

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**BÁO CÁO ĐỒ ÁN**

**KHO DỮ LIỆU VÀ OLAP**

**ĐỀ TÀI**

**PHÂN TÍCH DỮ LIỆU GLOBAL SUPERSTORE 2011-2014**

Giảng viên hướng dẫn:

**ThS. Đỗ Thị Kim Phụng**

Sinh viên thực hiện:

Lê Nguyễn Bá Duy 20521232

**TP. Hồ Chí Minh, tháng 5 nă**

**NHẬN XÉT CỦA GIẢNG VIÊN**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**…………………………………………………………………………………………**

**MỤC LỤC**

[DANH SÁCH BIỂU ĐỒ 6](#_Toc169183523)

[CHƯƠNG 1: GIỚI THIỆU ĐỒ ÁN 7](#_Toc169183524)

[1.1. Lý do chọn đề tài 7](#_Toc169183525)

[1.2. Dataset 7](#_Toc169183526)

[1.3. Xây dựng kho dữ liệu 9](#_Toc169183527)

[CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU 11](#_Toc169183528)

[2.1. Tạo project SSIS và thiết lập kết nối 11](#_Toc169183529)

[2.2. Nạp dữ liệu từ CSV vào cơ sở dữ liệu 16](#_Toc169183530)

[CHƯƠNG 3: QUÁ TRÌNH SSAS 21](#_Toc169183531)

[3.1. Tạo project SSAS và thiết lập kết nối 21](#_Toc169183532)

[3.2. Xây dựng khối dữ liệu 25](#_Toc169183533)

[3.3. Truy vấn dữ liệu 31](#_Toc169183534)

[Chương 4: Report với Power BI 42](#_Toc169183535)

[4.1. Tạo file PowerBI và thiết lập kết nối 42](#_Toc169183536)

[4.2. Các truy vấn được đồ thị hóa trên PowerBI 46](#_Toc169183537)

[CHƯƠNG 5: Data Mining 48](#_Toc169183538)

[5.1. Tiền xử lý dữ liệu và EDA 48](#_Toc169183539)

[5.2. Áp dụng thuật toán và đánh giá 49](#_Toc169183540)

## Danh sách hình ảnh

[Hình 1: Cấu trúc kho dữ liệu 11](#_Toc169183575)

[Hình 2: Tạo project SSIS (Bước 1) 12](#_Toc169183576)

[Hình 3: Tạo project SSIS (Bước 2) 12](#_Toc169183577)

[Hình 4: Tạo project SSIS (Bước 3) 13](#_Toc169183578)

[Hình 5: Tạo project SSIS (Bước 4) 13](#_Toc169183579)

[Hình 6: Tạo cơ sở dữ liệu (Bước 1) 14](#_Toc169183580)

[Hình 7: Tạo cơ sở dữ liệu (Bước 2) 14](#_Toc169183581)

[Hình 8: Thiết lập kết nối (Bước 1) 15](#_Toc169183582)

[Hình 9: Thiết lập kết nối (Bước 2) 15](#_Toc169183583)

[Hình 10: Thiết lập kết nối (Bước 3) 16](#_Toc169183584)

[Hình 11: Thiết lập kết nối (Bước 4) 16](#_Toc169183585)

[Hình 12: Tạo Data Flow (Bước 1) 17](#_Toc169183586)

[Hình 13: Tạo Data Flow (Bước 2) 17](#_Toc169183587)

[Hình 14: Tạo Data Flow (Bước 3) 17](#_Toc169183588)

[Hình 15: Tạo Data Flow (Bước 4) 18](#_Toc169183589)

[Hình 16: Tạo Data Flow (Bước 5) 18](#_Toc169183590)

[Hình 17: Tạo Control Flow (Bước 1) 19](#_Toc169183591)

[Hình 18: Tạo Control Flow (Bước 2) 19](#_Toc169183592)

[Hình 19: Tạo Control Flow (Bước 3) 20](#_Toc169183593)

[Hình 20: Nạp dữ liệu vào cơ sở dữ liệu (Bước 1) 20](#_Toc169183594)

[Hình 21: Nạp dữ liệu vào cơ sở dữ liệu (Bước 1) 20](#_Toc169183595)

[Hình 22: Nạp dữ liệu vào cơ sở dữ liệu (Bước 2) 21](#_Toc169183596)

[Hình 23: Tạo project SSAS và thiết lập kết nối (Bước 1) 22](#_Toc169183597)

[Hình 24: Tạo project SSAS và thiết lập kết nối (Bước 2) 22](#_Toc169183598)

[Hình 25: Tạo project SSAS và thiết lập kết nối (Bước 3) 23](#_Toc169183599)

[Hình 26: Tạo project SSAS và thiết lập kết nối (Bước 4) 23](#_Toc169183600)

[Hình 27: Tạo project SSAS và thiết lập kết nối (Bước 5) 24](#_Toc169183601)

[Hình 28: Tạo project SSAS và thiết lập kết nối (Bước 6) 24](#_Toc169183602)

[Hình 29: Tạo project SSAS và thiết lập kết nối (Bước 7) 25](#_Toc169183603)

[Hình 30: Tạo project SSAS và thiết lập kết nối (Bước 8) 25](#_Toc169183604)

[Hình 31: Xây dựng khối dữ liệu (Bước 1) 26](#_Toc169183605)

[Hình 32: Xây dựng khối dữ liệu (Bước 2) 26](#_Toc169183606)

[Hình 33: Xây dựng khối dữ liệu (Bước 3) 27](#_Toc169183607)

[Hình 34: Xây dựng khối dữ liệu (Bước 4) 27](#_Toc169183608)

[Hình 35: Xây dựng khối dữ liệu (Bước5 ) 28](#_Toc169183609)

[Hình 36: Xây dựng khối dữ liệu (Bước 6) 28](#_Toc169183610)

[Hình 37: Xây dựng khối dữ liệu (Bước 7) 29](#_Toc169183611)

[Hình 38: Xây dựng khối dữ liệu (Bước 8) 29](#_Toc169183612)

[Hình 39: Xây dựng khối dữ liệu (Bước 9) 30](#_Toc169183613)

[Hình 40: Xây dựng khối dữ liệu (Bước 10) 30](#_Toc169183614)

[Hình 41: Xây dựng khối dữ liệu (Bước 11) 31](#_Toc169183615)

[Hình 42: Truy vấn dữ liệu sử dụng MDX (Câu 1) 32](#_Toc169183616)

[Hình 43: Truy vấn dữ liệu sử dung Pivot Table (Câu 1) 32](#_Toc169183617)

[Hình 44: Truy vấn dữ liệu sử dụng MDX (Câu 2) 33](#_Toc169183618)

[Hình 45: Truy vấn dữ liệu sử dung Pivot Table (Câu 2) 33](#_Toc169183619)

[Hình 46: Truy vấn dữ liệu sử dụng MDX (Câu 3) 34](#_Toc169183620)

[Hình 47: Truy vấn dữ liệu sử dung Pivot Table (Câu 3) 34](#_Toc169183621)

[Hình 48: Truy vấn dữ liệu sử dụng MDX (Câu 4) 35](#_Toc169183622)

[Hình 49: Truy vấn dữ liệu sử dung Pivot Table (Câu 4) 35](#_Toc169183623)

[Hình 50: Truy vấn dữ liệu sử dụng MDX (Câu 5) 36](#_Toc169183624)

[Hình 51: Truy vấn dữ liệu sử dung Pivot Table (Câu 5) 36](#_Toc169183625)

[Hình 52: Truy vấn dữ liệu sử dụng MDX (Câu 6) 37](#_Toc169183626)

[Hình 53: Truy vấn dữ liệu sử dung Pivot Table (Câu 6) 37](#_Toc169183627)

[Hình 54: Truy vấn dữ liệu sử dụng MDX (Câu 7) 38](#_Toc169183628)

[Hình 55: Truy vấn dữ liệu sử dung Pivot Table (Câu 7) 38](#_Toc169183629)

[Hình 56: Truy vấn dữ liệu sử dụng MDX (Câu 8) 39](#_Toc169183630)

[Hình 57: Truy vấn dữ liệu sử dung Pivot Table (Câu 8) 39](#_Toc169183631)

[Hình 58: Truy vấn dữ liệu sử dụng MDX (Câu 9) 40](#_Toc169183632)

[Hình 59: Truy vấn dữ liệu sử dung Pivot Table (Câu 9) 40](#_Toc169183633)

[Hình 60: Truy vấn dữ liệu sử dụng MDX (Câu 10) 41](#_Toc169183634)

[Hình 61: Truy vấn dữ liệu sử dung Pivot Table (Câu 10) 41](#_Toc169183635)

[Hình 62: Truy vấn dữ liệu sử dụng MDX (Câu 11) 42](#_Toc169183636)

[Hình 63: Truy vấn dữ liệu sử dung Pivot Table (Câu 11) 42](#_Toc169183637)

[Hình 64: Tạo file PowerBI và thiết lập kết nối (Bước 1) 43](#_Toc169183638)

[Hình 65: Tạo file PowerBI và thiết lập kết nối (Bước 2) 43](#_Toc169183639)

[Hình 66: Tạo file PowerBI và thiết lập kết nối (Bước 3) 44](#_Toc169183640)

[Hình 67: Tạo file PowerBI và thiết lập kết nối (Bước 4) 44](#_Toc169183641)

[Hình 68: Tạo file PowerBI và thiết lập kết nối (Bước 5) 45](#_Toc169183642)

[Hình 69: Tạo file PowerBI và thiết lập kết nối (Bước 6) 45](#_Toc169183643)

[Hình 70: Tạo file PowerBI và thiết lập kết nối (Bước 7) 46](#_Toc169183644)

[Hình 71: Tạo file PowerBI và thiết lập kết nối (Bước 8) 47](#_Toc169183645)

[Hình 72: Truy vấn 1-5 trên PowerBI 47](#_Toc169183646)

[Hình 73: Truy vấn 6-9 trên PowerBI 48](#_Toc169183647)

[Hình 74: Truy vấn 10-11 trên PowerBI 48](#_Toc169183648)

[Hình 75: Tiền xử lý dữ liệu và EDA (Bước 1) 49](#_Toc169183649)

[Hình 76: Tiền xử lý dữ liệu và EDA (Bước 2) 49](#_Toc169183650)

[Hình 77: Tiền xử lý dữ liệu và EDA (Bước 3) 49](#_Toc169183651)

[Hình 78: Tiền xử lý dữ liệu và EDA (Bước 4) 50](#_Toc169183652)

[Hình 79: Tiền xử lý dữ liệu và EDA (Bước 5) 50](#_Toc169183653)

[Hình 80: Kết quả của thuật toán SES 51](#_Toc169183654)

[Hình 81: Kết quả thuật toán Holt 51](#_Toc169183655)

[Hình 82: Kết quả thuật toán Holt’s Winter 52](#_Toc169183656)

[Hình 83: Kết quả thuật toán SARIMAX 52](#_Toc169183657)

## 

## CHƯƠNG 1: GIỚI THIỆU ĐỒ ÁN

### 1.1. Lý do chọn đề tài

Phân tích doanh số bán hàng là một bước then chốt đối với các dự án hướng tới doanh nghiệp kinh doanh vì việc này có mối quan hệ trực tiếp đối với hiệu suất và sự ổn định của doanh nghiệp. Sau đây là một số lý do tại sao nó lại quan trọng đến vậy.

* **Đối với hiệu suất doanh nghiệp:** Việc phân tích doanh số bán hàng giúp các công ty hiểu được họ đang thực hiện tốt ở đâu và chỗ nào cần cải thiện.
* **Các cơ hội bị bỏ lỡ:** từ các dữ liệu có sẵn công ty có thể xác định được lý do một số cơ hội tại sao lại thất bại và liệu có thể khôi phục được không.
* **Đưa ra quyết định trong tương lai:** những dữ liệu sẽ ảnh hưởng đến các quyết định quan trọng liên quan đến quản lý tồn kho, hoạt động tiếp thị và quy trình sản xuất như tiếp tục hay ngưng sản xuất một sản phẩm, liệu có nên đầu tư vào một công ty hay không.
* **Xu hướng thị trường:** Phân tích doanh số bán hàng tiết lộ các xu hướng thị trường hiện tại ví dụ như sự tăng doanh số đột ngột của một sản phẩm liên quan đến một sự kiện cụ thể. Từ đó giúp công ty điều chỉnh chiến lược và tận dụng động lực thị trường.

Tóm lại, phân tích doanh số bán hàng giúp cung cấp các thông tin hữu ích, hướng dẫn quyết định và đảm bảo sự thành công lâu dài của một công ty.

### 1.2. Dataset

#### 1.2.1. Một số thông tin về dataset

* **Tên dataset:** Global Superstore Dataset
* **Tác giả:** Fatih İlhan, Anandaram Ganapathi.
* **Thời gian cập nhật**: 6 tháng trước.
* **Link**: <https://www.kaggle.com/datasets/fatihilhan/global-superstore-dataset>
* **Thông tin:** Bộ dữ liệu bao gồm 51290 dòng là các sản phầm được bán ra và 26 cột chứa các thông tin liên quan tới đơn hàng.
* **Mô tả chi tiết**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Ý nghĩa** | **Kiểu dữ liệu** |
| 1 | RowID | ID của các dòng trong bộ dữ liệu | Int |
| 2 | OrderID | ID của các đơn hàng | String |
| 3 | OrderPriority | Độ ưu tiên của đơn hàng | String |
| 4 | OrderDate | Thời gian đơn hàng được đặt | DateTime |
| 5 | ShipMode | Phương thức giao hàng | String |
| 6 | CustomerID | ID của khách hàng | String |
| 7 | CustomerName | Tên khách hàng | String |
| 8 | Segment | Phân loại khách hàng | String |
| 9 | ProductID | ID của sản phẩm | String |
| 10 | Category | Danh mục của sản phẩm | String |
| 11 | SubCategory | Danh mục phụ của sản phẩm | String |
| 12 | City | Thành phố của cửa hàng. | String |
| 13 | State | Bang của cửa hàng. | String |
| 14 | Country | Đất nước của cửa hàng | String |
| 15 | Region | Khu vực của cửa hàng | String |
| 16 | Market | Thị trường của cửa hàng | String |
| 17 | Profit | Lợi nhuận thu được từ đơn hàng | Float |
| 18 | Quantity | Số lượng sản phẩm có trong đơn hàng | Int |
| 19 | Sales | Doanh thu từ đơn hàng | Float |
| 20 | Discount | Tỉ lệ giảm giá của đơn hàng | Float |
| 21 | Shipping cost | Chi phí giao hàng | Float |

### 1.3. Xây dựng kho dữ liệu

#### 1.3.1 Các bảng trong kho dữ liệu

**Đối với bộ dữ liệu này em chia ra làm các bảng sau:**

**a. Bảng Dim\_Customer**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | CustomerID | String |
|  | CustomerName | String |
|  | Segment | String |

**b. Bảng Dim\_Product**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | ProductID | String |
|  | Category | String |
|  | ProductName | String |
|  | SubCategory | String |

**c. Bảng Dim\_Location**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | LocationID | String |
|  | City | String |
|  | Country | String |
|  | Market | String |
|  | Region | String |
|  | State | String |

**c. Bảng Dim\_Order**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | OrderID | String |
|  | OrderDate | DateTime |
|  | OrderPriority | String |
|  | ShipMode | String |

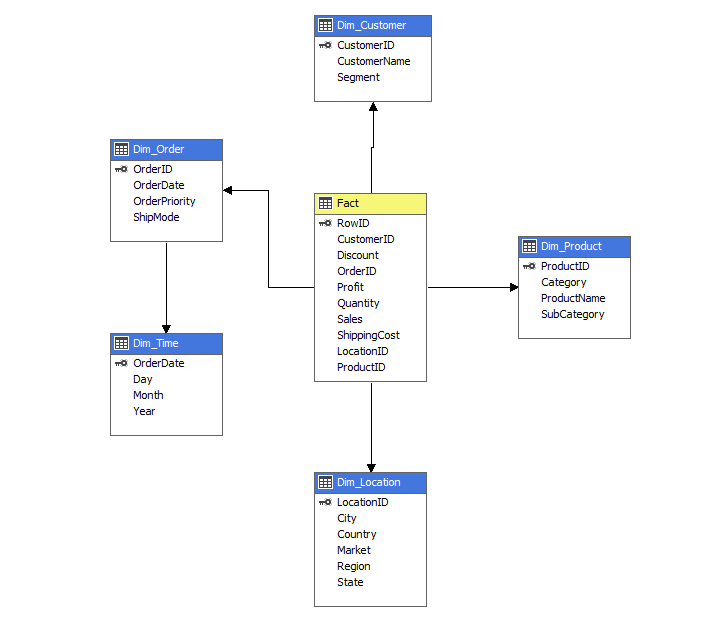
**d. Bảng Dim\_Time**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | OrderDate | DateTime |
|  | Day | Int |
|  | Month | Int |
|  | Year | Int |

**e. Bảng Fact**

|  |  |  |
| --- | --- | --- |
| **Khóa chính** | **Tên thuộc tính** | **Kiểu dữ liệu** |
| 🔑 | RowID | Int |
|  | CustomerID | String |
|  | OrderID | String |
|  | LocationID | String |
|  | ProductID | String |
|  | Discount | Float |
|  | Profit | Float |
|  | Sales | Float |
|  | Quatity | Int |
|  | ShippingCost | Float |

