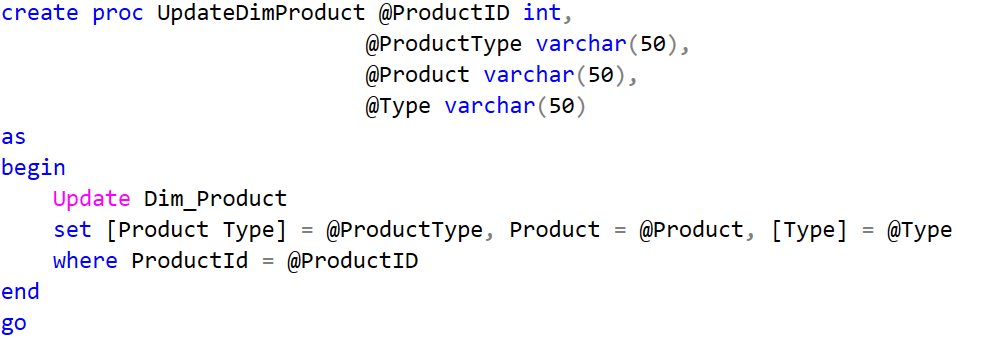
Bước 6: Nếu có thì thực hiện kiểm tra tương thích tất cả các cột trong bảng. Nếu không tương thích hết thì thực hiện Update Dim\_Product



Hình 3.5. *Mô hình ETL của Dim\_Product*

****

*Hình 3.6. Mô hình ETL của Dim\_Product sau khi chạy thành công*

**

*Hình 3.7. Thủ tục Update Dim\_Product*

**2.2. Dim\_Time**

Bước 1: Thực hiện chọn các thuộc tính của Dim\_Product

Bước 2: Thực hiện Group by các thuộc tính trên để loại bỏ những giá trị trùng lặp

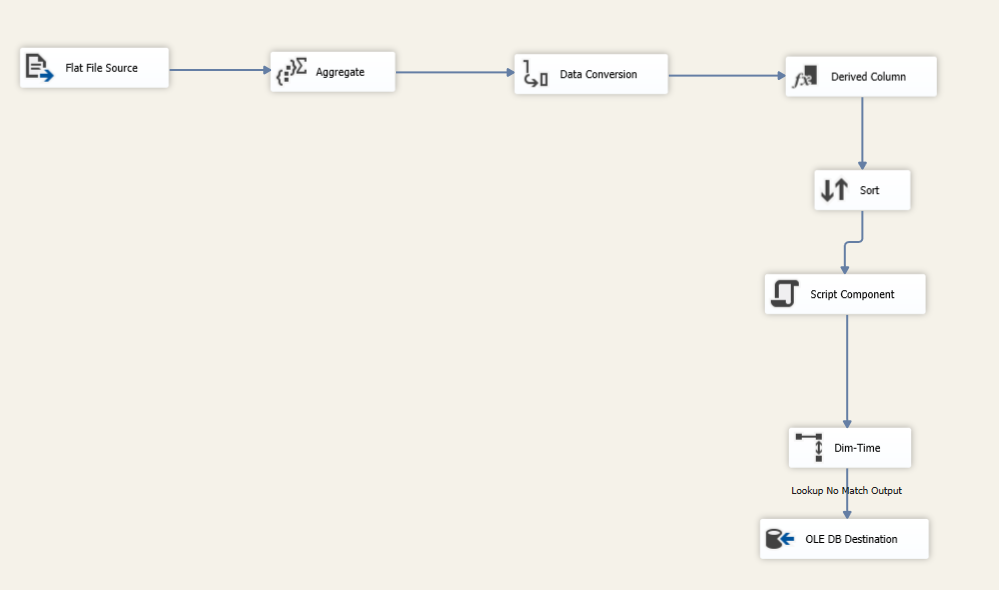
Bước 3: Thực hiện chuyển đổi kiểu dữ liệu

Bước 4: Tạo các cột Day, Month, Year, Quarter, Day of Week

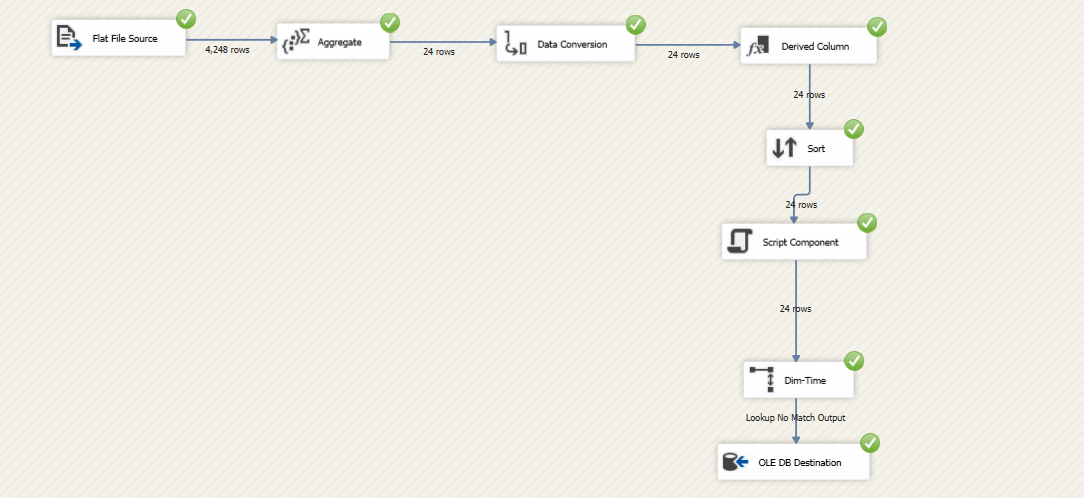
Bước 5: Sắp xếp theo Date

Bước 6: Tạo khóa chính DateID

Bước 7: Kiểm tra dữ liệu có tương thích với ProductId trong cơ sở dữ liệu đích không. Nếu không thì thực hiện đổ dữ liệu vào bảng Dim\_Product.

