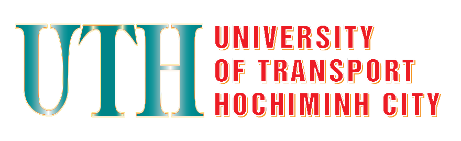
TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI TP. HỒ CHÍ MINH

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ ĐIỆN, ĐIỆN TỬ**



**ĐỒ ÁN MÔN HỌC**

**KHO DỮ LIỆU VÀ HỆ THỐNG RA QUYẾT ĐỊNH**

**TÊN ĐỀ TÀI**

**PHÂN TÍCH TÂM TRẠNG KHÁCH HÀNG ĐỂ ĐÁNH GIÁ THƯƠNG HIỆU**

**GVHD:** Ths. Lê Huỳnh Long

**Nhóm SVTH:**

072305004043\_Ngô Thị Kim Ngân

066305011537\_Nguyễn Lê Quỳnh

054305000332\_Nguyễn Thị Huyền Trân

079305001883\_Huỳnh Thị Ngọc Loan

075305019523\_Nguyễn Thị Kim Ngọc

Tp. Hồ Chí Minh, ngày 23 tháng 11 năm 2025

**NHẬN XÉT CỦA GIẢNG VIÊN**

*Nhận xét*:

Chữ ký của giảng viên

(Ký tên)

**Lê Huỳnh Long**

**BẢNG PHÂN CÔNG CÔNG VIỆC**

|  |  |  |
| --- | --- | --- |
| Họ và tên | Công việc | Mức độ hoàn thành công việc (%) |
| Ngô Thị Kim Ngân | Trưởng nhóm, Chương 2.2, Chương 3.4 & 3.5, Dashboard 1 & Dashboard 4 | 100 |
| Nguyễn Lê Quỳnh | Chương 2.1, Trình bày báo cáo | 75 |
| Nguyễn Thị Huyền Trân | Chương 1.1 + Trình bày báo cáo, Chương 3.2.5-3.3.3, 4.3 & 4.4 | 100 |
| Huỳnh Thị Ngọc Loan | Chương 1.2, 3.3.4-3.3.5, Dashboard 2 | 100 |
| Nguyễn Thị Kim Ngọc | Chương 1.3, 3.1 & 3.2.1-3.2.4 , 4.1 | 100 |

**MỤC LỤC**

[LỜI MỞ ĐẦU 1](#_Toc215040903)

[LỜI CẢM ƠN 2](#_Toc215040904)

[CHƯƠNG 1: TỔNG QUAN 3](#_Toc215040905)

[1.1  Giới thiệu về data warehouse (dw) và business intelligence (bi) 3](#_Toc215040906)

[1.1.1  Data warehouse 3](#_Toc215040907)

[1.1.2 Business intelligence 3](#_Toc215040908)

[1.2 Các đặc trưng về mặt dữ liệu của một hệ thống DW và BI 5](#_Toc215040909)

[1.2.1 Các đặc trưng về mặt dữ liệu của một hệ thống DW. 5](#_Toc215040910)

[1.2.2 Các đặc trưng về mặt dữ liệu của một hệ thống BI. 6](#_Toc215040911)

[1.3. CÁC PHƯƠNG PHÁP LUẬN XÂY DỰNG 7](#_Toc215040912)

[1.3.1. Phương pháp luận Top-down 7](#_Toc215040913)

[1.3.2. Phương pháp luận Bottom-up 8](#_Toc215040914)

[1.3.3. Phương pháp luận Spiral 11](#_Toc215040915)

[1.3.4. Cơ sở lựa chọn phương pháp 12](#_Toc215040916)

[CHƯƠNG 2: PHÂN TÍCH THIẾT KẾ KHO DỮ LIỆU 15](#_Toc215040917)

[2.1 Giới thiệu về dữ liệu dataset 15](#_Toc215040918)

[2.1.1 Nguồn dữ liệu: 15](#_Toc215040919)

[2.1.2 Mô tả chi tiết dữ liệu: 15](#_Toc215040920)

[2.1.3 Chuẩn bị dữ liệu: 16](#_Toc215040921)

[2.2 Thiết kế kho dữ liệu 24](#_Toc215040922)

[2.2.1 Fact table 24](#_Toc215040923)

[2.2.2 Bảng Dimension Tables: 25](#_Toc215040924)

[CHƯƠNG 3: TÍCH HỢP DỮ LIỆU 27](#_Toc215040925)

[3.1 Giới thiệu SSIS: 27](#_Toc215040926)

[3.2 Quá trình đổ dữ liệu vào kho: 28](#_Toc215040927)

[3.2.1  Chuẩn bị các công cụ 28](#_Toc215040928)

[3.2.2  Chuẩn bị cơ sở dữ liệu 29](#_Toc215040929)

[3.2.3 Tạo mới project SSIS 35](#_Toc215040930)

[3.2.4 Tạo bảng Dim và bảng Fact 36](#_Toc215040931)

[3.2.5 Chạy dự án SSIS 55](#_Toc215040932)

[3.2.6 Kiểm tra dữ liệu các bảng 64](#_Toc215040933)

[3.2.7 Lược đồ sau khi hoàn thành 66](#_Toc215040934)

[3.3    Phân tích dữ liệu SSAS 67](#_Toc215040935)

[3.3.1 Xác định dữ liệu nguồn (Data Sources) 67](#_Toc215040936)

[3.3.2 Xác định khung nhìn dữ liệu nguồn (Data Source Views) 70](#_Toc215040937)

[3.3.4  Xây dựng các khối (Cubes) 74](#_Toc215040938)

[3.3.5  Xác định các chiều (Dimensions) 77](#_Toc215040939)

[3.4 Xác định độ đo (Measures) 79](#_Toc215040940)

[3.4.1 Thuộc tính độ đo ban đầu 79](#_Toc215040941)

[3.4.2 Tổng kết độ đo 80](#_Toc215040942)

[3.5 Phân cấp trong bảng chiều (Dimensions) 81](#_Toc215040943)

[3.5.1 Phân cấp bảng Dim\_Time 81](#_Toc215040944)

[3.5.2 Phân cấp bảng Dim\_Product 84](#_Toc215040945)

[CHƯƠNG 4: POWER BI TRỰC QUAN VÀ RA QUYẾT ĐỊNH 87](#_Toc215040946)

[4.1 Kết nối dữ liệu với POWER BI 87](#_Toc215040947)

[4.1.1 Bảng tổng lượt đánh giá theo từng năm của từng thương hiệu 92](#_Toc215040948)

[4.2  Xây dựng các biểu đồ trực quan 94](#_Toc215040949)

[4.2.1 Biểu đồ phân phối cảm xúc theo thương hiệu 94](#_Toc215040950)

[4.2.2 Biểu đồ xu hướng cảm xúc theo tháng 97](#_Toc215040951)

[4.2.3 Từ khóa phổ biến trong bài đánh giá của Maybelline 99](#_Toc215040952)

[4.2.4 Biểu đồ phân bố điểm đánh giá 101](#_Toc215040953)

[4.2.5 Xây dựng Card KPI 104](#_Toc215040954)

[4.2.6 Dashboard cảm xúc khách hàng 106](#_Toc215040955)

[4.2.7 Biểu đồ thể hiện sản phẩm có điểm đánh giá cao nhất 106](#_Toc215040956)

[4.2.8 Biểu đồ quan hệ giữa cảm xúc và điểm đánh giá 108](#_Toc215040957)

[4.2.9 Biểu đồ so sánh brand Maybelline và các brand còn lại 110](#_Toc215040958)

[4.2.10 Biểu đồ thể hiện mức độ quan tâm đến thương hiệu 112](#_Toc215040959)

[4.2.11 Biểu đồ nhóm sản phẩm ảnh hưởng lớn đến thương hiệu 113](#_Toc215040960)

[4.2.12  Dashboard Hiệu quả sản phẩm và thương hiệu 116](#_Toc215040961)

[4.2.13 Biểu đồ xu hướng chỉ số BHI theo thời gian 116](#_Toc215040962)

[4.2.14 Biểu đồ dự báo BHI 119](#_Toc215040963)

[4.2.15 Biểu đồ phân rã nguyên nhân gây sự biến động BHI 121](#_Toc215040964)

[4.2.15 Xây dựng KPI Card tổng hợp của Maybelline New York 123](#_Toc215040965)

[4.2.16 Dashboard sức khỏe thương hiệu và chiến lược 125](#_Toc215040966)

[4.3 Đánh giá kết quả 125](#_Toc215040967)

[4.4 Hướng phát triển trong tương lai 126](#_Toc215040968)

**DANH MỤC HÌNH ẢNH**

[Hình 2. 1 Làm sạch dữ liệu 16](#_Toc215040969)

[Hình 2. 2 Làm sạch dữ liệu 17](#_Toc215040970)

[Hình 2. 3 Làm sạch dữ liệu 18](#_Toc215040971)

[Hình 2. 4 Làm sạch dữ liệu 18](#_Toc215040972)

[Hình 2. 5 Làm sạch dữ liệu 19](#_Toc215040973)

[Hình 2. 6 Làm sạch dữ liệu 19](#_Toc215040974)

[Hình 2. 7 Làm sạch dữ liệu 20](#_Toc215040975)

[Hình 2. 8 Làm sạch dữ liệu 20](#_Toc215040976)

[Hình 2. 9 Làm sạch dữ liệu 21](#_Toc215040977)

[Hình 2. 10 Làm sạch dữ liệu 21](#_Toc215040978)

[Hình 2. 11 Làm sạch dữ liệu 21](#_Toc215040979)

[Hình 2. 12 Làm sạch dữ liệu 22](#_Toc215040980)

[Hình 2. 13 Làm sạch dữ liệu 22](#_Toc215040981)

[Hình 2. 14 Làm sạch dữ liệu 23](#_Toc215040982)

[Hình 2. 15 Làm sạch dữ liệu 23](#_Toc215040983)

[Hình 2. 16 Làm sạch dữ liệu 23](#_Toc215040984)

[Hình 3. 1 Mở SQL Server 29](#_Toc215040985)

[Hình 3. 2 Tạo cơ sở dữ liệu 29](#_Toc215040986)

[Hình 3. 3 Khởi tạo bảng: Dim\_Brand 30](#_Toc215040987)

[Hình 3. 4 Khởi tạo bảng: Dim\_Product 30](#_Toc215040988)

[Hình 3. 5 Khởi tạo bảng: Dim\_Customer 31](#_Toc215040989)

[Hình 3. 6 Khởi tạo bảng: Dim\_Time 31](#_Toc215040990)

[Hình 3. 7 Khởi tạo bảng: Fact\_review 32](#_Toc215040991)

[Hình 3. 8 Khởi tạo bảng: FactBrandSummary 33](#_Toc215040992)

[Hình 3. 9 Kết quả 34](#_Toc215040993)

[Hình 3. 10 Mở Visual Studio 2022 35](#_Toc215040994)

[Hình 3. 11 Tạo Project 35](#_Toc215040995)

[Hình 3. 12 Tạo các kết nối Connection Managers 36](#_Toc215040996)

[Hình 3. 13 Bảng Add SSIS Connection Manager 36](#_Toc215040997)

[Hình 3. 14 Bảng Excel Connection Manager 37](#_Toc215040998)

[Hình 3. 15 Kết quả 37](#_Toc215040999)

[Hình 3. 16 Bảng Connection Manager 38](#_Toc215041000)

[Hình 3. 17 Chọn cơ sở dữ liệu KhoDL 38](#_Toc215041001)

[Hình 3. 18 Chọn Test Connection 39](#_Toc215041002)

[Hình 3. 19 Bảng Control Flow 40](#_Toc215041003)

[Hình 3. 20 Tải dữ liệu từ sheet vào trong SQL Server 40](#_Toc215041004)

[Hình 3. 21 Kết quả 41](#_Toc215041005)

[Hình 3. 22 Cấu hình Excel Source 42](#_Toc215041006)

[Hình 3. 23 Cấu hình bảng Dim\_Brand 43](#_Toc215041007)

[Hình 3. 24 Cấu hình SQL Server Destination 44](#_Toc215041008)

[Hình 3. 25 Ánh xạ từ các cột 45](#_Toc215041009)

[Hình 3. 26 Cấu hình SQL Server Destination 46](#_Toc215041010)

[Hình 3. 27 Ánh xạ từ các cột 47](#_Toc215041011)

[Hình 3. 28 Cấu hình SQL Server Destination 48](#_Toc215041012)

[Hình 3. 29 Ánh xạ từ các cột 49](#_Toc215041013)

[Hình 3. 30 Cấu hình SQL Server Destination 50](#_Toc215041014)

[Hình 3. 31 Ánh xạ từ các cột 51](#_Toc215041015)

[Hình 3. 32 Cấu hình SQL Server Destination 51](#_Toc215041016)

[Hình 3. 33 Ánh xạ từ các cột 52](#_Toc215041017)

[Hình 3. 34 Cấu hình SQL Server Destination 53](#_Toc215041018)

[Hình 3. 35 Ánh xạ từ các cột 54](#_Toc215041019)

[Hình 3. 36 Load\_DimBrand 55](#_Toc215041020)

[Hình 3. 37 Bảng Sort 55](#_Toc215041021)

[Hình 3. 38 Load\_DimCustomer 56](#_Toc215041022)

[Hình 3. 39 Bảng Sort 56](#_Toc215041023)

[Hình 3. 40 Load\_DimTime 57](#_Toc215041024)

[Hình 3. 41 Bảng Sort 57](#_Toc215041025)

[Hình 3. 42 Derived Column Transformation Editor 58](#_Toc215041026)

[Hình 3. 43 Load\_DimProduct 58](#_Toc215041027)

[Hình 3. 44 Bảng Sort 59](#_Toc215041028)

[Hình 3. 45 Load\_Fact 59](#_Toc215041029)

[Hình 3. 46 Lookup\_Brand 60](#_Toc215041030)

[Hình 3. 47 Lookup\_Customer 61](#_Toc215041031)

[Hình 3. 48 Lookup\_Product 62](#_Toc215041032)

[Hình 3. 49 Lookup\_Time 63](#_Toc215041033)

[Hình 3. 50 Kiểm tra dữ liệu bảng Dim\_product 64](#_Toc215041034)

[Hình 3. 51 Kiểm tra dữ liệu bảng Dim\_FactBrand 65](#_Toc215041035)

[Hình 3. 52 Lược đồ sau khi hoàn thành 66](#_Toc215041036)

[Hình 3. 53 Tạo Project mới 67](#_Toc215041037)

[Hình 3. 54 Data Source Wizard 68](#_Toc215041038)

[Hình 3. 55 Create a data source based on an existing or new connection 68](#_Toc215041039)

[Hình 3. 56 Hộp thoại Connection Manager 69](#_Toc215041040)

[Hình 3. 57 Select Tables and Views 71](#_Toc215041041)

[Hình 3. 58 Hoàn tất quá trình tạo Data Source View 72](#_Toc215041042)

[Hình 3. 59 Kết quả 73](#_Toc215041043)

[Hình 3. 60 Solution Explorer 74](#_Toc215041044)

[Hình 3. 61 Hộp thoại Cube Wizard 74](#_Toc215041045)

[Hình 3. 62 Chọn use existing tables 75](#_Toc215041046)

[Hình 3. 63 Select Measure Group Tables 75](#_Toc215041047)

[Hình 3. 64 Select Measures 76](#_Toc215041048)

[Hình 3. 65 Completing the Wizard 76](#_Toc215041049)

[Hình 3. 66 Dimensions 77](#_Toc215041050)

[Hình 3. 67 Chọn Dim Customer.dim 77](#_Toc215041051)

[Hình 3. 68 Chọn Dim Date.dim 78](#_Toc215041052)

[Hình 3. 69 Chọn Dim Product.dim 78](#_Toc215041053)

[Hình 3. 70 Data Source View 79](#_Toc215041054)

[Hình 3. 71 Chọn Show Measures Grid 79](#_Toc215041055)

[Hình 3. 72 Attributes 81](#_Toc215041056)

[Hình 3. 73 Attribute Relationships 81](#_Toc215041057)

[Hình 3. 74 Properties 82](#_Toc215041058)

[Hình 3. 75 Bảng Key Columns 83](#_Toc215041059)

[Hình 3. 76 Chọn Name Columns 83](#_Toc215041060)

[Hình 3. 77 Bảng Attributes 84](#_Toc215041061)

[Hình 3. 78 Attribute Relationships 84](#_Toc215041062)

[Hình 3. 79 Bảng Key Columns 85](#_Toc215041063)

[Hình 3. 80 Product Title 86](#_Toc215041064)

[Hình 4. 1 Mở Power BI 88](#_Toc215041065)

[Hình 4. 2 Chọn SQL Server 89](#_Toc215041066)

[Hình 4. 3 Bảng Navigator 90](#_Toc215041067)

[Hình 4. 4 Thiết lập quan hệ giữa các bảng theo sơ đồ hình sao 91](#_Toc215041068)

[Hình 4. 5 Bảng tổng lượt đánh giá theo từng năm của từng thương hiệu 92](#_Toc215041069)

[Hình 4. 6 Biểu đồ phân phối cảm xúc theo thương hiệu 94](#_Toc215041070)

[Hình 4. 7 Biểu đồ xu hướng cảm xúc theo tháng 97](#_Toc215041071)

[Hình 4. 8 Từ khóa phổ biến trong bài đánh giá của Maybelline 99](#_Toc215041072)

[Hình 4. 9 Biểu đồ phân bố điểm đánh giá 101](#_Toc215041073)

[Hình 4. 10 Xây dựng Card KPI 104](#_Toc215041074)

[Hình 4. 11 Dashboard Cảm xúc khách hàng 106](#_Toc215041075)

[Hình 4. 12 Biểu đồ thể hiện sản phẩm có điểm đánh giá cao nhất 106](#_Toc215041076)

[Hình 4. 13 Biểu đồ quan hệ giữa cảm xúc và điểm đánh giá 108](#_Toc215041077)

[Hình 4. 14 Biểu đồ so sánh Brand Maybelline và các Brand còn lại 110](#_Toc215041078)

[Hình 4. 15 Biểu đồ thể hiện mức độ quan tâm đến thương hiệu 112](#_Toc215041079)

[Hình 4. 16 Biểu đồ nhóm sản phẩm ảnh hưởng lớn đến thương hiệu 114](#_Toc215041080)

[Hình 4. 17 Dashboard hiệu quả sản phẩm và thương hiệu 116](#_Toc215041081)

[Hình 4. 18 Biểu đồ xu hướng chỉ số BHI theo thời gian 116](#_Toc215041082)

[Hình 4. 19 Biểu đồ dự báo BHI 119](#_Toc215041083)

[Hình 4. 20 Biểu đồ phân rã nguyên nhân gây sự biến động BHI 121](#_Toc215041084)

[Hình 4. 21 Xây dựng KPI Card tổng hợp của Maybelline New York 123](#_Toc215041085)

[Hình 4. 22 Dashboard sức khỏe thương hiệu và chiến lược 125](#_Toc215041086)

# **LỜI MỞ ĐẦU**

Trong thời đại công nghệ số, người tiêu dùng không chỉ là người mua hàng mà còn là nguồn dữ liệu quý giá thông qua những chia sẻ, bình luận và đánh giá trên các nền tảng trực tuyến. Việc nắm bắt và phân tích cảm xúc của họ giúp doanh nghiệp hiểu rõ hơn cách thương hiệu được nhìn nhận, từ đó đưa ra những chiến lược phù hợp để nâng cao hình ảnh và uy tín.

Đề tài “Phân tích tâm trạng để đánh giá thương hiệu” được thực hiện với mục tiêu xây dựng một hệ thống có khả năng thu thập, xử lý và phân tích dữ liệu từ các nguồn trực tuyến như mạng xã hội hay trang thương mại điện tử. Thông qua việc ứng dụng các kỹ thuật phân tích cảm xúc (Sentiment Analysis), hệ thống giúp nhận diện xu hướng tâm trạng của người tiêu dùng và cung cấp thông tin hữu ích cho việc quản lý, phát triển thương hiệu.

Báo cáo sẽ trình bày các bước từ thu thập và xử lý dữ liệu đến phân tích kết quả, đồng thời đưa ra những đề xuất cải tiến để hoàn thiện hệ thống trong tương lai. Đề tài hướng đến việc hỗ trợ doanh nghiệp hiểu khách hàng hơn và xây dựng thương hiệu dựa trên cảm xúc thực tế của người tiêu dùng.

# **LỜI CẢM ƠN**

Nhóm xin gửi lời tri ân sâu sắc đến thầy ThS. Lê Huỳnh Long, giảng viên phụ trách môn “Xây dựng kho dữ liệu và hệ thống hỗ trợ ra quyết định”*.* Trong suốt quá trình học tập và thực hiện đề tài “Phân tích tâm trạng để đánh giá thương hiệu”**,** thầy đã luôn tận tâm hướng dẫn, định hướng phương pháp tiếp cận, đồng thời chia sẻ nhiều kinh nghiệm quý báu giúp nhóm có thể hoàn thành bài nghiên cứu một cách hiệu quả hơn.

Trong quá trình thực hiện, nhóm đã cố gắng vận dụng kiến thức được học trên lớp, kết hợp với việc tìm hiểu thêm tài liệu tham khảo và dữ liệu thực tế để xây dựng nội dung đề tài. Nhóm xin chân thành cảm ơn thầy vì sự tận tâm, nhiệt huyết và những định hướng quý giá trong suốt thời gian thực hiện đề tài.

# **CHƯƠNG 1: TỔNG QUAN**

**1.1**  **Giới thiệu về data warehouse (dw) và business intelligence (bi)**

### 1.1.1  Data warehouse

Data Warehouse (DW) hay còn gọi là Kho dữ liệu, là một hệ thống lưu trữ tập trung dùng để thu thập và quản lý dữ liệu từ nhiều nguồn khác nhau trong tổ chức. Mục tiêu của DW là cung cấp một môi trường dữ liệu thống nhất, ổn định và có cấu trúc, phục vụ cho việc khai thác, phân tích và hỗ trợ quá trình ra quyết định. Dữ liệu trong kho thường được lưu trữ theo thời gian, mang tính chất tích hợp và ít thay đổi, giúp người dùng dễ dàng truy xuất và so sánh các thông tin trong dài hạn.

Data Warehouse đóng vai trò như nền tảng trung tâm cho các hoạt động phân tích trong doanh nghiệp. Dữ liệu được trích xuất từ các hệ thống giao dịch, xử lý qua quy trình ETL (Extract – Transform – Load) trước khi được đưa vào kho. Sau đó, người dùng có thể truy cập và khai thác thông tin thông qua các công cụ Business Intelligence, báo cáo hoặc dashboard để đưa ra các quyết định chiến lược hiệu quả.

### 1.1.2 Business intelligence

Business Intelligence (BI) là tập hợp các công cụ, công nghệ và quy trình được sử dụng để phân tích và trực quan hóa dữ liệu, nhằm hỗ trợ doanh nghiệp đưa ra quyết định dựa trên dữ liệu thực tế. BI giúp chuyển đổi dữ liệu thô từ kho dữ liệu (Data Warehouse) thành thông tin có ý nghĩa, cung cấp cái nhìn toàn diện về hoạt động của tổ chức trong quá khứ, hiện tại và dự đoán xu hướng tương lai.

Hệ thống BI hỗ trợ tạo báo cáo, dashboard và các mô hình phân tích, giúp nhà quản lý dễ dàng theo dõi hiệu suất kinh doanh, xác định cơ hội hoặc rủi ro, và đưa ra các quyết định chiến lược chính xác hơn. Nhờ vào BI, doanh nghiệp có thể tối ưu hóa hoạt động, nâng cao năng lực cạnh tranh và định hướng phát triểndựa trên cơ sở dữ liệu được phân tích một cách khoa học.

**1.2 Các đặc trưng về mặt dữ liệu của một hệ thống DW và BI**

**1.2.1 Các đặc trưng về mặt dữ liệu của một hệ thống DW.**

*Subject-Oriented*: Dữ liệu được tổ chức và sắp xếp xung quanh các chủ đề quan trọng của doanh nghiệp như khách hàng, sản phẩm, bán hàng, thay vì các quy trình hoạt động hàng ngày. Giúp người dùng dễ dàng truy cập và phân tích dữ liệu theo các khía cạnh cụ thể để hỗ trợ việc ra quyết định.Ví dụ: Một kho dữ liệu có thể tập trung vào các chủ đề như bán hàng, tài chính, sản phẩm, khách hàng.