#### 1.3.2. Cấu trúc kho dữ liệu



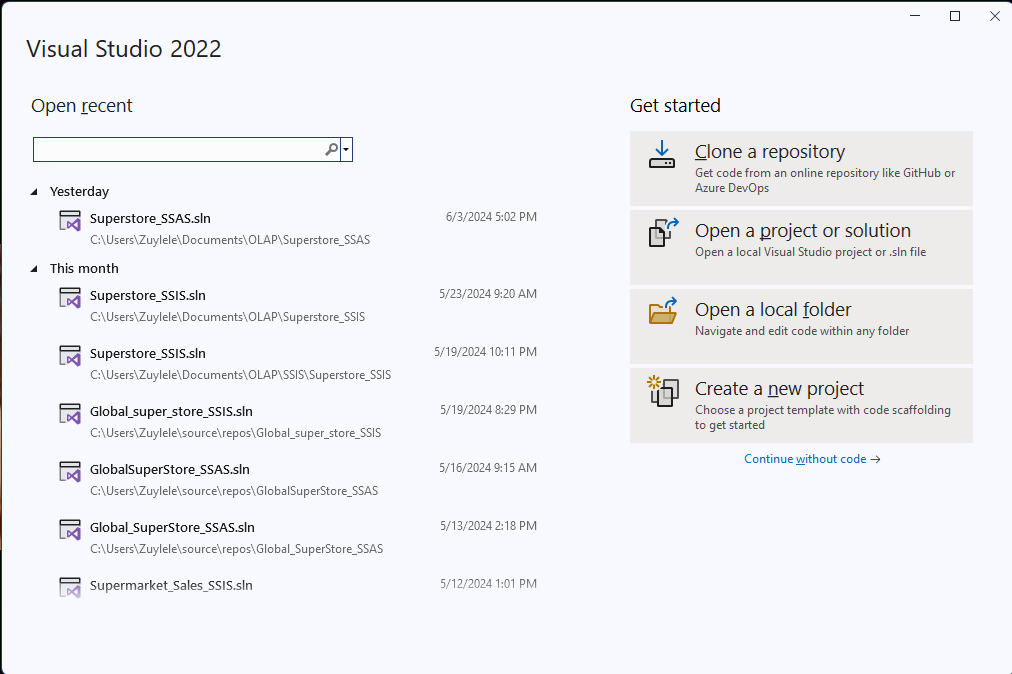
Hình 1: Cấu trúc kho dữ liệu

## CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU

### 2.1. Tạo project SSIS và thiết lập kết nối

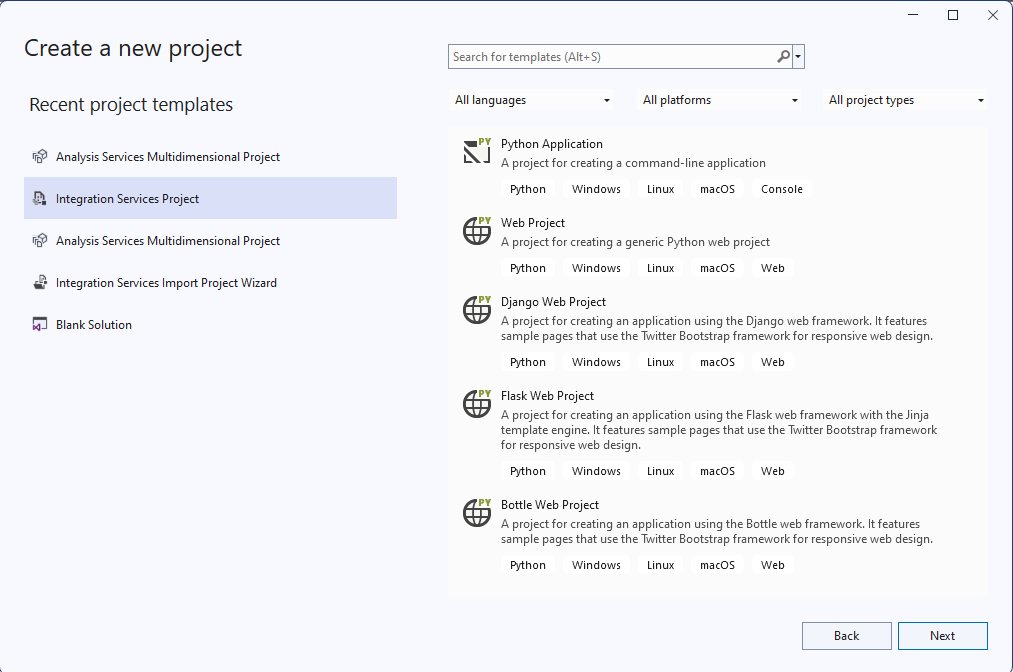
#### 2.1.1. Tạo project SSIS

Bước 1: Khởi động Visual Studio và chọn vào Create a new project



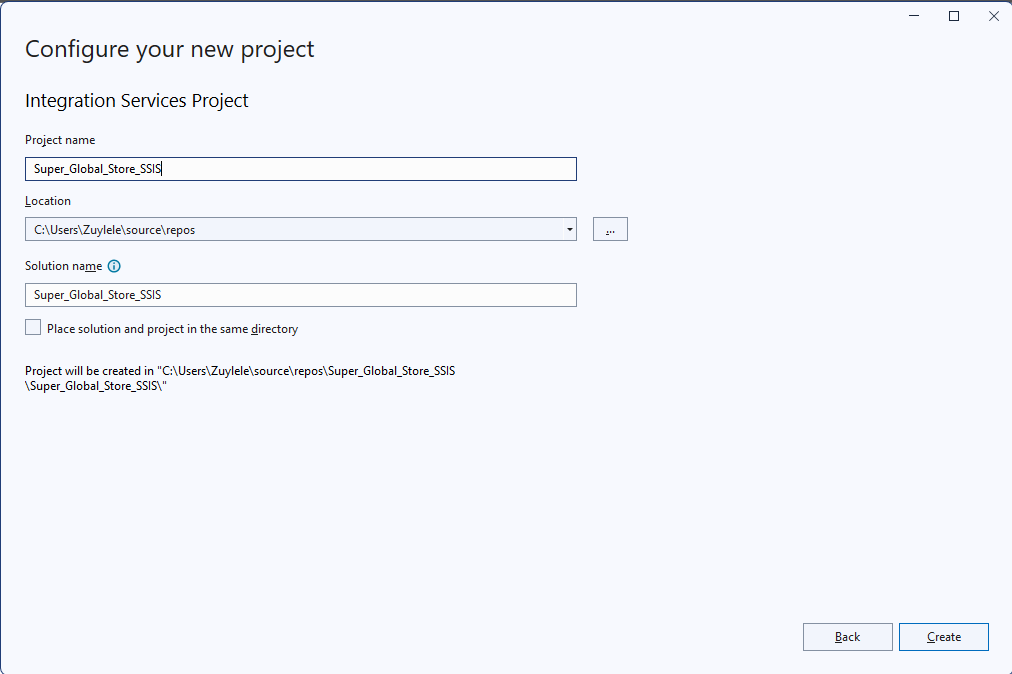
Hình 2: Tạo project SSIS (Bước 1)

Bước 2: Tiếp theo chọn vào Integration Services Project



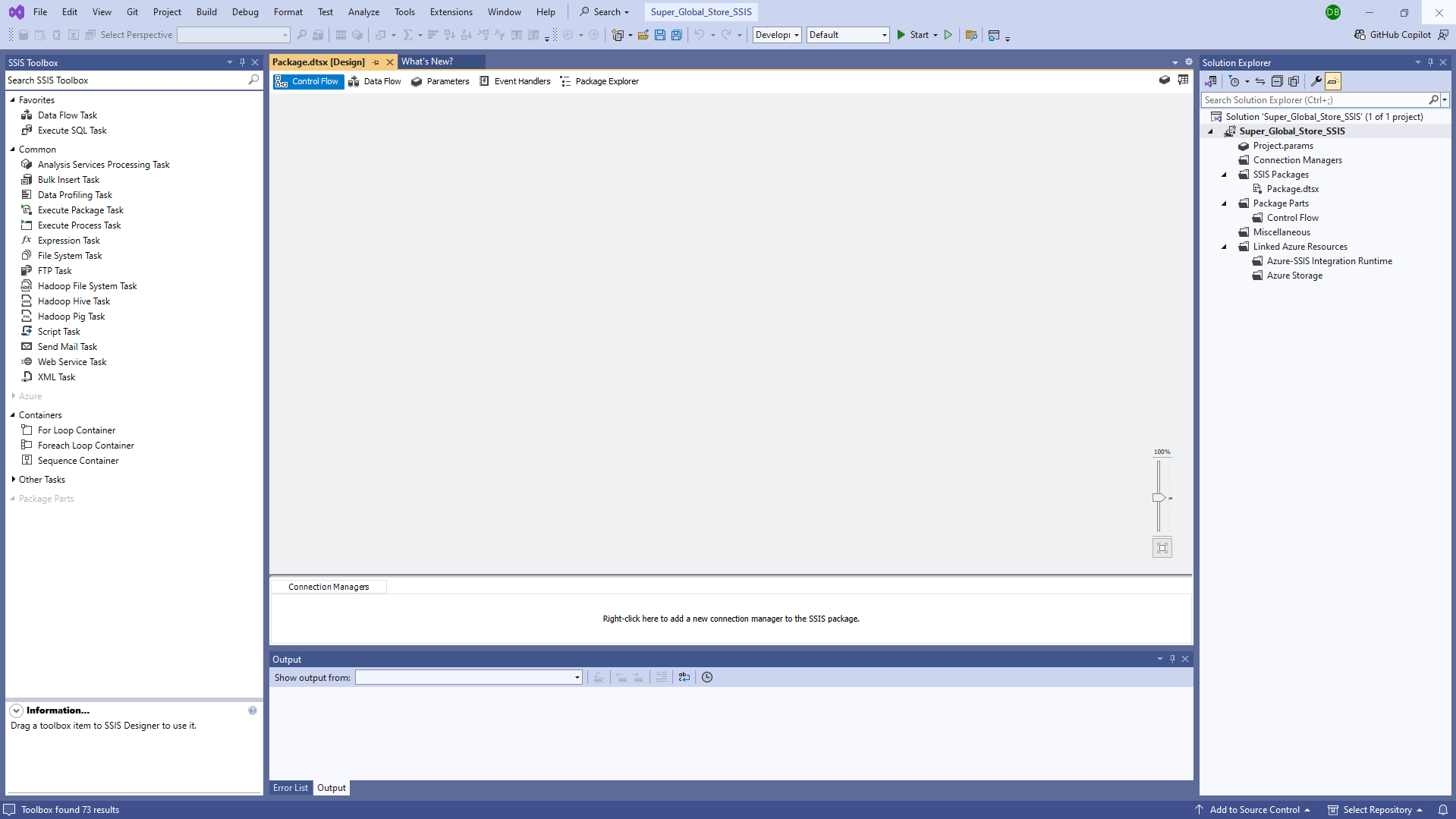
Hình 3: Tạo project SSIS (Bước 2)

Bước 3: Đặt tên và chọn nơi lưu trữ sau đó chọn Create



Hình 4: Tạo project SSIS (Bước 3)

Bước 4: Giao diện của SSIS

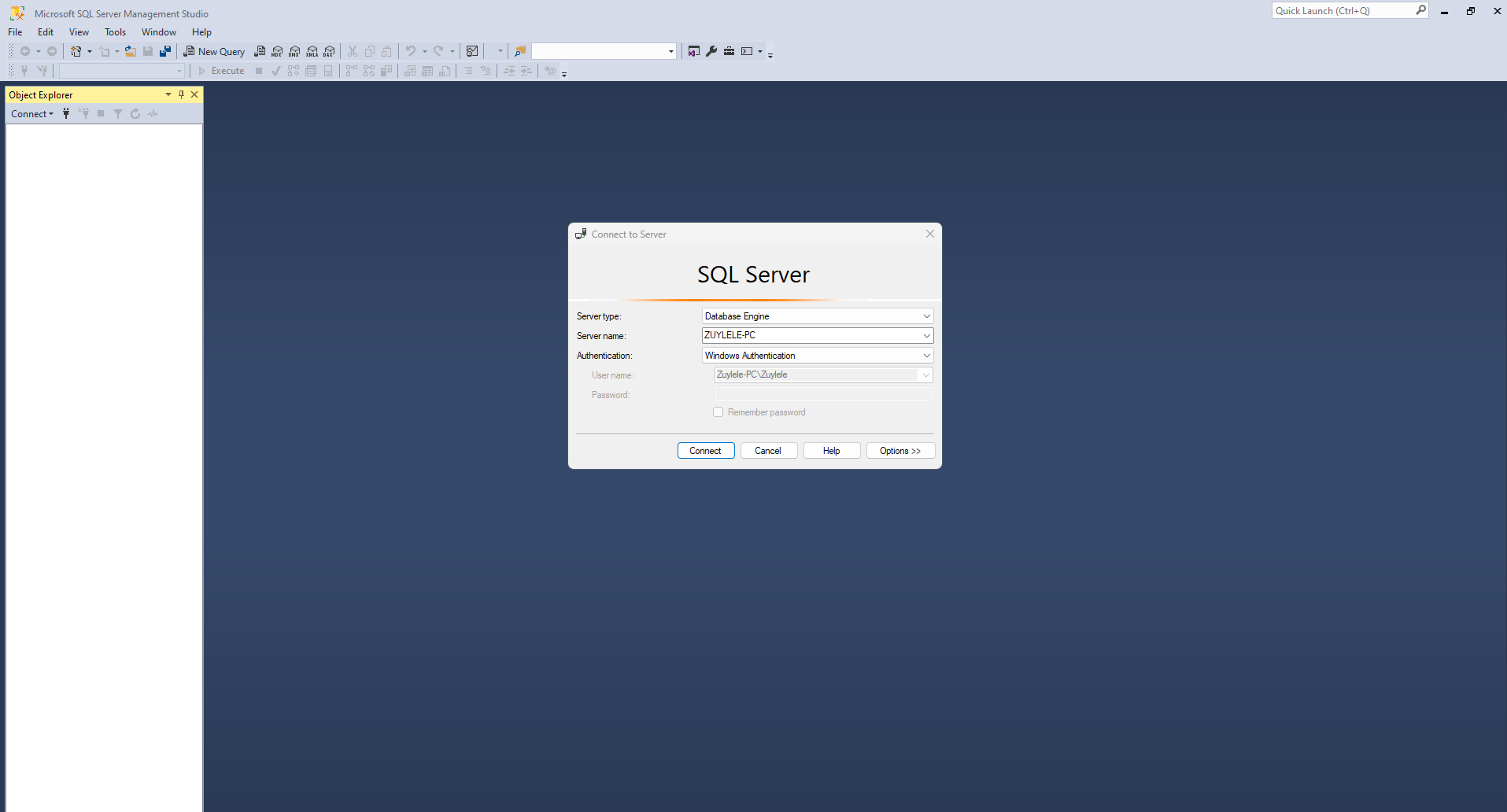


Hình 5: Tạo project SSIS (Bước 4)

#### 2.1.2. Tạo cơ sở dữ liệu và thiết lập kết nối

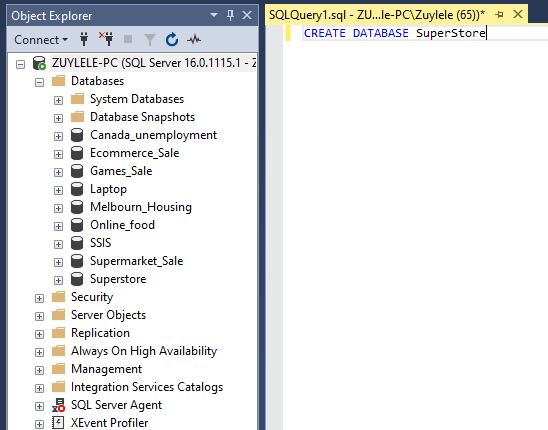
**2.1.2.1. Tạo cơ sở dữ liệu**

Bước 1: Khởi động Microsoft SQL Server Management Studio và kết nối đến cơ sở dữ liệu.



Hình 6: Tạo cơ sở dữ liệu (Bước 1)

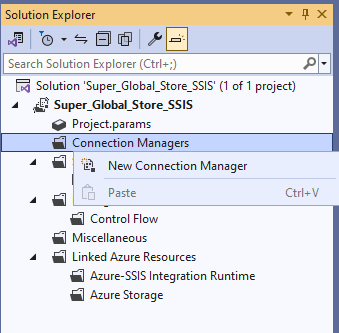
Bước 2: Tạo một query mới và chạy câu lệnh CREATE DATABASE SuperStore để tạo một cơ sở dữ liệu mới.



Hình 7: Tạo cơ sở dữ liệu (Bước 2)

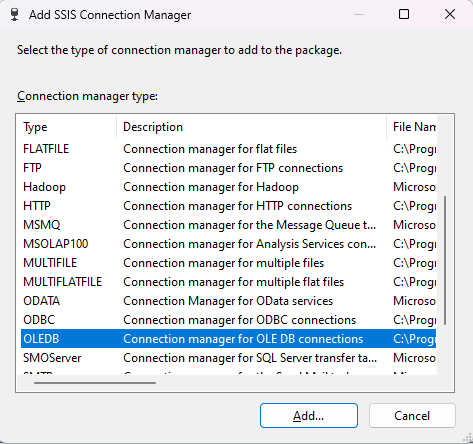
**2.1.2.2. Thiết lập kết nối**

Bước 1: Quay lại giao diện SSIS chuột phải vào Connection Managers → New Connection Managers.