****

*Hình 3.8.**Mô hình ETL của Dim\_Time*

****

*Hình 3.9. Mô hình ETL của Dim\_Time sau khi chạy thành công*

**2.3. Dim\_Market**

Bước 1: Thực hiện chọn các thuộc tính của Dim\_Product

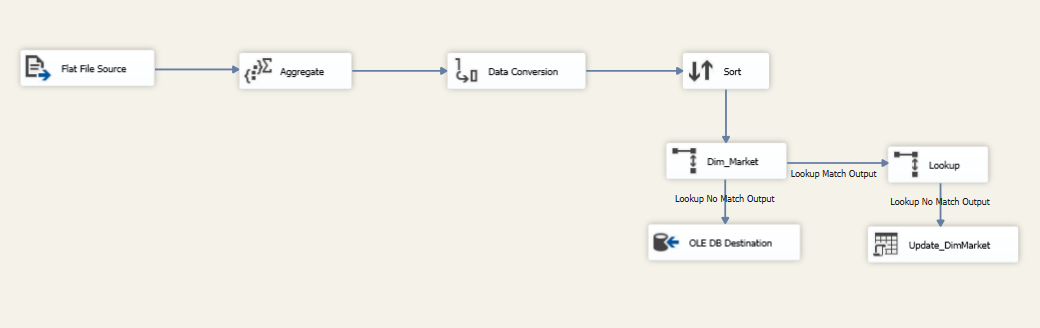
Bước 2: Thực hiện Group by các thuộc tính trên để loại bỏ những giá trị trùng lặp

Bước 3: Thực hiện chuyển đổi kiểu dữ liệu

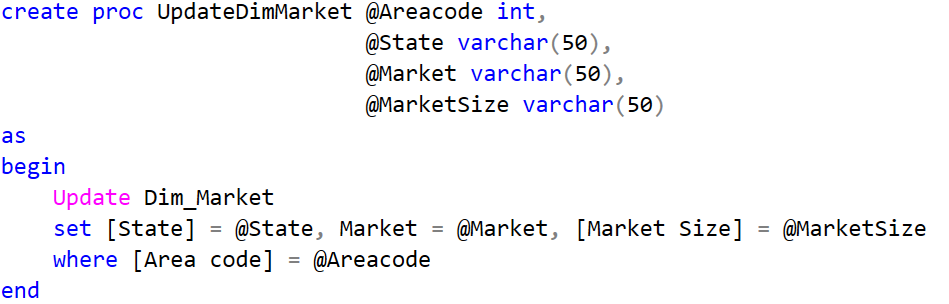
Bước 4: Sắp xếp theo Area Code

Bước 5: Kiểm tra dữ liệu có tương thích với ProductId trong cơ sở dữ liệu đích không. Nếu không thì thực hiện đổ dữ liệu vào bảng Dim\_Market.

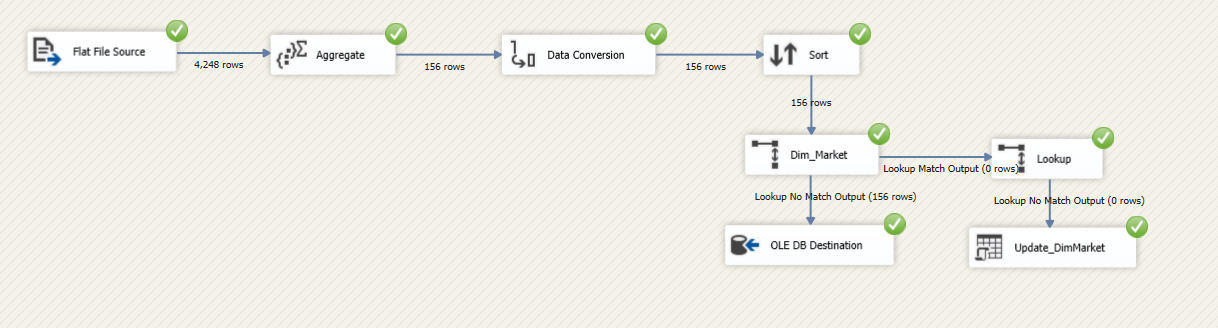
Bước 6: Nếu có thì thực hiện kiểm tra tương thích tất cả các cột trong bảng. Nếu không tương thích hết thì thực hiện Update Dim\_Market.

****

*Hình 3.10. Mô hình ETL của Dim\_Market*

**

*Hình 3.11. Thủ tục Update Dim\_Market*

****

*Hình 3.12. Mô hình ETL của Dim\_Market sau khi chạy thành công*

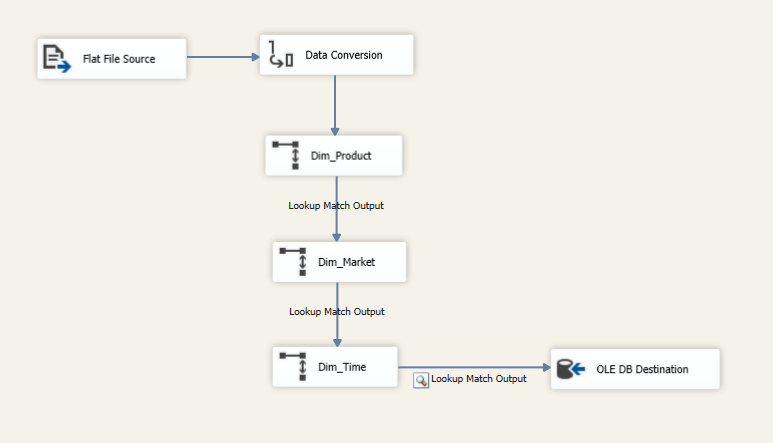
**2.4. Thực hiện quy trình ETL tạo bảng Fact**