*Integrated*: Dữ liệu được thu thập từ nhiều nguồn khác nhau (như RDBMS, tệp phẳng, hệ thống CRM, v.v.) và được làm sạch, chuẩn hóa để tạo ra một định dạng nhất quán. Loại bỏ các xung đột về tên gọi, mã hóa, định dạng dữ liệu và đảm bảo tính nhất quán để phân tích hiệu quả hơn.

*Time-Variant*: Dữ liệu trong doanh nghiệp thường thay đổi liên tục theo thời gian, do đó chúng cần được gán nhãn thời gian tương ứng tại thời điểm nhập liệu. Việc gắn thời gian cho dữ liệu còn giúp người dùng dễ so sánh dữ liệu với nhau, phân biệt những thay đổi theo mặt tích cực hoặc tiêu cực. Ví dụ: Thông qua đặc tính có gắn nhãn thời gian, ban quản lý có thể so sánh doanh số bán hàng từ quý này với quý trước hoặc cùng kỳ năm trước.

*Non-Volatile*: Khi dữ liệu đã được tải vào data warehouse, nó sẽ không bị sửa đổi hoặc xóa bỏ. Dữ liệu chỉ có hai thao tác chính là tải vào và đọc ra (chỉ đọc - read-only). Sự bất biến này đảm bảo tính ổn định và nhất quán của dữ liệu lịch sử, rất quan trọng cho việc phân tích và báo cáo tin cậy.

### 1.2.2 Các đặc trưng về mặt dữ liệu của một hệ thống BI.

Dữ liệu được thu thập và tích hợp từ nhiều nguồn khác nhau: bao gồm hệ thống giao dịch (OLTP), ERP, CRM, website, mạng xã hội, file Excel, POS,… giúp tổng hợp đầy đủ bức tranh hoạt động của doanh nghiệp.

Dữ liệu được làm sạch và chuẩn hóa trước khi sử dụng: loại bỏ dữ liệu trùng lặp, thiếu giá trị, sai định dạng; chuẩn hóa đơn vị đo, format ngày tháng, mã sản phẩm,… để đảm bảo tính tương thích và so sánh được giữa các nguồn.

Đảm bảo tính nhất quán và độ tin cậy của dữ liệu: quy định các ràng buộc, điều kiện và chuẩn định danh (keys) để tránh sai lệch; dữ liệu trong kho BI phải là nguồn tin cậy duy nhất cho phân tích.

Mang tính lịch sử và có khả năng truy vết theo thời gian: lưu lại dữ liệu theo từng mốc thời gian (daily/weekly/monthly snapshot) phục vụ phân tích xu hướng, so sánh các kỳ, dự báo.

Dữ liệu có cấu trúc theo mô hình dữ liệu đa chiều (Star Schema, Snowflake Schema):  các bảng fact ghi nhận số đo kinh doanh; các bảng dimension mô tả ngữ cảnh (sản phẩm, khách hàng, thời gian, khu vực…) giúp thực hiện các phép phân tích OLAP (drill-down, roll-up, slice, dice).

Tối ưu hóa cho việc truy vấn và phân tích, không phải cho giao dịch ngay lập tức:  cấu trúc dữ liệu, chỉ mục, phân vùng được thiết kế để giảm thời gian trả lời truy vấn phân tích, thay vì xử lý các thao tác ghi nhanh như OLTP.

Hỗ trợ trực quan hóa và khai thác dữ liệu linh hoạt: dữ liệu được phục vụ cho dashboard, báo cáo tự phục vụ (self-service BI), biểu đồ, KPI,… giúp ban lãnh đạo nhìn thấy xu hướng và bất thường nhanh chóng.

Dữ  an toàn, bảo mật và phân quyền chi tiết**:** mỗi người dùng chỉ được truy cập dữ liệu phù hợp với quyền hạn; dữ liệu nhạy cảm được mã hóa hoặc ẩn danh theo chính sách quản trị dữ liệu.

Có khả năng cập nhật định kỳ và mở rộng theo thời gian:hỗ trợ cập nhật hàng ngày/giờ/real-time tùy thiết kế; hệ thống có thể mở rộng khi quy mô dữ liệu hoặc yêu cầu phân tích tăng lên.

**1.3. CÁC PHƯƠNG PHÁP LUẬN XÂY DỰNG**

Trong đề tài *“Phân tích tâm trạng khách hàng để đánh giá thương hiệu”*, việc lựa chọn phương pháp luận phù hợp đóng vai trò rất quan trọng nhằm bảo đảm hệ thống vừa chính xác, vừa linh hoạt.

Quy trình xây dựng kho dữ liệu (Data Warehouse – DW) và hệ thống phân tích cảm xúc (Sentiment Analysis) thường được thực hiện theo các phương pháp luận chính: Top-down, Bottom-up, và Spiral.

**1.3.1. Phương pháp luận Top-down**

Phương pháp Top-down (Tiếp cận từ trên xuống), gắn liền với Bill Inmon, người được xem là "cha đẻ của kho dữ liệu" (father of the data warehouse), bắt đầu từ tầm nhìn tổng thể của tổ chức.

Khái niệm gốc: Theo Inmon, kho dữ liệu (Enterprise Data Warehouse - EDW) phải được xây dựng đầu tiên. Đây là một kho lưu trữ trung tâm, tích hợp, và mang tính lịch sử, đóng vai trò là "nguồn chân lý duy nhất" (single source of truth) cho toàn doanh nghiệp. Dữ liệu trong EDW thường được chuẩn hóa (ví dụ: Cấu trúc 3NF - Dạng chuẩn 3) để đảm bảo tính nhất quán và toàn vẹn. Sau khi EDW trung tâm được xây dựng, các Data Mart (kho dữ liệu con, phục vụ các phòng ban cụ thể) mới được trích xuất (derived) từ EDW.

Trong bối cảnh đề tài, doanh nghiệp sẽ xác định mục tiêu chiến lược như: “Phân tích cảm xúc của khách hàng về thương hiệu X trên mạng xã hội nhằm cải thiện hình ảnh thương hiệu và chất lượng dịch vụ.”

**Ví dụ thực tế:**

* Coca-Cola sử dụng phương pháp Top-down để thiết kế hệ thống phân tích phản hồi khách hàng toàn cầu. Dữ liệu được thu thập từ Twitter, Facebook, YouTube và các khảo sát khách hàng. Sau đó, công ty tạo Data Warehouse trung tâm, lưu trữ hàng triệu phản hồi, cho phép ban lãnh đạo xem biểu đồ cảm xúc theo từng khu vực, sản phẩm, hoặc chiến dịch marketing.
* Ở Việt Nam, Thế Giới Di Động đã xây dựng hệ thống kho dữ liệu nội bộ để tổng hợp hàng triệu phản hồi từ website và fanpage, giúp công ty đánh giá thái độ khách hàng sau mỗi đợt khuyến mãi lớn.

**Ưu điểm:**

* Cấu trúc dữ liệu rõ ràng, nhất quán trên toàn doanh nghiệp.
* Dễ bảo trì và mở rộng về lâu dài.
* Thích hợp cho doanh nghiệp có quy mô lớn, dữ liệu đa nguồn.

**Nhược điểm:**

* Thời gian xây dựng ban đầu rất dài, cần đầu tư nhiều nguồn lực.
* Nếu mục tiêu chiến lược ban đầu thay đổi, hệ thống khó điều chỉnh linh hoạt.
* Mang lại giá trị kinh doanh chậm.

**1.3.2. Phương pháp luận Bottom-up**

Ngược lại, phương pháp Bottom-up (Tiếp cận từ dưới lên) được khởi xướng bởi Ralph Kimball. Cách tiếp cận này tập trung vào việc cung cấp giá trị kinh doanh nhanh chóng thông qua việc xây dựng các phần nhỏ, cụ thể trước.

Khái niệm gốc: Kimball cho rằng nên bắt đầu bằng cách xây dựng các Data Mart riêng lẻ, độc lập, tập trung vào các quy trình nghiệp vụ cụ thể (ví dụ: Data Mart về Marketing, Data Mart về Bán hàng). Các Data Mart này được thiết kế theo mô hình chiều (dimensional model), điển hình là mô hình sao (star schema), để tối ưu hóa tốc độ truy vấn và phân tích. Kho dữ liệu doanh nghiệp (EDW) sẽ được hình thành dần dần bằng cách *tích hợp* các Data Mart này lại với nhau (Kimball gọi đây là kiến trúc "data warehouse bus").

Trong đề tài này, nhóm có thể bắt đầu phân tích cảm xúc từ một nguồn duy nhất (ví dụ: bình luận Facebook) để xây dựng một Data Mart đầu tiên, trước khi mở rộng sang Shopee Review, Google Maps.

**Ví dụ thực tế:**

* Một nhà hàng địa phương có thể bắt đầu thu thập và phân tích bình luận của khách hàng trên Facebook (Data Mart 1) để xác định tỷ lệ hài lòng. Khi hệ thống ổn định, nhà hàng mở rộng thêm dữ liệu từ GrabFood (Data Mart 2), Foody (Data Mart 3) và tích hợp chúng để so sánh tâm trạng khách hàng giữa các nền tảng.
* Highlands Coffee cũng đã áp dụng quy trình tương tự, khởi đầu với dữ liệu phản hồi Facebook, sau đó tích hợp dữ liệu đánh giá từ Google Maps và hệ thống phản hồi tại quầy.

**Ưu điểm:**

* Thời gian triển khai nhanh, mang lại kết quả (ROI) sớm.
* Linh hoạt, dễ triển khai thử nghiệm (pilot).
* Chi phí ban đầu thấp, phù hợp với nhóm nghiên cứu hoặc doanh nghiệp vừa và nhỏ.

**Nhược điểm:**

* Dữ liệu có thể thiếu nhất quán giữa các Data Mart nếu không được thiết kế cẩn thận (ví dụ: định nghĩa "khách hàng" ở Data Mart Marketing khác với ở Data Mart Bán hàng).
* Dễ gặp khó khăn khi mở rộng quy mô nếu không có kiến trúc "bus" chuẩn.

**1.3.3. Phương pháp luận Spiral**

Phương pháp Spiral (Mô hình Xoắn ốc) là một mô hình phát triển lặp lại, kết hợp các yếu tố của cả Top-down và Bottom-up. Mô hình này ban đầu được đề xuất bởi **Barry Boehm** (1986) cho lĩnh vực phát triển phần mềm, nhưng rất phù hợp với các dự án phức tạp như kho dữ liệu và học máy.

Khái niệm gốc: Khái niệm cốt lõi của mô hình Spiral là quản lý rủi ro (risk management). Dự án được chia thành nhiều vòng lặp (xoắn ốc). Mỗi vòng lặp bao gồm 4 giai đoạn chính:

1. Lập kế hoạch (Planning): Xác định mục tiêu, các phương án, và ràng buộc.
2. Phân tích rủi ro (Risk Analysis): Đánh giá các rủi ro (kỹ thuật, quản lý) và đưa ra giải pháp.
3. Phát triển và Thử nghiệm (Engineering & Test): Xây dựng một phiên bản (prototype) của sản phẩm.
4. Đánh giá (Evaluation): Khách hàng/người dùng đánh giá sản phẩm và lập kế hoạch cho vòng lặp tiếp theo. Mỗi vòng lặp sẽ tạo ra một sản phẩm hoàn thiện hơn vòng lặp trước.

Ví dụ thực tế:

Shopee Việt Nam khi triển khai hệ thống phân tích phản hồi khách hàng đã áp dụng mô hình Spiral.

*Vòng 1:* Chỉ tập trung xây dựng mô hình phát hiện phản hồi cực kỳ tiêu cực (đánh giá 1–2 sao) để cảnh báo cho đội chăm sóc khách hàng.

*Vòng 2:* Cải tiến mô hình để phân loại thêm cảm xúc trung tính và tích cực.

*Vòng 3:* Mở rộng mô hình để phân tích sâu hơn (aspect-based sentiment), ví dụ: khách hàng phàn nàn về "giao hàng" hay "chất lượng sản phẩm".

Ưu điểm:

* Quản lý rủi ro tốt, đặc biệt là rủi ro về kỹ thuật (ví dụ: mô hình AI không chính xác) hoặc yêu cầu thay đổi.
* Linh hoạt, cho phép điều chỉnh liên tục.
* Rất phù hợp với các dự án học máy (như phân tích cảm xúc), nơi việc hiệu chỉnh mô hình là liên tục.

Nhược điểm:

* Phức tạp trong quản lý dự án (cần theo dõi nhiều vòng lặp).
* Cần nhân sự có chuyên môn kỹ thuật cao (Data Science, ML).
* Mất thời gian trong quá trình tinh chỉnh và thử nghiệm nhiều vòng.

**1.3.4. Cơ sở lựa chọn phương pháp**

Dựa trên tính chất của đề tài "Phân tích tâm trạng khách hàng để đánh giá thương hiệu", phương pháp luận Top-down là không phù hợp do tính cứng nhắc, thời gian phát triển lâu và rủi ro cao khi các nguồn dữ liệu mạng xã hội (Facebook, TikTok, Shopee...) thay đổi liên tục.

Hai phương pháp Bottom-up và Spiral là phù hợp hơn, trong đó phương pháp Spiral (Xoắn ốc) thể hiện tính ưu việt rõ rệt.

Ví dụ cụ thể về áp dụng phương pháp Spiral cho đề tài: Đề tài tập trung vào sản phẩm "Maybeline" với nguồn dữ liệu chính là từ sàn thương mại điện tử Nykaa, một trong những thương hiệu bán lẻ mỹ phẩm lớn nhất tại Ấn Độ.

Vòng lặp 1 (Spiral 1): Xây dựng Data Mart cơ bản và Mô hình Prototype

* Kế hoạch: Dữ liệu được thu thập từ bộ dữ liệu *“Nykaa – Popular Brands Cosmetics & Beauty Products”* trên Kaggle. Từ đó, nhóm chọn lọc các đánh giá (reviews) liên quan đến sản phẩm Maybelline trên sàn thương mại điện tử Nykaa. Mục tiêu của giai đoạn đầu là xây dựng một mô hình phân tích cảm xúc 3 nhãn (*Tích cực / Tiêu cực / Trung tính*) để hiểu phản hồi chung của khách hàng về sản phẩm.
* Phân tích rủi ro: Rủi ro chính là chất lượng dữ liệu đánh giá không đồng nhất, có thể chứa nhiều ngôn ngữ, lỗi chính tả, hoặc các từ viết tắt, emoji. Bên cạnh đó, dữ liệu có thể thiếu thông tin mô tả chi tiết, gây khó khăn trong quá trình huấn luyện mô hình.
* Phát triển: Xây dựng một Data Mart dạng Star Schema với bảng sự kiện là *FactReview* (chứa nội dung đánh giá, điểm sao, mã sản phẩm, mã thương hiệu...). Tiến hành tiền xử lý văn bản (text preprocessing): chuẩn hóa chữ viết, loại bỏ ký tự đặc biệt, xử lý từ đồng nghĩa và các biểu tượng cảm xúc. Huấn luyện mô hình PhoBERT hoặc BERT multilingual để phân loại cảm xúc.
  + - Đánh giá: Mô hình đạt độ chính xác 70%. Nhóm quyết định qua vòng 2 để cải thiện độ chính xác và mở rộng nguồn.

Vòng lặp 2 (Spiral 2): Mở rộng nguồn dữ liệu và Tinh chỉnh mô hình

* Kế hoạch: Trong vòng thứ hai, nhóm nghiên cứu mở rộng nguồn dữ liệu sang các sàn thương mại khác có phân phối sản phẩm Maybelline, chẳng hạn như Shopee, Amazon, hoặc Flipkart. Dự kiến thu thập thêm 10.000 đánh giá mới nhằm tăng tính đa dạng và khái quát cho mô hình.
* Mục tiêu: nâng độ chính xác lên 80% và tích hợp dữ liệu từ nhiều sàn vào Data Warehouse để có góc nhìn toàn diện hơn.
* Phân tích rủi ro: Các sàn khác nhau có cấu trúc dữ liệu không đồng nhất (ví dụ: Shopee có trường “số sao”, Amazon có “verified purchase”). Rủi ro trong việc mapping và đồng bộ schema dữ liệu khi tích hợp vào DW. Dữ liệu từ sàn khác có thể chứa ngôn ngữ hoặc văn phong khác, ảnh hưởng đến hiệu năng của mô hình gốc.
* Phát triển: Cập nhật mô hình dữ liệu trong DW để tiếp nhận nhiều nguồn dữ liệu. Áp dụng Transfer Learning để huấn luyện lại mô hình với dữ liệu từ cả Nykaa và các sàn khác. Xây dựng pipeline ETL tự động hóa để đồng bộ dữ liệu định kỳ.
  + - Đánh giá: Sau khi tinh chỉnh, mô hình đạt độ chính xác trung bình 78%, DW cho phép truy vấn chéo dữ liệu giữa các sàn, hỗ trợ phân tích xu hướng cảm xúc theo từng kênh phân phối.

Vòng lặp 3 (Spiral 3): Phát triển tính năng Phân tích Chi tiết (Aspect-Based)

* + Kế hoạch: Thay vì chỉ nói "tiêu cực", mô hình cần chỉ ra "tiêu cực *về điều gì*?". Mục tiêu: Phân tích sâu vào các khía cạnh (aspects) như "Giao hàng", "Giá cả", "Chăm sóc khách hàng".
  + Phân tích rủi ro: Đây là bài toán kỹ thuật phức tạp, đòi hỏi mô hình ML cao cấp hơn và dữ liệu gán nhãn chi tiết.
  + Phát triển: Xây dựng mô hình *Aspect-Based Sentiment Analysis* (ABSA). Cập nhật Data Mart để lưu trữ các khía cạnh này.
  + Đánh giá: Hệ thống cho phép ban lãnh đạo xem: "Trong tháng 10, 30% phản hồi tiêu cực tập trung vào 'Giao hàng chậm' trên Shopee, trong khi Facebook chủ yếu phàn nàn về 'Thái độ nhân viên'".

# **CHƯƠNG 2: PHÂN TÍCH THIẾT KẾ KHO DỮ LIỆU**

## 2.1 Giới thiệu về dữ liệu dataset

### 2.1.1 Nguồn dữ liệu:

Nguồn dữ liệu được sử dụng trong nghiên cứu này được thu thập từ trang Kaggle, cụ thể là bộ dữ liệu “Nykaa – Popular Brands Cosmetics & Beauty Products”. Đây là tập dữ liệu liên quan đến các sản phẩm mỹ phẩm và làm đẹp được phân phối trên nền tảng thương mại điện tử Nykaa, một trong những thương hiệu bán lẻ mỹ phẩm lớn nhất tại Ấn Độ.

Bộ dữ liệu cung cấp thông tin chi tiết về các thương hiệu phổ biến, tên sản phẩm, giá cả, đánh giá của khách hàng, xếp hạng sao, và phân loại sản phẩm (như skincare, makeup, haircare, v.v.). Dữ liệu này được công bố công khai trên Kaggle nhằm hỗ trợ cho các mục đích phân tích dữ liệu, nghiên cứu hành vi người tiêu dùng, và ứng dụng học máy trong lĩnh vực thương mại điện tử mỹ phẩm.

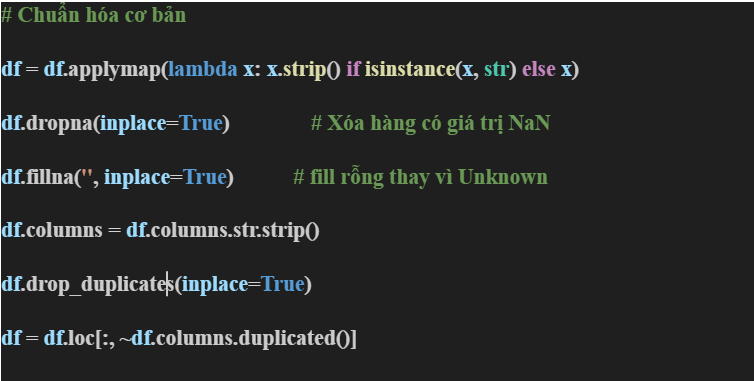
Nguồn: <https://www.kaggle.com/datasets/jithinanievarghese/nykaa-popular-brands-cosmetics-beauty-products>

### 2.1.2 Mô tả chi tiết dữ liệu:

|  |  |  |
| --- | --- | --- |
| **Tên thuộc tính** | **Mô tả** | **Ví dụ** |
| Brand\_name | Tên thương hiệu | Olay |
| Review\_id | Mã đánh giá | 16752142 |
| Review\_title | Tiêu đề đánh giá | Worth buying 50g one |
| Review\_text | Nội dung đánh giá | Works as it claims. Could see the difference from the first day. Use it with Olay cleanser for best results |
| Customer\_name | Tên khách hàng đánh giá | Ashton Dsouza |
| Review\_date | Thời gian đăng bài đánh giá | 1/23/2021  3:17:00 PM |
| Review\_rating | Điểm đánh giá của người dùng | 5 |
| IsBuyer | Xác định người viết có mua sản phẩm hay không | True |
| Pro\_user | Xác định người dùng là chuyên nghiệp (pro reviewer) hay không | False |
| Review\_label | Nhãn xác thực người mua | Verified Buyer |
| Product\_title | Tên sản phẩm được đánh giá | Olay Ultra Lightweight Moisturiser: Luminous Whip Day Cream (non SPF) |
| Price | Giá sản phẩm | 1599 |
| Product\_rating | Điểm trung bình của sản phẩm | 4.1 |

**2.1.3 Chuẩn bị dữ liệu:**

**2.1.3.1 Làm sạch dữ liệu:**

****

Hình 2. 1 Làm sạch dữ liệu

Applymap():  áp dụng một hàm cho từng ô trong bảng.

Lambda x: x.strip() if isinstance(x, str) else x: nếu ô dữ liệu là chuỗi (str) thì loại bỏ khoảng trắng ở đầu và cuối, nếu không phải chuỗi thì giữ nguyên.

Dropna() xóa các hàng chứa giá trị bị thiếu (NaN).

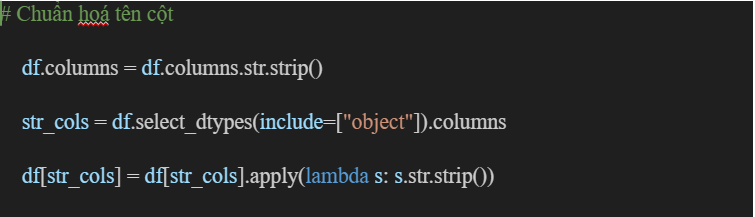
Inplace=True có nghĩa là sửa trực tiếp trên df mà không tạo bản sao.

Df.fillna('', inplace=True): thay các giá trị bị thiếu (NaN) còn sót lại bằng chuỗi rỗng (''), lúc này toàn bộ dữ liệu không còn giá trị null.

Df.columns = df.columns.str.strip(): loại bỏ khoảng trắng ở tên cột (header).

Df.drop\_duplicates(inplace=True): xóa các dòng trùng lặp hoàn toàn (tất cả cột giống nhau).

Df.columns.duplicated() trả về mảng Boolean cho biết cột nào bị trùng tên.  
 “~” là phủ định (ngược lại), nên chỉ giữ cột không trùng.  
 “.loc[:, ...]” chọn toàn bộ hàng (:) và chỉ các cột duy nhất.



Hình 2. 2 Làm sạch dữ liệu

Chọn các cột có kiểu dữ liệu là chuỗi (object).

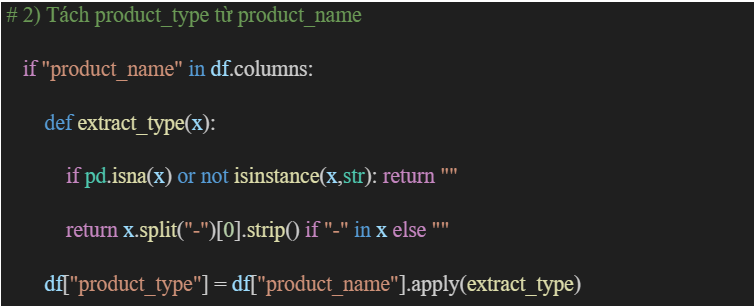
Áp dụng .str.strip() để loại bỏ khoảng trắng đầu/cuối trong từng ô.



Hình 2. 3 Làm sạch dữ liệu

Kiểm tra xem có cột brand hay không.

Nếu có, chuyển chữ cái đầu mỗi từ thành chữ hoa (.str.title()).