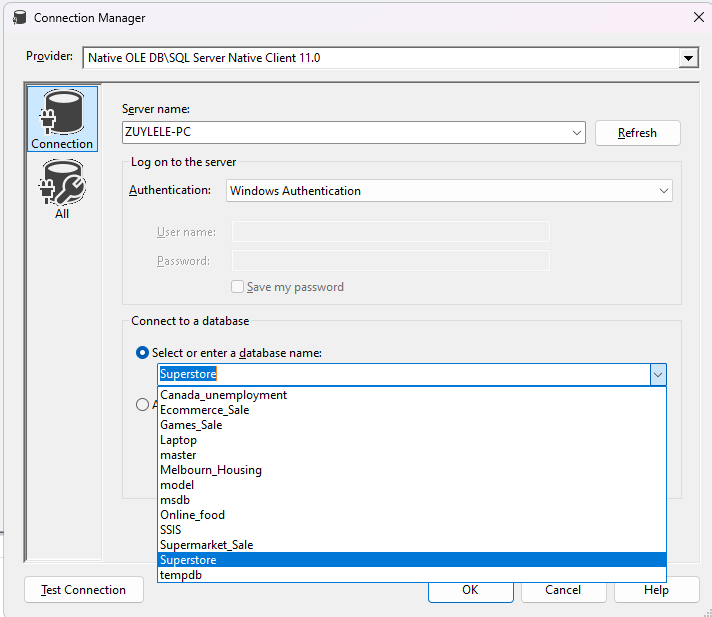
Hình 8: Thiết lập kết nối (Bước 1)

Bước 2: Chọn OLEDB → Add



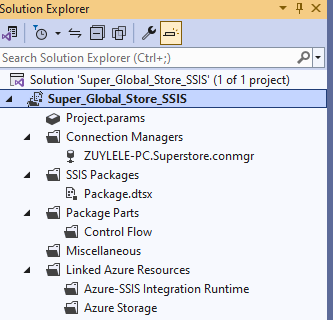
Hình 9: Thiết lập kết nối (Bước 2)

Bước 3: Ở màn hình tiếp theo chọn New sau đó nhập tên của SQL server rồi chọn cơ sở dữ liệu đã tạo trước đó rồi chọn OK.



Hình 10: Thiết lập kết nối (Bước 3)

Bước 4: Nhấn OK ở màn hình tiếp theo và đã kết nối được với cở sở dữ liệu.

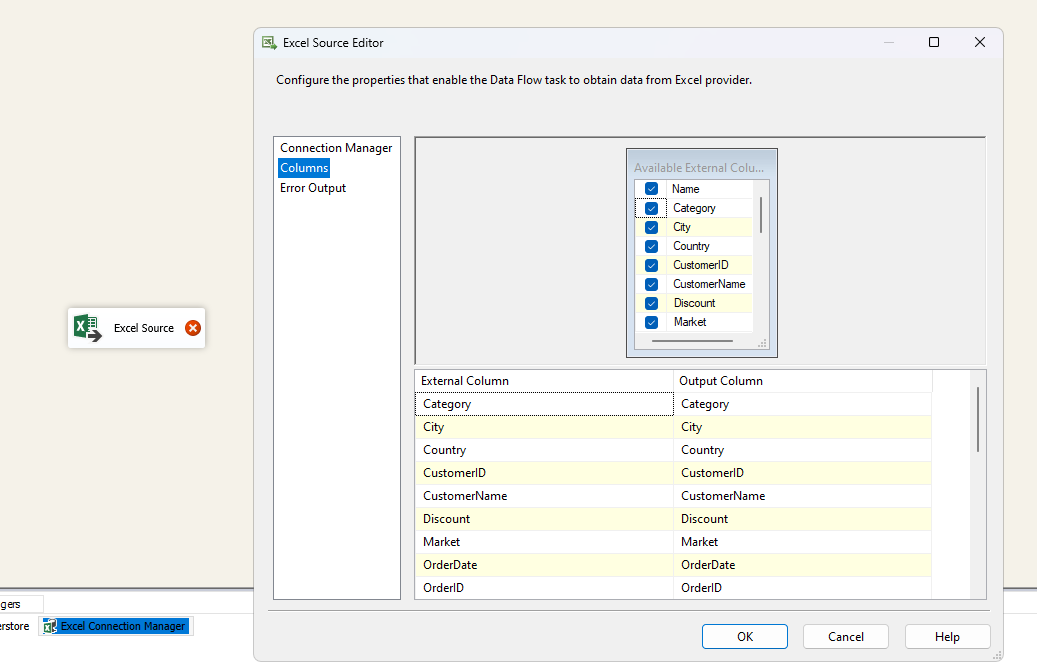


Hình 11: Thiết lập kết nối (Bước 4)

### 2.2. Nạp dữ liệu từ CSV vào cơ sở dữ liệu

**2.2.1. Tạo Data Flow**

Bước 1: Tạo một Toolbox Source tương ứng với dữ liệu muốn nạp vào và thêm dữ liệu đó vào sources.



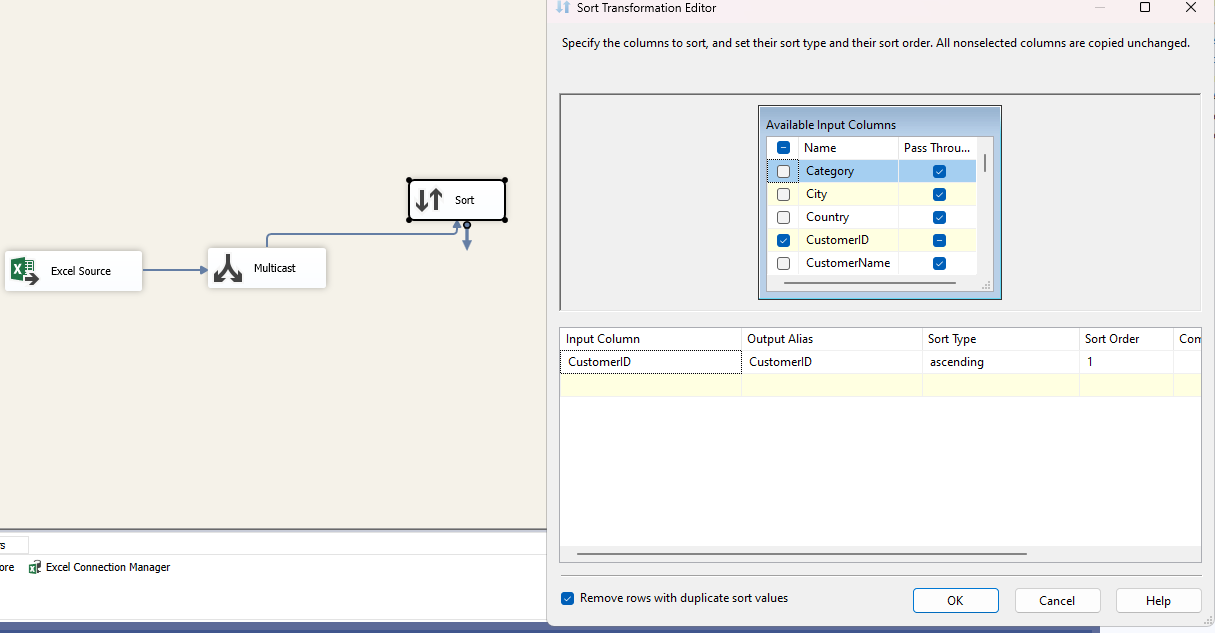
Hình 12: Tạo Data Flow (Bước 1)

Bước 2: Thêm một Multicast để phân luồng dữ liệu cho các bảng khác nhau.



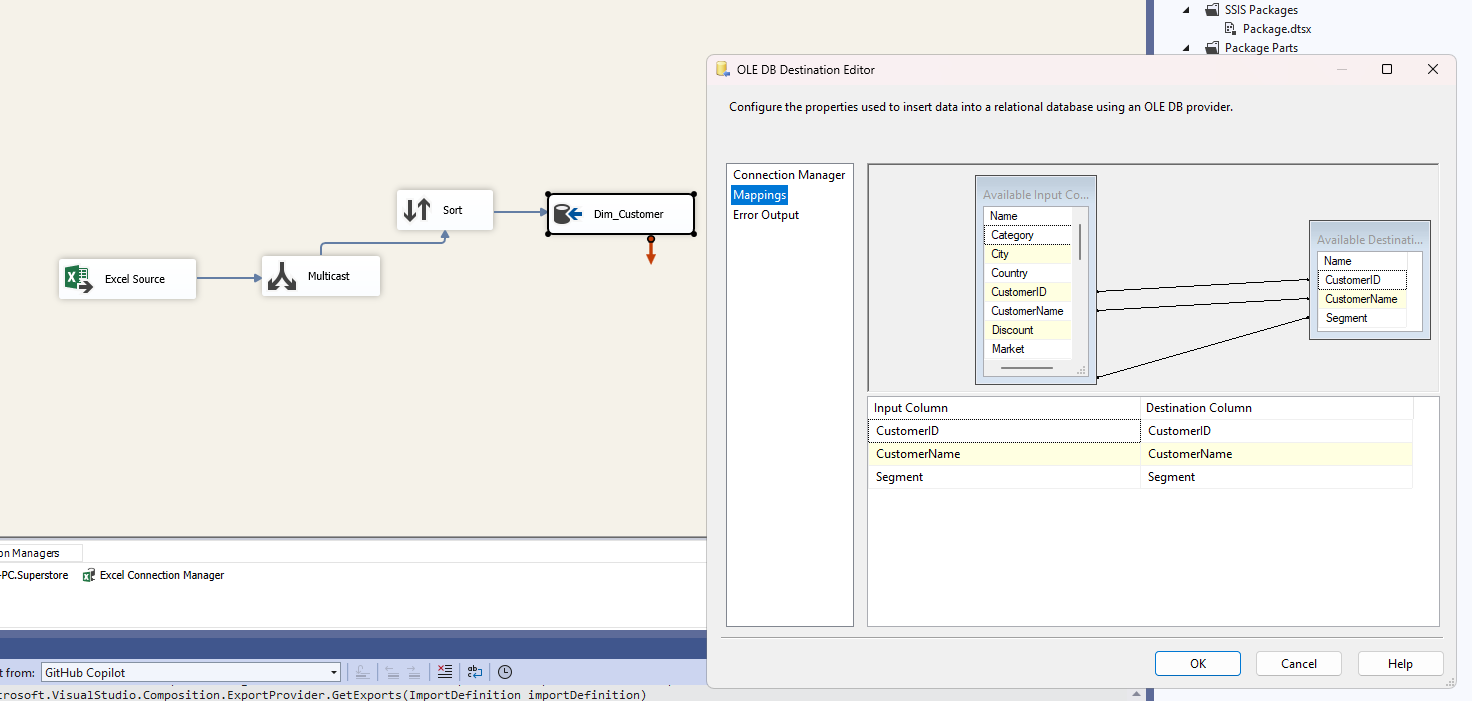
Hình 13: Tạo Data Flow (Bước 2)

Bước 3: Thêm các Sort và chọn trường dữ liệu đóng vai trò là khóa chính



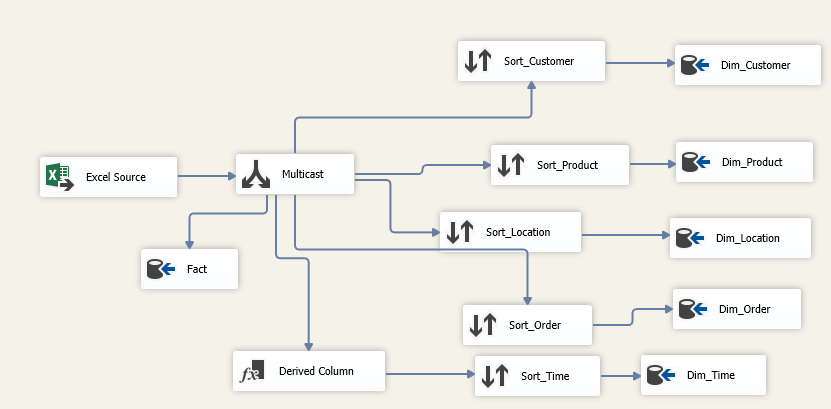
Hình 14: Tạo Data Flow (Bước 3)

Bước 4: Tạo các OLEDB Destination tương ứng với các bảng Dim ở trên và nối nó với các Sort tương ứng.



Hình 15: Tạo Data Flow (Bước 4)

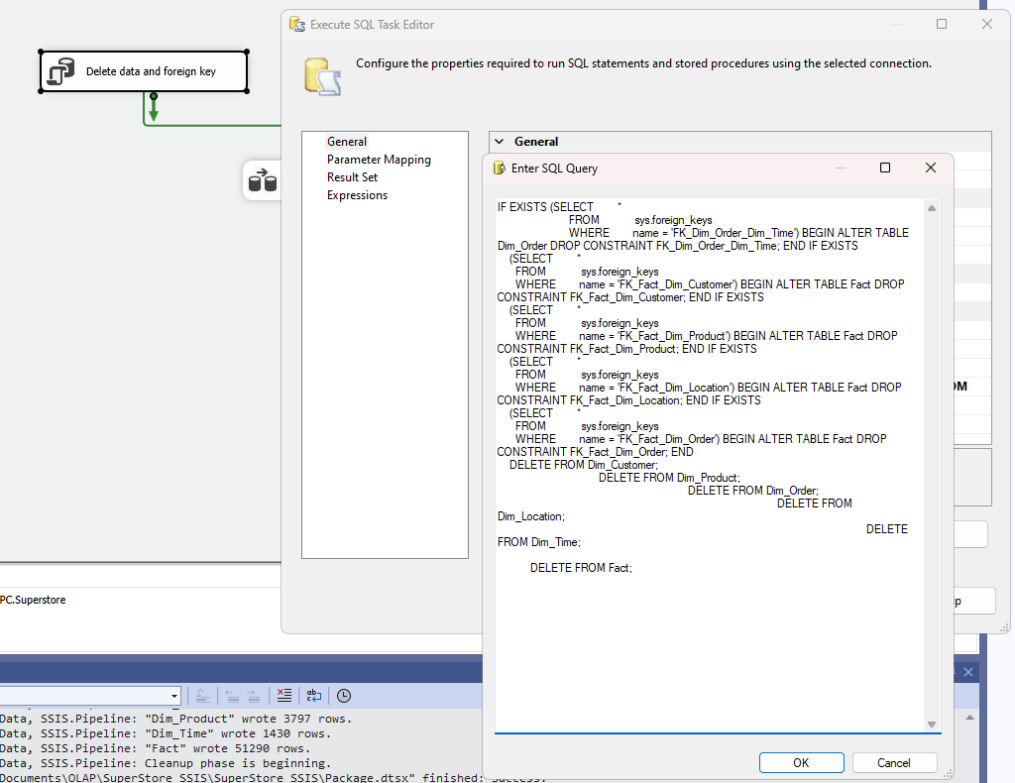
Bước 5: Làm tương tự với các bảng Dim còn lại.



Hình 16: Tạo Data Flow (Bước 5)

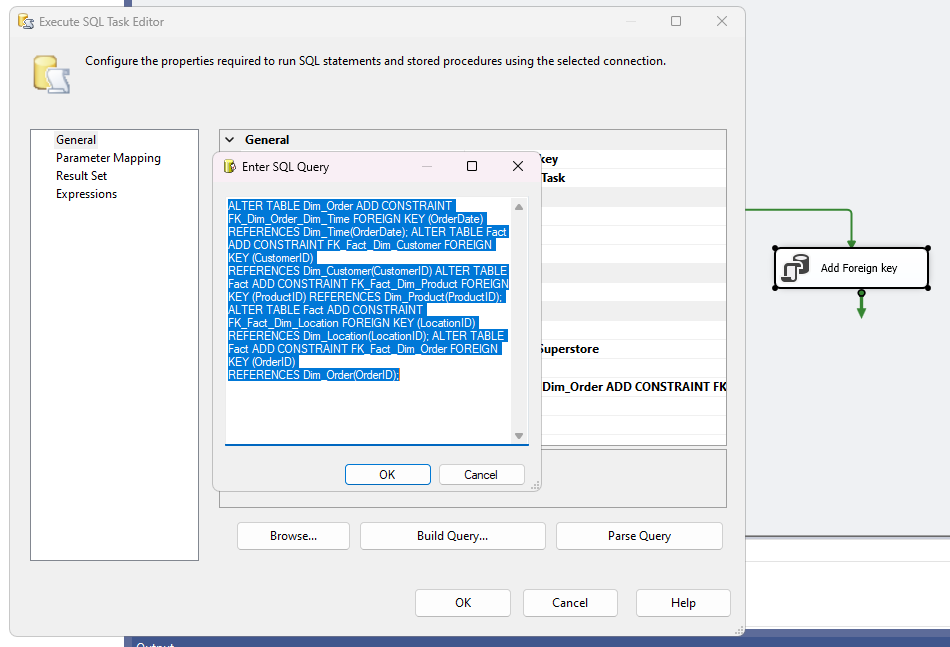
**2.2.2. Tạo Control Flow**

Bước 1: Tạo một Execute SQL task để xóa những dữ liệu và khóa ngoại có trong các bảng trước đó.