Bước 1: Chọn những thuộc tính của bảng Fact

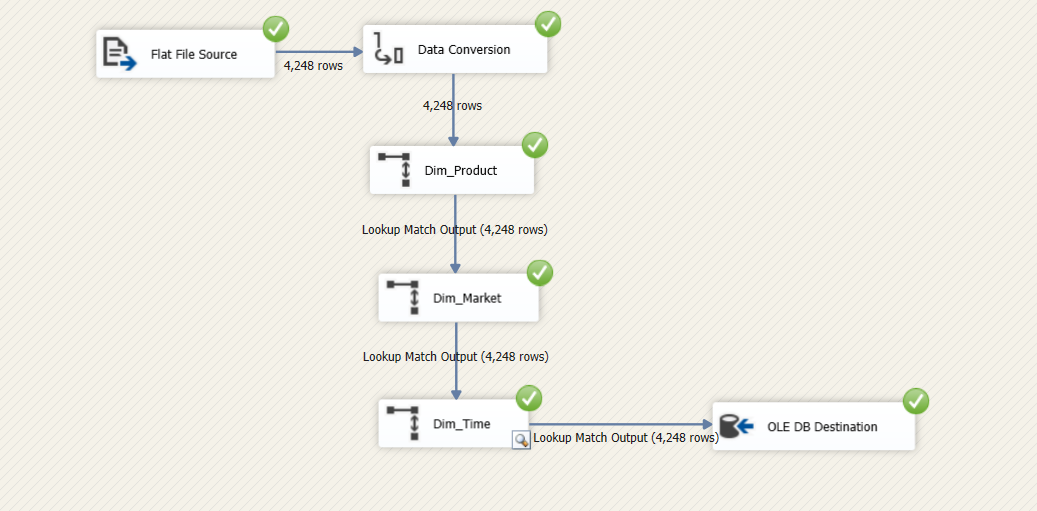
Bước 2: Chuyển dữ liệu về đúng với kiểu dữ liệu của các bảng trong cơ sở dữ liệu đích

Bước 3: Thực hiện kiểm tra tương thích của của bảng với các khóa chính của các bảng đó

Bước 4: Nếu tương thích thìthực hiện đổ dữ liệu vào bảng Fact.

****

*Hình 3.13. Mô hình ETL của Fact\_Table*

****

*Hình 3.14. Mô hình ETL của Fact\_Table sau khi chạy thành công*

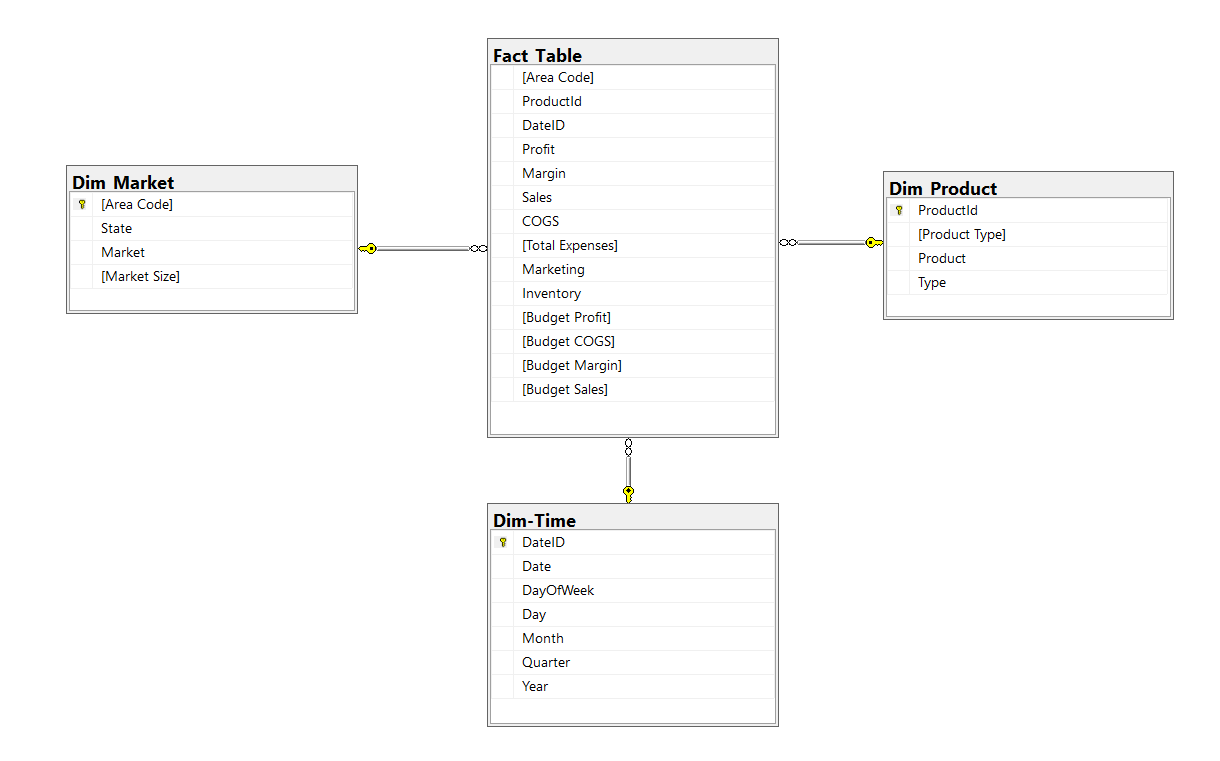
**Tạo các khóa ngoại cho Fact\_Table**

**A screenshot of a computer

Description automatically generated**

*Hình 3.16. Mô hình ETL tổng hợp sau khi chạy thành công*

# **OLAP:**

****

*Hình 4.1. Mô hình Star schema*

Sau khi có được database của mô hình star schema etl thì ta sẽ tiến hành làm OLAP.