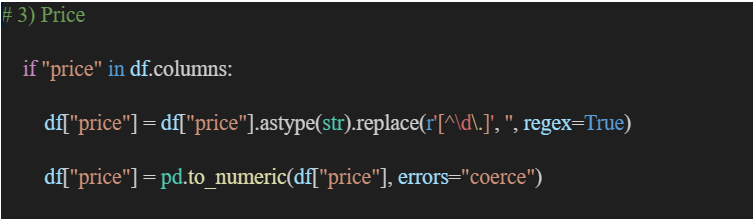


Hình 2. 4 Làm sạch dữ liệu

Tạo hàm extract\_type(x) để lấy loại sản phẩm từ tên.

Nếu tên có dấu -, lấy phần trước dấu trừ làm loại sản phẩm.

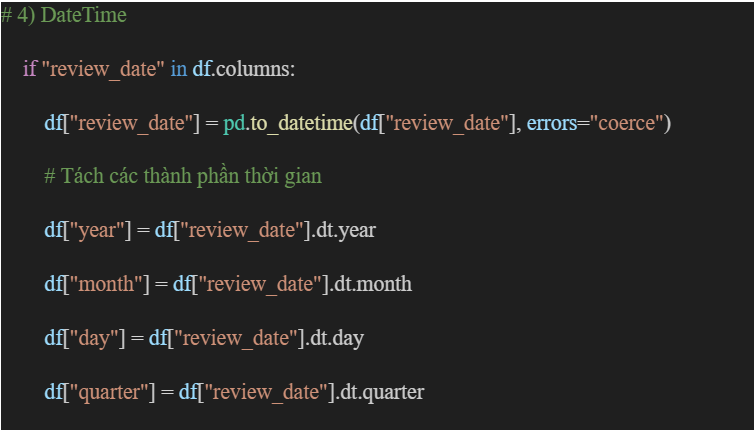
Nếu không có -, trả về chuỗi rỗng.



Hình 2. 5 Làm sạch dữ liệu

Astype(str) → chuyển giá trị sang chuỗi.  
 Replace(r'[^\d\.]', '', regex=True): Loại bỏ tất cả ký tự không phải số hoặc dấu chấm.

Pd.to\_numeric(..., errors="coerce"): Chuyển chuỗi sang dạng số (float). Nếu giá trị không hợp lệ thì gán NaN.



Hình 2. 6 Làm sạch dữ liệu

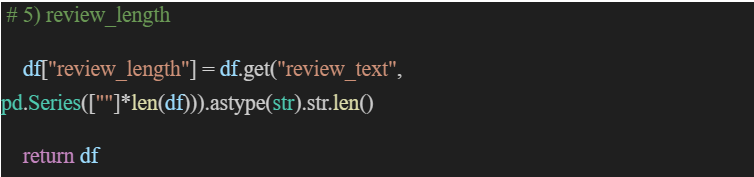
Df["review\_date"] = pd.to\_datetime(df["review\_date"], errors="coerce"): chuyển cột review\_date (hiện có thể là chuỗi hoặc các định dạng khác) thành kiểu thời gian datetime64[ns] của pandas.

Df["year"] = df["review\_date"].dt.year: truy cập accessor .dt để thao tác các thuộc tính thời gian trên cột datetime (vectorized), .dt.year lấy **giá trị năm** (ví dụ 2024) từ mỗi phần tử datetime.

Df["month"] = df["review\_date"].dt.month: lấy **tháng** (1–12) từ review\_date. Tương tự như year, nếu có nat thì giá trị sẽ là nan (thực ra numeric float).

Df["day"] = df["review\_date"].dt.day: lấy **ngày trong tháng** (1–31).

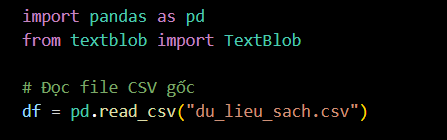
Df["quarter"] = df["review\_date"].dt.quarter: lấy **quý trong năm** (1–4)



Hình 2. 7 Làm sạch dữ liệu

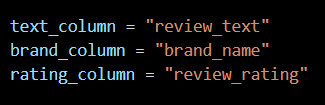
Lấy cột review\_text (nếu có); nếu không có thì tạo chuỗi rỗng cùng độ dài. Astype(str) chuyển về chuỗi. “.str.len()” tính độ dài của chuỗi (số ký tự).

**2.1.3.2 Tạo các cột phân tích cần thiết:**



Hình 2. 8 Làm sạch dữ liệu

Cập nhật thư viện cho dự án, sử dụng TextBlob để thực hiện phân tích cảm xúc từ bình luận khách hàng. Mở tệp dữ  ban đầu, nơi chứa thông tin về các bài đánh giá sản phẩm, điểm đánh giá,...



Hình 2. 9 Làm sạch dữ liệu

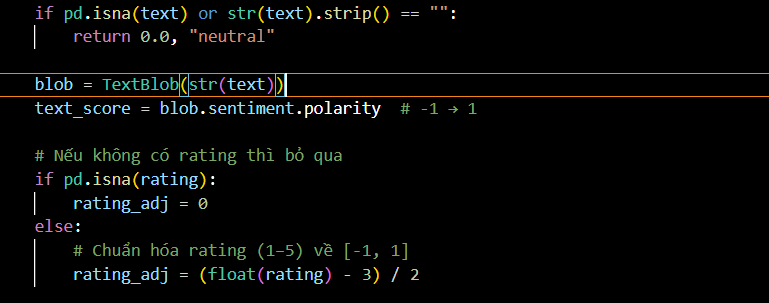
Sử dụng 3 cột chính trong file dữ liệu, gán vào tên biến tiện sử dụng khi cần thiết, mà không cần truy cập thông qua df.

C:\Users\Welcome\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\28970519.tmp

Hình 2. 10 Làm sạch dữ liệu

Sử dụng hàm với mục đích tạo ra 2 giá trị cốt lõi là:

* Sentiment\_core: điểm cảm xúc thực trong [-1, 1]
* Sentimen\_label: nhãn cảm xúc (“positive”, “negative”, “neutral”)

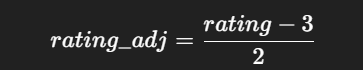


Hình 2. 11 Làm sạch dữ liệu

Kiểm tra dữ liệu tại ô chứa “text” nếu không có nội dung mặc định gắn là “neutral”

Dùng **TextBlob** để phân tích cảm xúc văn bản (chèn text vào), lấy giá trị “polarity”

Chuẩn hóa điểm “rating” (thang 1–5) về cùng khoảng [-1, 1] qua công thức:



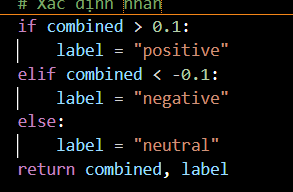
Kết hợp giữa điểm cảm xúc và điểm đánh giá (rating) của khách hàng, lấy 70% là từ điểm đánh giá và 30% sẽ là từ sao đánh giá. Để cân bằng mức độ cảm xúc của khách hàng.

Đôi khi có trường hợp khách hàng đánh giá cao nhưng có một số bất tiện với sản phẩm, để không bị quá phụ thuộc vào một trường cung cấp thông tin, nhóm em quyết định tích hợp lại để đưa ra kết quả khách quan hơn về bài đánh giá của sản phẩm thương hiệu.



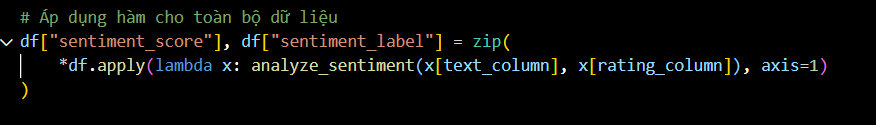
Hình 2. 12 Làm sạch dữ liệu

Gán nhãn cảm xúc cho từng bài đánh giá bằng số điểm vừa kết hợp. Sau đó trả về kết quả điểm và nhãn cảm xúc.



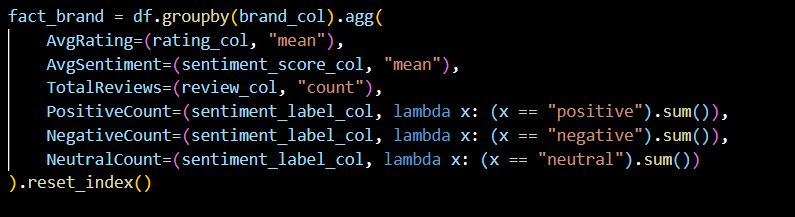
Hình 2. 13 Làm sạch dữ liệu

Sau đó kết quả sẽ được áp dụng cho toàn dữ liệu



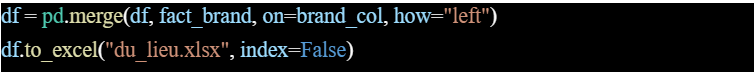
Hình 2. 14 Làm sạch dữ liệu

Ghi đè vào file dữ liệu csv ban đầu sau đó thực hiện gom nhóm sản phẩm tính trung bình các mục: AvgRating, AvgSentiment, TotalReviews, PositiveCount, NegativeCount, NeutralCount.



Hình 2. 15 Làm sạch dữ liệu

Gộp dữ liệu tổng hợp trở lại dataset chính, cùng với các điểm cảm xúc ở trên, sau đó xuất file sang excel để dễ dàng thao tác sau đó.



Hình 2. 16 Làm sạch dữ liệu

## 2.2 Thiết kế kho dữ liệu

### 2.2.1 Fact table

**2.2.1.1 Bảng Fact\_review:**

Đây là bảng trung tâm của kho dữ liệu, ghi lại từng đánh giá (review) của khách hàng đối với sản phẩm trên Nykaa.

* Review\_ID:        Mã duy nhất của đánh giá
* Product\_ID: Khóa ngoại đến Dim\_Product
* Brand\_ID:           Khóa ngoại đến Dim\_Brand
* Customer\_ID:     Khóa ngoại đến Dim\_Customer
* Time\_ID:            Khóa ngoại đến Dim\_Time
* Sentiment\_ID:     Khóa ngoại đến Dim\_Sentiment
* Review\_Rating:  Điểm người dùng đánh giá (1-5)
* Product\_Rating:  Điểm trung bình sản phẩm
* Price:        Giá sản phẩm
* Review\_Length:  Độ dài nội dung đánh giá
* IsBuyer:              Người đánh giá có phải người mua thật hay không
* Is\_MainBrand:    1 nếu là Maybelline New York
* Sentiment\_Score:    Điểm cảm xúc (TextBlob + Rating)

#### **2.2.2.2 Bảng Fact\_BrandSummary:**

FactBrandSummary là bảng fact tổng hợp, được dùng để lưu trữ thông tin thống kê của các thương hiệu (brand). Dùng để so sánh với thương hiệu chính từ đó hỗ trợ ra quyết định khách quan và bao quát.

* SummaryID: Mã định danh của bản ghi tổng hợp
* brand\_id: Mã thương hiệu
* AvgRating: Điểm đánh giá trung bình của tất cả sản phẩm thuộc thương hiệu
* AvgSentiment: Điểm cảm xúc trung bình
* TotalReviews: Tổng số lượt đánh giá của thương hiệu
* PositiveCount: Số lượng đánh giá có nhãn Positive
* NegativeCount: Số lượng đánh giá có nhãn Negative
* NeutralCount: Số lượng đánh giá có nhãn Neutral

### 2.2.2 Bảng Dimension Tables:

**2.2.2.1 Bảng Dim\_Brand:**

Chứa thông tin tổng hợp mô tả về các thương hiệu mỹ phẩm xuất hiện trong dữ liệu.  
 Mỗi dòng đại diện cho một thương hiệu (Olay, Maybelline, Nykaa Cosmetics, L’Oreal,…).

* Brand\_ID:           Mã thương hiệu
* Brand\_Name:      Tên thương hiệu
* Avg\_Rating: Điểm đánh giá trung bình
* Avg\_Sentiment:  Điểm cảm xúc trung bình
* Total\_Reviews:   Tổng số đánh giá
* Positive\_Count:   Số đánh giá tích cực
* Negative\_Count: Số đánh giá tiêu cực
* Neutral\_Count:    Số đánh giá trung lập

**2.2.2.2 Bảng Dim\_Product**

Lưu thông tin mô tả về từng sản phẩm cụ thể (ví dụ: “Olay Moisturiser 50g” hoặc “Kay Beauty Eyeshadow Pencil”).

* Product\_ID: Mã sản phẩm
* Product\_Title:     Tên sản phẩm
* Brand\_ID:           Thương hiệu sản phẩm
* Category:            Loại sản phẩm
* Price:        Giá niêm yết
* Product\_Rating:  Điểm đánh giá trung bình

**2.2.2.3 Bảng Dim\_Customer**

Chứa thông tin mô tả người dùng (người viết review).

* Customer\_ID:     Mã khách hàng
* Customer\_Name:    Tên khách hàng
* Is\_Pro\_User: Người dùng chuyên nghiệp hay không

**2.2.2.4 Bảng Dim\_time**

Giúp phân tích xu hướng theo thời gian.

* Time\_ID:            Mã thời gian
* Review\_Date:     Ngày đăng đánh giá
* Year:                  Năm
* Month:                Tháng
* Day:                   Ngày
* Weekday:           Thứ trong tuần

**CHƯƠNG 3: TÍCH HỢP DỮ LIỆU**

**3.1 Giới thiệu SSIS:**

SQL Server Integration Services (SSIS) là một thành phần cốt lõi của nền tảng cơ sở dữ liệu Microsoft SQL Server. Đây là một nền tảng mạnh mẽ được thiết kế chuyên biệt cho các tác vụ tích hợp dữ liệu (data integration) và xây dựng quy trình làm việc (workflow applications).

SSIS không phải là công cụ ETL (Extract, Transform, Load) đầu tiên của Microsoft. Nó là sự kế thừa và thay thế hoàn toàn cho công cụ tiền nhiệm là DTS (Data Transformation Services), vốn là công cụ đi kèm với SQL Server 7.0 (1998) và SQL Server 2000.

SSIS được ra mắt lần đầu tiên cùng với phiên bản SQL Server 2005, đánh dấu một bước nhảy vọt về kiến trúc. Nó được thiết kế lại hoàn toàn từ đầu để cung cấp hiệu suất cao hơn, tính linh hoạt vượt trội và khả năng mở rộng tốt hơn đáng kể so với DTS.

Mục đích cốt lõi của SSIS là giải quyết các bài toán phức tạp liên quan đến di chuyển và xử lý dữ liệu, được tóm gọn trong quy trình ETL (Extract - Transform - Load):

1. Extract (Trích xuất): SSIS có khả năng kết nối và trích xuất dữ liệu từ vô số nguồn khác nhau, ví dụ:

* Các tệp phẳng (Flat Filbes) như ( .txt, .csv )
* Tệp Excel ( .xls, .xlsx )
* Các hệ quản trị CSDL quan hệ khác (Oracle, DB2, MySQL...).
* Nguồn dữ liệu OLE DB, ODBC, ADO.NET.
* Các dịch vụ Web (Web services).

1. Transform (Biến đổi): Đây là nơi SSIS thể hiện sức mạnh lớn nhất. Sau khi trích xuất, dữ liệu thô thường cần được "làm sạch" và "chuẩn hóa" trước khi nạp vào đích. Các phép biến đổi (transformations) phổ biến bao gồm:

* Data Cleansing: Làm sạch dữ liệu (ví dụ: chuẩn hóa chữ hoa/thường, xóa ký tự đặc biệt).
* Aggregate: Tổng hợp dữ liệu (ví dụ: tính tổng doanh thu, đếm số lượng).
* Merge/Join: Kết hợp dữ liệu từ nhiều nguồn khác nhau.
* Lookup: Tra cứu thông tin (ví dụ: tra cứu tên sản phẩm từ ID sản phẩm).
* Derived Column: Tạo ra các cột dữ liệu mới dựa trên logic nghiệp vụ (ví dụ: tính cột "Lợi nhuận" = "Doanh thu" - "Chi phí").

1. Load (Nạp): Sau khi dữ liệu đã được biến đổi, SSIS sẽ nạp (load) dữ liệu này vào một hoặc nhiều hệ thống đích. Trong bối cảnh của môn học Kho dữ liệu, đích đến quan trọng nhất chính là Kho dữ liệu (Data Warehouse) hoặc các Data Mart để phục vụ cho việc phân tích và ra quyết định (Business Intelligence).

**3.2 Quá trình đổ dữ liệu vào kho:**

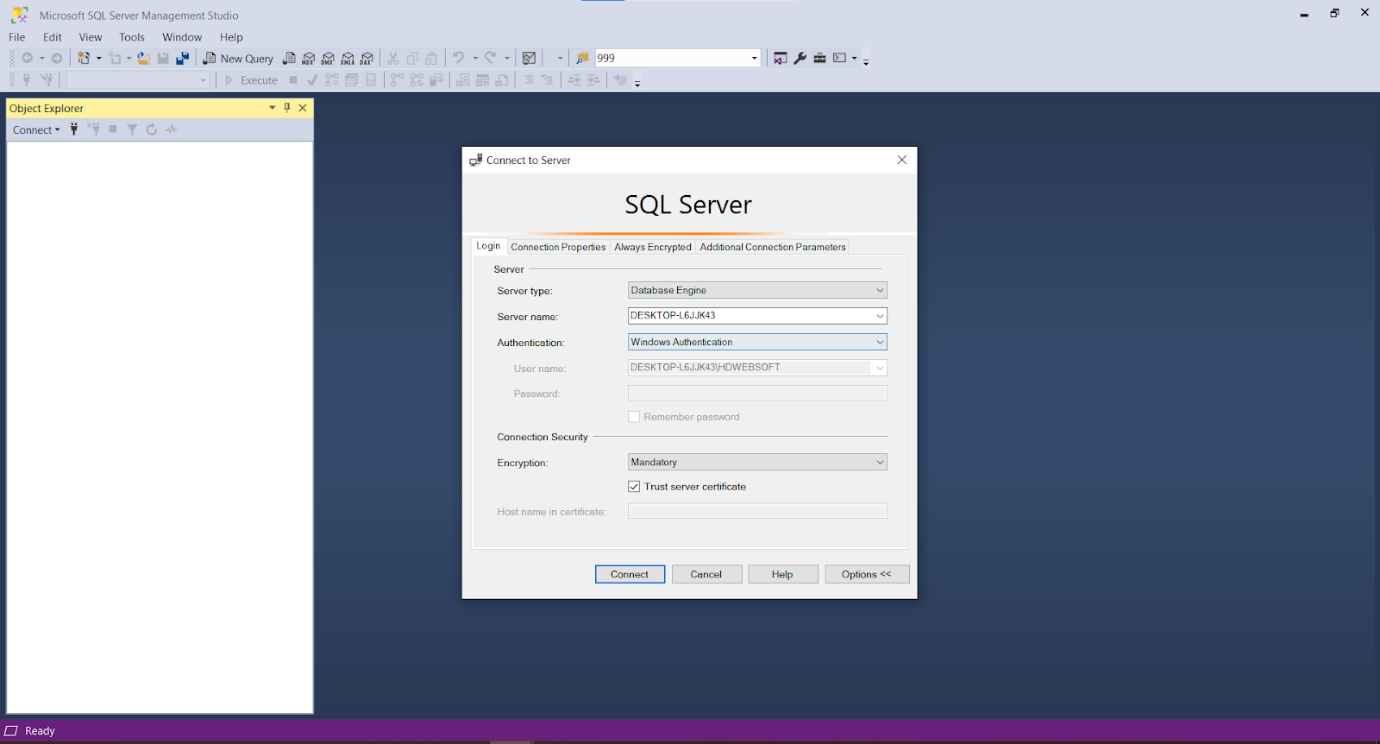
**3.2.1**  **Chuẩn bị các công cụ**

Để thực hiện được quá trình SSIS ta cần chuẩn bị và cài đặt các công cụ:

* Visual Studio Community 2022
* SQL Server Integration Services Project (SSISSIS)
* SQL Server

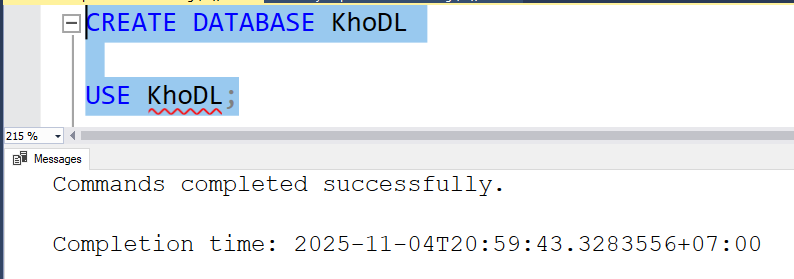
**3.2.2  Chuẩn bị cơ sở dữ liệu**

**Bước 1:** Mở SQL Server và kết nối với server bằng tài khoản user của window (Windows Authentication) hoặc SQL Server Authentication.



Hình 3. 1 Mở SQL Server

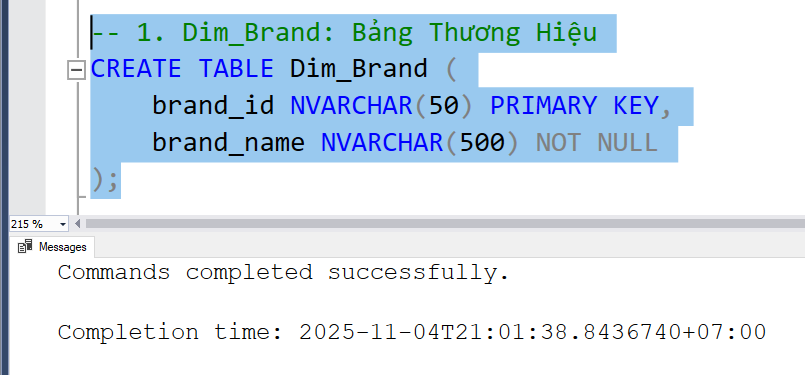
**Bước 2:** Khởi tạo một cơ sở dữ liệu có tên KhoDL



Hình 3. 2 Tạo cơ sở dữ liệu

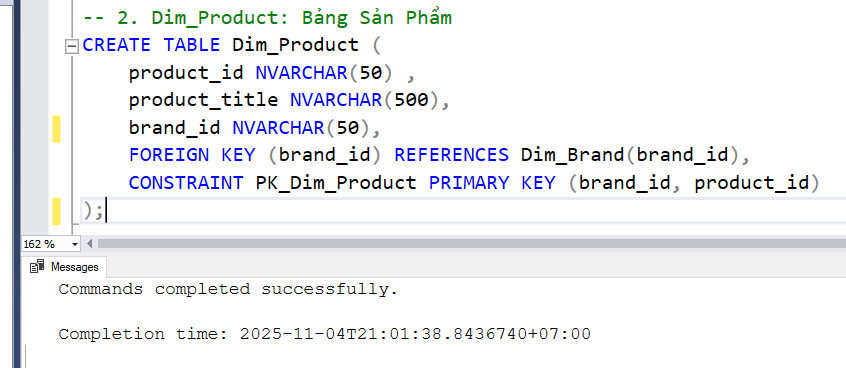
 Đây là nơi lưu các bảng Dim và bảng Fact cùng dữ liệu của các bảng đó.