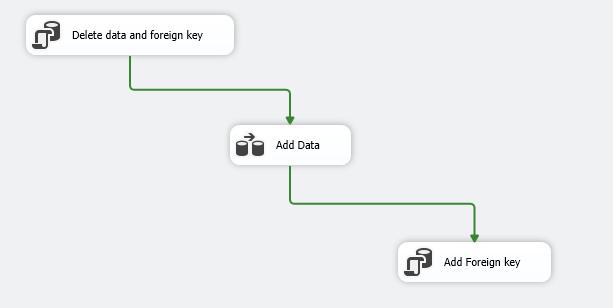
Hình 17: Tạo Control Flow (Bước 1)

Bước 2: Tạo một Execute SQL task để nối các khóa ngoại.



Hình 18: Tạo Control Flow (Bước 2)

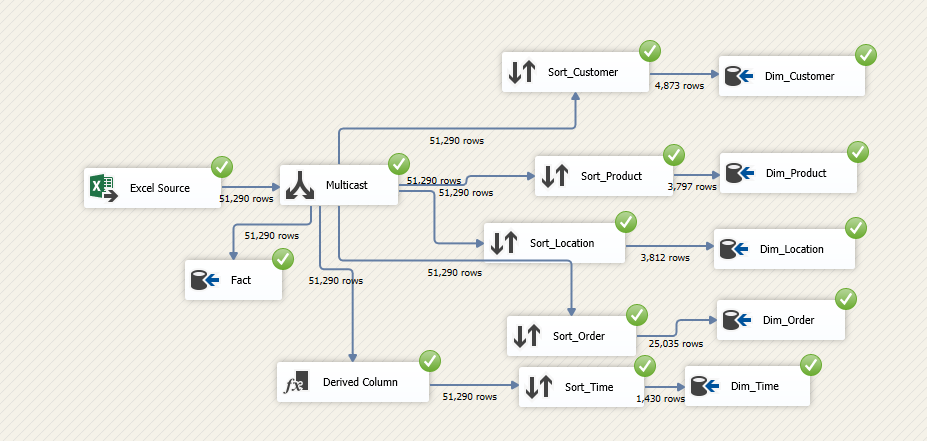
Bước 3: Nối các thành phần lại với nhau



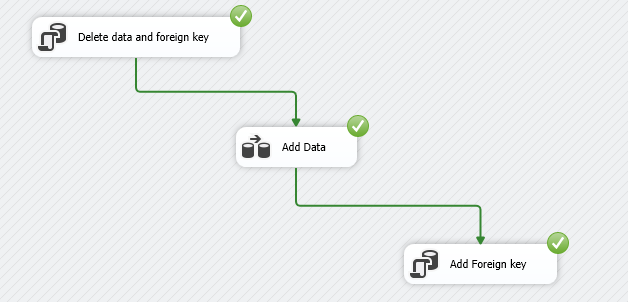
Hình 19: Tạo Control Flow (Bước 3)

**2.2.3. Nạp dữ liệu vào cơ sở dữ liệu**

Bước 1: Cho chạy Control Flow đã được tạo ở trên.

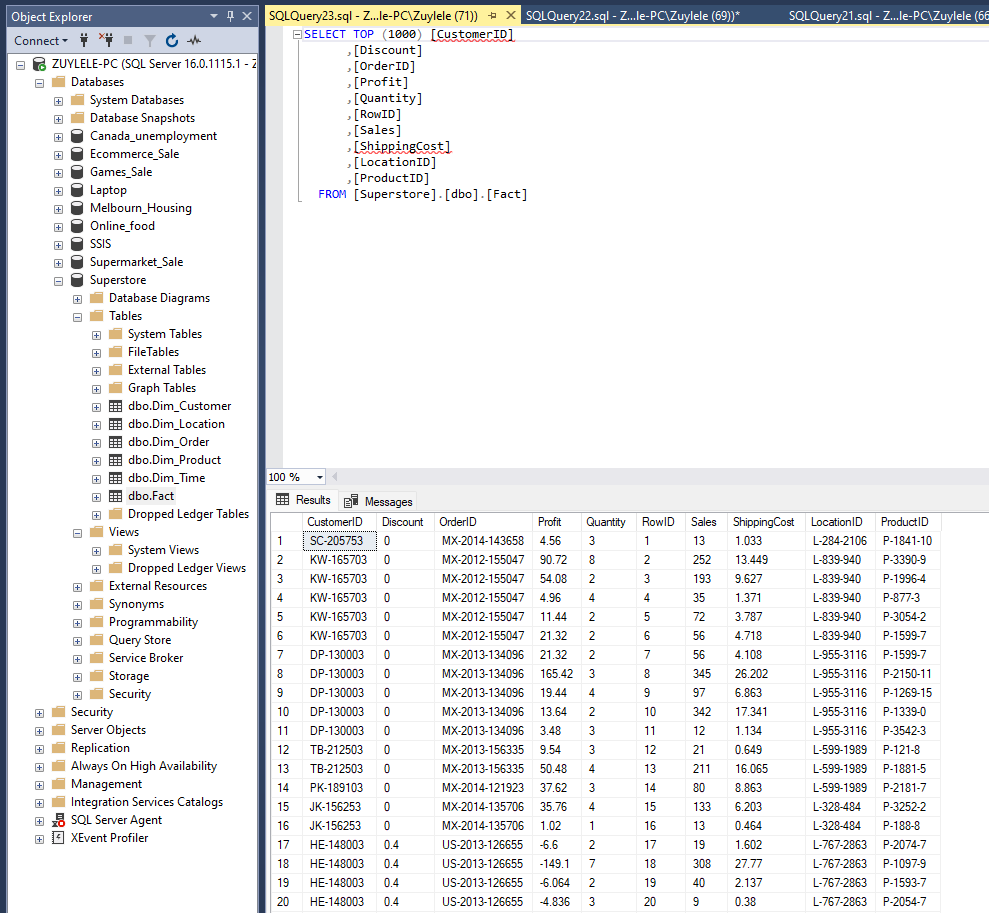


Hình 20: Nạp dữ liệu vào cơ sở dữ liệu (Bước 1)



Hình 21: Nạp dữ liệu vào cơ sở dữ liệu (Bước 1)

Bước 2: Kiểm tra dữ liệu đã được nạp vào cơ sở dữ liệu

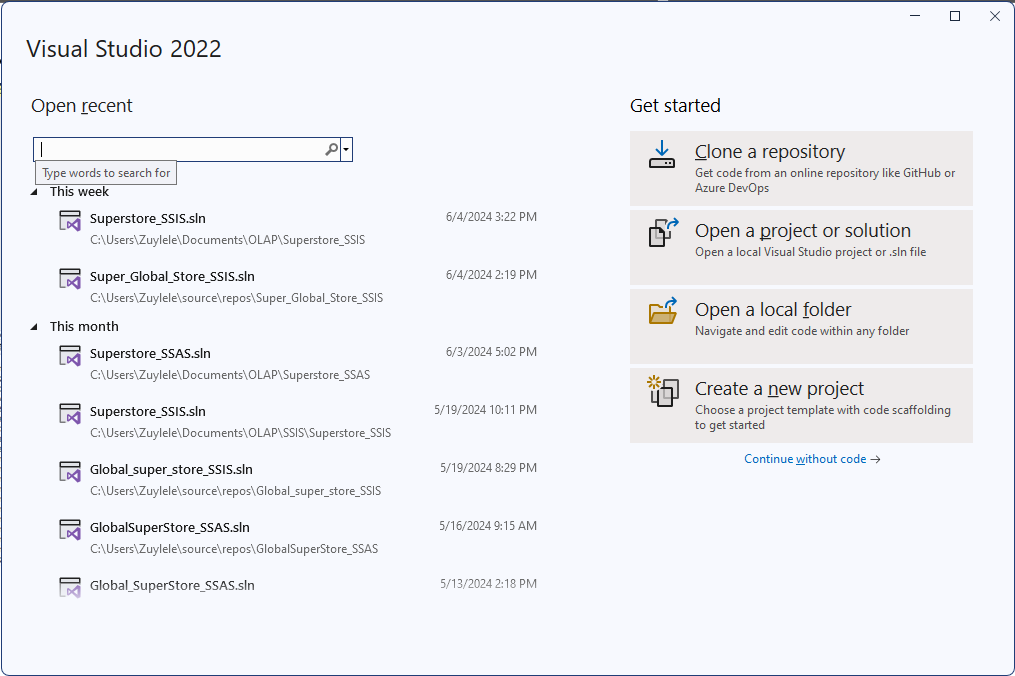


Hình 22: Nạp dữ liệu vào cơ sở dữ liệu (Bước 2)

## CHƯƠNG 3: QUÁ TRÌNH SSAS

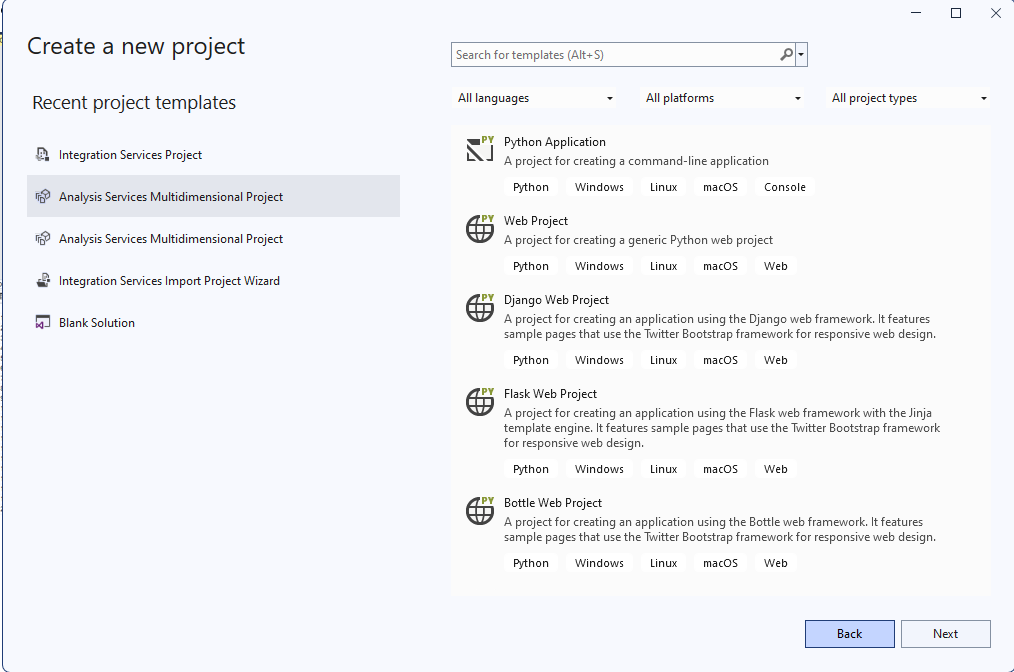
### 3.1. Tạo project SSAS và thiết lập kết nối

Bước 1: Vào Visual Studio và chọn Create a new project



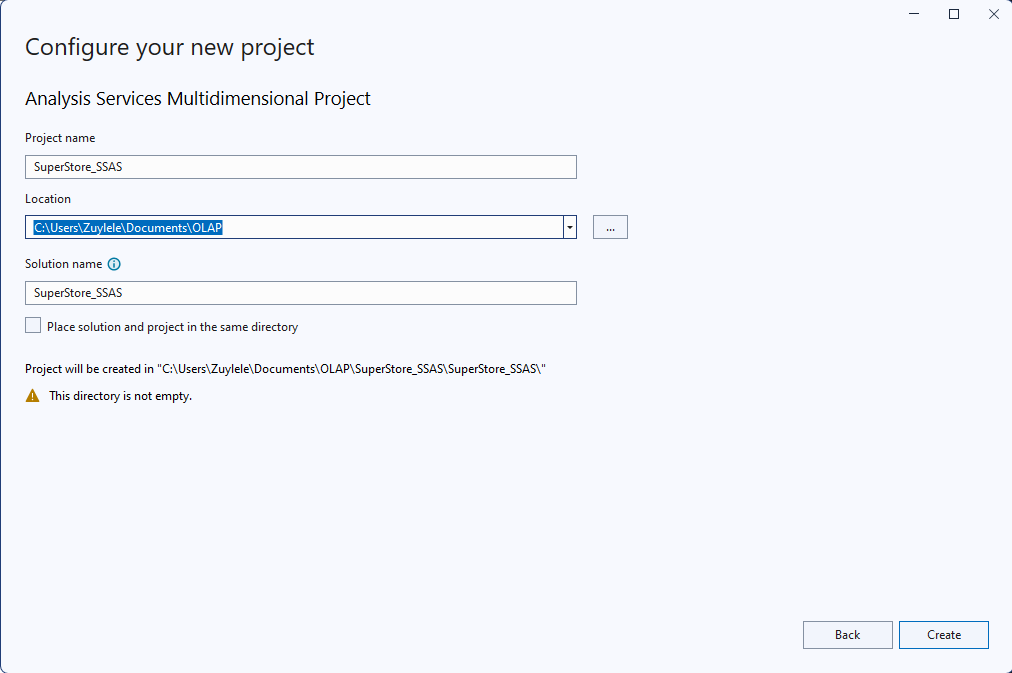
Hình 23: Tạo project SSAS và thiết lập kết nối (Bước 1)

Bước 2: Chọn Analysis Services Multidimensional Project và chọn Next



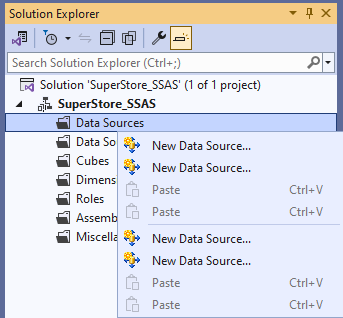
Hình 24: Tạo project SSAS và thiết lập kết nối (Bước 2)

Bước 3: Đặt tên và địa điểm lưu project sau đó chọn Create.



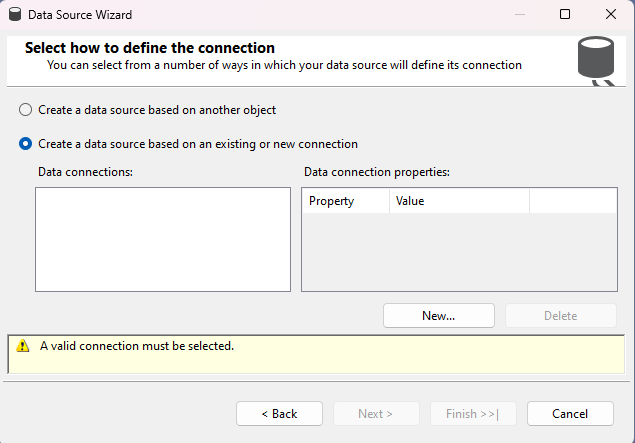
Hình 25: Tạo project SSAS và thiết lập kết nối (Bước 3)

Bước 4: Tại Solution Explorer, chuột phải vào Data Sources → chọn New Data Source



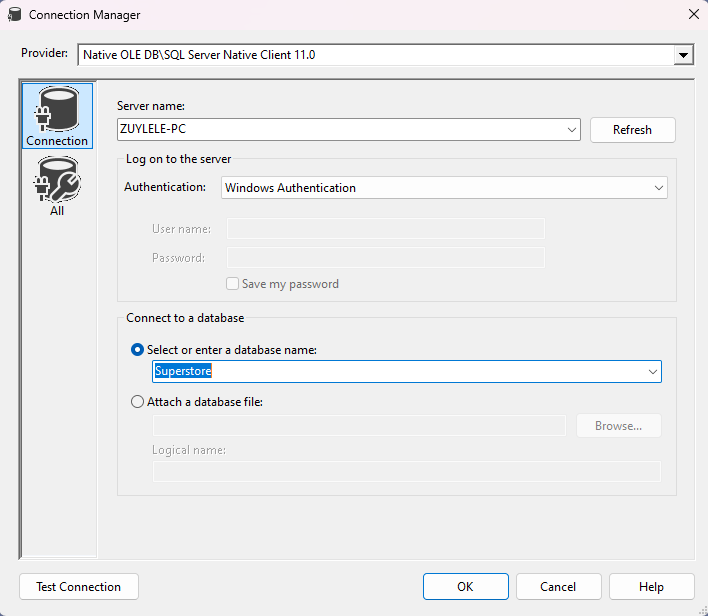
Hình 26: Tạo project SSAS và thiết lập kết nối (Bước 4)

Bước 5: Chọn Create a data source based on an existing or new connection sau đó chọn New…



Hình 27: Tạo project SSAS và thiết lập kết nối (Bước 5)

Bước 6: Nhập tên Server và chọn Database mới được tạo ở Chương 2 sau đó chọn OK rồi chọn Next.



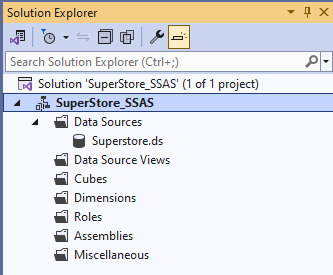
Hình 28: Tạo project SSAS và thiết lập kết nối (Bước 6)

Bước 7: Nhập tên người dùng và mật khẩu của tài khoản trên máy sau đó chọn Next sau đó chọn Finish.



