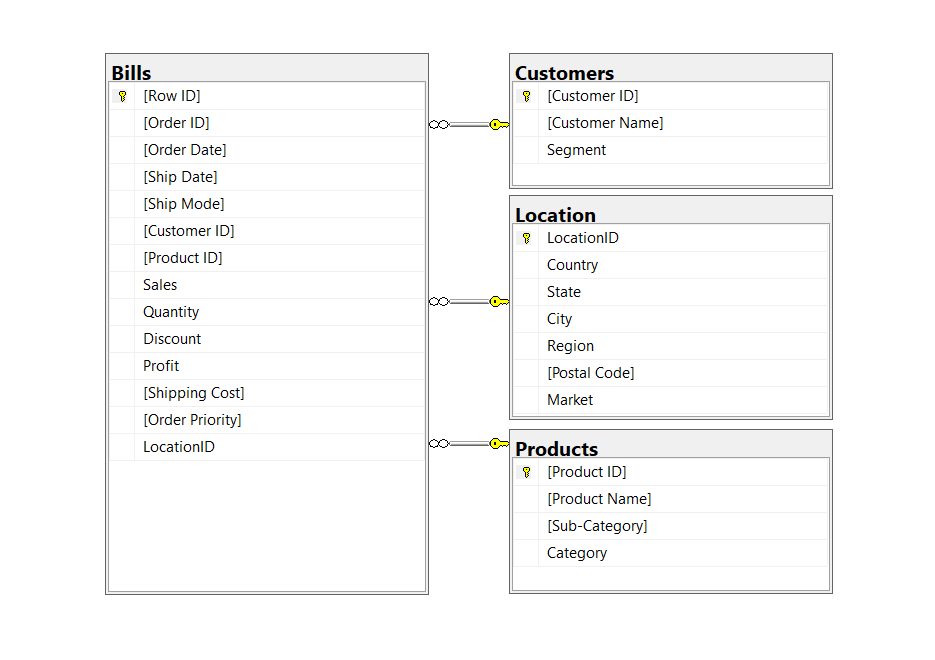
# **I) Xác lập mối quan hệ giữa các bảng (Tạo lược đồ hình sao)**



**Bảng Customers**: là bảng tổng hợp thông tin của khách hàng. Trong đó Customer ID đóng vai trò làm khóa chính.

**Bảng Location**: Là bảng tổng hợp các thông tin về các khu vực bán hàng.

Trong đó LocationID đóng vai trò làm khóa chính. Đặc biệt LocationID được tạo bằng cách cộng các chuỗi “Country + [Postal Code] + City + State’’ nhằm tạo tính duy nhất phù hợp cho việc thiết lập khóa chính.

Để làm được điều này, ta phải điều chỉnh kiểu dữ liệu của Postal Code sang kiểu chuỗi, những giá trị Null thì chuyển sang khoảng trắng ‘ ‘

**Bảng Products**: Là bảng tổng hợp các loại mặt hàng gồm các cột như trên. Trong đó Product ID đóng vai trò làm khóa chính.

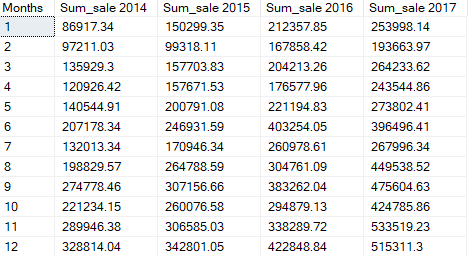
**Bảng Bills**: Là bảng tổng hợp các đơn bán hàng gồm các cột như trên.

Trong đó Row ID đóng vai trò làm khóa chính. Khóa ngoại liên kết tới bảng Customer là Customer ID, bảng Products là Product ID, bảng Location là Location ID.