Đầu tiên sẽ new data sources ở data source này thì chúng ta sẽ chọn nguồn là database của mô hình star schema trong sql và bước tiếp theo đó sẽ chọn cách đăng nhập tài khoản đăng nhập analyst. Sau đó data sources views thì ta sẽ chọn những bảng dimension và fact để đưa vào. Ở cube thì ta sẽ chọn measure là bảng Fact, sau đó chúng ta sẽ có được một mô hình như hình và chúng ta sẽ deploy nó lên. Sau khi deploy thành công chúng ta sẽ dùng những dữ liệu này để phân tích các độ đo.

Dựa vào OLAP thì chúng ta còn có thể đưa ra những bảng tổng hợp (pivottable) phân tích và các đồ thị để thấy rõ được xu hướng của các cửa hàng theo năm, tháng, quý,...

* **Quy trình:**

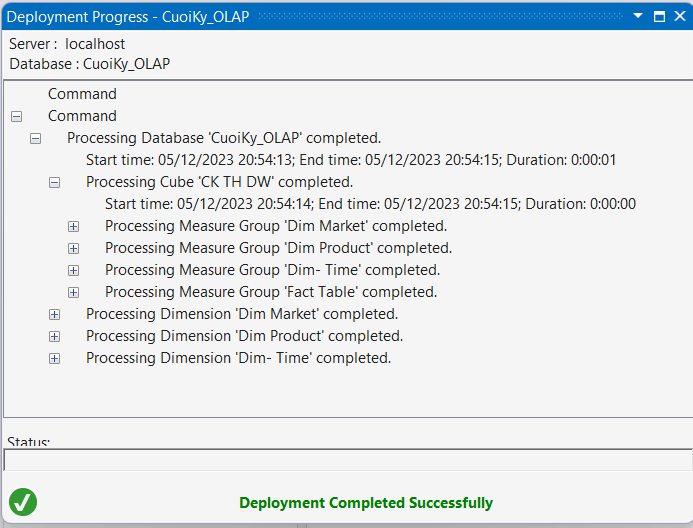
Bước 1: Connect Data Source

Bước 2: Tạo Data Source View

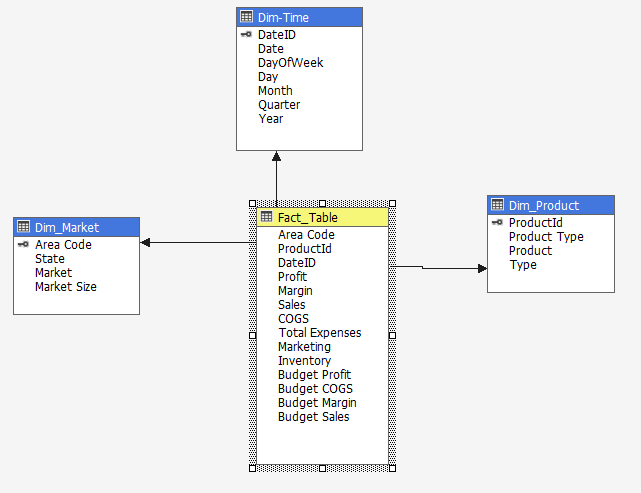
Bước 3: Tạo New Cube

Bước 4: Deploy

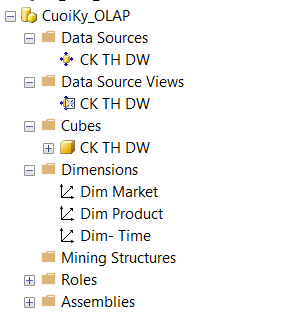
Bước 5: Kéo các thuộc tính cần để phân tích dữ liệu