Hình 29: Tạo project SSAS và thiết lập kết nối (Bước 7)

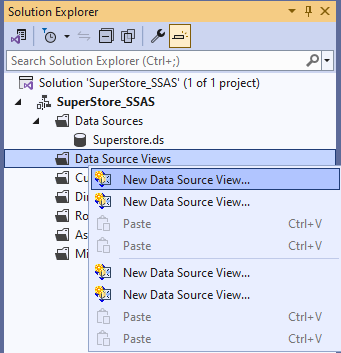
Bước 8: Đã kết nối thành công đến cơ sở dữ liệu.



Hình 30: Tạo project SSAS và thiết lập kết nối (Bước 8)

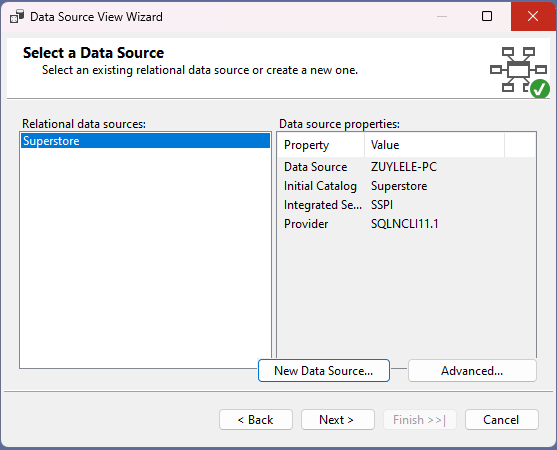
### 3.2. Xây dựng khối dữ liệu

Bước 1: Tại tab Solution Explorer, chuột phải vào Data Source Views và chọn New Data Source View.



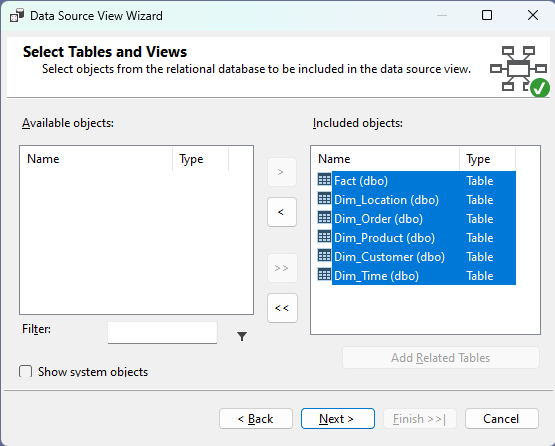
Hình 31: Xây dựng khối dữ liệu (Bước 1)

Bước 2: Chọn Data Source mới được thêm và chọn Next.



Hình 32: Xây dựng khối dữ liệu (Bước 2)

Bước 3: Thêm các bảng cần thiết vào Included objects sau đó chọn Next → Finish.



Hình 33: Xây dựng khối dữ liệu (Bước 3)

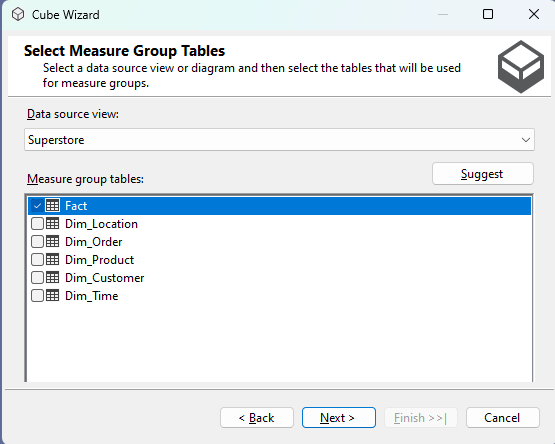
Bước 4: Tại tab Solution Explorer, chuột phải vào Cubes → New Cubes.

A screenshot of a computer

Description automatically generated

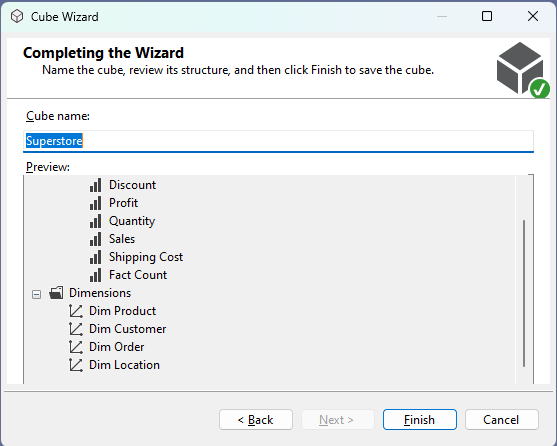
Hình 34: Xây dựng khối dữ liệu (Bước 4)

Bước 5: Chọn Next, chọn vào bảng Fact và chọn Next.



Hình 35: Xây dựng khối dữ liệu (Bước5 )

Bước 6: Tiếp tục chọn Next sau đó chọn Finish.



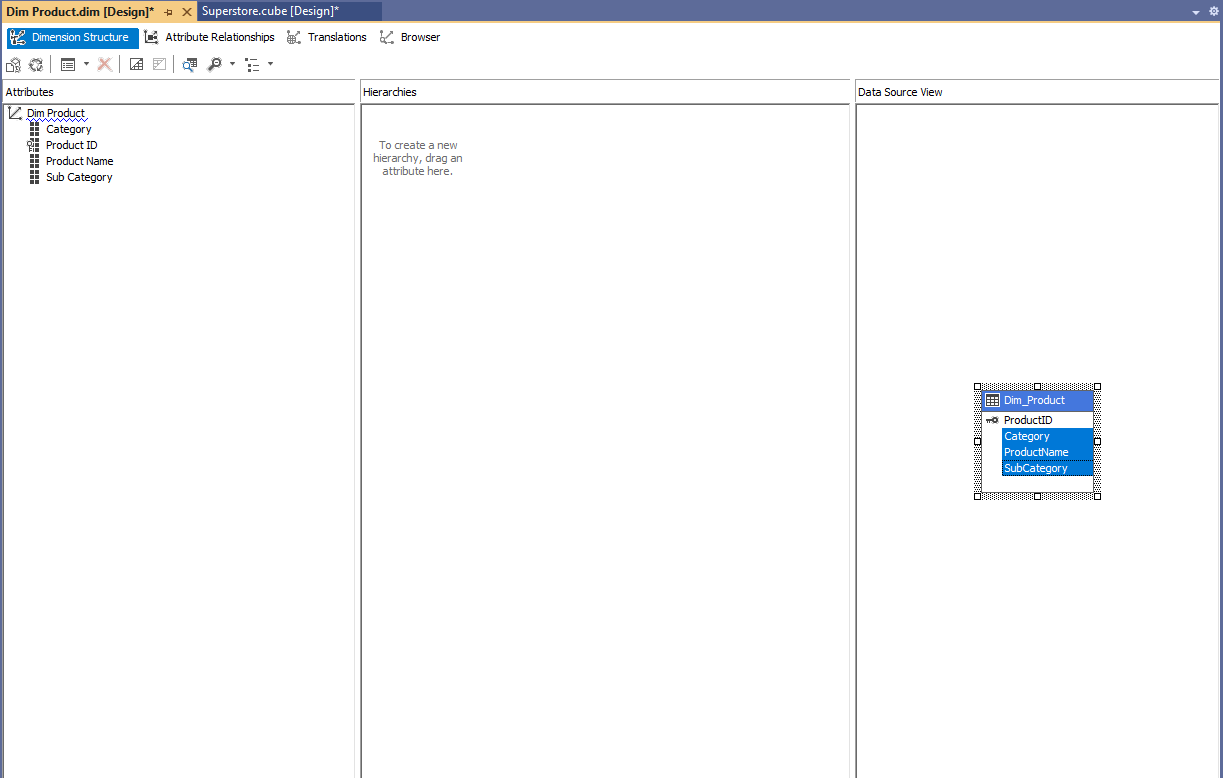
Hình 36: Xây dựng khối dữ liệu (Bước 6)

A screenshot of a computer

Description automatically generatedBước 7: Nhấn đúp vào các dimension tại tab Solution Explorer.

Hình 37: Xây dựng khối dữ liệu (Bước 7)

Bước 8: Kéo các cột còn thiếu từ Data Source View vào Attributes ( Làm vậy với tất cả các mục trong Dimension.



Hình 38: Xây dựng khối dữ liệu (Bước 8)

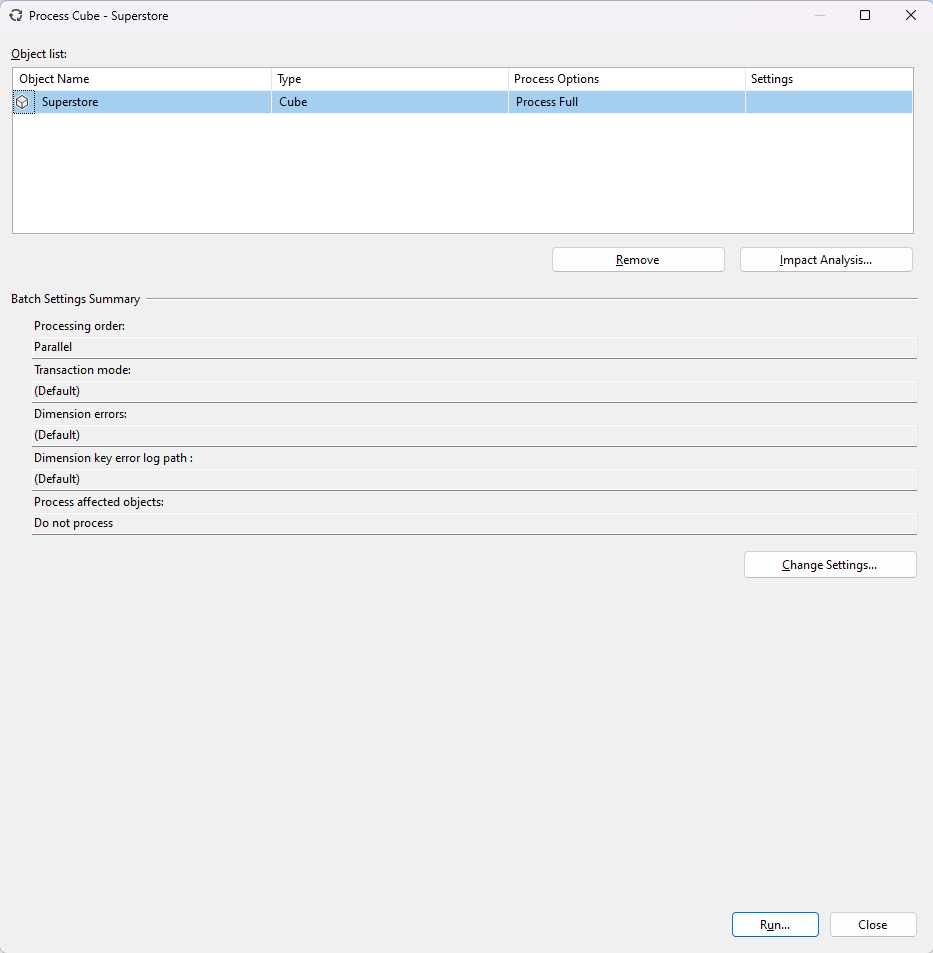
Bước 9: Sau khi đã lưu các thao tác trước chuột phải vào Cubes đã tạo tại tab Solution Explorer và chọn Process.

A screenshot of a computer

Description automatically generated

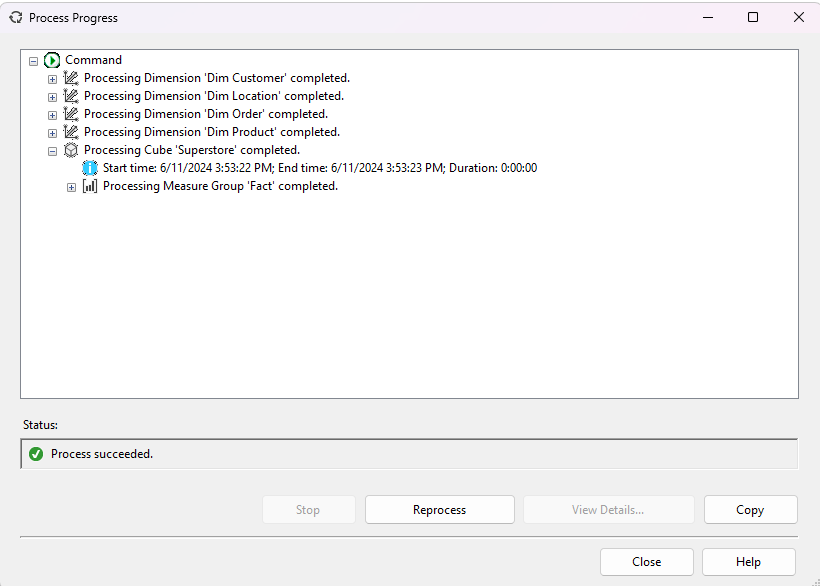
Hình 39: Xây dựng khối dữ liệu (Bước 9)

Bước 10: Sau khi Chọn OK thì chọn Run…



Hình 40: Xây dựng khối dữ liệu (Bước 10)

Bước 11: Xây dựng khối dữ liệu thành công



Hình 41: Xây dựng khối dữ liệu (Bước 11)

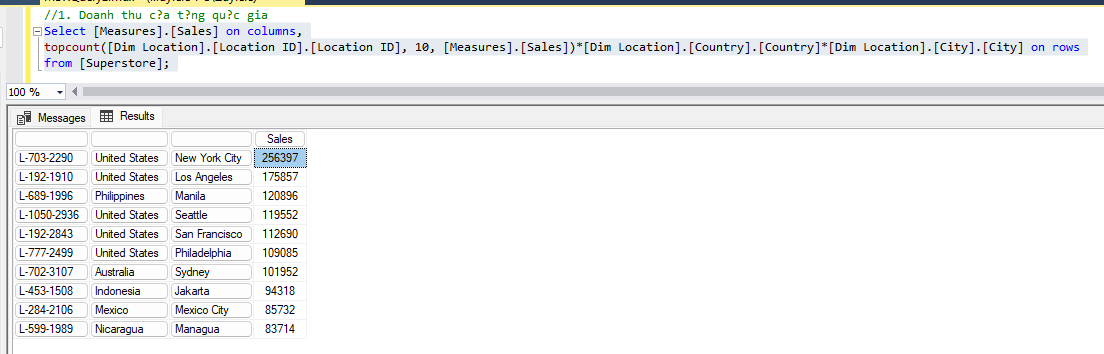
**![A white background with black dots

Description automatically generated](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAEwAAABxCAYAAAB/c0VYAAAAAXNSR0IArs4c6QAAAARnQU1BAACxjwv8YQUAAAAJcEhZcwAADsMAAA7DAcdvqGQAAAEBSURBVHhe7dCxAYAwDMCw0P9/BoYeEO/S4t3P+xvWzi1LhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYZFhkWGRYYlMx92JATeoapDAAAAAABJRU5ErkJggg==)**

### 3.3. Truy vấn dữ liệu

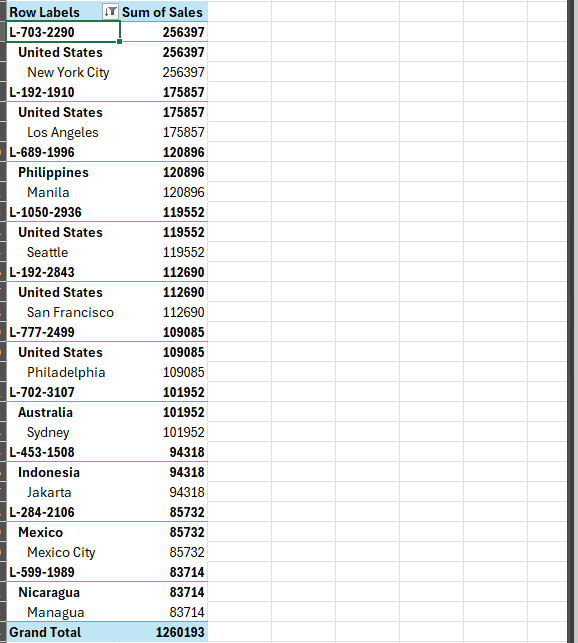
**Câu 1.** Top 10 thành phố có doanh thu lớn nhất thuộc về các đất nước nào

* Sử dụng MDX



Hình 42: Truy vấn dữ liệu sử dụng MDX (Câu 1)

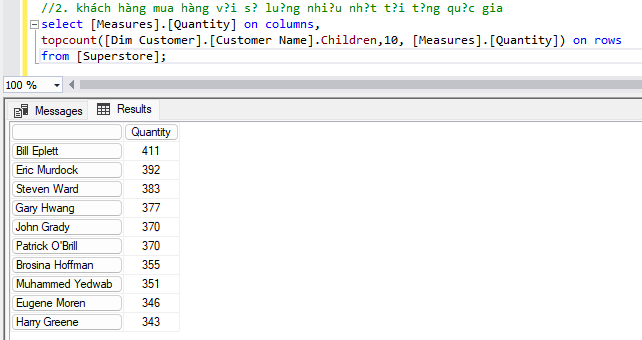
* Sử dụng pivot table



Hình 43: Truy vấn dữ liệu sử dung Pivot Table (Câu 1)

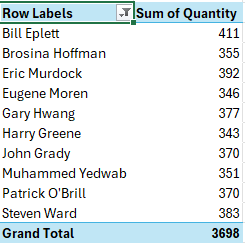
Câu 2. Top 10 khách hàng có đặt nhiều hàng nhất

* Sử dụng MDX



Hình 44: Truy vấn dữ liệu sử dụng MDX (Câu 2)

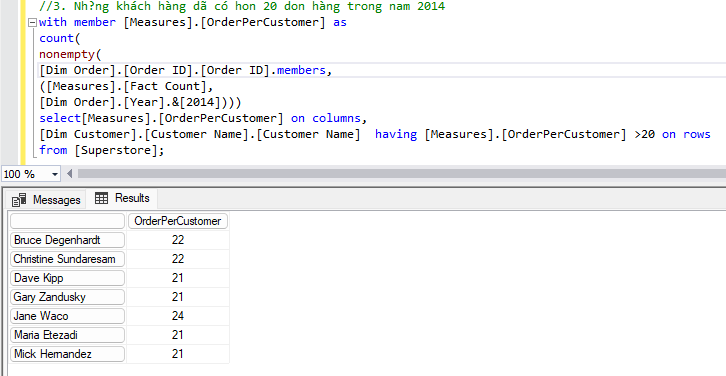
* Sử dụng pivot table



Hình 45: Truy vấn dữ liệu sử dung Pivot Table (Câu 2)

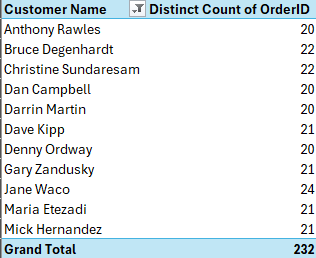
Câu 3. Những khách hàng có hơn 20 đơn hàng trong năm 2014.