# **II) Các truy vấn tổng hợp dữ liệu để mô tả hoạt động bán hàng**

1. **Tính doanh thu của cửa hàng theo tháng trong 4 năm:**

**- Bảng doanh thu của cửa hàng theo tháng trong 4 năm**



**Các bước thực hiện:**

* Bước 1: Tạo 1 bảng Revenue (Doanh thu) gồm các cột dữ liệu được lấy từ bảng Bills đã xây dựng. Các cột bao gồm:

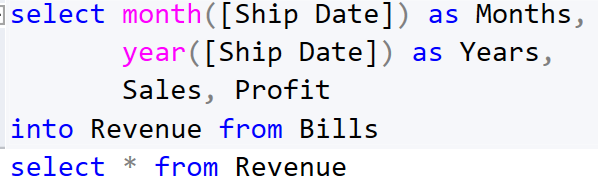
Months: là giá trị tháng được tách ra từ cột [Ship Date]

Years: là giá trị năm được tách ra từ cột [Ship Date]

Sales là doanh thu

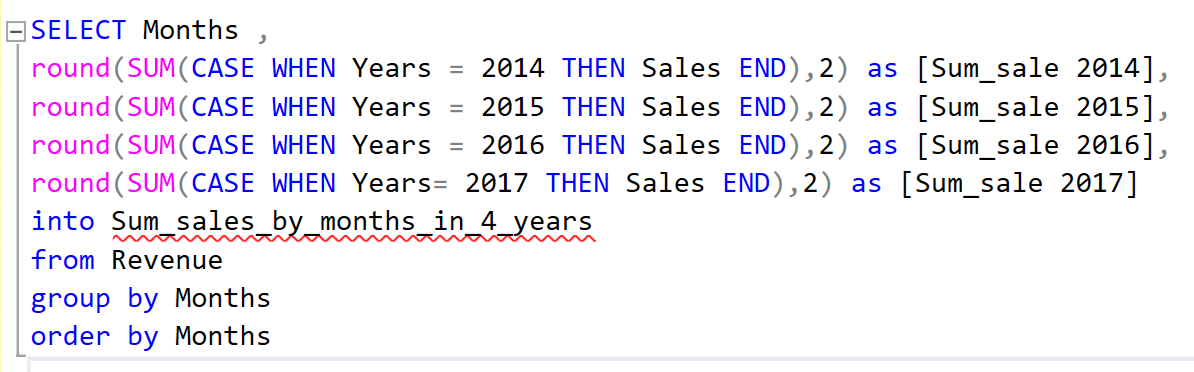
Profit là lợi nhuận.

Code:



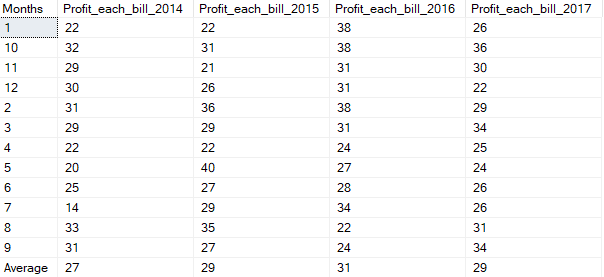
* Bước 2: Tạo bảng doanh thu của cửa hàng theo tháng trong 4 năm, bằng cách tính tổng Doanh thu (cột Sales) theo mỗi năm tương ứng và nhóm theo từng tháng.

Code:



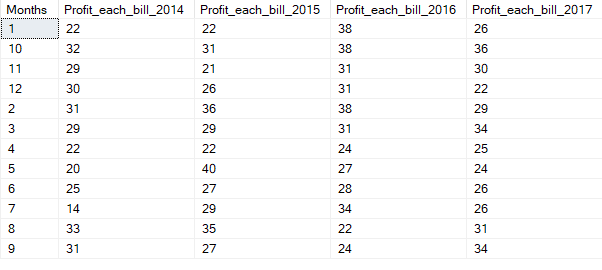
1. **Lợi nhuận trung bình trên một đơn hàng theo tháng và năm**

* Bảng lợi nhuận trung bình trên 1 đơn hàng theo tháng và năm.