****

*Hình 4.2. Deploy thành công*

****

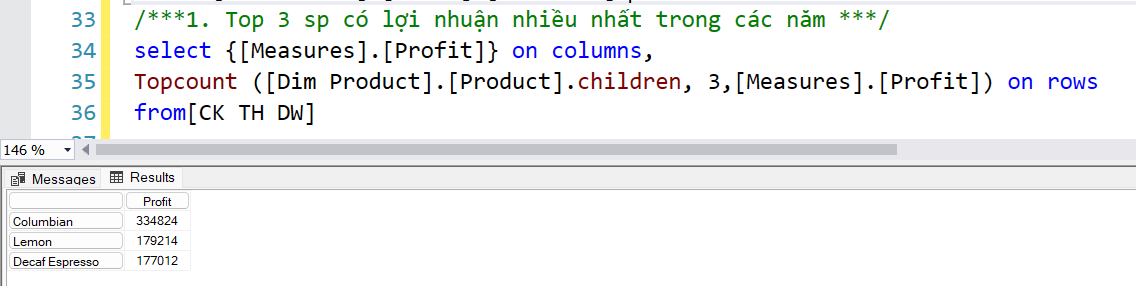
*Hình 4.3. Data Source View*

****

*Hình 4.4. Các thư mục ở cơ sở dữ liệu đích*

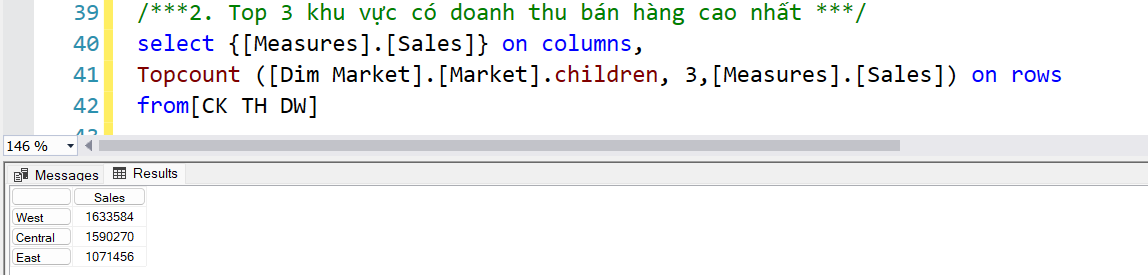
# **MDX:**

## **Top 3 sản phẩm có lợi nhuận nhiều nhất:**

****

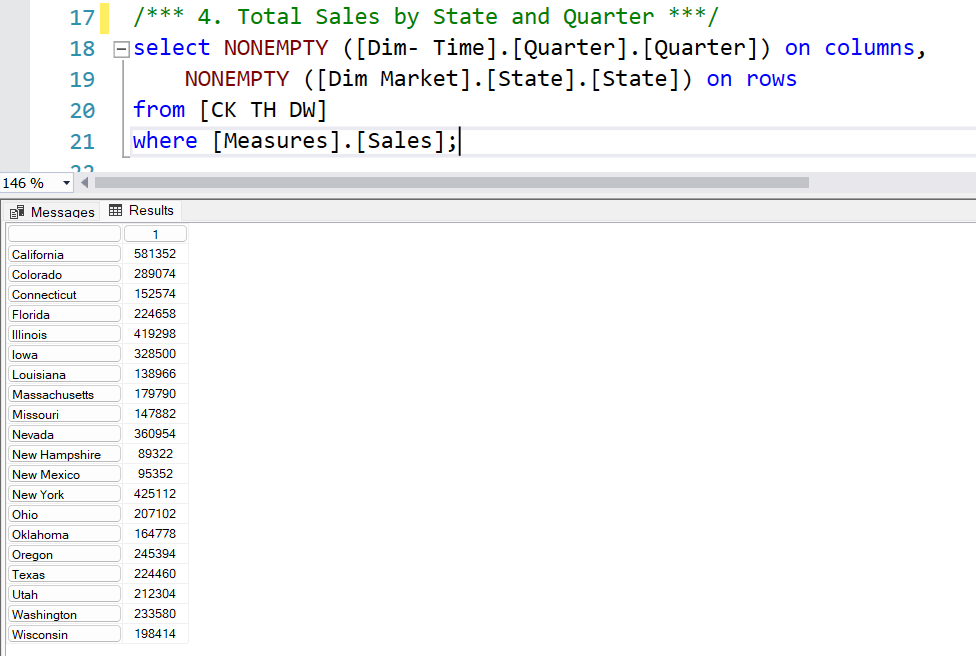
*Hình 5.1. Top 3 sản phẩm có lợi nhuận nhiều nhất*

## **Top 3 khu vực có doanh thu bán hàng cao nhất:**

****

*Hình 5.2. Top 3 khu vực có doanh thu bán hàng cao nhất*

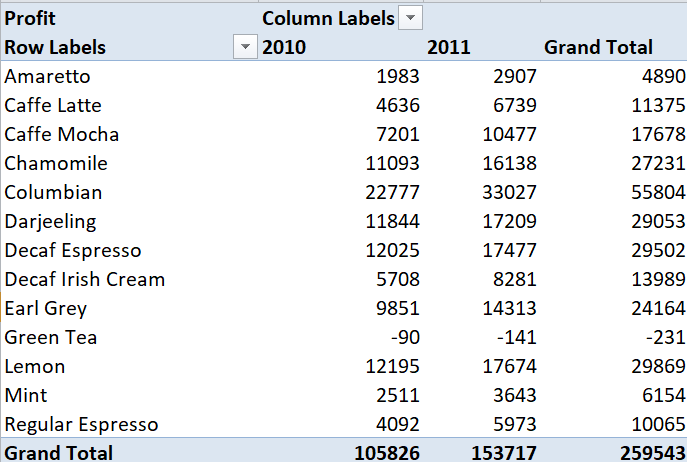
## **Tổng doanh thu của các tiểu bang theo quý:**

****

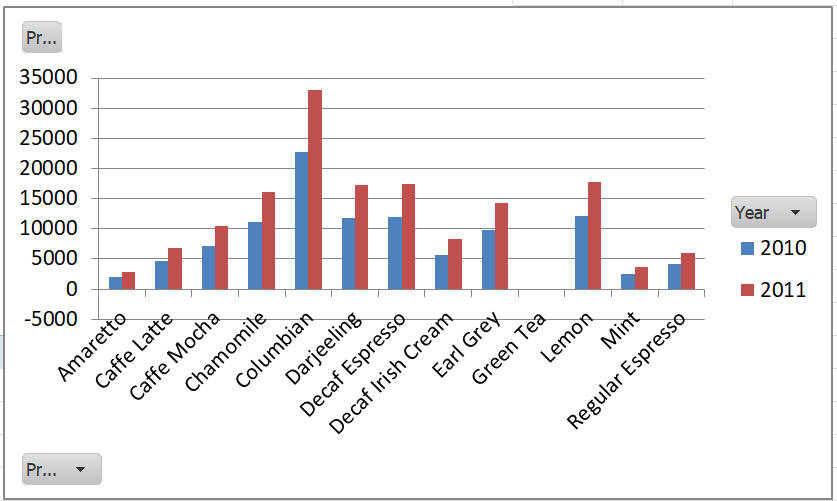
*Hình 5.3. Tổng doanh thu của các tiểu bang theo quý*