* Sử dụng MDX



Hình 46: Truy vấn dữ liệu sử dụng MDX (Câu 3)

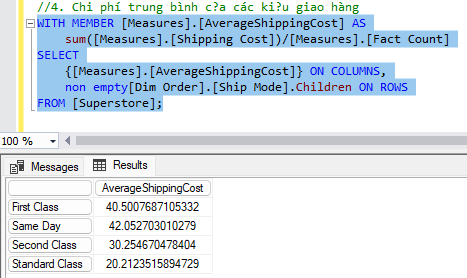
* Sử dụng pivot table



Hình 47: Truy vấn dữ liệu sử dung Pivot Table (Câu 3)

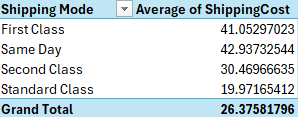
Câu 4. Chi phí giao hàng trung bình của các phương thức giao hàng.

* Sử dụng MDX



Hình 48: Truy vấn dữ liệu sử dụng MDX (Câu 4)

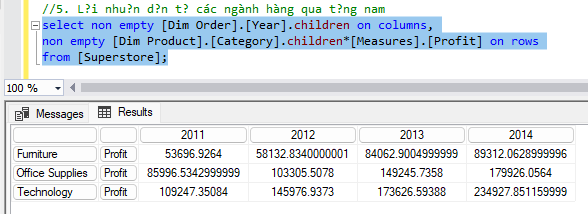
* Sử dụng pivot table



Hình 49: Truy vấn dữ liệu sử dung Pivot Table (Câu 4)

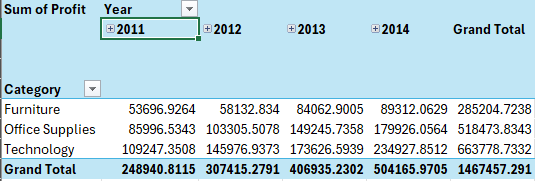
Câu 5. Lợi nhuận của của các nghành hàng qua các năm.

* Sử dụng MDX



Hình 50: Truy vấn dữ liệu sử dụng MDX (Câu 5)

* Sử dụng pivot table

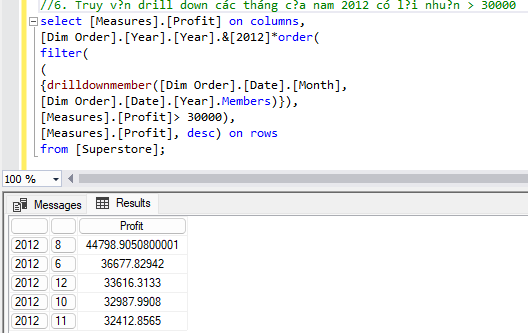


Hình 51: Truy vấn dữ liệu sử dung Pivot Table (Câu 5)

Câu 6. Truy vấn drilldown nhưng tháng có lợi nhuận lớn hơn 40000 đô trong năm

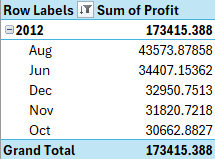
2014.

* Sử dụng MDX



Hình 52: Truy vấn dữ liệu sử dụng MDX (Câu 6)

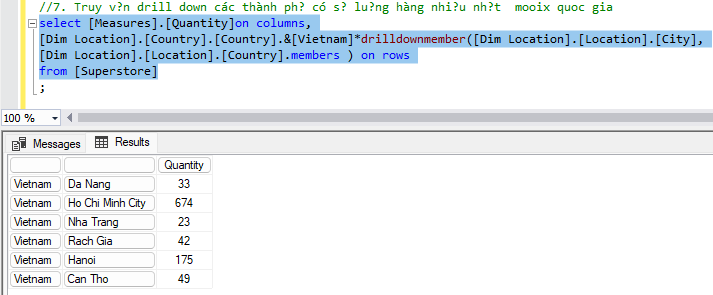
* Sử dụng pivot table



Hình 53: Truy vấn dữ liệu sử dung Pivot Table (Câu 6)

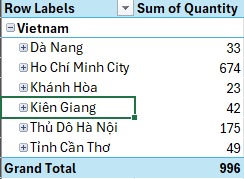
Câu 7. Truy vấn drilldown số lượng hàng hóa tại các tỉnh/ thành phố của Việt Nam

* Sử dụng MDX



Hình 54: Truy vấn dữ liệu sử dụng MDX (Câu 7)

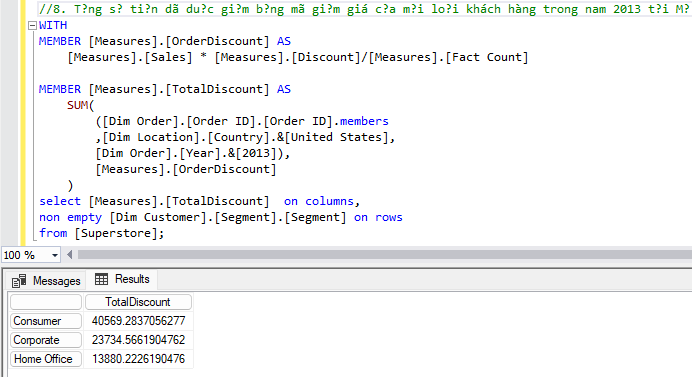
* Sử dụng pivot table



Hình 55: Truy vấn dữ liệu sử dung Pivot Table (Câu 7)

Câu 8. Tổng số tiền đã được trừ thông qua giảm giá theo các phân loại khách hàng tại Mỹ năm 2013.

* Sử dụng MDX



Hình 56: Truy vấn dữ liệu sử dụng MDX (Câu 8)

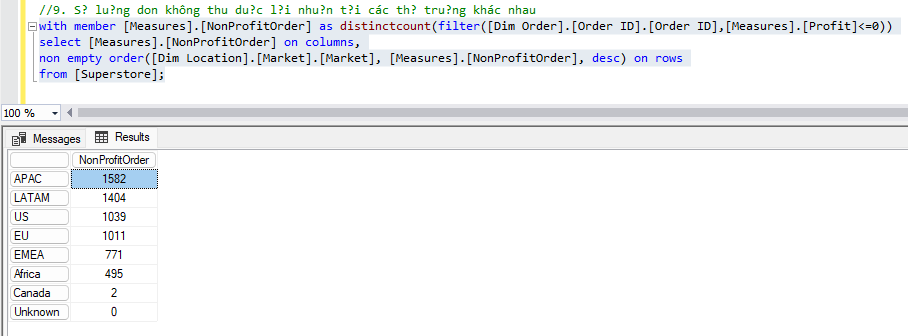
* Sử dụng pivot table



Hình 57: Truy vấn dữ liệu sử dung Pivot Table (Câu 8)

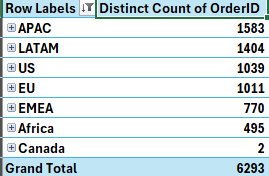
Câu 9. Số lượng đơn hàng không thu về lợi nhuận tại các thị trường khác nhau.

* Sử dụng MDX



Hình 58: Truy vấn dữ liệu sử dụng MDX (Câu 9)

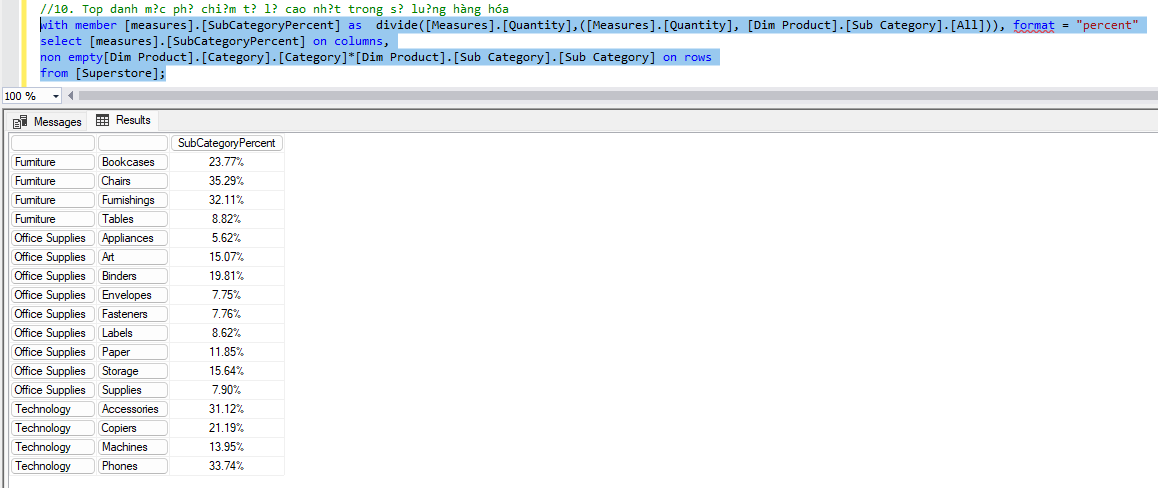
* Sử dụng pivot table



Hình 59: Truy vấn dữ liệu sử dung Pivot Table (Câu 9)

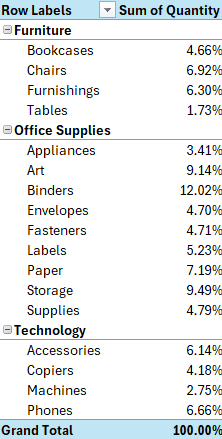
Câu 10. Tỉ lệ hàng hóa thuộc các nghành hàng .

* Sử dụng MDX



Hình 60: Truy vấn dữ liệu sử dụng MDX (Câu 10)

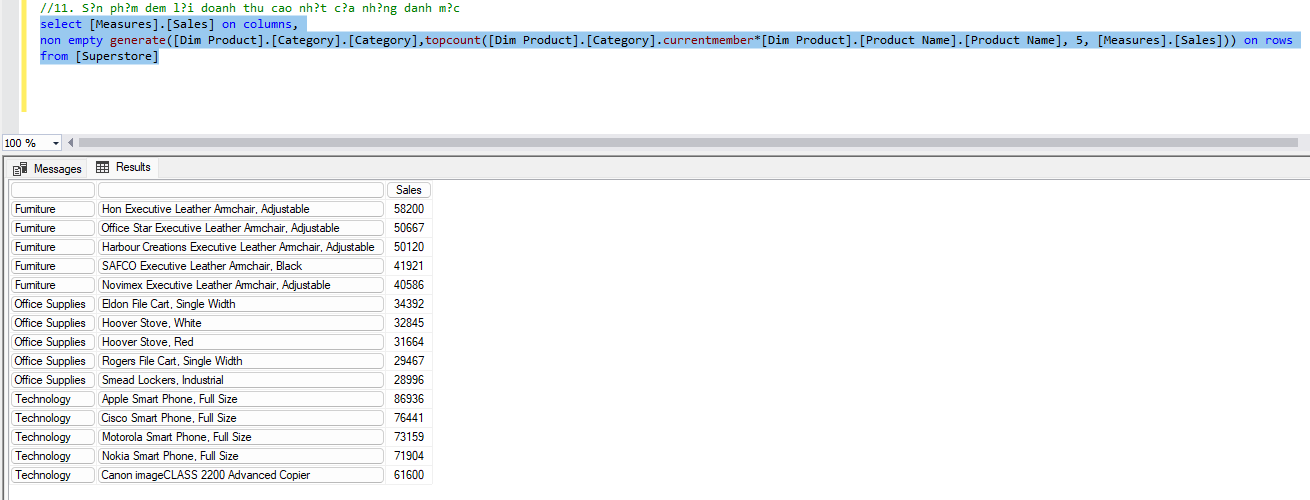
* Sử dụng pivot table



Hình 61: Truy vấn dữ liệu sử dung Pivot Table (Câu 10)

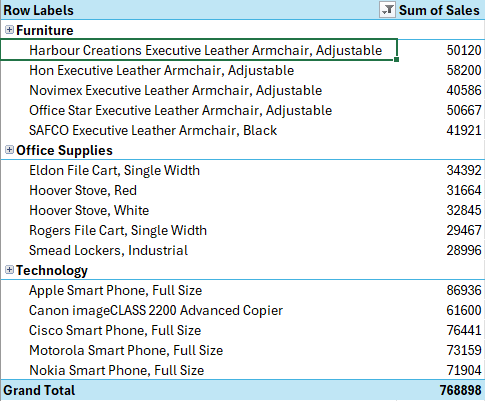
Câu 11. Top 5 những sản phẩm có doanh thu cao nhất trong từng nghành hàng.

* Sử dụng MDX



Hình 62: Truy vấn dữ liệu sử dụng MDX (Câu 11)

* Sử dụng pivot table

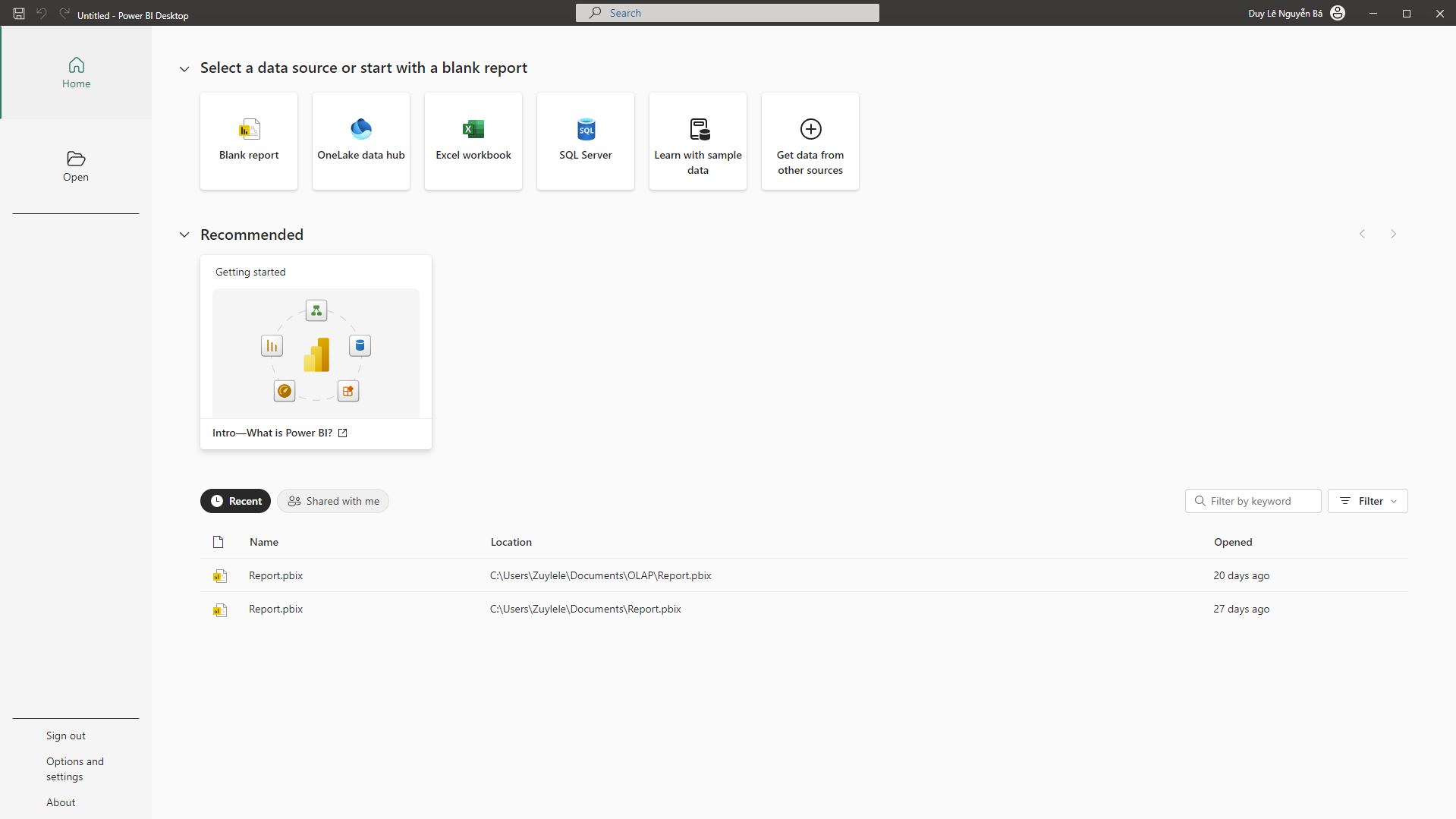


Hình 63: Truy vấn dữ liệu sử dung Pivot Table (Câu 11)