****

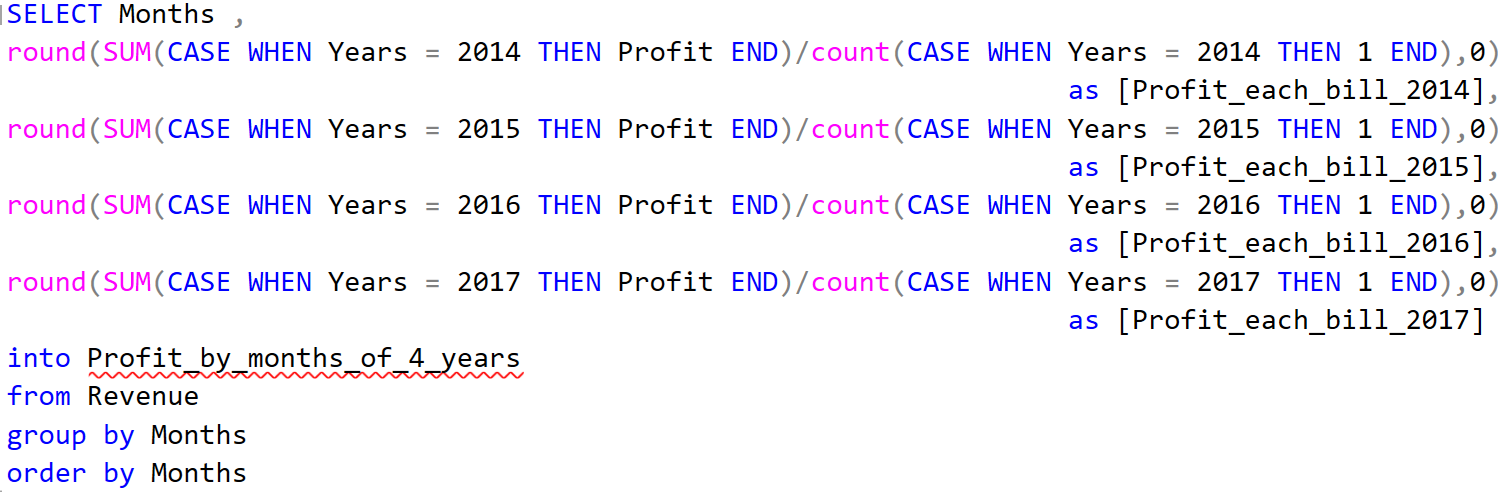
**Các bước thực hiện:**

* **Bước 1:** Dựa vào bảng Revenue đã tạo ở bước 1 phần truy vấn 1, ta tiến hành tạo bảng *Profit\_by\_months\_of\_4\_year* có cấu trúc và kết quả hiển thị như sau:



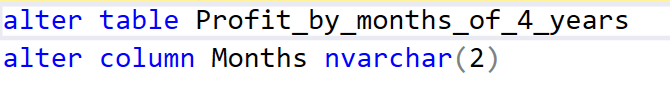
Mỗi cột như Profit\_each\_bill\_2014 được tính bằng cách tính tổng những lợi nhuận sau đó chia cho số đơn bán hàng trong năm tương ứng theo từng tháng.

Code:

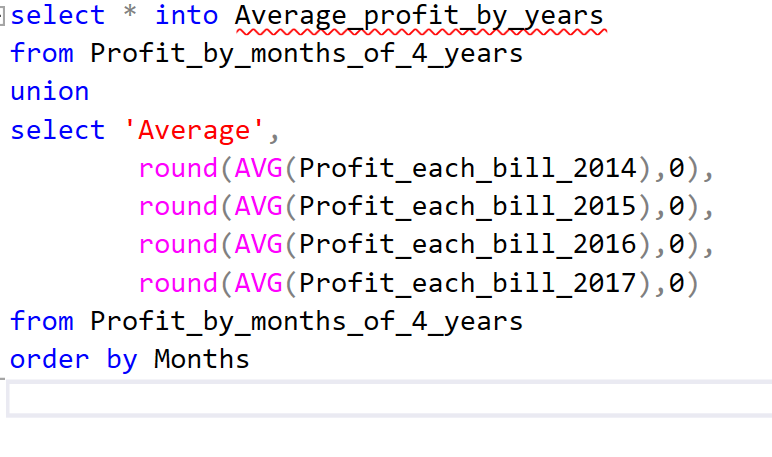
****

* **Bước 2**: Từ bảng đã tạo là *Profit\_by\_months\_of\_4\_years* ta kết hợpthêm cột Average để tính lợi nhuận trung bình trên 1 đơn hàng của mỗi năm, bằng cách tính trung bình các tháng đã tính ở bảng *Profit\_by\_months\_of\_4\_years.*