# **PIVOT TABLE:**

## **Lợi nhuận của từng sản phẩm theo năm:**

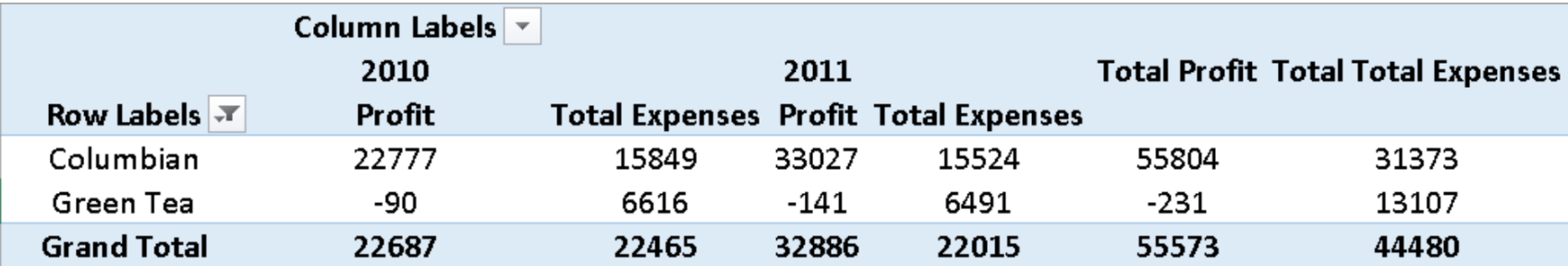
****

*Hình 6.1. Lợi nhuận của từng sản phẩm theo năm*



*Hình 6.2. Biểu đồ Lợi nhuận của từng sản phẩm theo năm*

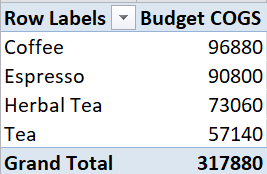
Hình trên là tổng lợi nhuận của các sản phẩm theo năm. Nhìn vào bảng tổng hợp ta có thể thấy được rõ ràng là tổng lợi nhuận là gần 259.543 USD, trong đó lợi nhuận của Columbian là cao nhất (55.804 USD) trong khi lợi nhuận của Green Tea ở mức âm (-231 USD) trong cả 2 năm 2010 và 2011, các loại sản phẩm có sự thay đổi lợi nhuận lớn nhất là Lemon (5.479 USD) và Mint (4.032 USD). Do đó, cần phải có chiến lược bán hàng và quảng cáo phù hợp để thúc đẩy lợi nhuận của dòng sản phẩm Green Tea cũng như các dòng sản phẩm khác có lợi nhuận thấp.



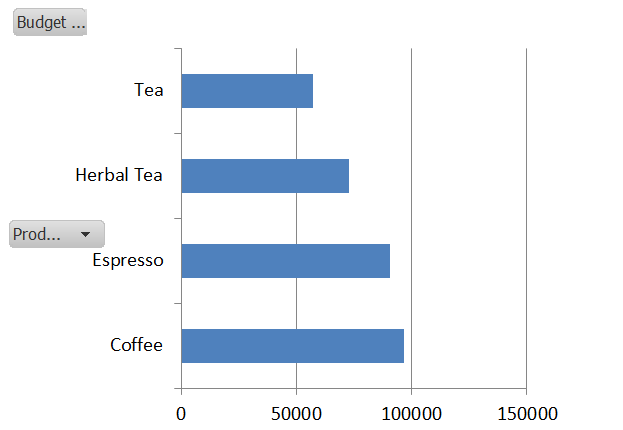
*Hình 6.3. Doanh số bán hàng của hai sản phẩm Columbian và Green Tea*

Qua bảng trên chúng ta có thể thấy chi phí của Green Tea cao hơn lợi nhuận vì vậy kết luận Green Tea đang là một sản phẩm khó khăn trong việc tiêu thụ.

## **Ngân sách giá vốn cho từng loại sản phẩm:**



*Hình 6.4. Ngân sách giá vốn cho từng loại sản phẩm*

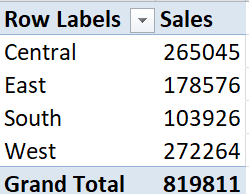


*Hình 6.5. Biểu đồ Ngân sách giá vốn cho từng loại sản phẩm*

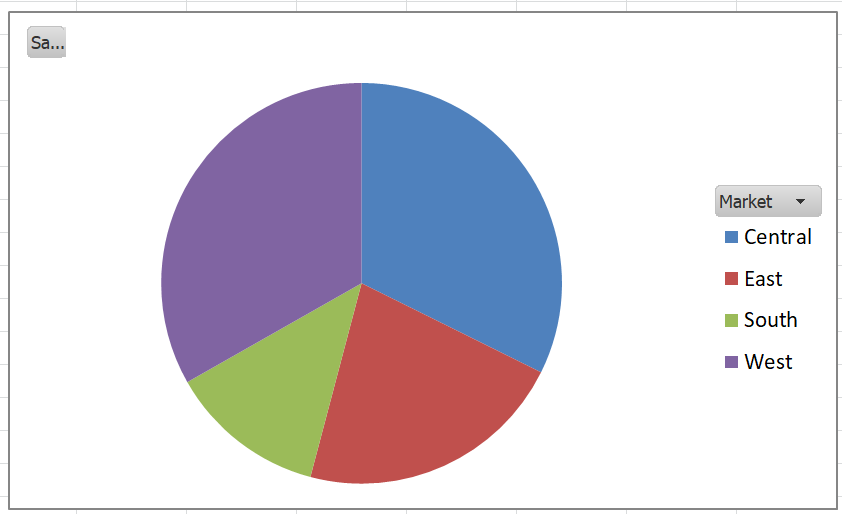
Nhìn vào biểu đồ ta có thể thấy tổng ngân sách giá vốn là 317.880 USD. Các loại sản phẩm có ngân sách giá vốn cao nhất là Coffee (96.880 USD) và Espresso (90.800 USD). Hai loại sản phẩm có ngân sách giá vốn thấp hơn là Herbal Tea (73.060 USD) và Tea (57.140 USD).

Từ biểu đồ lợi nhuận của từng sản phẩm theo năm, đối với sản phẩm mang lại lợi nhuận cao như cà phê thì các cửa hàng sẽ đầu tư ngân sách giá vốn cho sản phẩm đó cao hơn. Ngược lại với những loại sản phẩm ít bán chạy, lợi nhuận thấp thì ngân sách giá vốn thấp.