## Chương 4: Report với Power BI

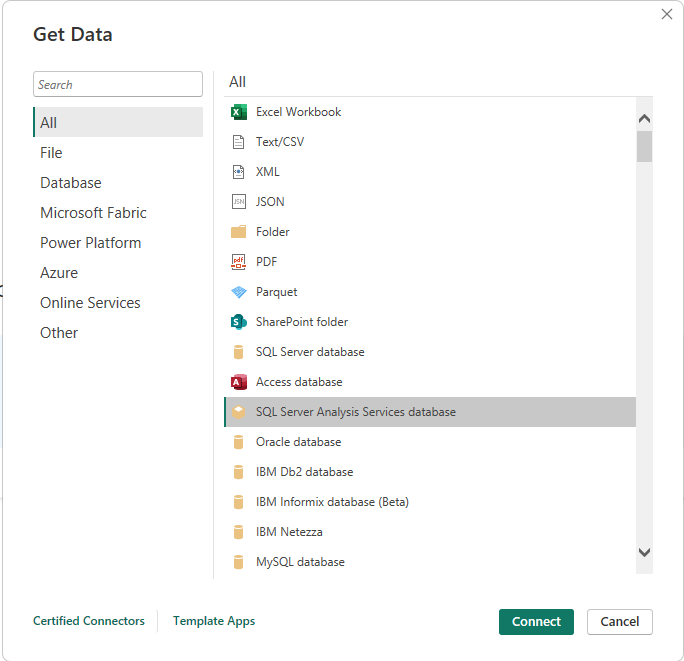
### 4.1. Tạo file PowerBI và thiết lập kết nối

Bước 1: Khởi động PowerBI và chọn Get data from other sources



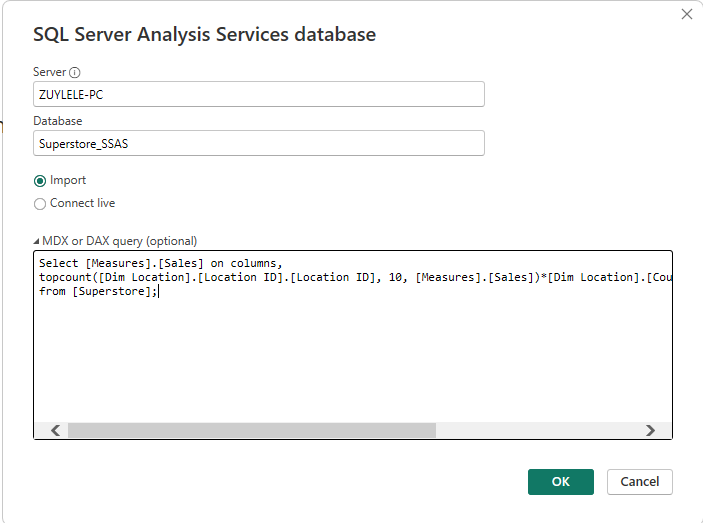
Hình 64: Tạo file PowerBI và thiết lập kết nối (Bước 1)

Bước 2: Chọn SQL Server Analysis Services database sau đó nhấn ok



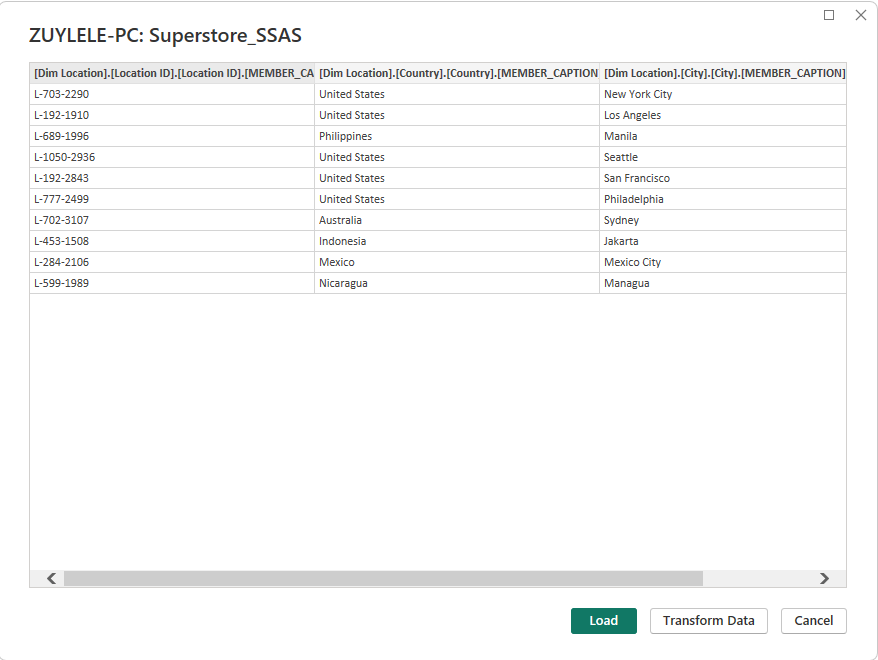
Hình 65: Tạo file PowerBI và thiết lập kết nối (Bước 2)

Bước 3: Nhập tên server, tên cơ sở dữ liệu chọn trường Import, nhập câu truy vấn đã được sử dụng ở chương 3 sau đó chọn OK.



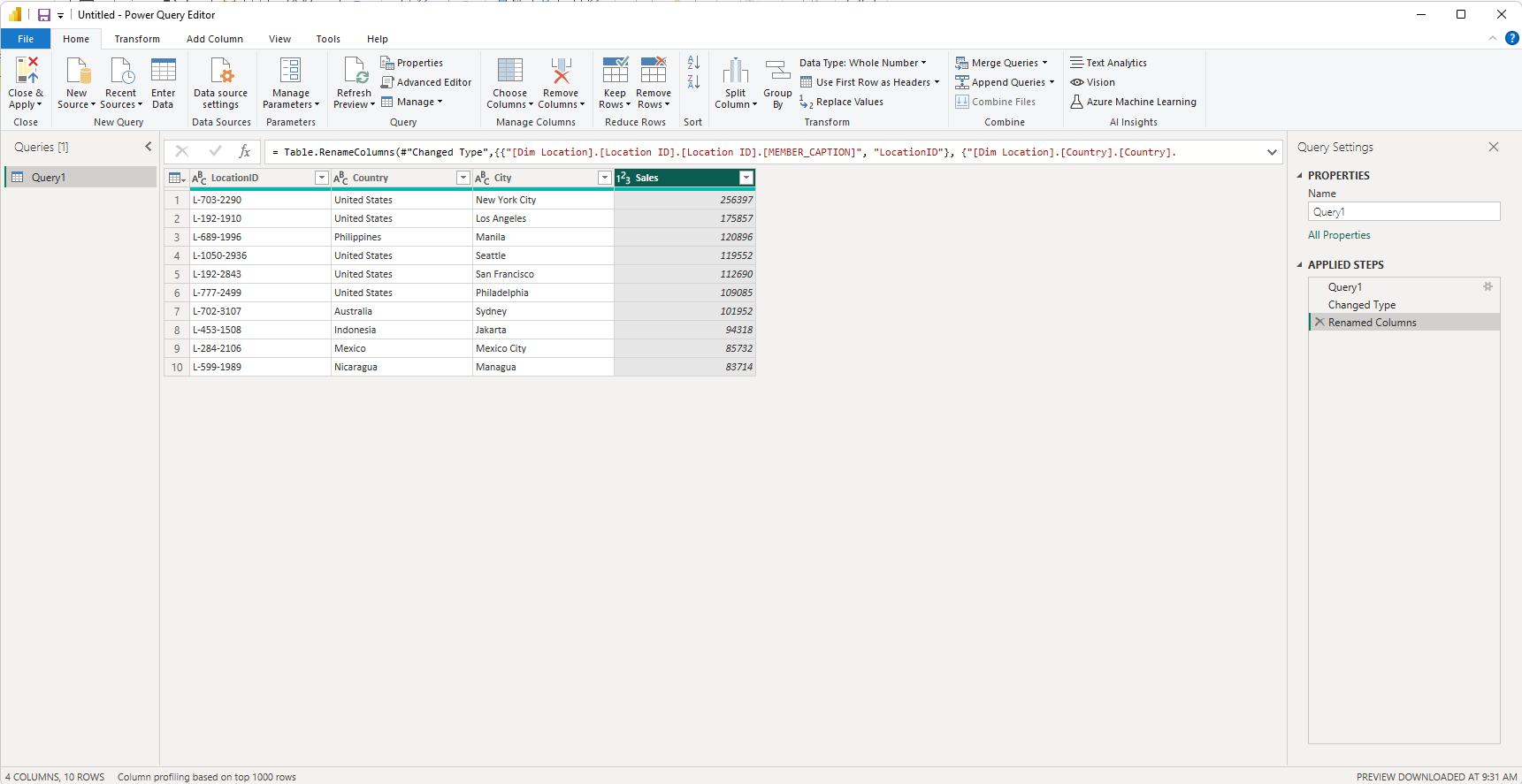
Hình 66: Tạo file PowerBI và thiết lập kết nối (Bước 3)

Bước 4: Tại đây chọn Transform Data



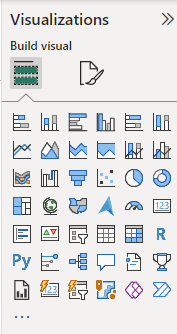
Hình 67: Tạo file PowerBI và thiết lập kết nối (Bước 4)

Bước 5: Tiến hành thực hiện một số biến đổi như đổi tên cột, đổi kiểu dữ liệu phù hợp sau đó chọn Close&apply



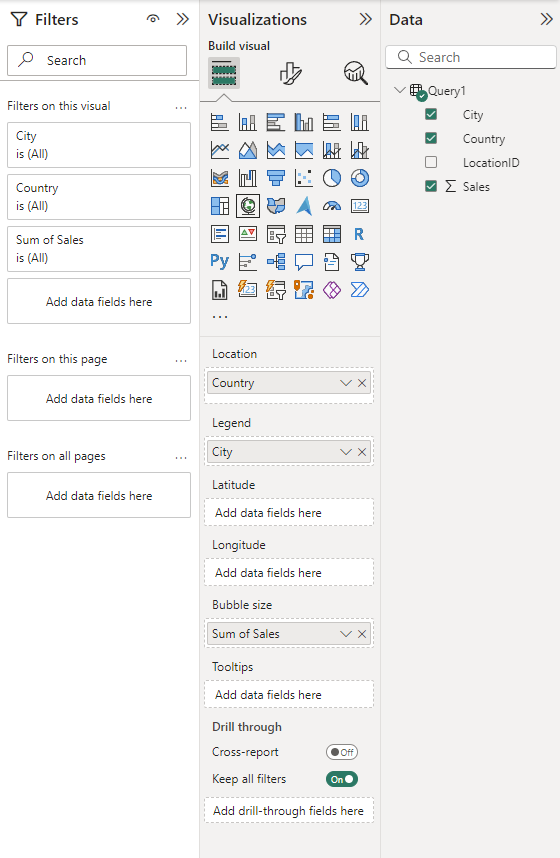
Hình 68: Tạo file PowerBI và thiết lập kết nối (Bước 5)

Bước 6: Tại mục Visualizations, chọn một biểu đồ phù hợp với dữ liệu.



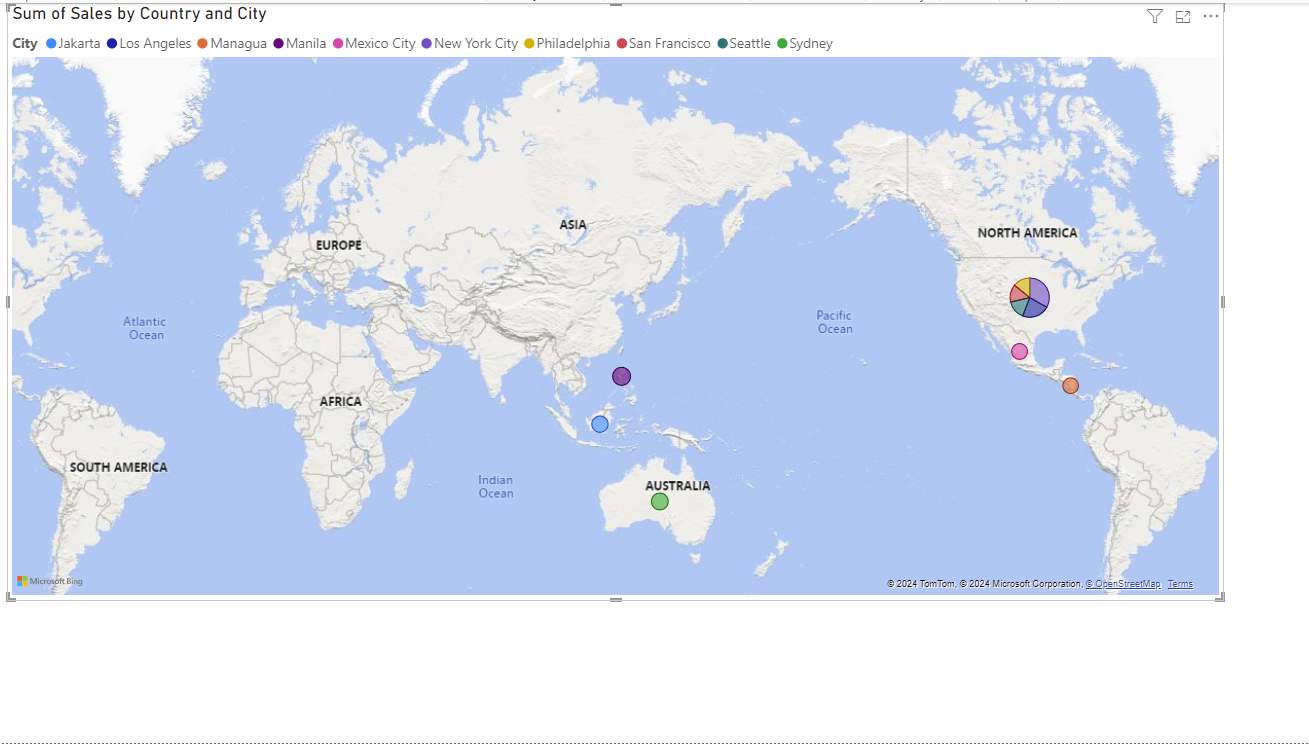
Hình 69: Tạo file PowerBI và thiết lập kết nối (Bước 6)

Bước 7: Thêm các mục dữ liệu từ tab Data



Hình 70: Tạo file PowerBI và thiết lập kết nối (Bước 7)

Bước 8: Biểu đồ sẽ được hiển thị ở phần bên trái



Hình 71: Tạo file PowerBI và thiết lập kết nối (Bước 8)

### 4.2. Các truy vấn được đồ thị hóa trên PowerBI

A screenshot of a computer

Description automatically generated

Hình 72: Truy vấn 1-5 trên PowerBI

**A screenshot of a graph

Description automatically generated**

Hình 73: Truy vấn 6-9 trên PowerBI

**A screenshot of a computer

Description automatically generated**

Hình 74: Truy vấn 10-11 trên PowerBI

## CHƯƠNG 5: Data Mining

### 5.1. Tiền xử lý dữ liệu và EDA

Bước 1: Loại bỏ các biến không cần thiết



Hình 75: Tiền xử lý dữ liệu và EDA (Bước 1)

Bước 2: Chuyển đổi dữ liệu thời gian sang DateTime

A black screen with white text

Description automatically generated

Hình 76: Tiền xử lý dữ liệu và EDA (Bước 2)

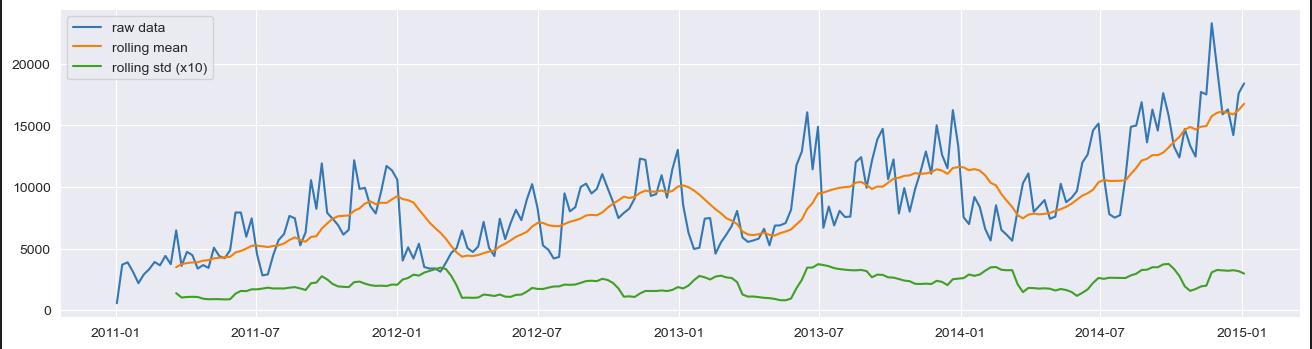
Bước 3: Kiểm tra xu hướng, tính mùa vụ của dữ liệu dựa trên thời gian.