**Bước 2.1** Khởi tạo bảng: Dim\_Brand (Bảng Thương Hiệu)



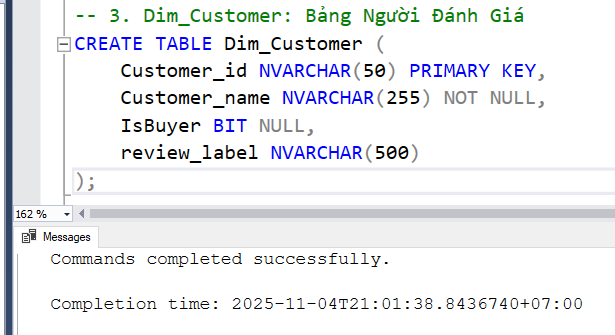
Hình 3. 3 Khởi tạo bảng: Dim\_Brand

**Bước 2.2** Khởi tạo bảng: Dim\_Product (Bảng Sản Phẩm)



Hình 3. 4 Khởi tạo bảng: Dim\_Product

**Bước 2.3** Khởi tạo bảng: Dim\_Customer (Bảng Người Đánh Giá)



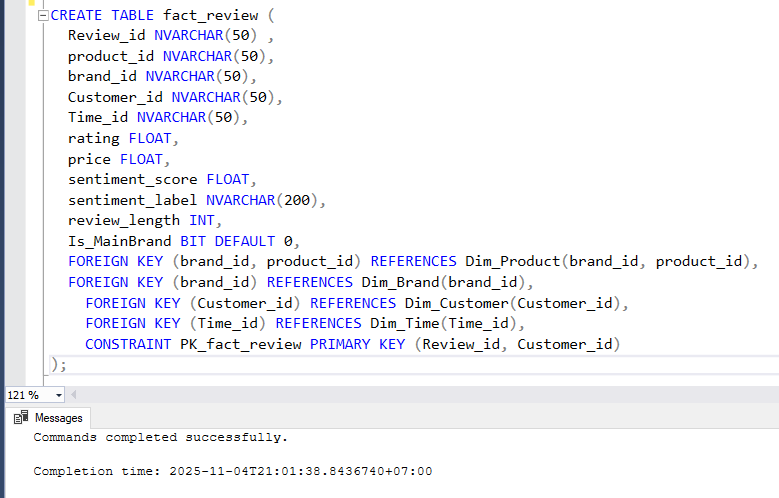
Hình 3. 5 Khởi tạo bảng: Dim\_Customer

**Bước 2.4** Khởi tạo bảng: Dim\_Time (Bảng Thời Gian)



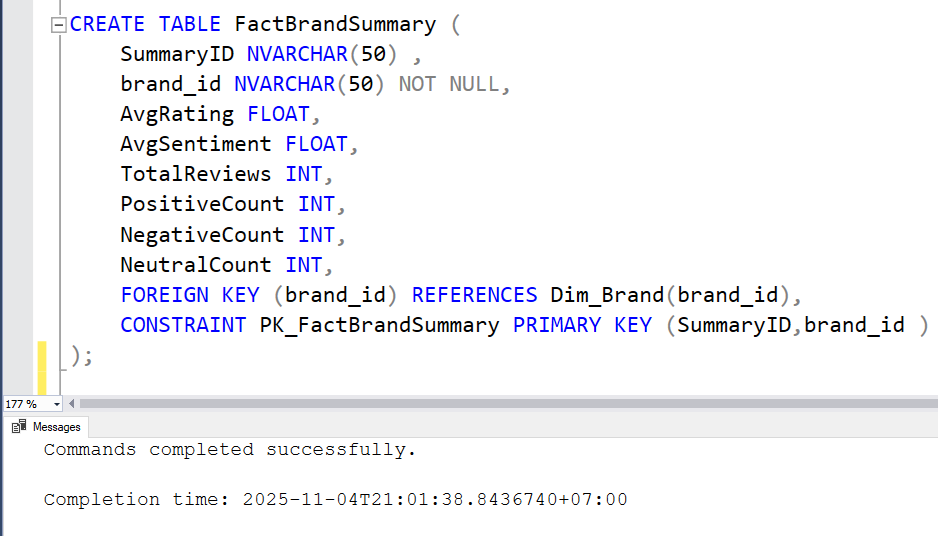
Hình 3. 6 Khởi tạo bảng: Dim\_Time

**Bước 2.5** Khởi tạo bảng: fact\_review



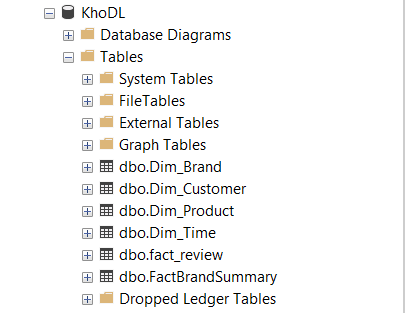
Hình 3. 7 Khởi tạo bảng: Fact\_review

**Bước 2.6** Khởi tạo bảng: FactBrandSummary



Hình 3. 8 Khởi tạo bảng: FactBrandSummary

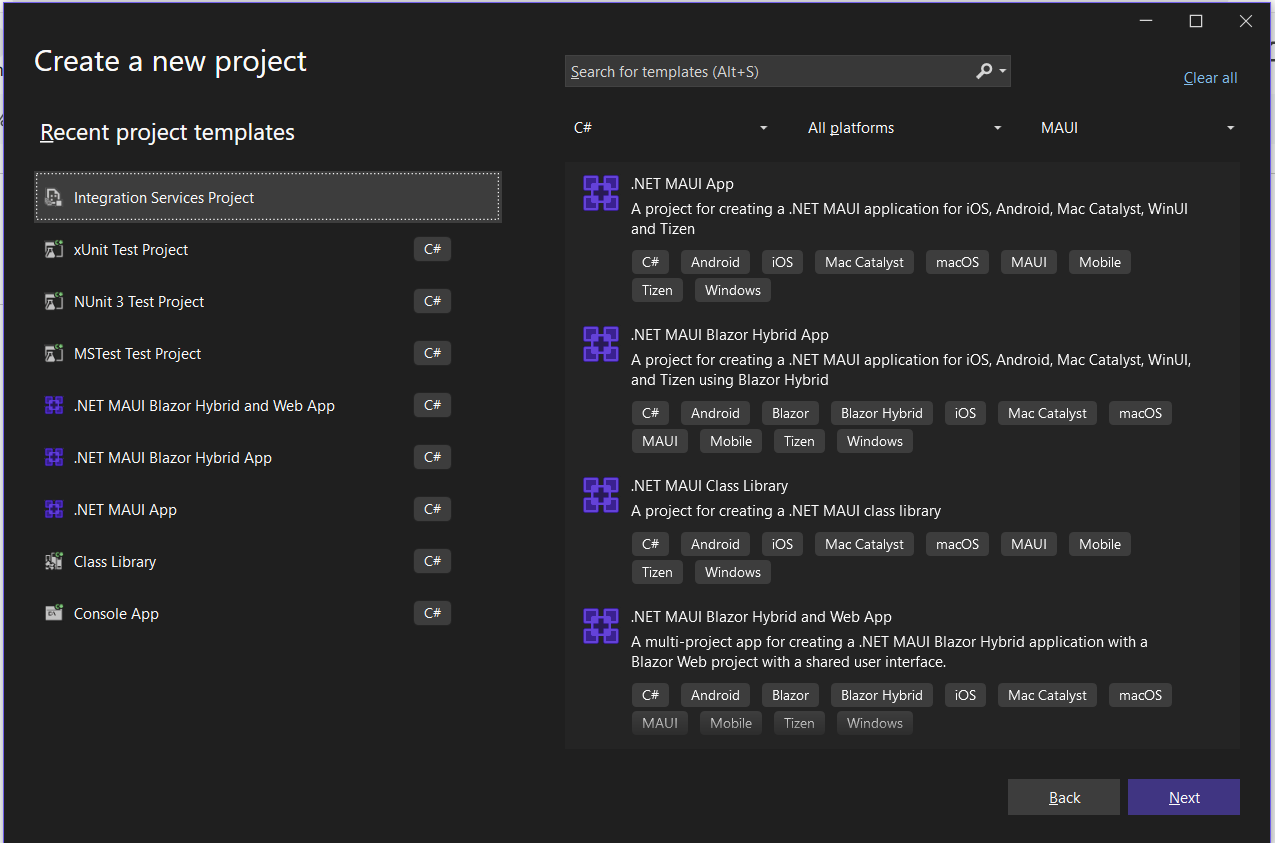
Kết quả:



Hình 3. 9 Kết quả

**3.2.3 Tạo mới project SSIS**

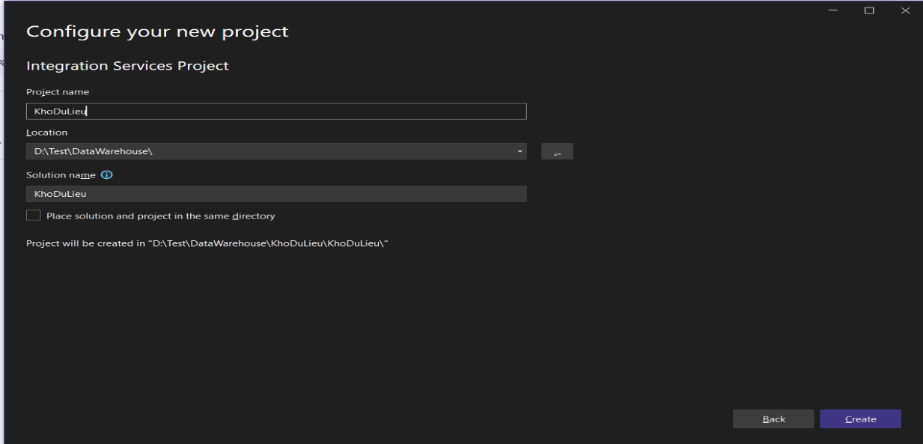
Bước 1: Mở Visual Studio 2022 và chọn “Create a new project”.



Hình 3. 10 Mở Visual Studio 2022

Bước 2: Chọn Integration Services Project và chọn Next.

Bước 3: Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn Create.



Hình 3. 11 Tạo Project

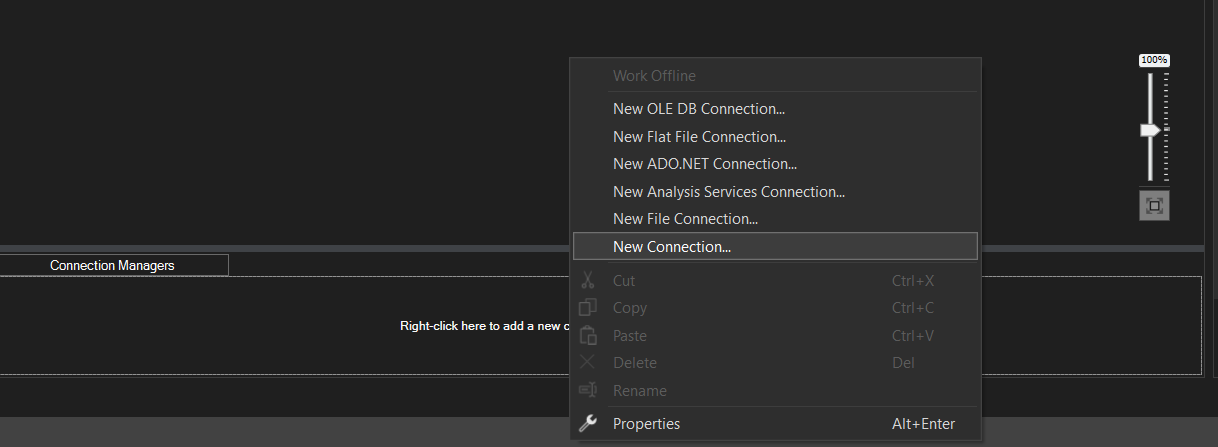
**3.2.4 Tạo bảng Dim và bảng Fact**

*Bước 1: Tạo các kết nối Connection Managers*

Excel Connection Manager

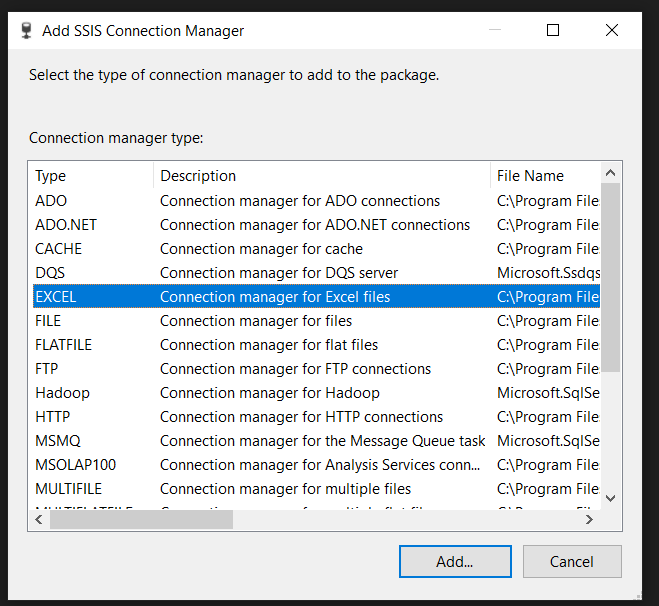
Connection Managers (dưới cùng giao diện), nhấp chuột phải > New

Connection Manager > chọn New Connection



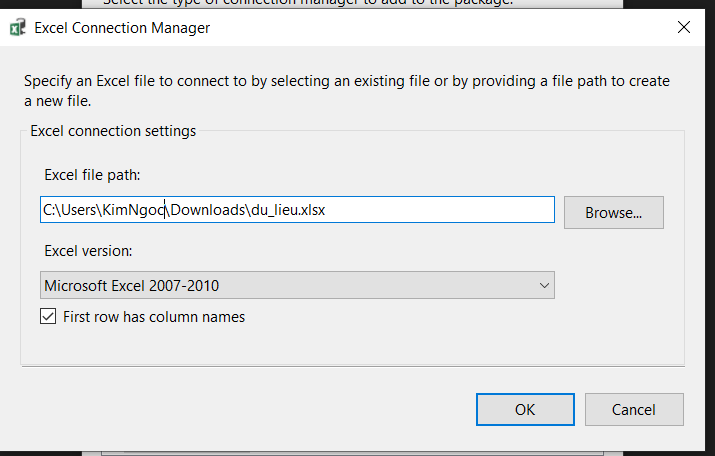
Hình 3. 12 Tạo các kết nối Connection Managers

Chọn Excel và nhấn Add.



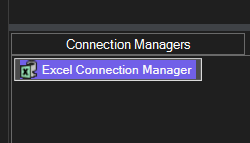
Hình 3. 13 Bảng Add SSIS Connection Manager

Nhấn Browse và dẫn đường dẫn tới file Excel chứa dữ liệu



Hình 3. 14 Bảng Excel Connection Manager

Kết quả: Thêm thành công Excel Connection



Hình 3. 15 Kết quả

OLE DB Connection Manager

* Nhấp chuột phải trong Connection Managers > New Connection Manager.
* Chọn loại OLE DB.
* Nhấn New: Provider: Chọn "SQL Server Native Client".

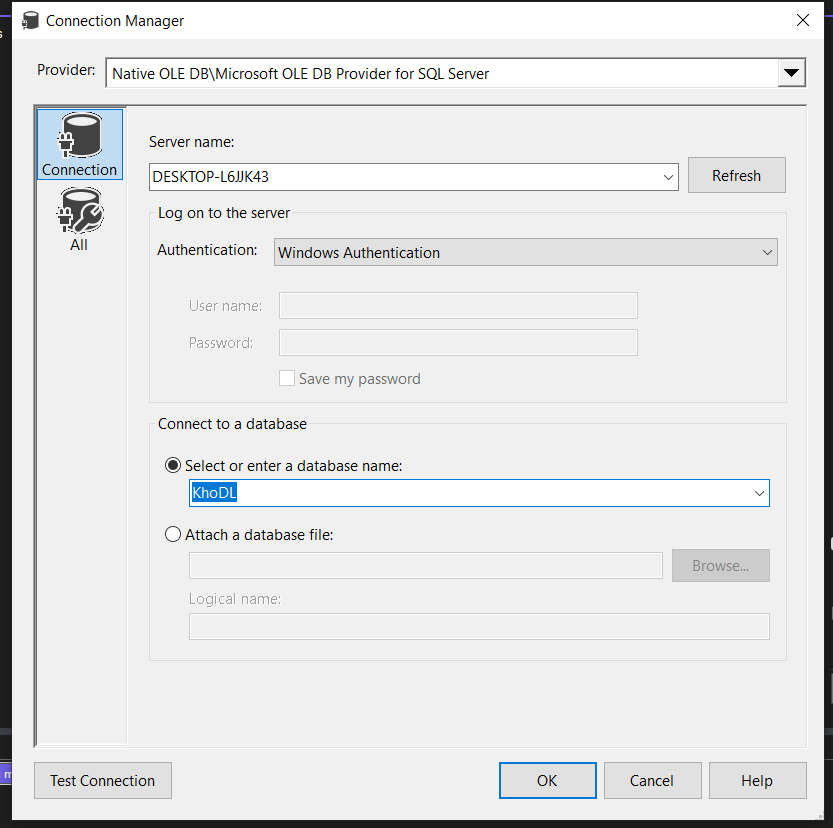


Hình 3. 16 Bảng Connection Manager

Server name: Nhập tên server

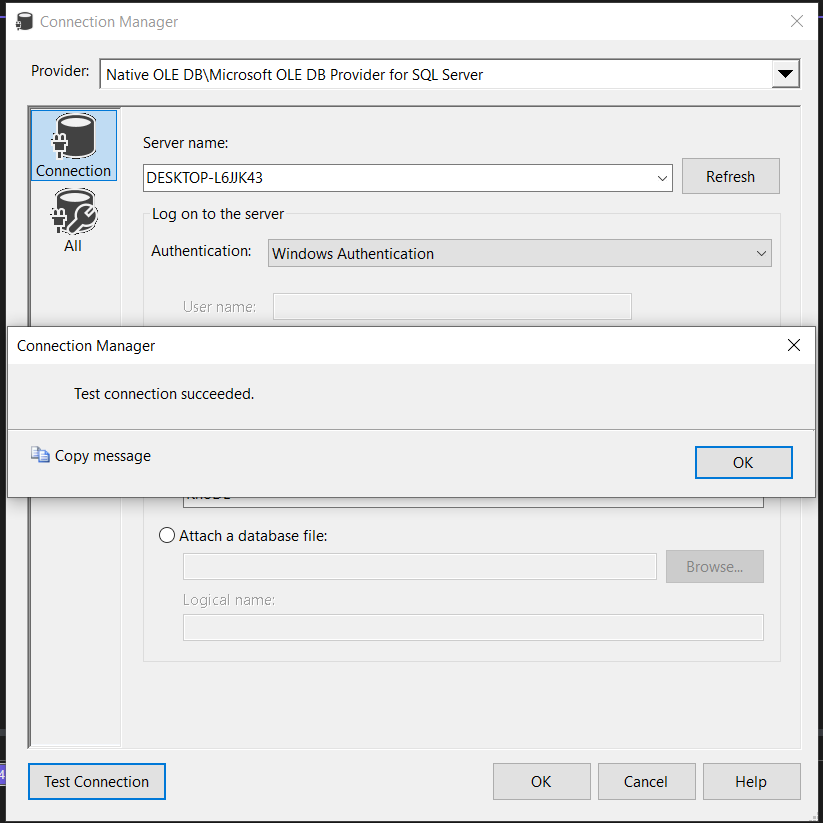
Authentication: Chọn Windows Authentication hoặc SQL Server Authentication.

Database name: Chọn cơ sở dữ liệu KhoDL đã tạo từ 3.2.2.



Hình 3. 17 Chọn cơ sở dữ liệu KhoDL

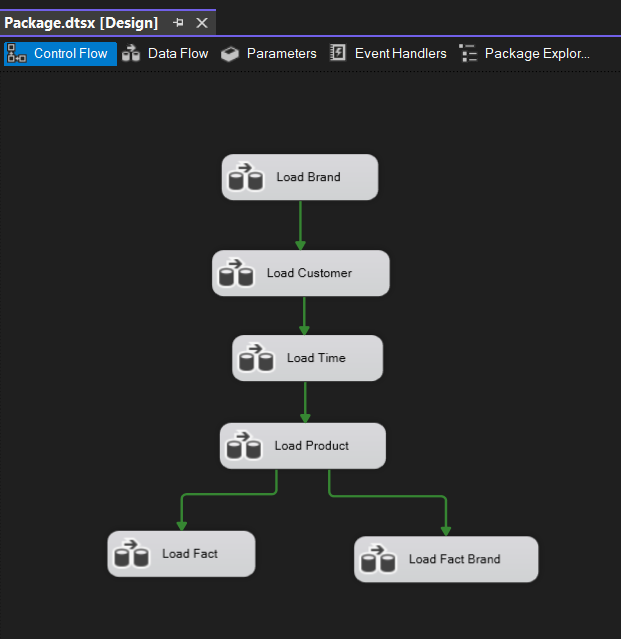
Nhấn Test Connection để kiểm tra kết nối thành công hay chưa, rồi nhấn OK.



Hình 3. 18 Chọn Test Connection

*Bước 2: Thiết kế Control Flow và Data Flow*

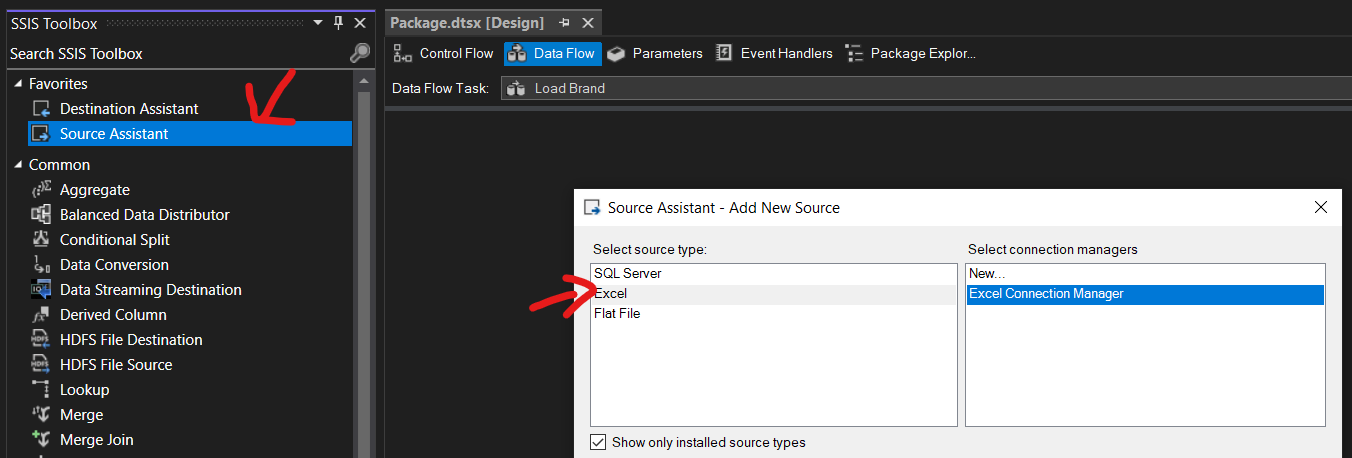
Thêm Data Flow Task. Trong tab Control Flow, kéo một Data Flow Task từ SSIS Toolbox vào khu vực thiết kế. Đặt tên : Load Brand, Load Customer, Load Time, Load Product, Load Fact, Load Fact Brand.



Hình 3. 19 Bảng Control Flow

*Bước 3: Trích xuất và tải dữ liệu từ sheet vào trong SQL Server*

Thêm Excel Source. Kéo Excel Source từ Toolbox vào khu vực thiết kế.



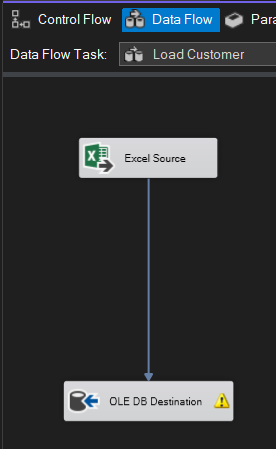
Hình 3. 20 Tải dữ liệu từ sheet vào trong SQL Server

Nhấp đúp để cấu hình

* OLE DB Connection Manager: Chọn ExcelConnection.
* Data access mode: Chọn "Table or view".
* Name of the Excel sheet: chọn bảng source (bảng nguồn) để đổ dữ liệu. Thêm OLE DB Destination
* Kéo OLE DB Destination từ Toolbox và kết nối mũi tên xanh từ Excel Source tới nó.