*Code:*

**

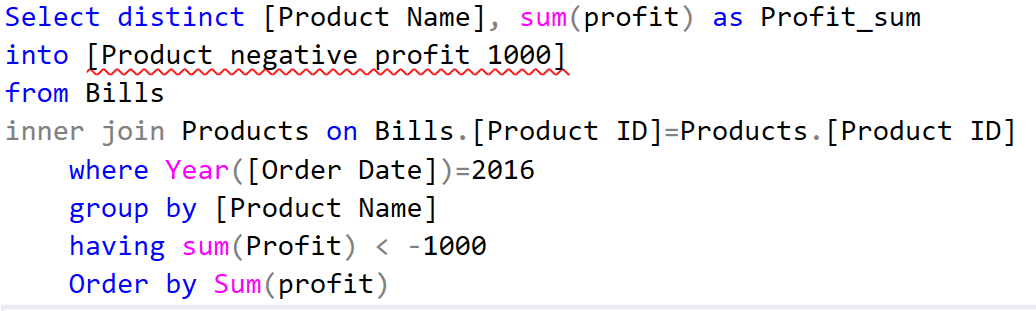
(Vì Months đang có kiểu dữ liệu Int nên chuyển sang kiểu chuỗi để có thể thêm dòng ‘Average’)

**

1. **Sản phẩm có lợi nhuận âm hơn 1000$ năm 2016**

* **Những sản phẩm có lợi nhuận âm: [Product negative profit 1000]**



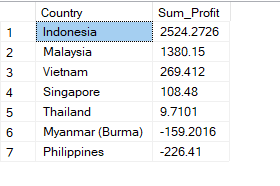


**Cách thực hiện:**

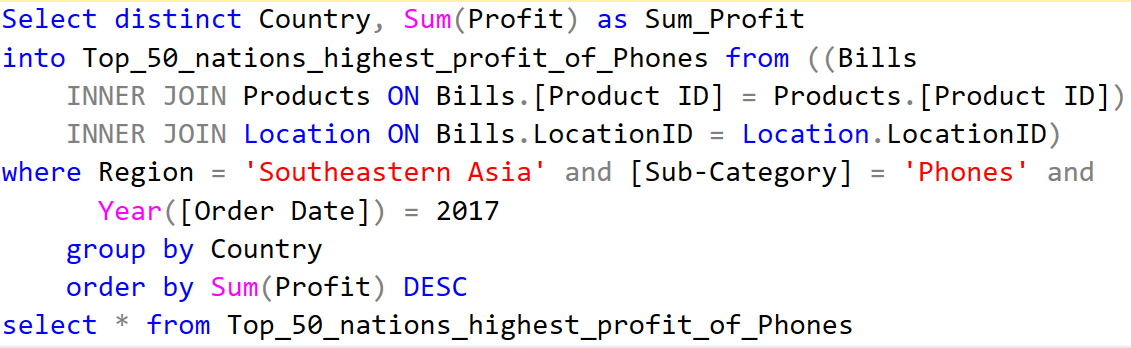
Sử dụng bảng Bills và nối khóa ngoại [Product ID] để lấy giá trị cột tên sản phẩm ([Products Name]) kèm thêm tùy chọn distinct để lấy giá trị Products Name là duy nhất. Cột thứ 2 ta lấy tổng của lợi nhuận có giá trị âm hơn 1000 với điều kiện theo năm tương ứng cụ thể ở truy vấn này chọn năm 2016. Nhóm dữ liệu theo tên sản phẩm ([Product Name]) ta được kết quả như trên.