A graph of a graph of a graph

Description automatically generated with medium confidence

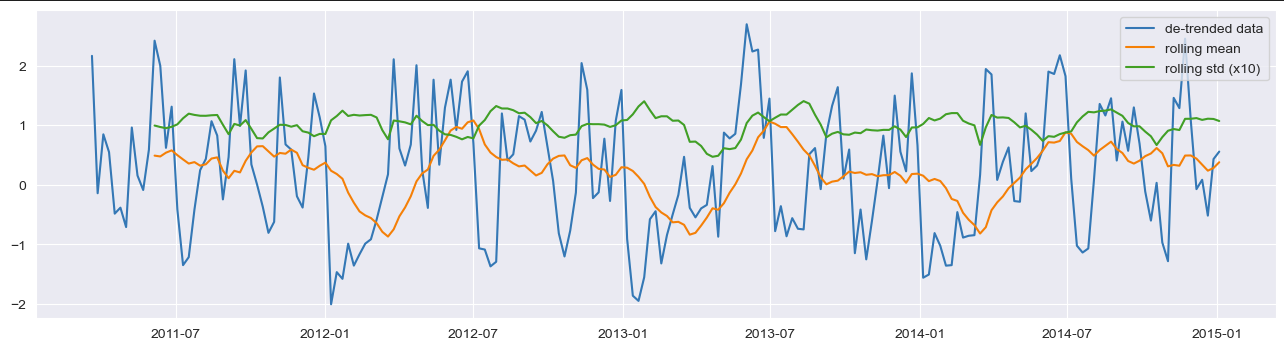
Hình 77: Tiền xử lý dữ liệu và EDA (Bước 3)

Bước 4: Kiểm tra tính dừng (stationary) của bộ dữ liệu (Bộ dữ liệu vẫn chưa có tính dừng)



Hình 78: Tiền xử lý dữ liệu và EDA (Bước 4)

Bước 5: Chuyển đổi dữ liệu thành dữ liệu có tính dừng.



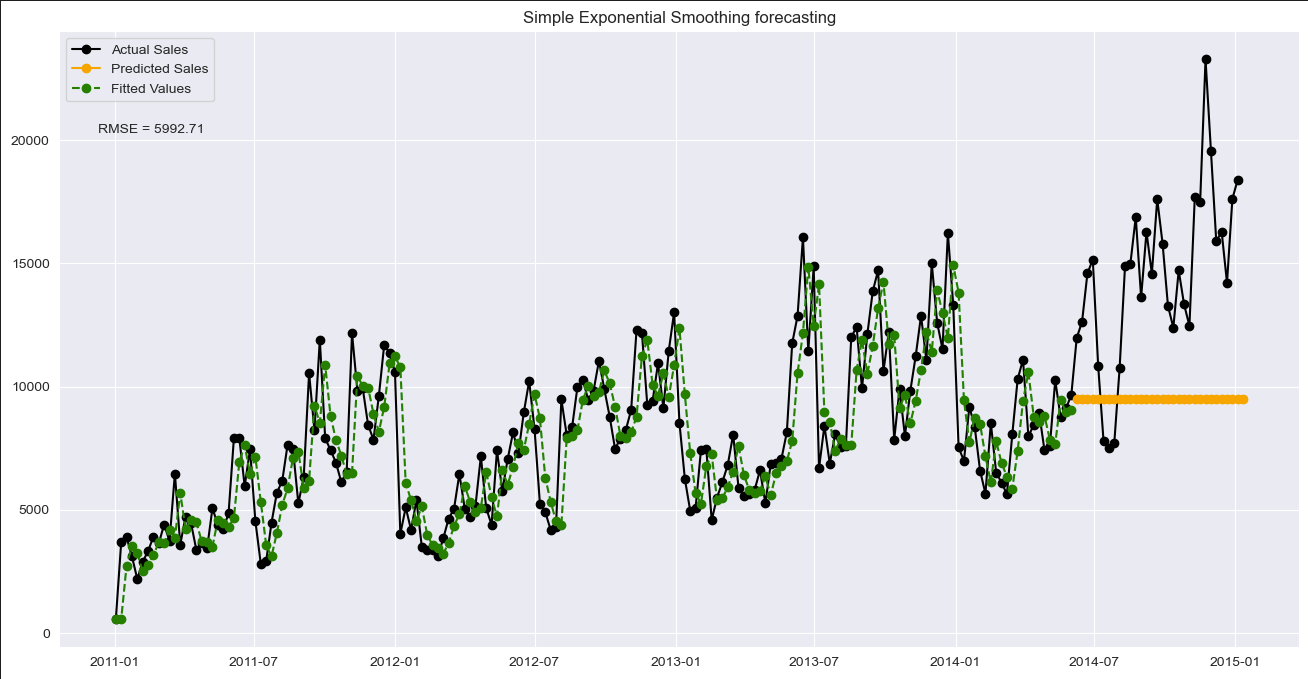
Hình 79: Tiền xử lý dữ liệu và EDA (Bước 5)

### 5.2. Áp dụng thuật toán và đánh giá

Đối với bộ dữ liệu về doanh số bán hàng em đã quyết định sử dụng chủ để dự đoán doanh thu dựa trên thời gian với 4 thuật toán.

#### Simple Exponential Smoothing (SES)

* Kết quả:

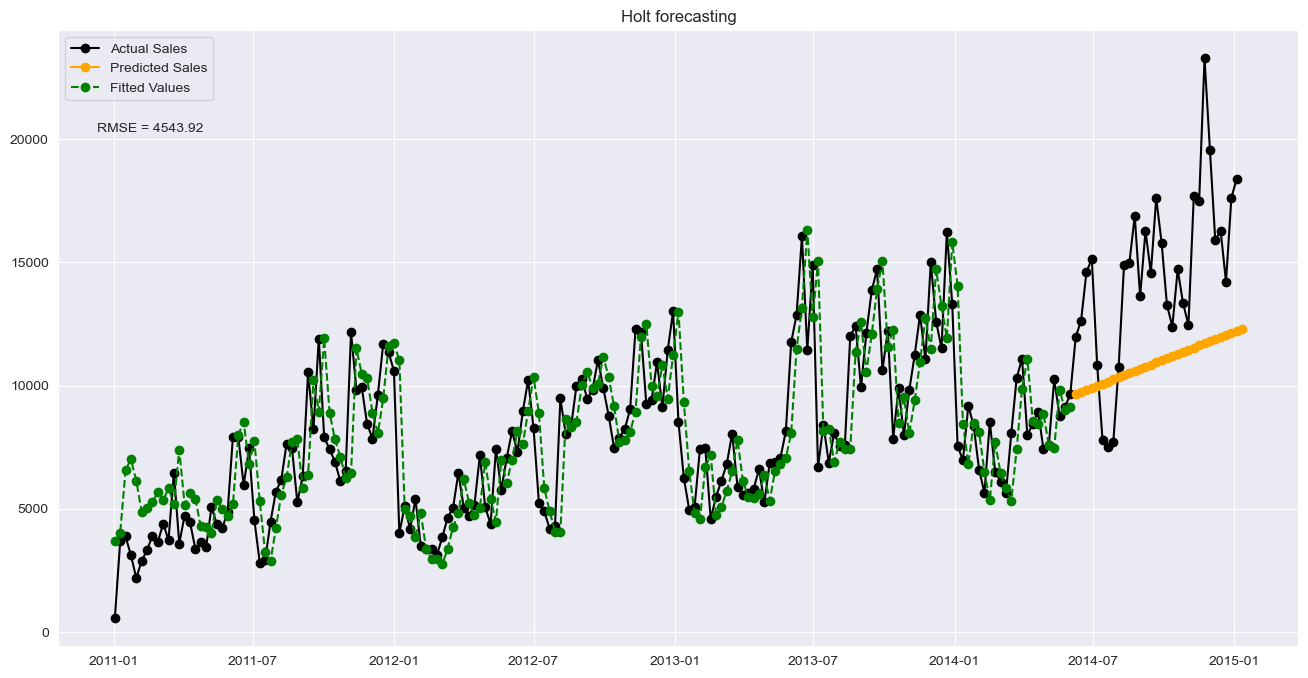


Hình 80: Kết quả của thuật toán SES

* Nhận xét: Thuật toán này chủ yếu sẽ dự đoán dựa trên tổng trọng lượng của của dãy dữ liệu để dự đoán mà không bao gồm xu hướng và tính lặp lại của bộ dữ liệu vì vậy đối với bộ dữ liệu trên SES tỏ ra kém hiệu quả khi chỉ đưa ra được những con số trung bình của bộ huấn luyện.

#### Holt

* Kết quả:

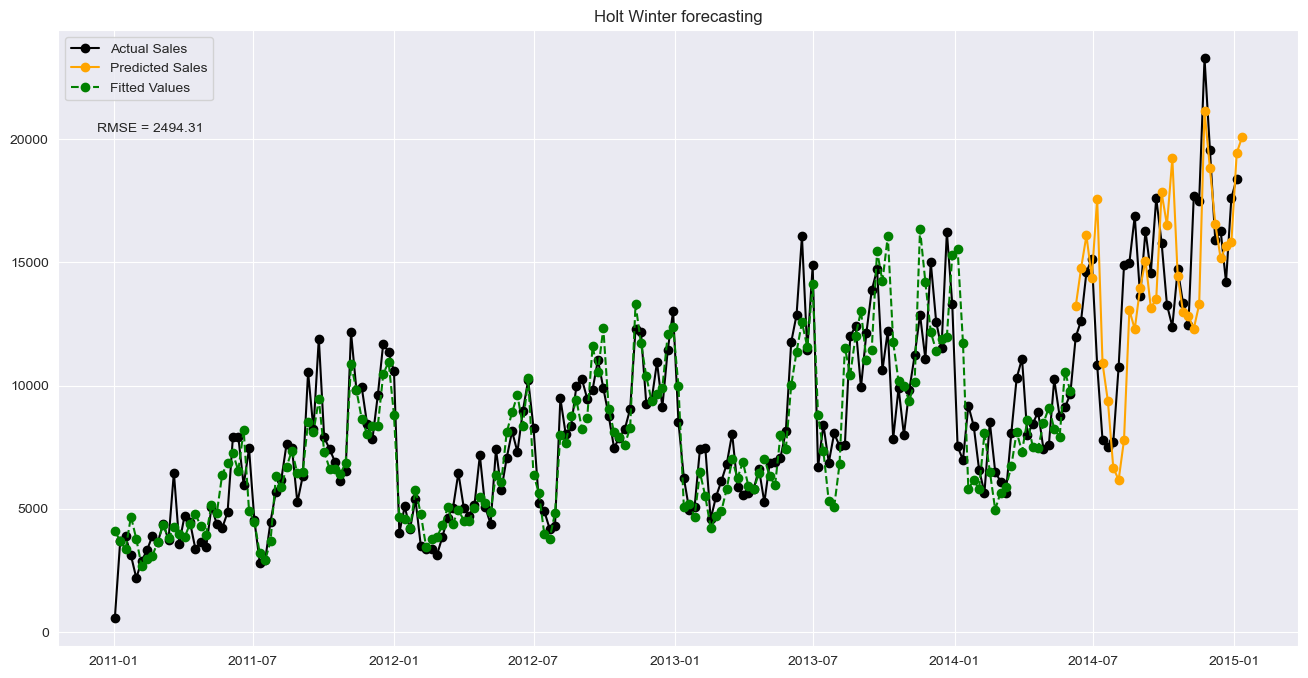


Hình 81: Kết quả thuật toán Holt

* Nhận xét: Đối với thuật toán Holt là một biến thể của SES mà trong đó xung hướng của dữ liệu đã được bao gồm và từ đó có thể thấy rằng kết quả dự đoán đã thể hiện được xu hướng của bộ dữ liệu nhưng những giá trị đó vẫn chưa thực sự có các dự đoán đúng với thực tế.

#### Holt’s Winter

* Kết quả:

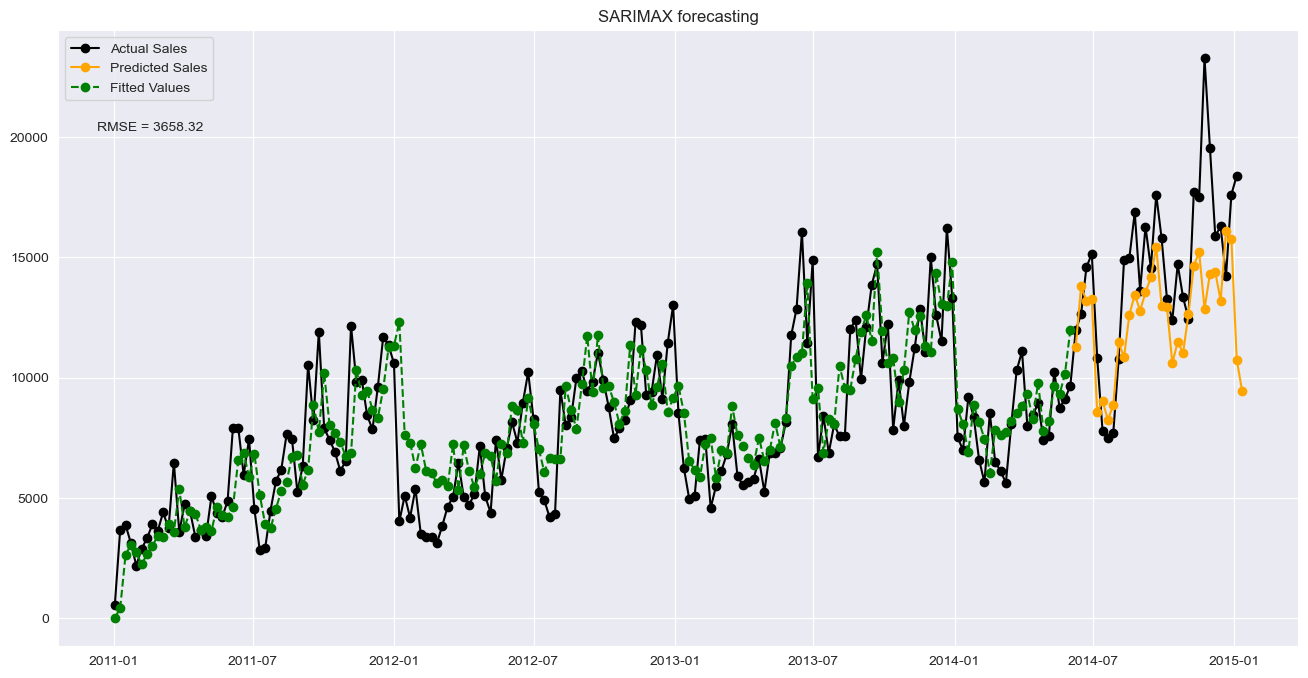


Hình 82: Kết quả thuật toán Holt’s Winter

* Nhận xét: Khi sử dụng thuật toán Holt’s Winter, nó đã có thể dự đoán tương đối chính xác những doanh thu trong bộ test đó là vì đối với Holt’s Winter, thuật toán nãy đã bao gồm tính mùa vụ khi đã đọc được sự lặp lại của dữ liệu qua các năm.

#### SARIMAX

* Kết quả:



Hình 83: Kết quả thuật toán SARIMAX

* Nhận xét: Ngoài ra em cũng đã sử dụng một thuật toán khác cũng sẽ đọc được tính mùa vụ và xu hướng và đã cho ra được kết quả khả quan.

#### Đánh giá

* Đối với mô hình:

|  |  |
| --- | --- |
| **Mô hình** | **RMSE** |
| SES | 5992.71 |
| Holt | 4543.92 |
| Holt’s Winter | 2494.31 |
| Sarimax | 3658.32 |

Thông qua bảng trên thì có thể thấy được 2 thuật toán có bao gồm tính mùa vụ và xu hướng tỏ ra vượt trội hơn hẳn so với 2 thuật toán còn lại. Đổi với thuật toán Holt’s Winter với việc đặt trọng số lớn hơn đối với các dữ liệu gần nhất đã tỏ ra phù hợp với bộ dữ liệu này hơn so với Sarimax

* Đối với bộ dữ liệu:

Thông qua quá trình EDA và Data mining, em đã có thể thấy được rằng doanh thu của chuổi cửa hàng này có xu hướng tăng lên theo từng năm cùng với đó có thể thấy được tính mùa vụ của doanh thu trong thời gian 1 năm khi mà sẽ có xu hướng tăng nhẹ từ tháng 1 đến tháng 6 sau đó sẽ có một sự đi xuống đáng kể tại các thời điểm tháng 8 nhưng lại tăng vọt vào khoảng thời gian cuối năm.