Nhấp đúp để cấu hình

* OLE DB Connection Manager: Chọn server\tên databases.
* Name of the table or view: Chọn bảng vừa chọn trong “Excel source”
* Nhấn Mappings để kiểm tra ánh xạ cột đảm bảo các cột khớp đúng, rồi nhấn OK. Kết quả: Ta được Source là Excel và Destination là SQL Server



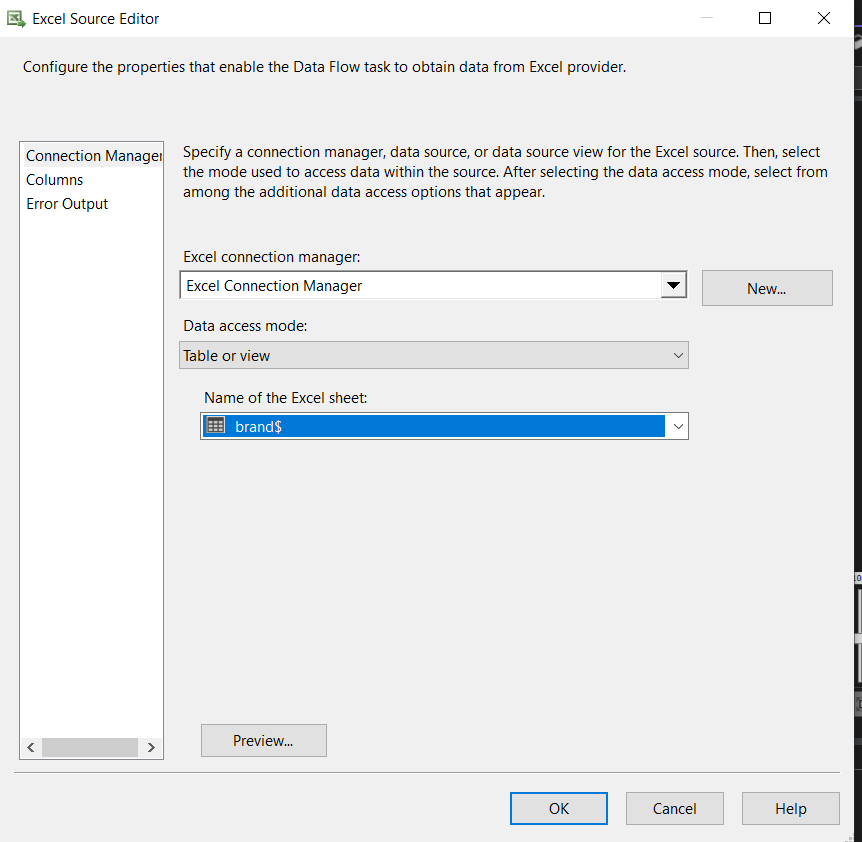
Hình 3. 21 Kết quả

Thêm các công cụ:

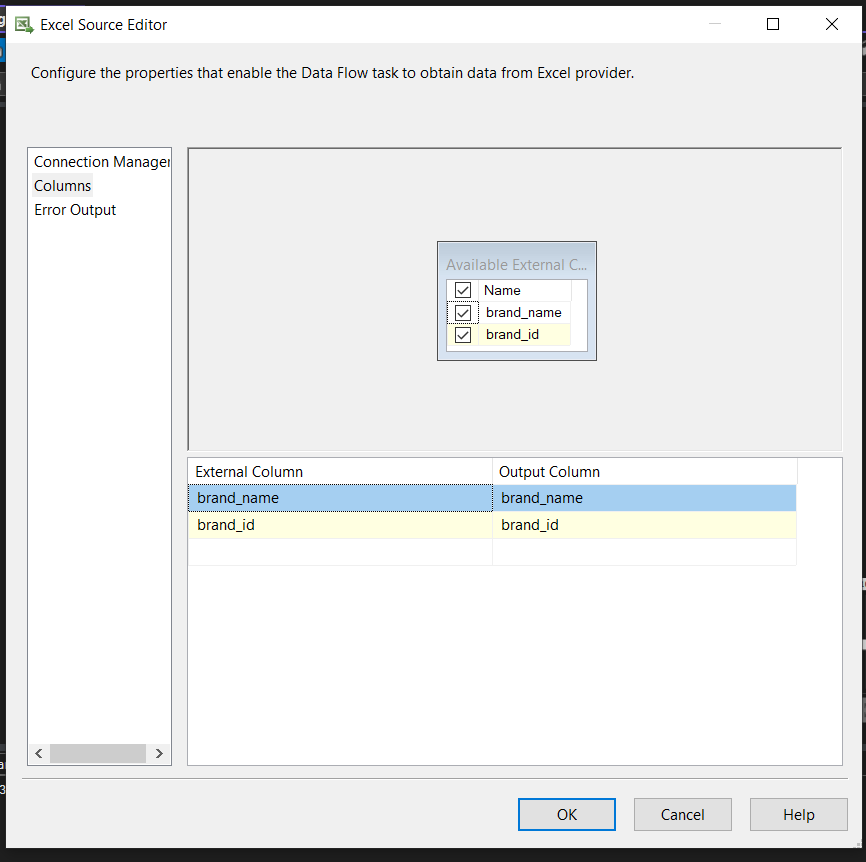
* Sort: Công cụ này được sử dụng để sắp xếp dữ liệu theo một hoặc nhiều tiêu chí nhất định, như theo thứ tự tăng dần hoặc giảm dần.
* Lookup: Công cụ này dùng để tìm kiếm và lấy thông tin từ một bảng dữ liệu khác dựa trên một khóa xác định.
* Derrived Column: Thường được sử dụng để xác định và điều chỉnh dữ liệu trong một cột cụ thể, giúp người dùng truy cập và phân tích thông tin dễ dàng hơn.
* Aggregate: Công cụ này cho phép thực hiện các phép toán tổng hợp trên một tập hợp dữ liệu, như tính tổng, trung bình, đếm số lượng.

*Bước 3.1: Cấu hình bảng brand và table DIM\_Brand*

Cấu hình Excel Source

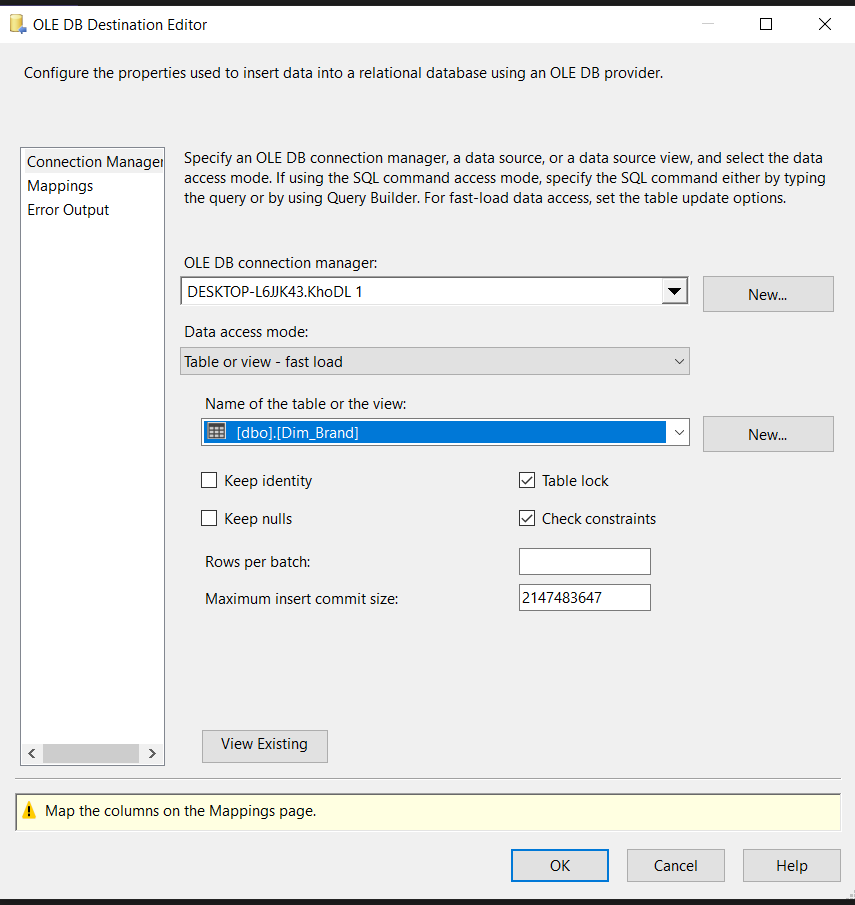


Hình 3. 22 Cấu hình Excel Source



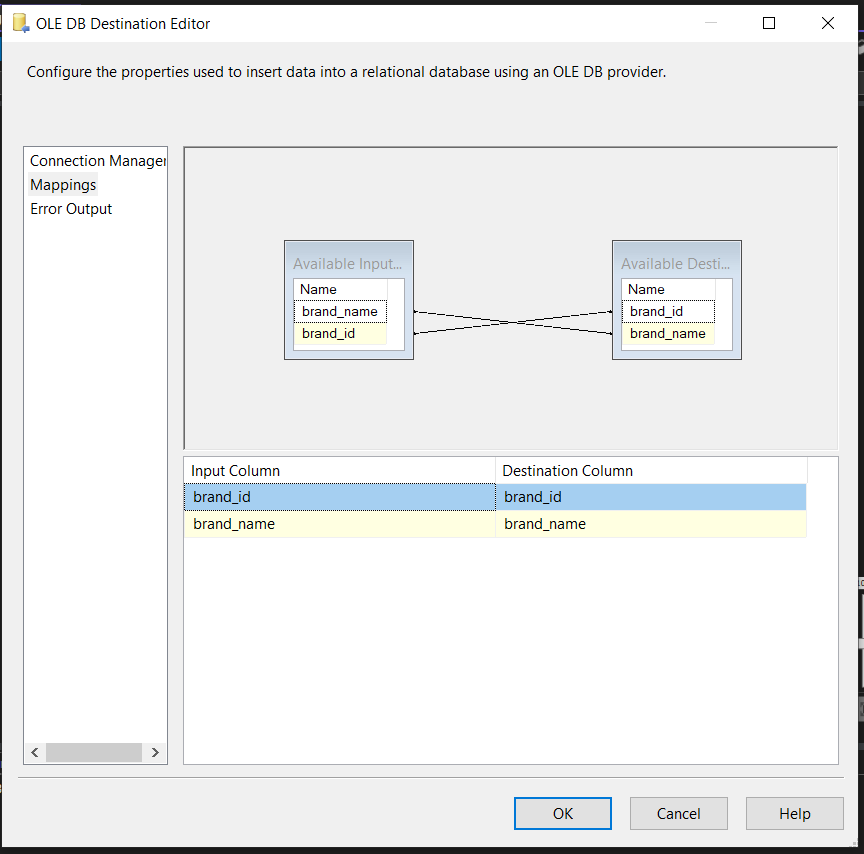
Hình 3. 23 Cấu hình bảng Dim\_Brand

Cấu hình SQL Server Destination



Hình 3. 24 Cấu hình SQL Server Destination

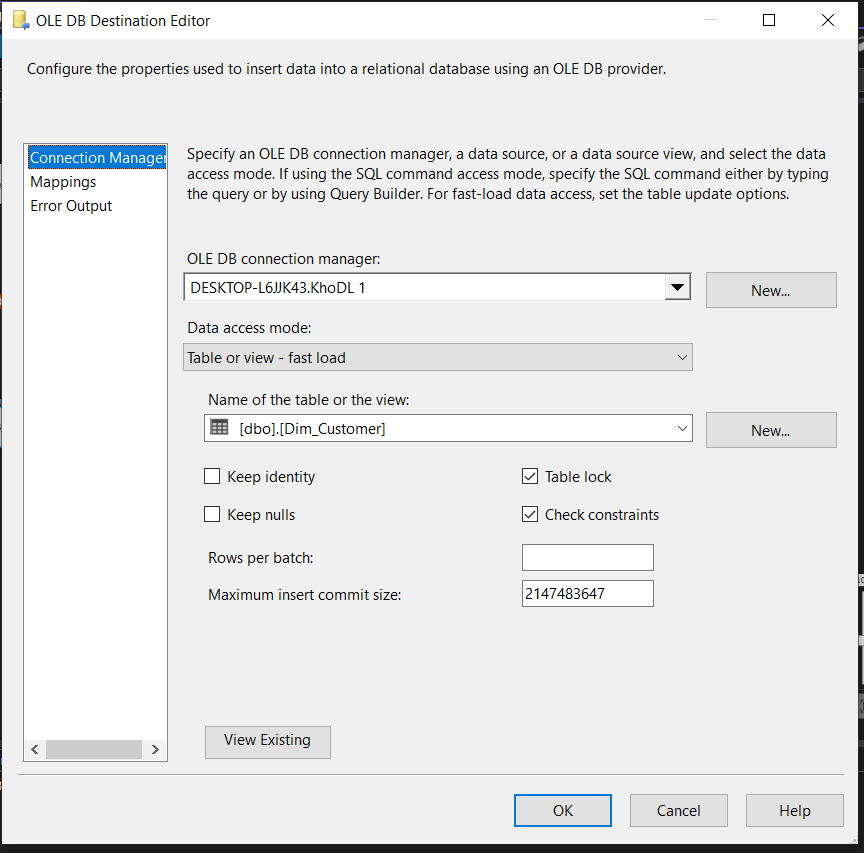
Ánh xạ các cột từ Excel Source vào bảng



Hình 3. 25 Ánh xạ từ các cột

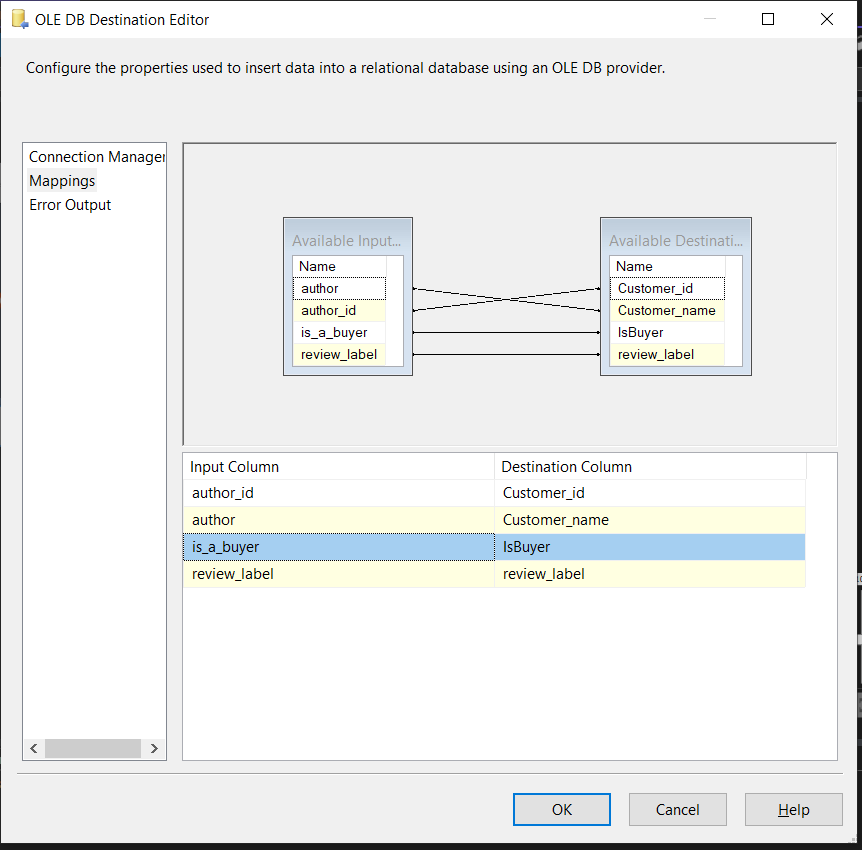
*Bước 3.2: Cấu hình bảng customer và table DIM\_Customer*

Cấu hình SQL Server Destination



Hình 3. 26 Cấu hình SQL Server Destination

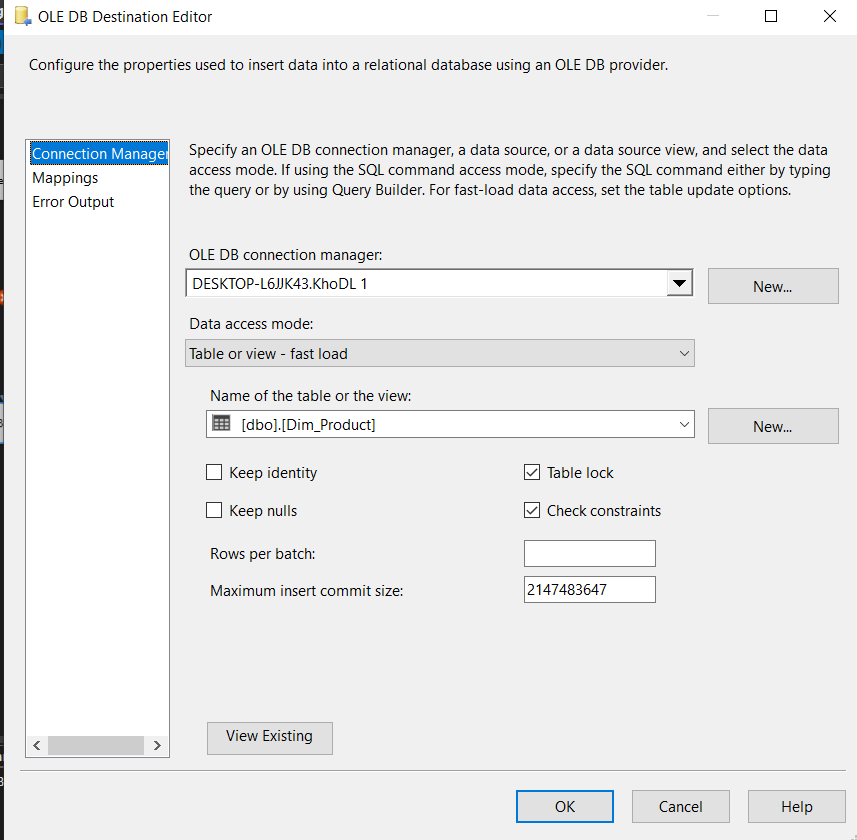
Do tên cột ở Excel Source và Database Table khác nhau tên ta phải cấu hình mapping các cột thủ công cho bảng DIM\_Customer



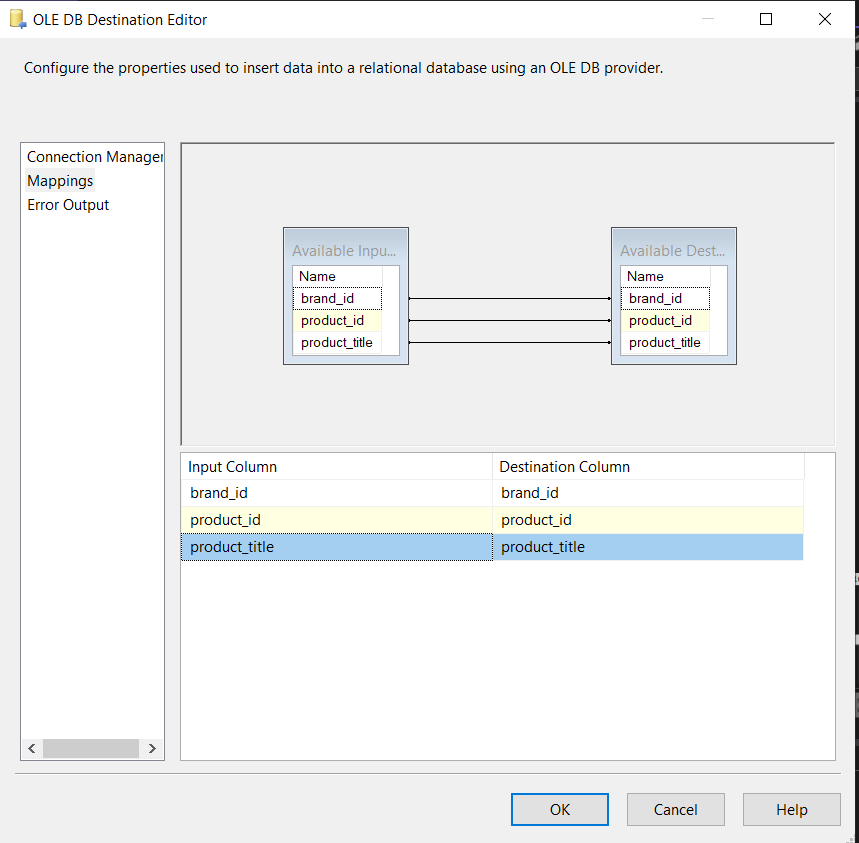
Hình 3. 27 Ánh xạ từ các cột

*Bước 3.3: Cấu hình bảng product và table DIM\_Product*

Cấu hình SQL Server Destination



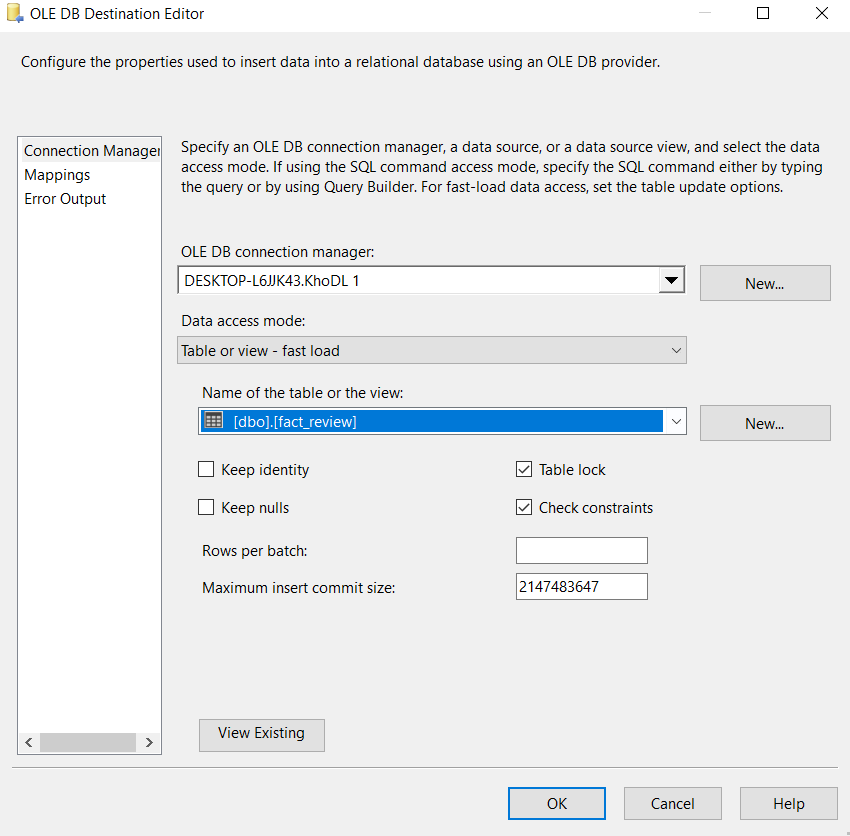
Hình 3. 28 Cấu hình SQL Server Destination