1. **Top 50 quốc gia có lợi nhuận từ điện thoại cao nhất Đông Nam Á năm 2017**

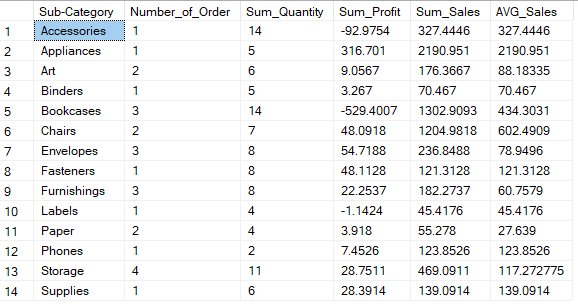
**Top\_50\_nations\_highest\_profit\_in\_Phones**

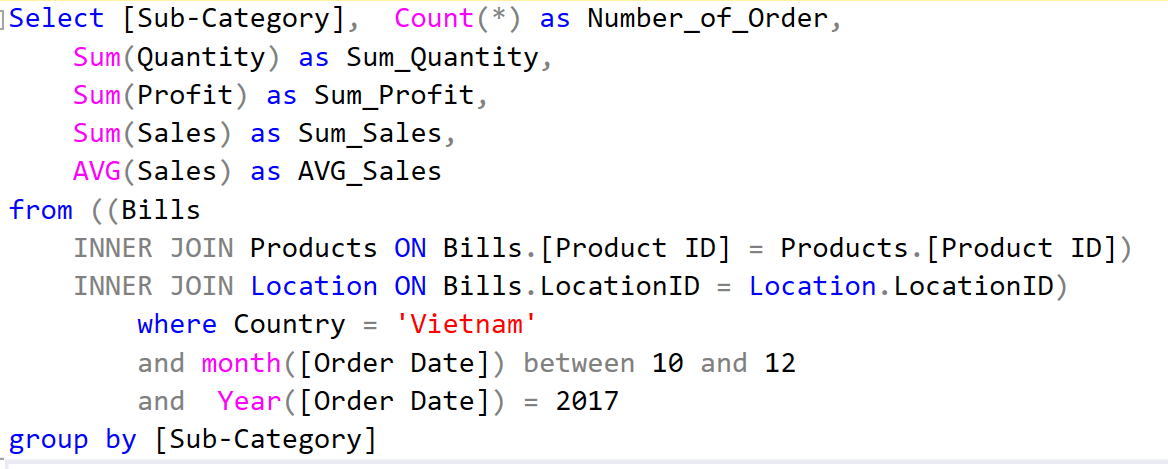
****

Code:

****

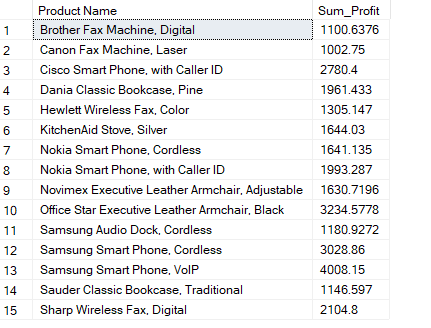
1. **Số lượng đơn đặt hàng, tổng số lượng, lợi nhuận, doanh số bán hàng, doanh thu trung bình theo danh mục phụ trong Quý 4 năm 2017 tại Việt Nam.**

**

**

* Đếm số đơn đặt hàng, tính tổng sản phẩm đơn hàng, lợi nhuận, doanh thu dựa vào hàm Sum. Tính trung bình doanh thu dựa vào hàm tính trung bình AVG dựa trên bảng Bills, bảng Location và bảng Products dựa trên tiêu thức Sub-Category
* Nối các bảng Location, Products đến bảng Bills dựa trên các khóa ngoại [Location ID], [Product ID].
* Đặt điều kiện từng dòng Country là ‘VietNam’, thời gian là quý 4, năm 2017 để khảo sát theo giả thuyết

1. **Sản phẩm có lợi nhuận > 1000 vào năm 2014 tại thị trường Châu Á Thái Bình Dương**