## **Doanh thu theo thị trường:**



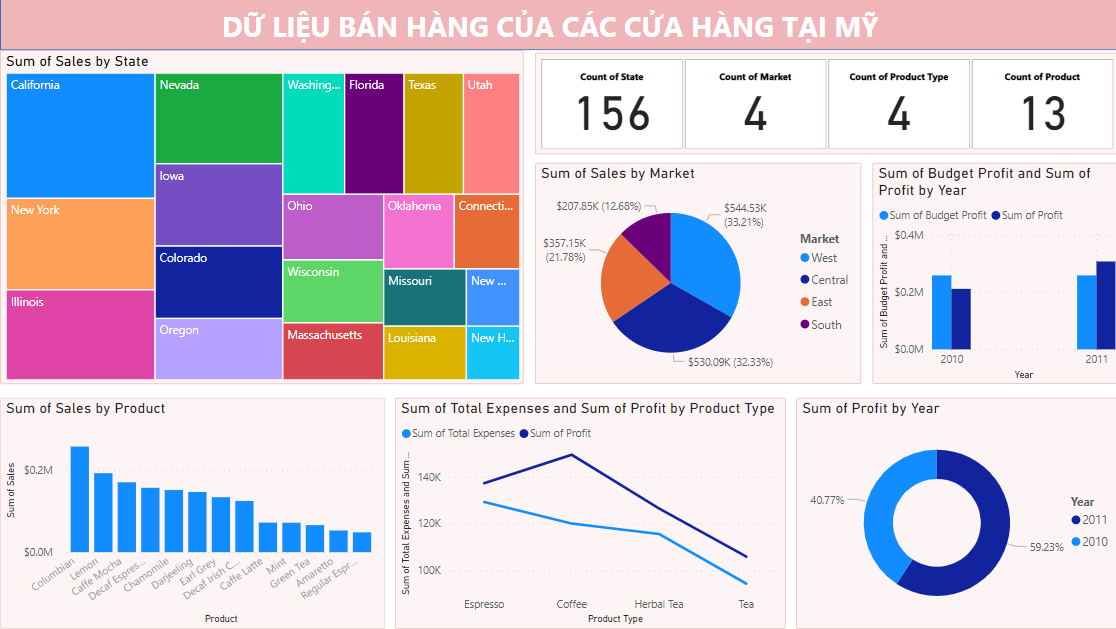
*Hình 6.6. Doanh thu theo thị trường*



*Hình 6.7. Biểu đồ Doanh thu theo thị trường*

Ở đây chúng ta có thể thấy doanh thu thu được trên thị trường Tây Mỹ đạt mức cao nhất với hơn 272.264 USD, theo sát đó là doanh thu ở vùng Trung tâm với hơn 265.045 USD. Điều này có thể được lý giải là do 2 vùng này là nơi tập trung các bang lớn nên lượng tiêu thụ sản phẩm đồ uống cao, nhu cầu cao hơn, từ đó mang lại doanh thu nhiều hơn so với khu vực Nam Mỹ và Đông Mỹ.

# **DASHBOARD:**



*Hình 7.1. Tổng quan về dữ liệu bán hàng*

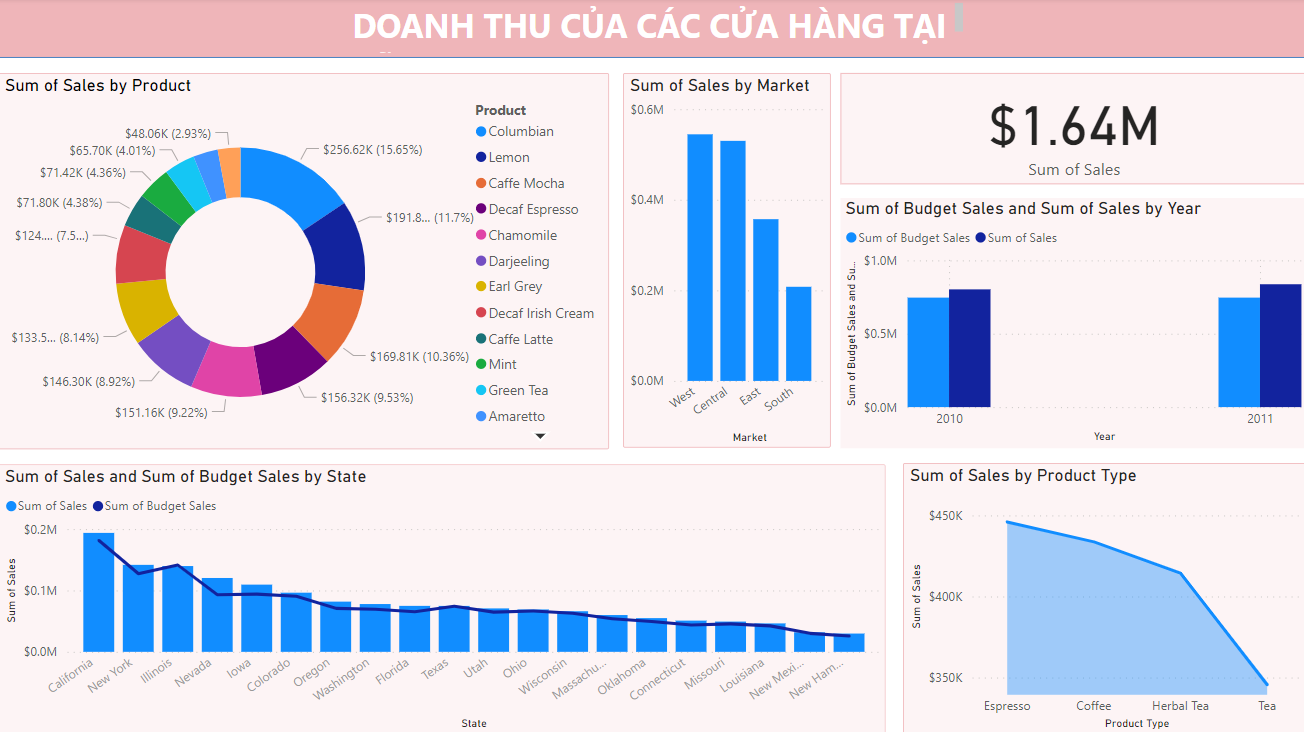
Phân tích dữ liệu bán hàng tại 156 cửa hàng, 4 khu vực (Đông, Tây, Nam, và Trung tâm), 4 loại sản phẩm (Coffee, Espresso, Herbal Tea và Tea) và 13 sản phẩm khác nhau tại 20 tiêu bang ở US.