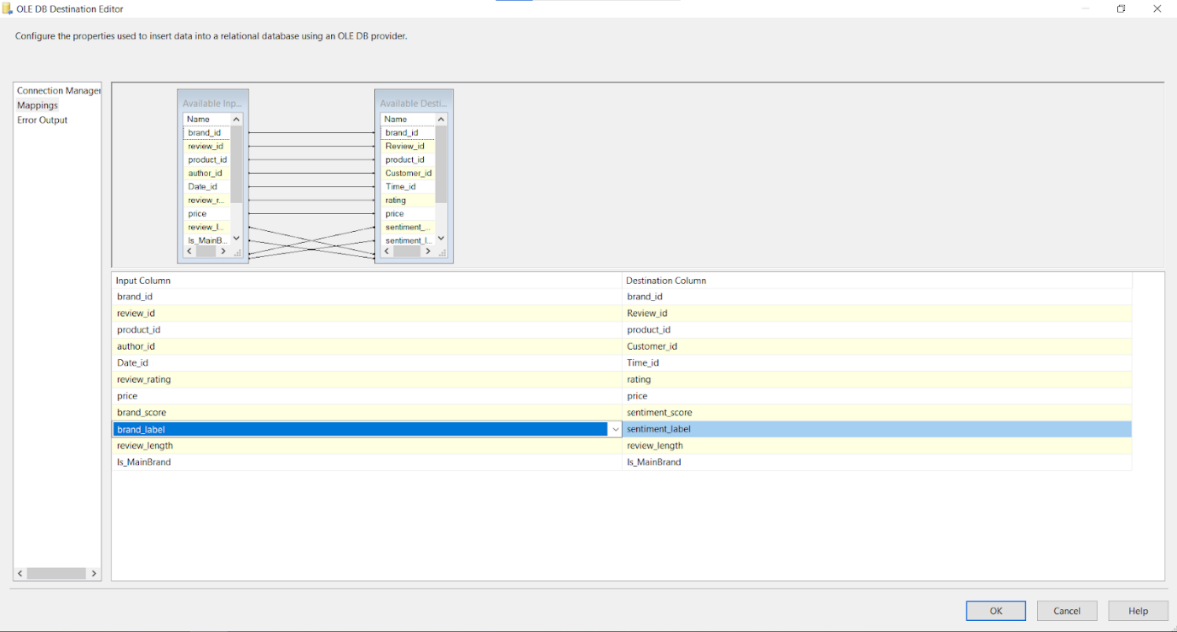
Hình 3. 29 Ánh xạ từ các cột

*Bước 3.4: Cấu hình bảng fact và table fact\_review*

Cấu hình SQL Server Destination



Hình 3. 30 Cấu hình SQL Server Destination

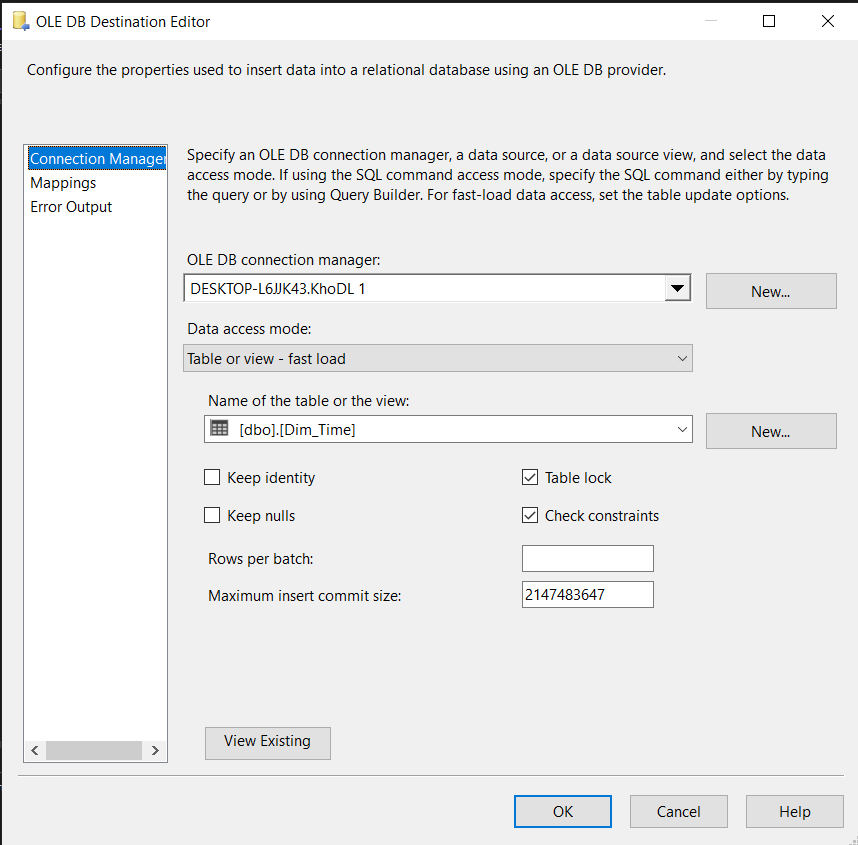


Hình 3. 31 Ánh xạ từ các cột

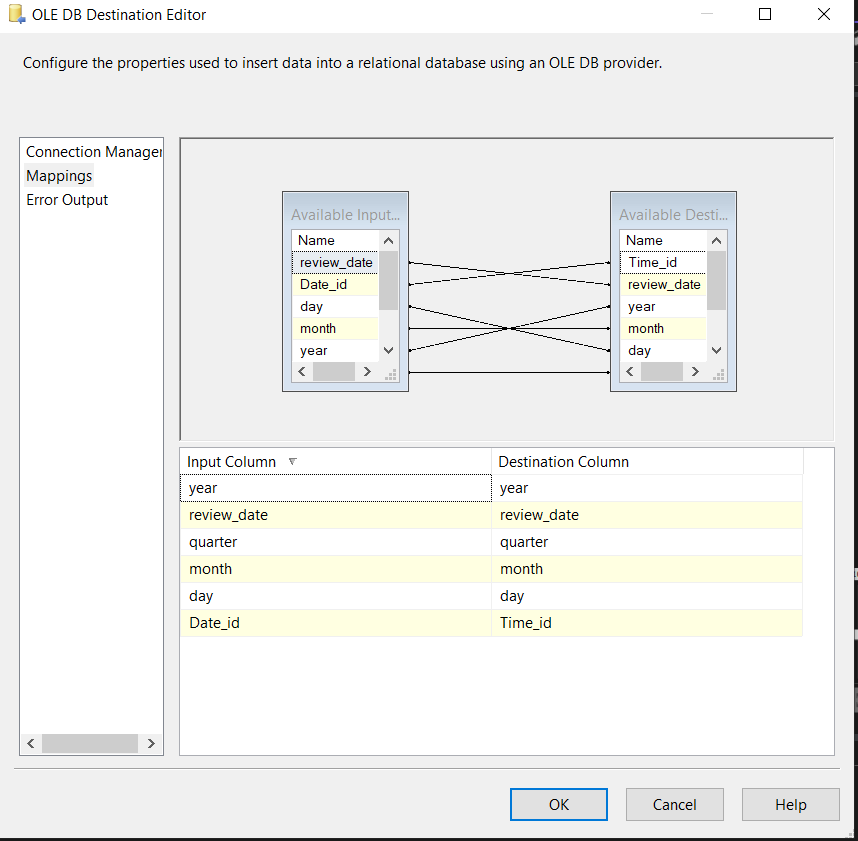
Ánh xạ các cột từ Excel Source vào bảng

*Bước 3.5: Cấu hình bảng time và table Dim\_Time*

Cấu hình SQL Server Destination



Hình 3. 32 Cấu hình SQL Server Destination

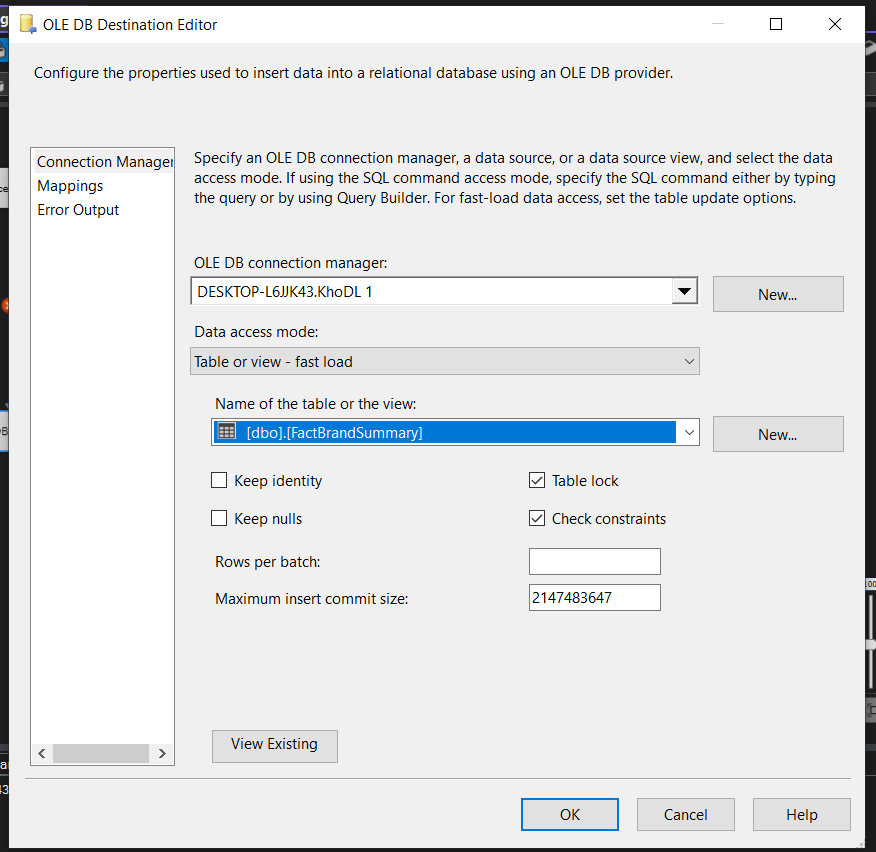


Hình 3. 33 Ánh xạ từ các cột

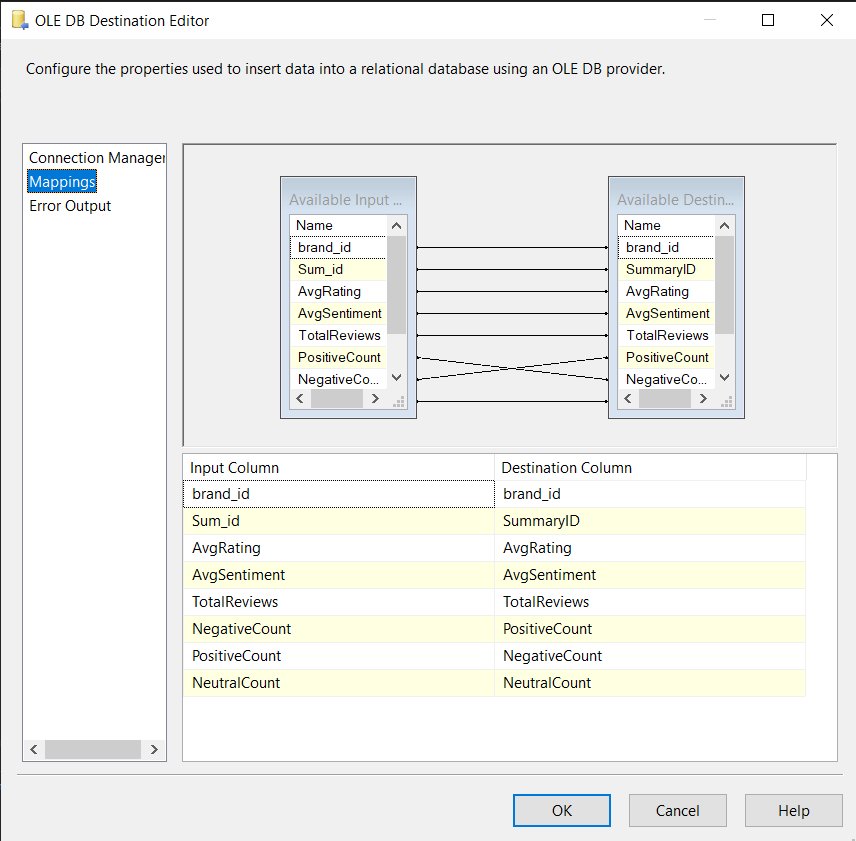
Ánh xạ các cột từ Excel Source vào bảng

*Bước 3.6: Cấu hình bảng factbrand và table FactBrandSummary*

Cấu hình SQL Server Destination



Hình 3. 34 Cấu hình SQL Server Destination

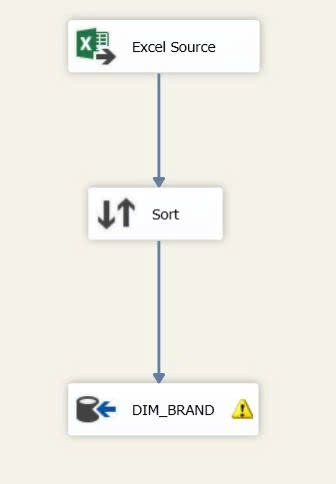


Hình 3. 35 Ánh xạ từ các cột

Ánh xạ các cột từ Excel Source vào bảng

**3.2.5 Chạy dự án SSIS**

Load\_DimBrand. Dùng Sort để xắp sếp: BrandID. Đồng thời loại bỏ các dòng trùng nhau dựa trên toàn bộ các cột được sort.



Hình 3. 36 Load\_DimBrand

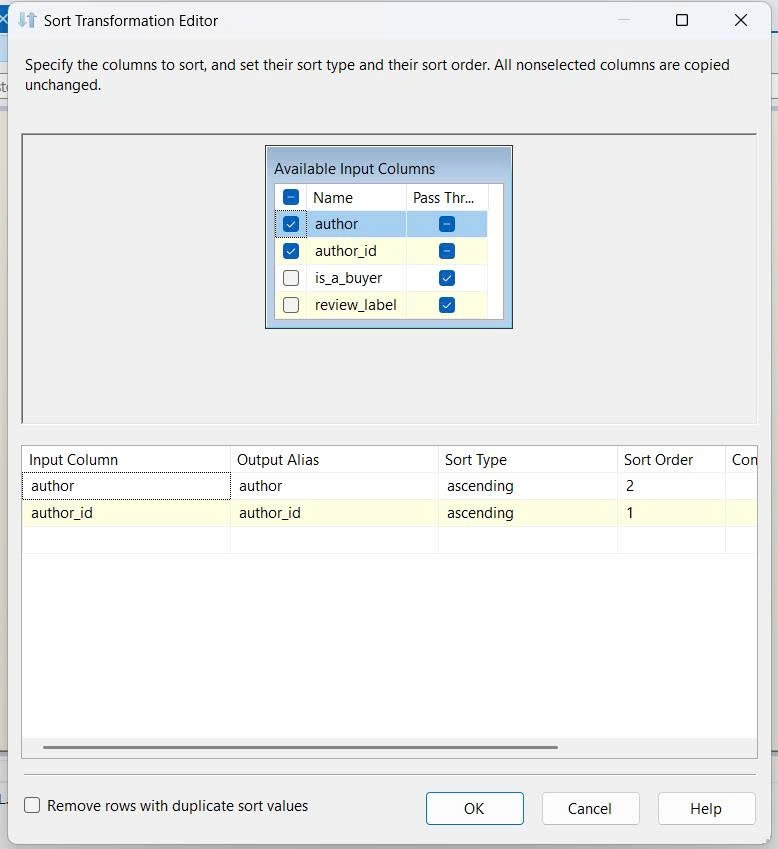


Hình 3. 37 Bảng Sort

Load\_DimCustomer. Dùng Sort để sắp xếp: Customer ID. Đồng thời loại bỏ các dòng trùng nhau dựa trên toàn bộ các cột được sort.

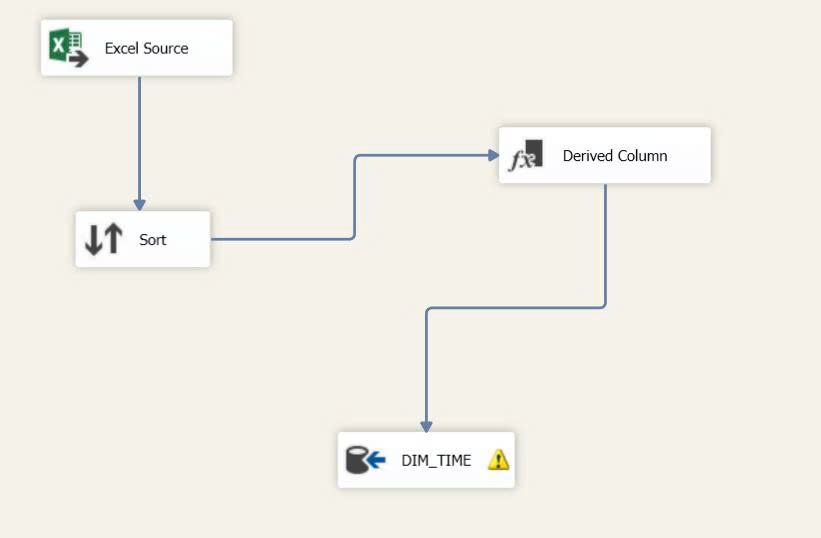


Hình 3. 38 Load\_DimCustomer

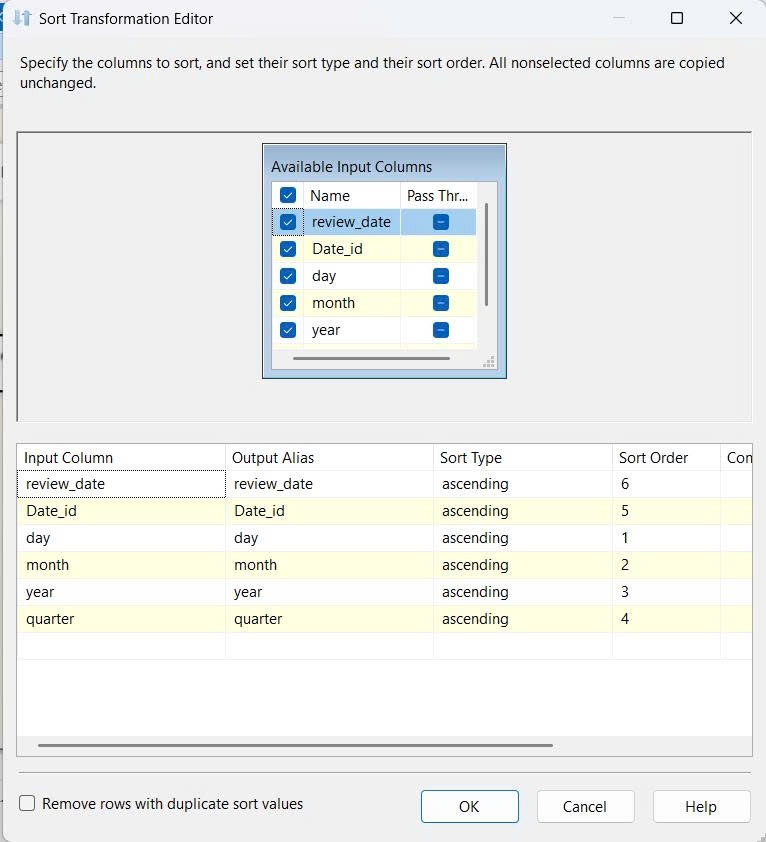


Hình 3. 39 Bảng Sort

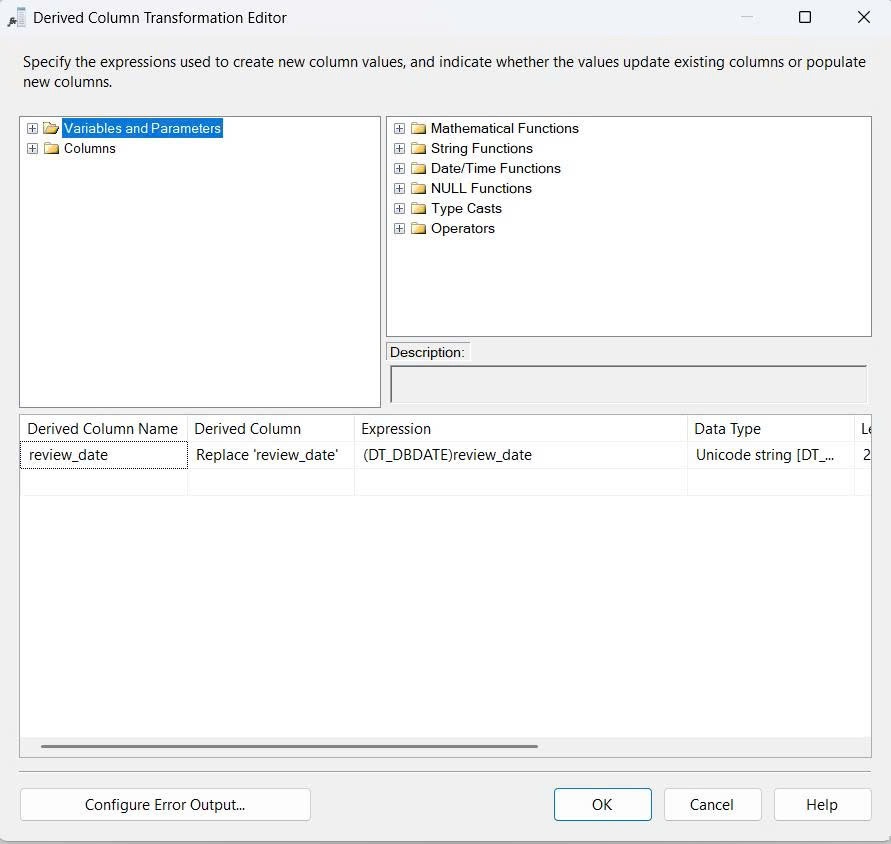
Load\_DimTime. Dùng Sort để xắp xếp: Date\_ID. Dùng Derived Column chuyển đổi review\_date từ dạng chuỗi → dạng ngày (DT\_DBDATE). Đồng thời loại bỏ các dòng trùng nhau dựa trên toàn bộ các cột được sort



Hình 3. 40 Load\_DimTime

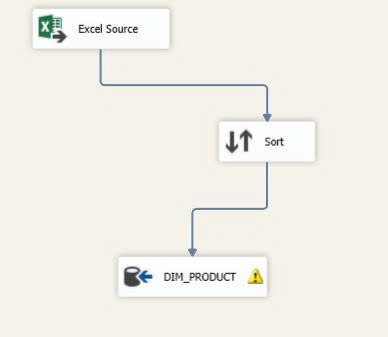


Hình 3. 41 Bảng Sort

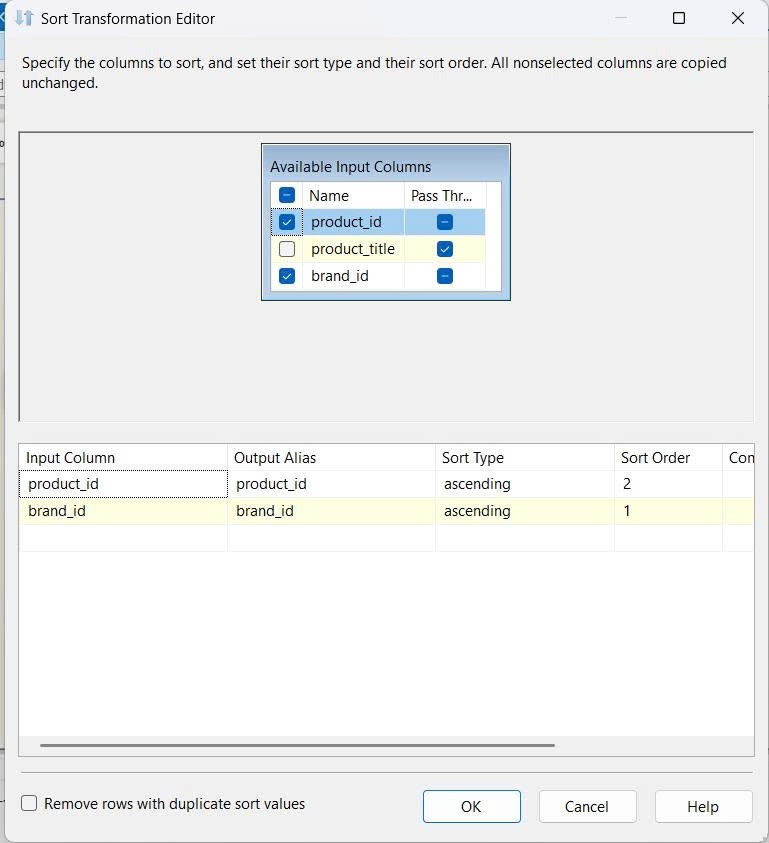


Hình 3. 42 Derived Column Transformation Editor

Load\_DimProduct. Dùng Sort để xắp xếp: ProductID. Đồng thời loại bỏ các dòng trùng nhau dựa trên toàn bộ các cột dùng được sort.

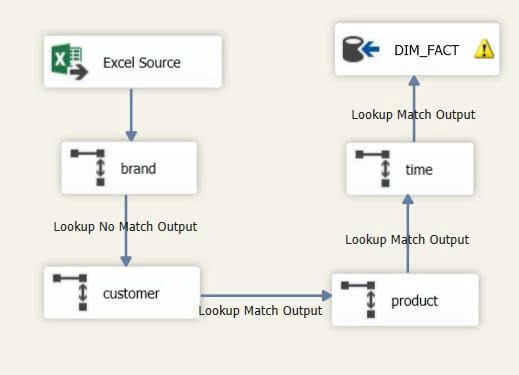


Hình 3. 43 Load\_DimProduct



Hình 3. 44 Bảng Sort

Load\_Fact



Hình 3. 45 Load\_Fact

* Lookup-Brand

Dùng để đối chiếu brand\_id từ excel với bảng DimBrand để lấy Brand\_id (Khóa chính). Lookup Match Output: nếu tìm thấy khớp sẽ trả lại Brand\_id và đẩy tiếp dữ liệu qua các bước sau.