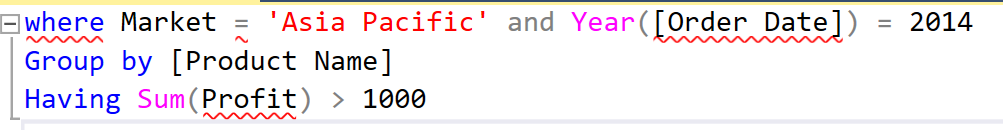
****

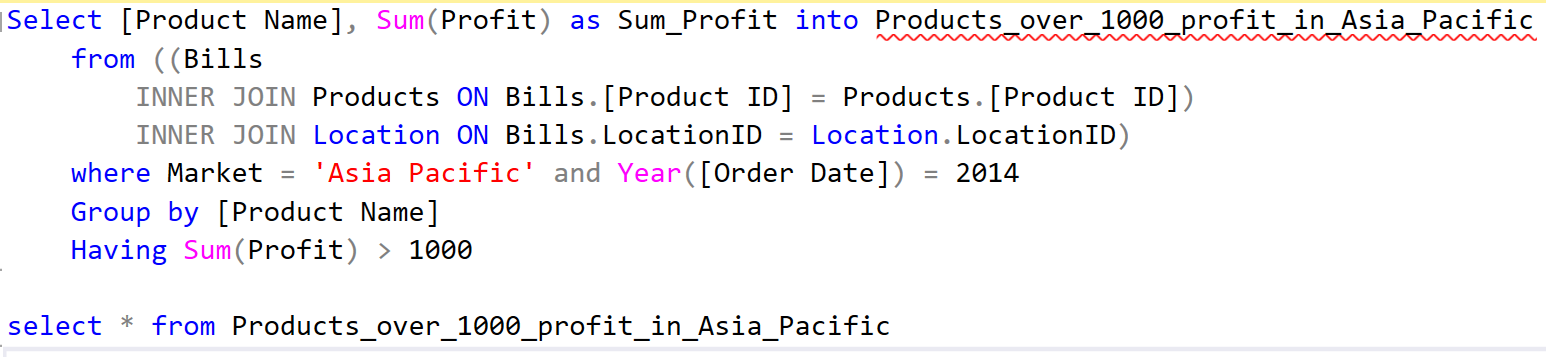
**Các bước thực hiện**

- Đầu tiên trích tên của các mặt hàng (sản phẩm) [Product Name], sau đó tính tổng các khoản lợi nhuận (Profit) theo từng mặt hàng tương ứng đưa vào cột Sum\_Profit (tổng lợi nhuận).

- Ở đây ta cần truy xuất thông tin ở 2 bảng Products và Location nên ta phải sử dụng inner join để liên kết thông tin 2 bảng thông qua cột LocationID.

**Điều kiện:**

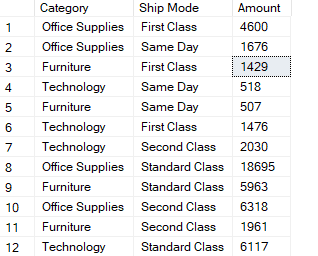
****

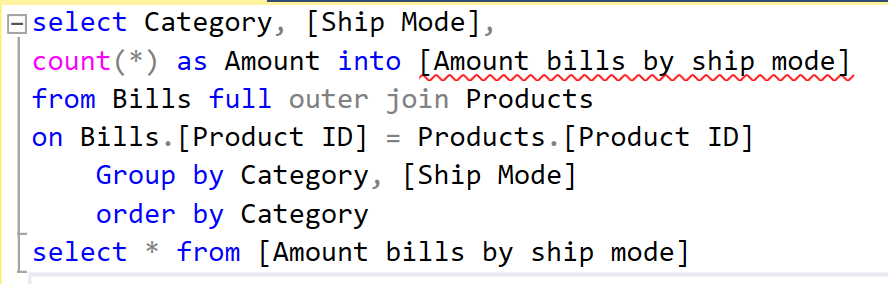
****

1. **Số lượng đơn hàng theo từng phương thức vận chuyển**

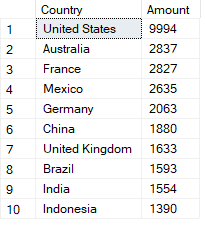
**Bảng số lượng đơn hàng theo từng phương thức vận chuyển**