Bang có doanh thu lớn nhất là Califonia tiếp theo là New York, Illinois, dòng sản phẩm mang loại doanh thu cao nhất là Columbian.

Khu vực có doanh số cao nhất là Tây Mỹ tiếp theo là trung tâm, Đông Mỹ và thấp nhất là Nam Mỹ.

Nhìn chung, lợi nhuận thực tế thu được luôn lớn hơn tổng chi phí phải bỏ ra để có được sản phẩm bán hàng, cho thấy sự ổn định của thị trường. Nhưng vẫn còn khá chênh lệch giữa các loại hàng, doanh nghiệp nên cân bằng hơn giữa các loại.

Lợi nhuận dự kiến năm 2010 cao hơn lợi nhuận thực tế, nhưng năm 2011 thì ngược và lợi nhuận thực tế năm 2011 đã tăng gần 20% cho thấy sự tiến bộ vượt bậc.

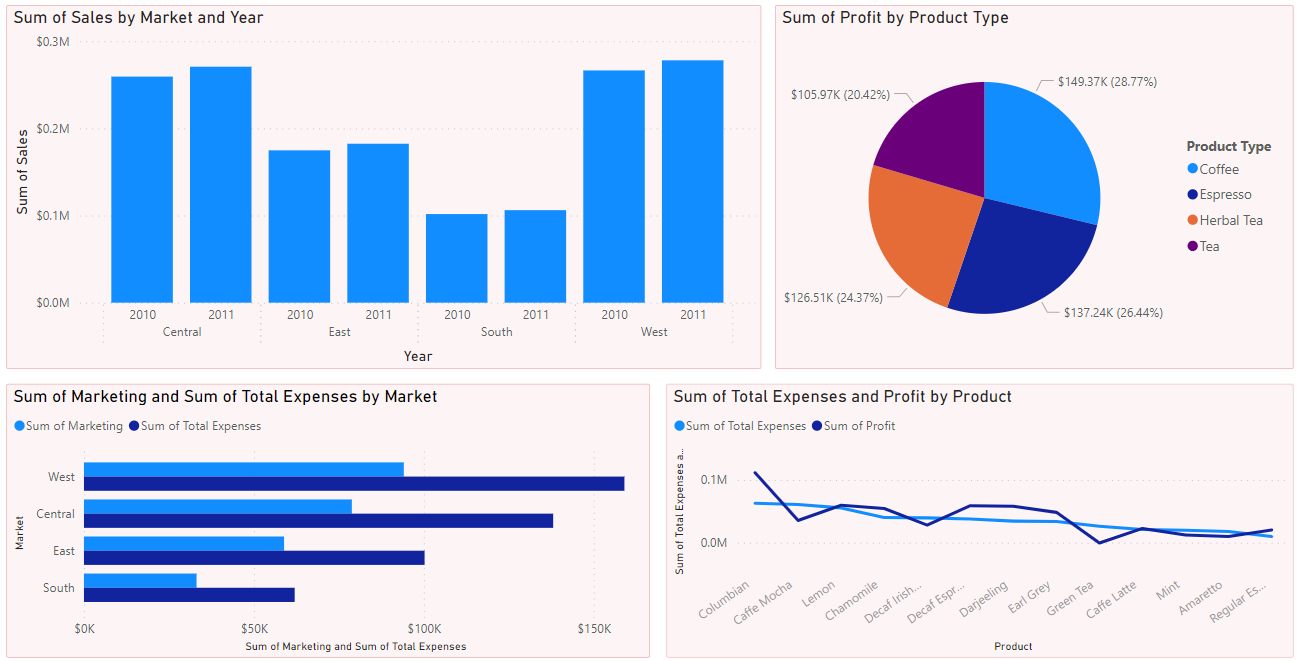


*Hình 7.2. Tổng quan về doanh thu bán hàng*

Nhìn chung cả 2 năm doanh thu đều vượt mức dự kiến ban đầu, năm 2011 có tăng so với 2010 nhưng không đáng kể.

Top 5 sản phẩm đem lại doanh thu lớn nhất: Columbian, Lemon, Cafe Mocha, Decaf Espresso, Chamomile.

Dòng sản phẩm Espresso và Herbal Tea được ưa chuộng khá đồng đều của tất cả các cửa hàng và mang lại doanh thu tốt.

****

*Hình 7.3. Tổng quan về hiệu suất kinh doanh*

Nhìn chung doanh thu các khu vực đều tăng so với năm ngoái.

Dòng sản phẩm có lợi nhuận cao nhất là Coffee.

Chi phí marketing chiếm khá cao, hơn 50% so với tổng chi phí bỏ ra để bán 1 mặt hàng.

Một số sản phẩm như Caffee Mocha hay Green Tea có lợi nhuận thấp hơn chi phí bỏ ra, doanh nghiệp cần tìm hiểu nguyên nhân và đưa ra chiến lược hợp lý cho sản phẩm.

# **REPORTS:**

***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.

Rất tốt khi một doanh nghiệp có khả năng dự kiến và quản lý các con số như chi phí, doanh thu và các chỉ số quan trọng khác. So với 2010 thì năm 2011các dự kiến đã sát hơn với thực tế, phần hơn các chỉ số đều tích cực hơn.

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, cân nhắc việc tăng ngân sách quảng cáo và tiếp thị tại các khu vực này để tăng cường nhận thức thương hiệu và kích thích nhu cầu tiêu dùng.

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. Có thể bao gồm các chương trình khuyến mãi, sự kiện đặc biệt, hoặc đối tác địa phương để tăng cường sự hiện diện của sản phẩm và tạo ra sự hứng thú từ phía khách hàng.