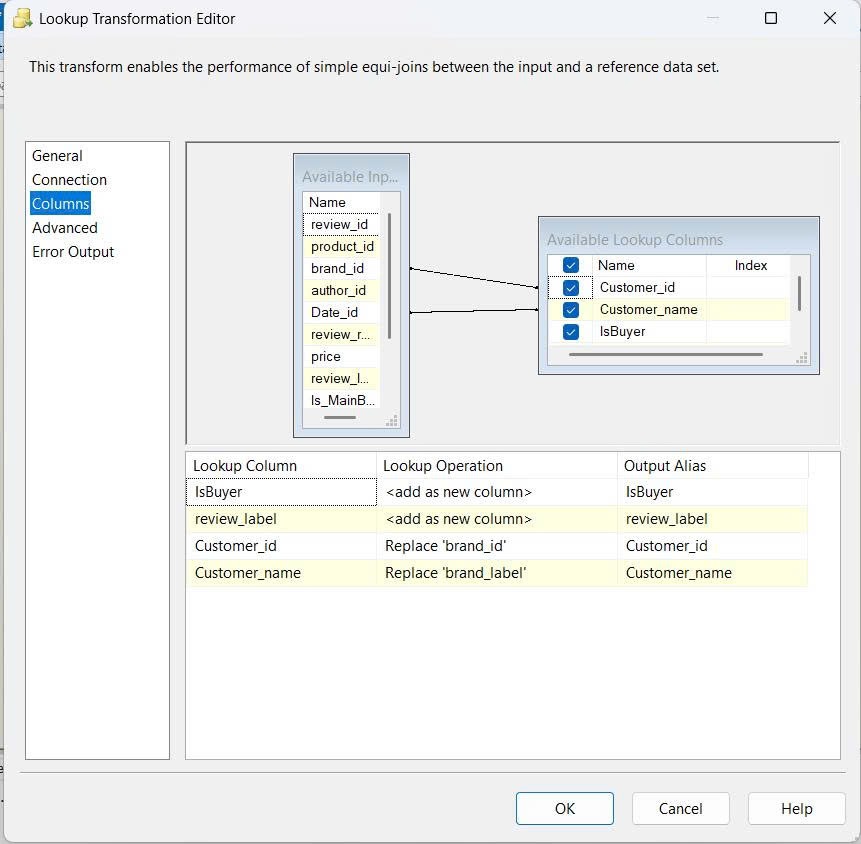


Hình 3. 46 Lookup\_Brand

* Lookup-Customer

Đối chiếu customer\_id với DimCustomer

Trả về Customer\_ID

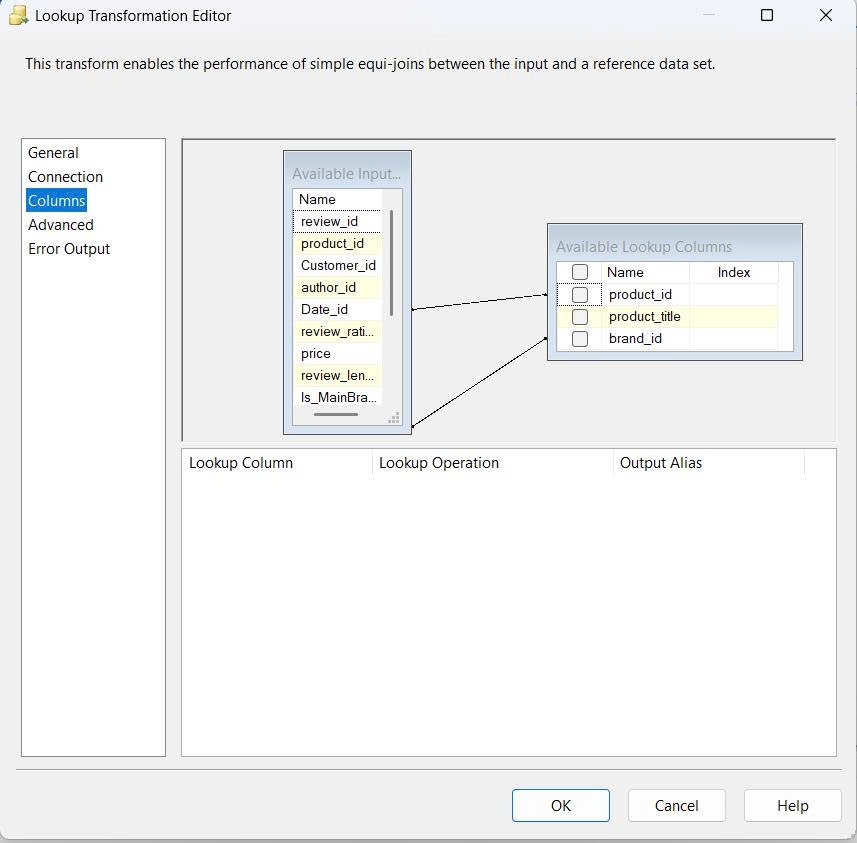


Hình 3. 47 Lookup\_Customer

* Lookup - Product

Đối chiếu product\_id hoặc product\_name với bảng DimProduct

Trả về Product\_ID



Hình 3. 48 Lookup\_Product

* Lookup - Time

Đối chiếu Time\_id với bảng DimTime

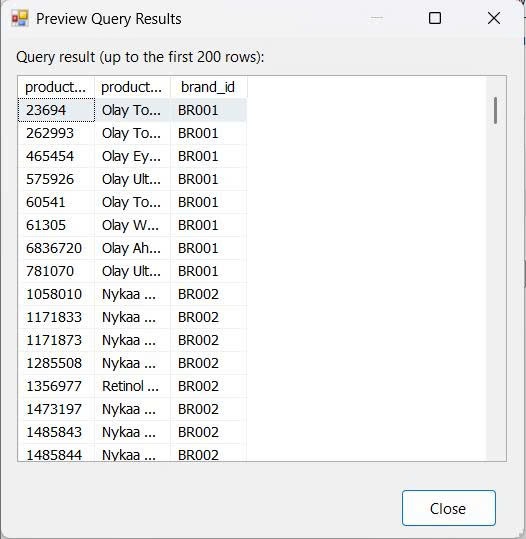
Trả về Time\_id



Hình 3. 49 Lookup\_Time

**3.2.6 Kiểm tra dữ liệu các bảng**

Kiểm tra dữ liệu bảng Dim\_product



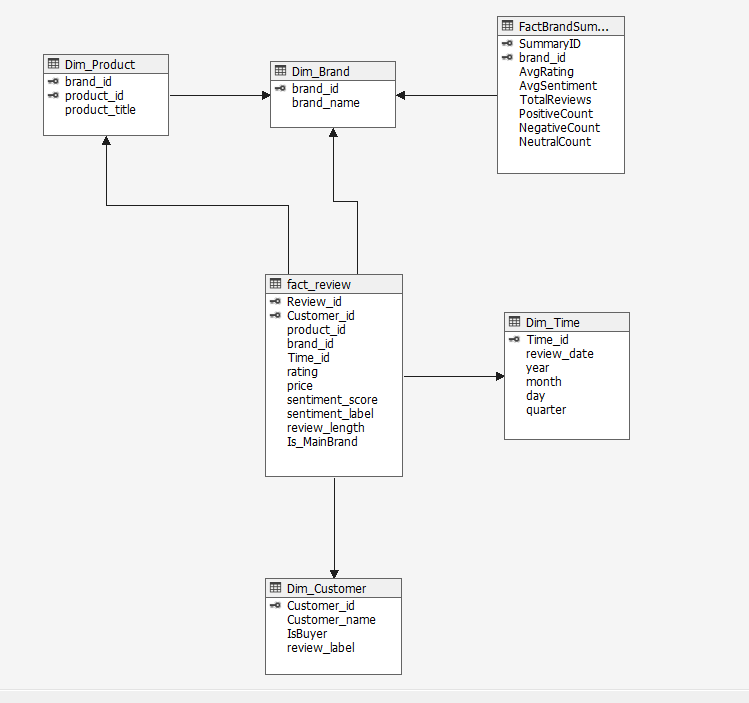
Hình 3. 50 Kiểm tra dữ liệu bảng Dim\_product

Kiểm tra dữ liệu bảng Dim\_FactBrand



Hình 3. 51 Kiểm tra dữ liệu bảng Dim\_FactBrand

**3.2.7 Lược đồ sau khi hoàn thành**



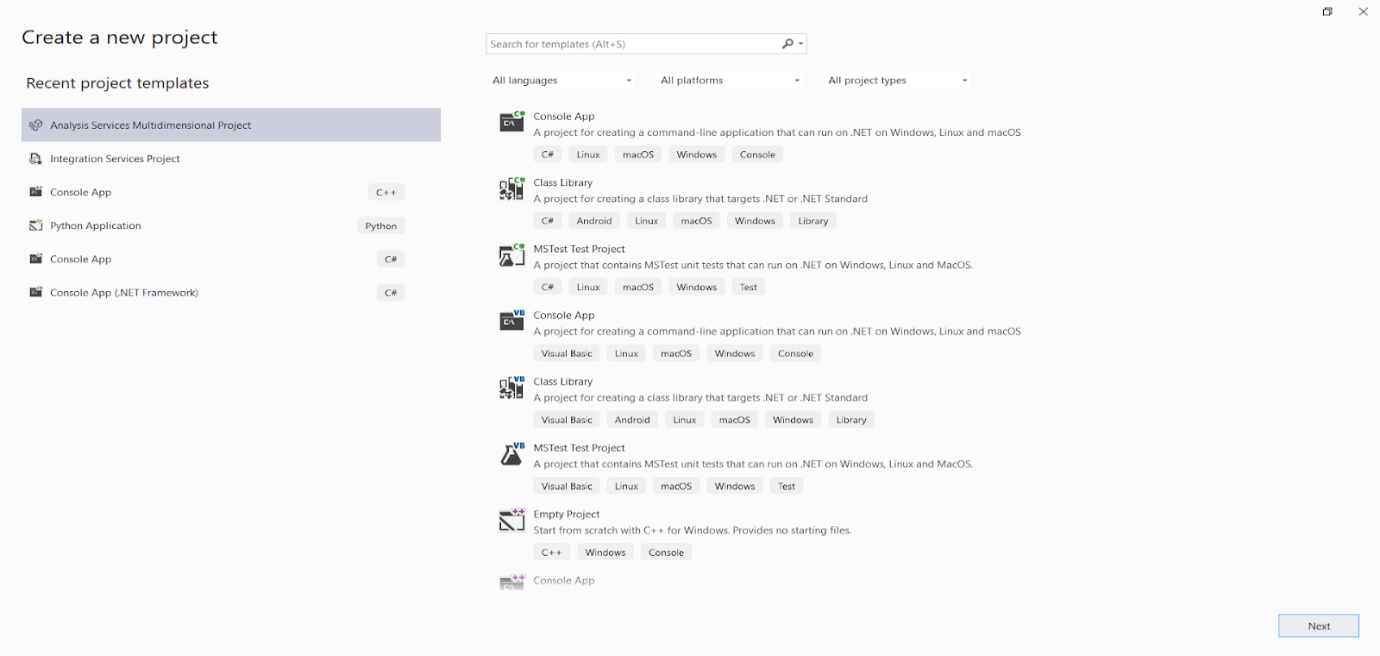
Hình 3. 52 Lược đồ sau khi hoàn thành

**3.3**    **Phân tích dữ liệu SSAS**

Mở Visual Studio 2022.

Vào menu File → New → Project.

Trong danh sách loại dự án, chọn Analysis Services Multidimensional and Data Mining Project. Đặt tên và vị trí lưu trữ phù hợp, sau đó nhấn Create để khởi tạo dự án.

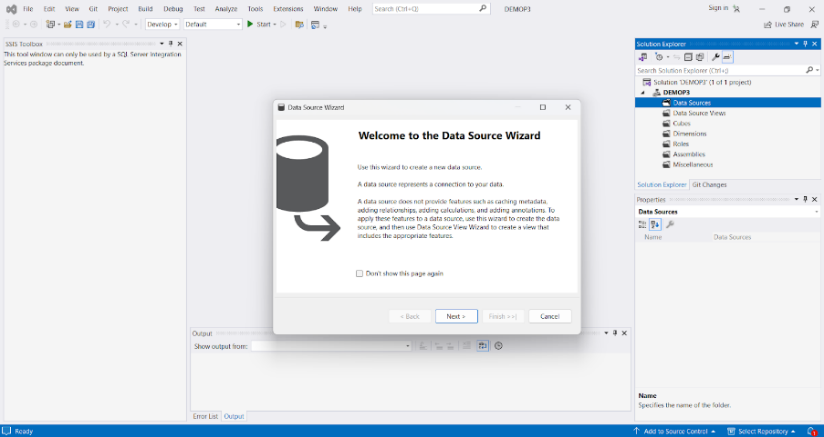


Hình 3. 53 Tạo Project mới

**3.3.1 Xác định dữ liệu nguồn (Data Sources)**

Bước 1:

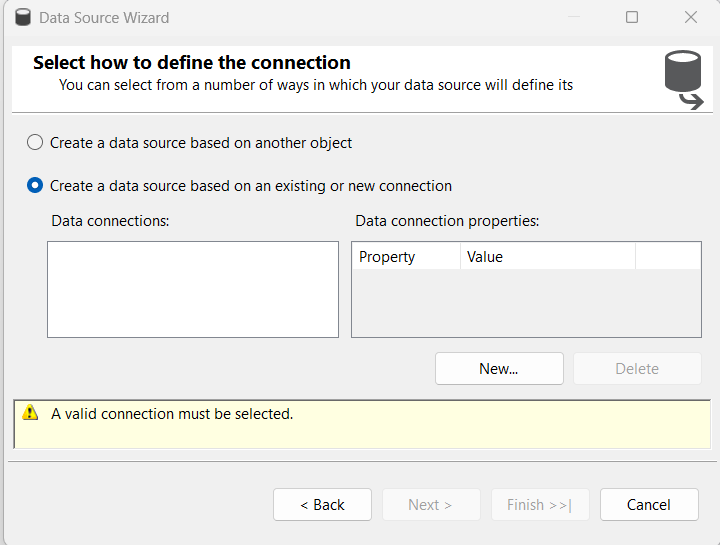
* Trong cửa sổ Solution Explorer, nhấn chuột phải vào thư mục Data Sources → chọn New Data Source.
* Hộp thoại Data Source Wizard mở ra, chọn Next để bắt đầu cấu hình.



Hình 3. 54 Data Source Wizard

Bước 2:

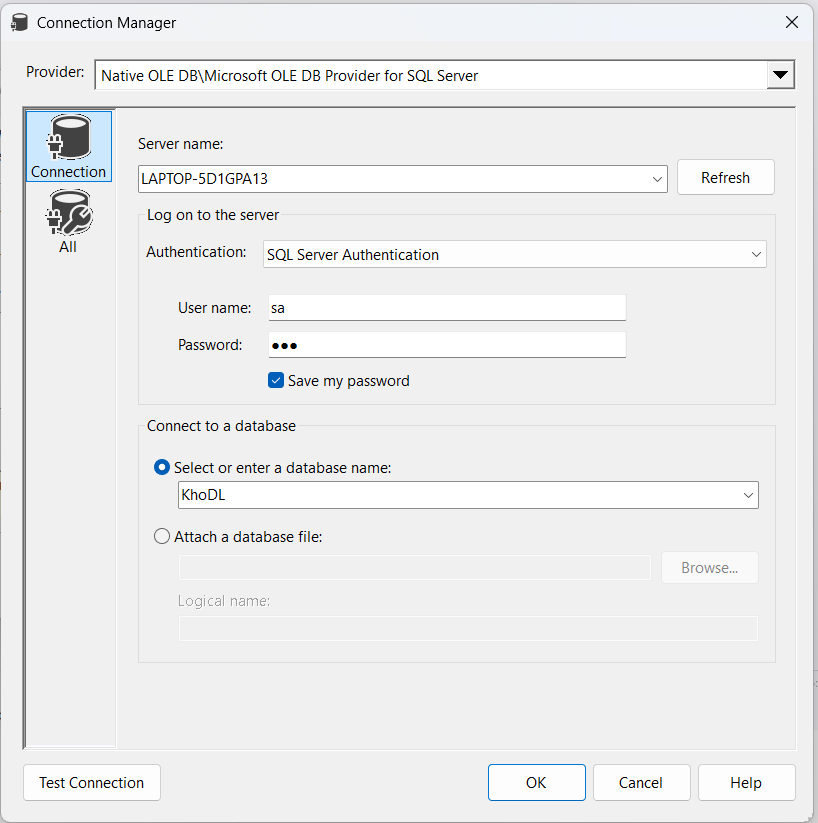
* Ở màn hình Select how to define the connection, chọn tùy chọn
* Create a data source based on an existing or new connection, sau đó nhấn New... để tạo kết nối tới cơ sở dữ liệu mà bạn đã xây dựng trong quá trình SSIS.



Hình 3. 55 Create a data source based on an existing or new connection

Bước 3: Hộp thoại Connection Manager xuất hiện:

* Nhập Server name (tên máy chủ SQL Server).
* Chọn Database name – chính là kho dữ liệu (Data Warehouse) bạn đã tạo ở bước ETL.
* Nhấn Test Connection để kiểm tra.
* Khi kết nối thành công, chọn OK để lưu.
* Sau đó, chọn lại Data Source vừa tạo → nhấn Next để tiếp tục.



Hình 3. 56 Hộp thoại Connection Manager

Bước 4:

* Tại bước Impersonation Information, chọn phương thức truy cập dữ liệu là “Use the service account” để SSAS sử dụng tài khoản dịch vụ của máy chủ khi truy cập dữ liệu.
* Nhấn Next.

Bước 5:

- Nếu tất cả chính xác, nhấn Finish để hoàn tất việc định nghĩa Data Source.

**3.3.2 Xác định khung nhìn dữ liệu nguồn (Data Source Views)**

Bước 1:

* Trong cửa sổ Solution Explorer, nhấp chuột phải vào thư mục Data Source Views → chọn New Data Source View.
* Hộp thoại Data Source View Wizard sẽ xuất hiện để hướng dẫn bạn cấu hình.

Bước 2:

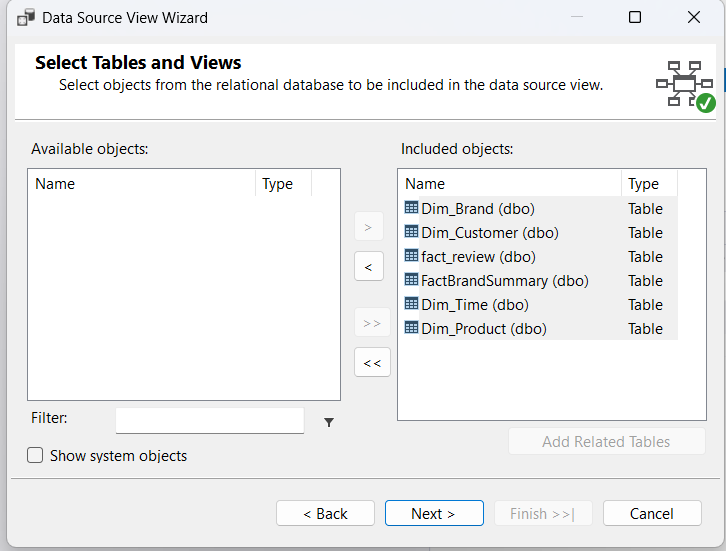
* Trong cửa sổ chào mừng của Data Source View Wizard, nhấn Next để bắt đầu quá trình thiết lập.

Bước 3:

* Chọn Data Source mà bạn đã tạo ở bước trước (trong phần 3.3.1),  
   sau đó nhấn Next để tiếp tục.

Bước 4:

* Màn hình Select Tables and Views hiển thị danh sách các bảng trong cơ sở dữ liệu.
* Tại đây, chọn các bảng cần sử dụng cho mô hình phân tích,  
   sau đó nhấn vào nút > để thêm các bảng này vào Included objects.



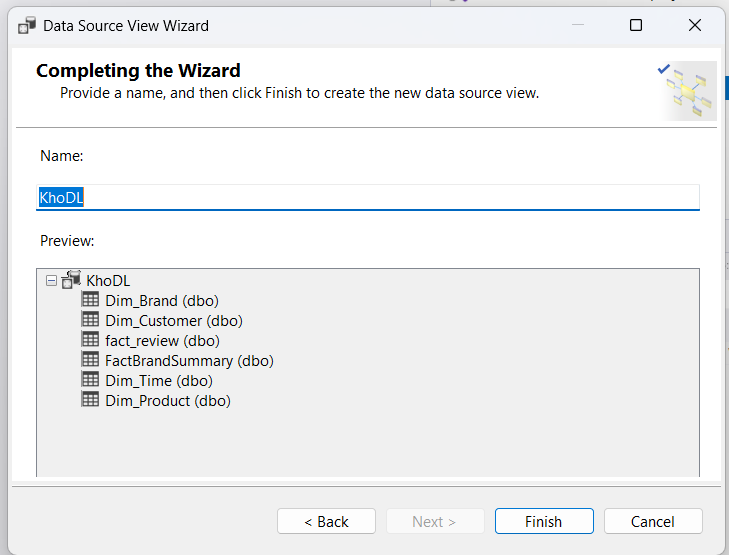
Hình 3. 57 Select Tables and Views

Bước 5:

* Nhấn nút Add Related Tables để tự động thêm tất cả các bảng liên quan — đặc biệt là các bảng chiều (Dim\_) có quan hệ với bảng sự kiện (Fact\_).
* Sau đó nhấn Next để tiếp tục.

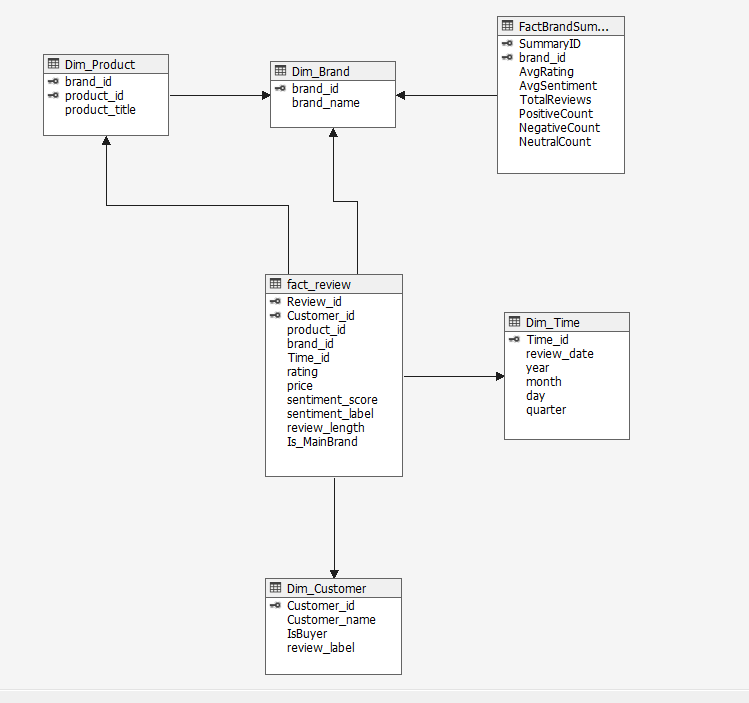
Bước 6:

* Kiểm tra lại danh sách các bảng đã chọn.
* Nếu mọi thứ chính xác, nhấn Finish để hoàn tất quá trình tạo Data Source View (DSV).