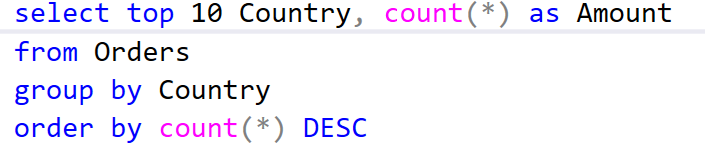
**[Amount bills by ship mode]**

****

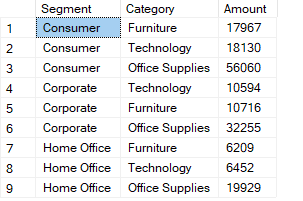
******

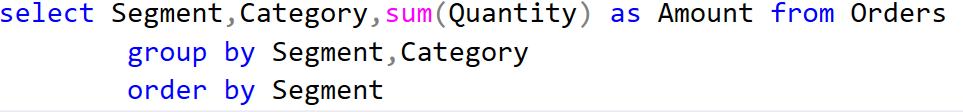
1. **Top 10 nước có số đơn hàng nhiều nhất**

****

****

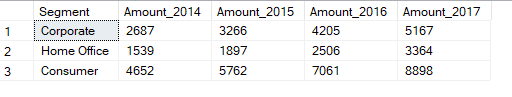
1. **Số lượng sản phẩm từng loại mặt hàng dựa trên phân khúc khách hàng (Thị hiếu mua hàng theo phân khúc khách hàng)**

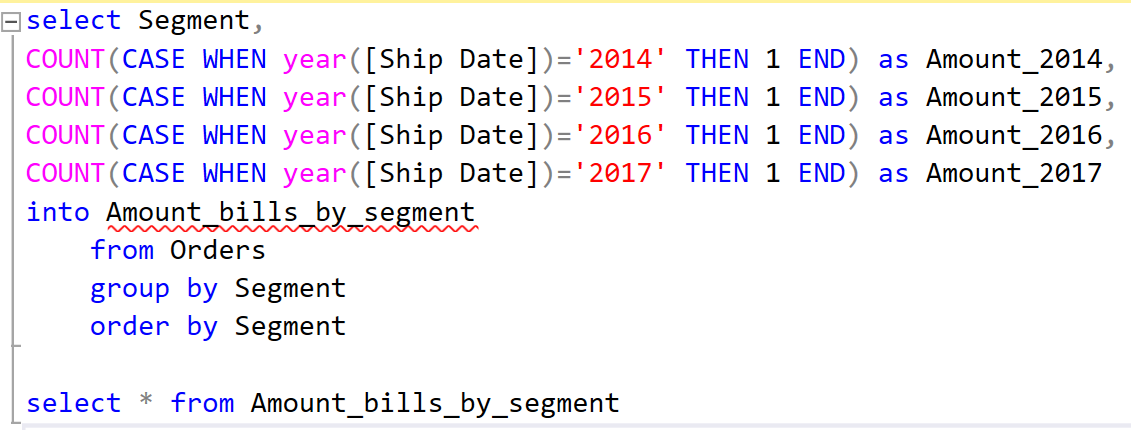
****



Đầu tiên trích ra các phân khúc khách hàng, các nhóm hàng, sau đó tính tổng số lượng theo các nhóm hàng tương ứng và nhóm dữ liệu theo Segment (phân khúc khách hàng) và Category (Loại mặt hàng).

1. **Số lượng đơn bán ra mỗi năm theo phân khúc khách hàng**





**Cách thực hiện:**

* Trích ra cột Segment (phân khúc khách hàng), sau đó dùng Count đếm có điều kiện (Case When) theo từng năm tương ứng rồi thêm vào 4 cột tương ứng.
* Nhóm dữ liệu theo phân khúc khách hàng.
* Cuối cùng ta được kết quả như trên.