Hình 3. 58 Hoàn tất quá trình tạo Data Source View

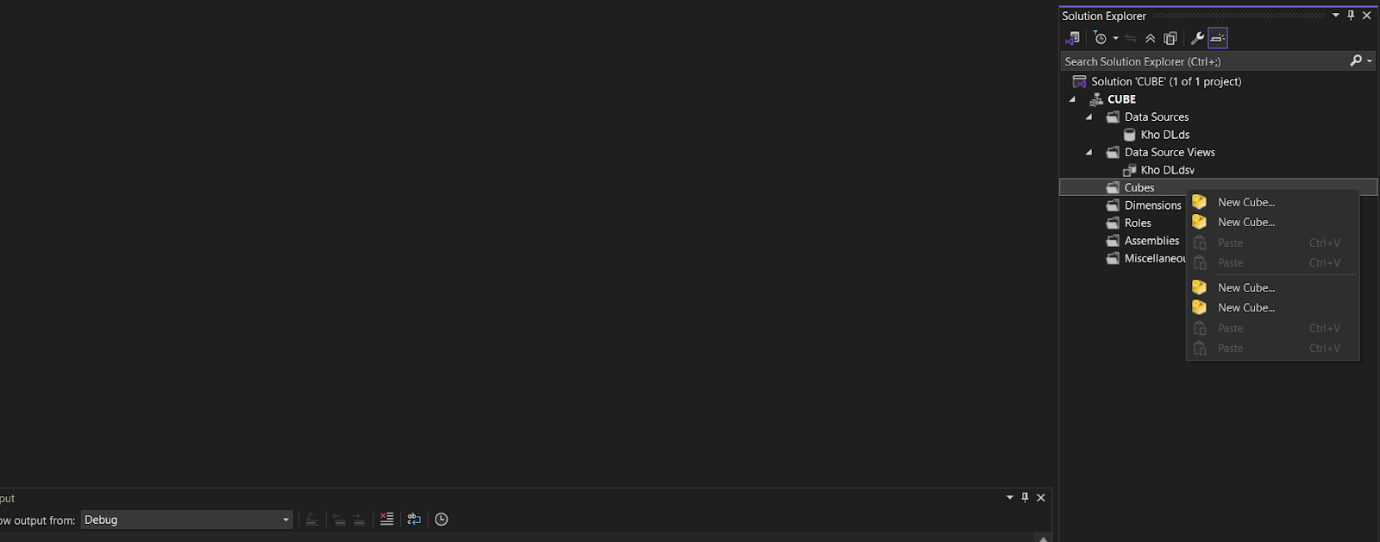
Sau khi kết thúc quá trình này, ta sẽ được data source view như hình sau



Hình 3. 59 Kết quả

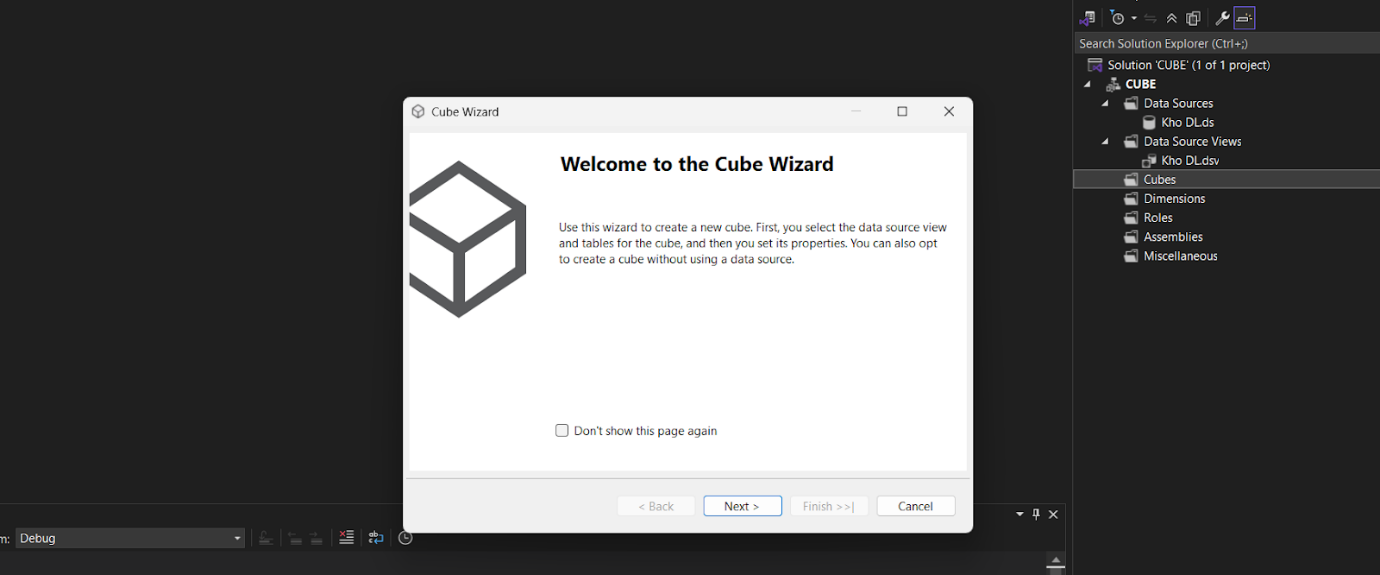
### 3.3.4  Xây dựng các khối (Cubes)

**Bước 1:** Tại Solution Explorer, ta click chuột phải vào thư mục Cubes và chọn New Cube.

****

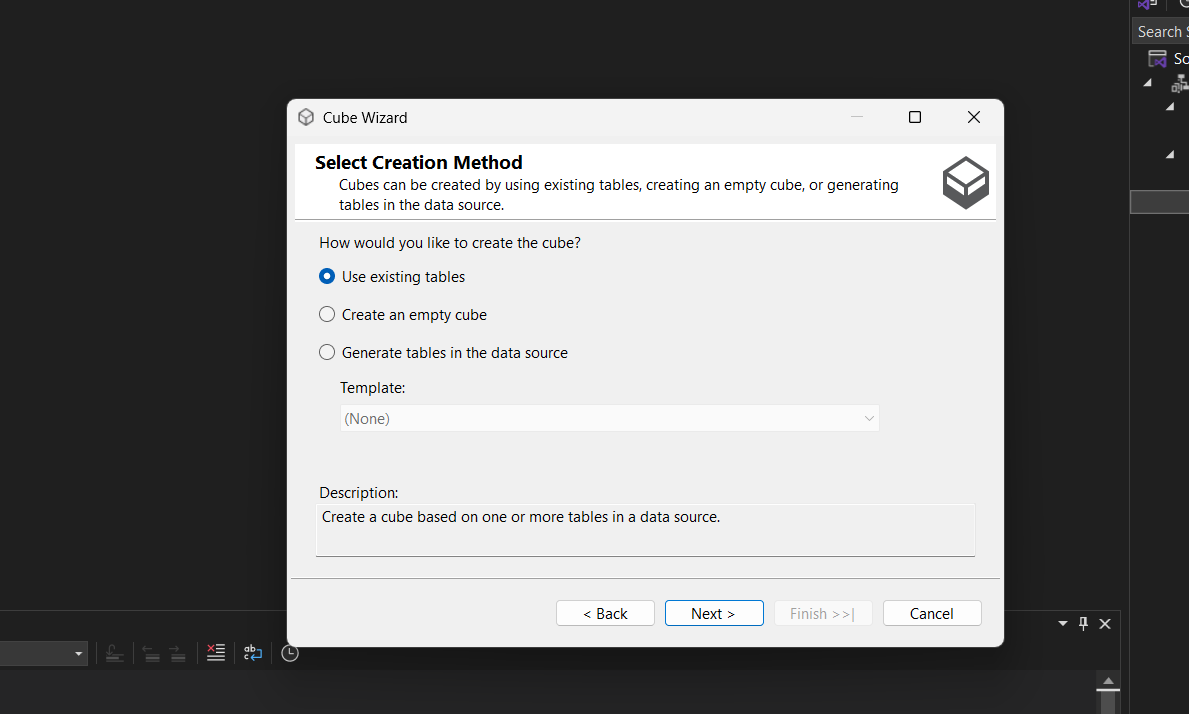
Hình 3. 60 Solution Explorer

**Bước 2**: Hộp thoại Cube Wizard xuất hiện, chọn Next để tiếp tục.

****

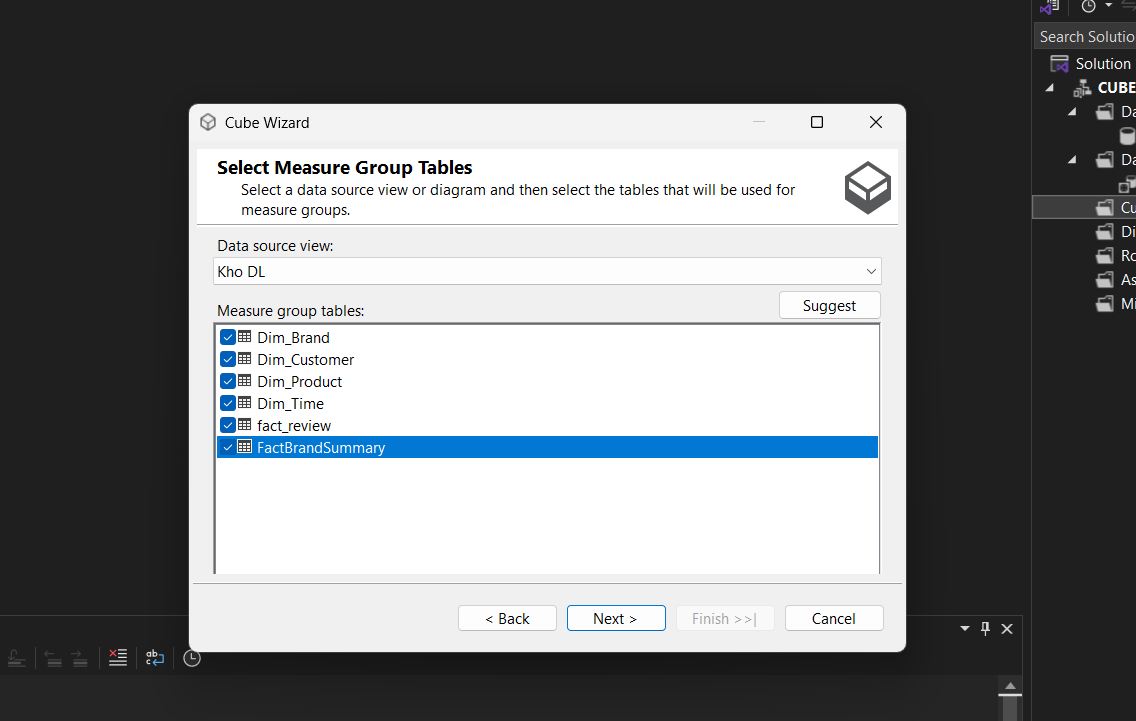
Hình 3. 61 Hộp thoại Cube Wizard

**Bước 3:** Chọn use existing tables, sau đó chọn Next để tiếp tục.



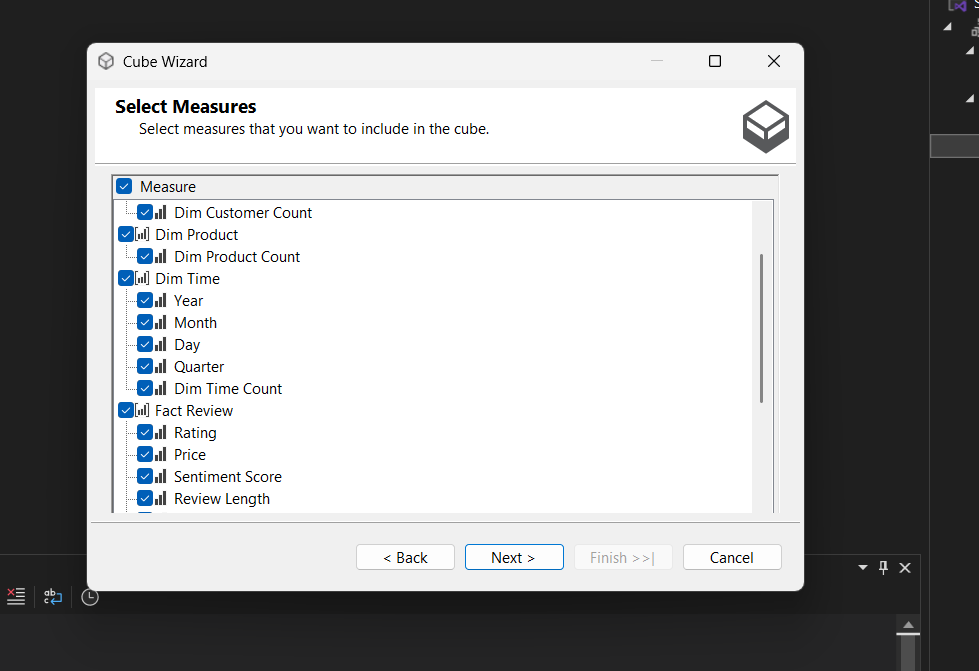
Hình 3. 62 Chọn use existing tables

**Bước 4:** Chọn bảng cần thêm.



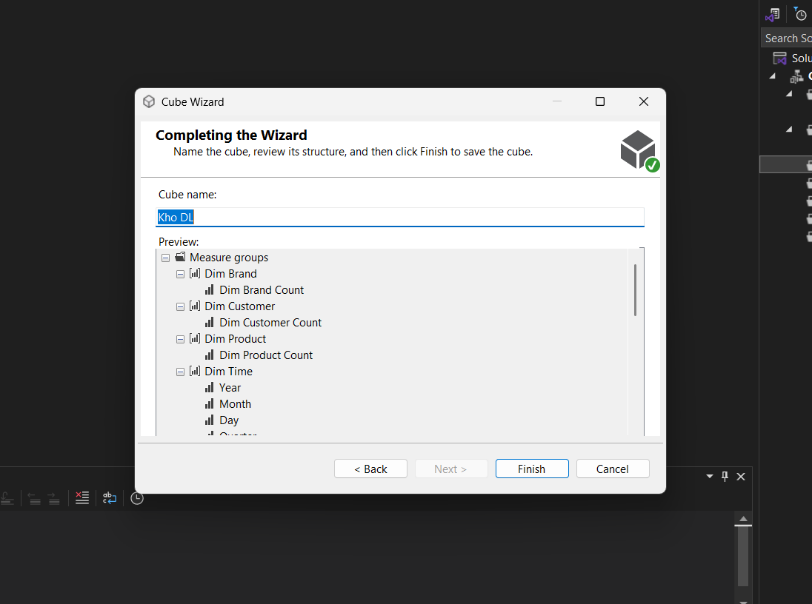
Hình 3. 63 Select Measure Group Tables

**Bước 5:** Chọn những độ đo đề xuất, sau đó chọn Next để tiếp tục.



Hình 3. 64 Select Measures

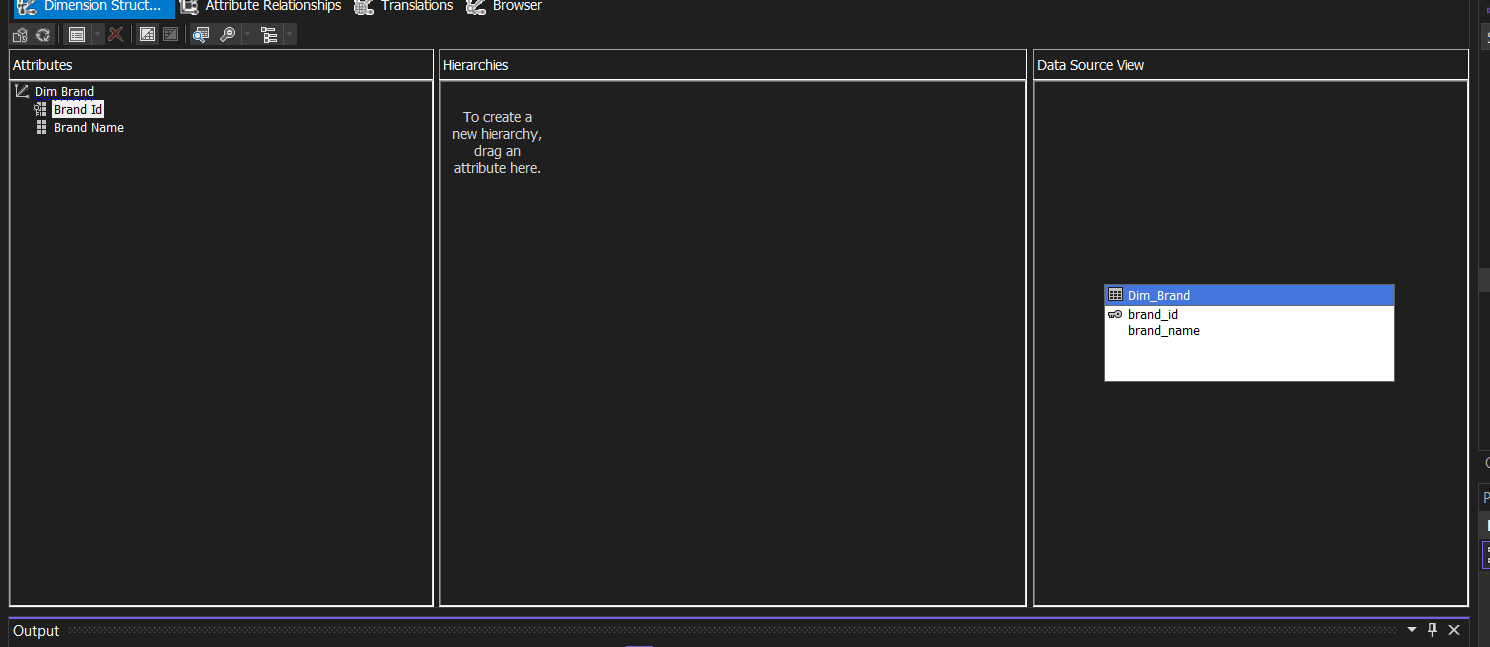
**Bước 6:** Chọn Finish để hoàn tất quy trình xây dựng các khối (Cubes) và xác định độ đo



Hình 3. 65 Completing the Wizard

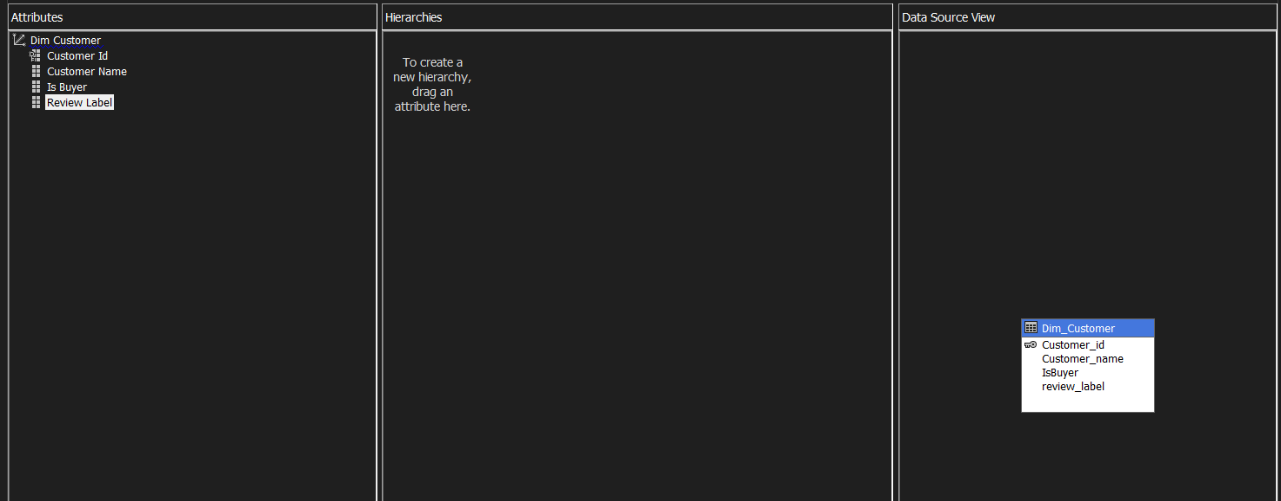
**3.3.5**  **Xác định các chiều (Dimensions)**

**Bước 1:** Tại folder Dimensions trong Solution Explorer, ta chọn Dim Brand.dim Sau đó kéo thả các thuộc tính Brand\_Id và Brand\_Name từ Data Source View vào Attributes.



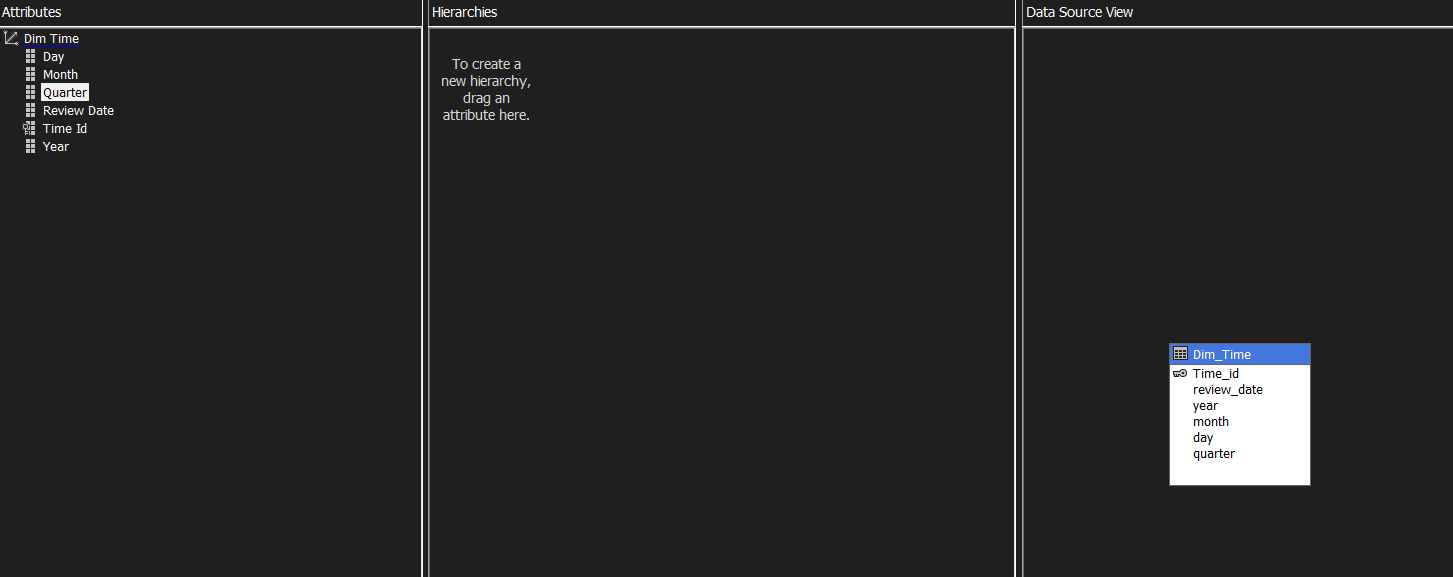
Hình 3. 66 Dimensions

**Bước 2:** Tại folder Dimensions trong Solution Explorer, ta chọn Dim Customer.dim Sau đó kéo thả các thuộc tính Customer\_name, Is\_buyer and Review\_label từ Data Source View vào Attributes.



Hình 3. 67 Chọn Dim Customer.dim

**Bước 3**: Tại folder Dimensions trong Solution Explorer, ta chọn Dim Date.dim Sau đó kéo thả các thuộc tính Review\_Date, Day, Month, Year, Quater từ Data Source View vào Attributes.



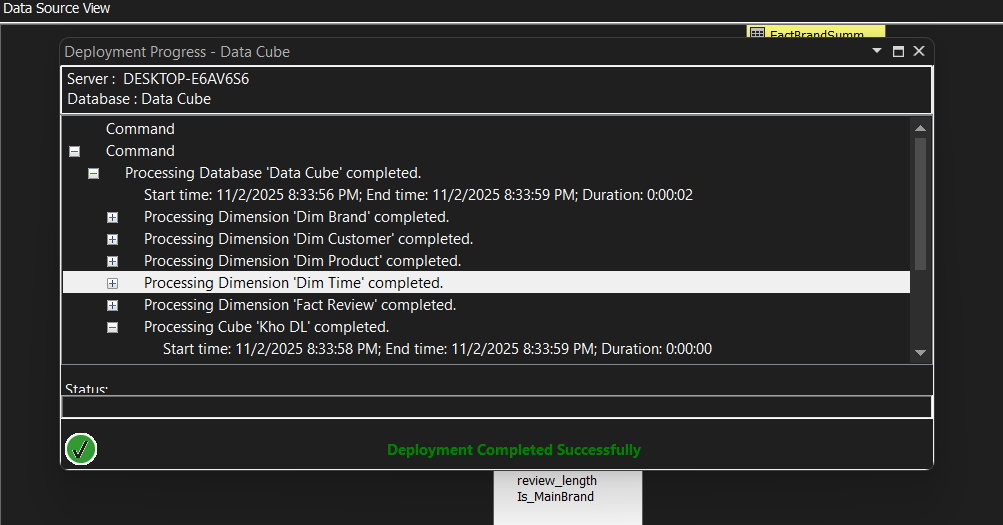
Hình 3. 68 Chọn Dim Date.dim

**Bước 4**: Tại folder Dimensions trong Solution Explorer, ta chọn Dim Product.dim Sau đó kéo thả các thuộc tính Product\_title từ Data Source View vào Attributes.



Hình 3. 69 Chọn Dim Product.dim

**Bước 5:** Ta chọn Start để deploy project. Khi deploy thành công ta sẽ nhận được kết quả như sau:

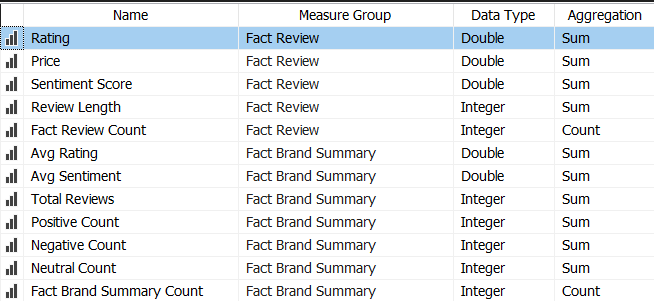


Hình 3. 70 Data Source View

**3.4 Xác định độ đo (Measures)**

**3.4.1 Thuộc tính độ đo ban đầu**

Tại khối vừa tạo, chọn Show Measures Grid để hiện thị chi tiết các độ đo.



Hình 3. 71 Chọn Show Measures Grid

**3.4.2 Tổng kết độ đo**

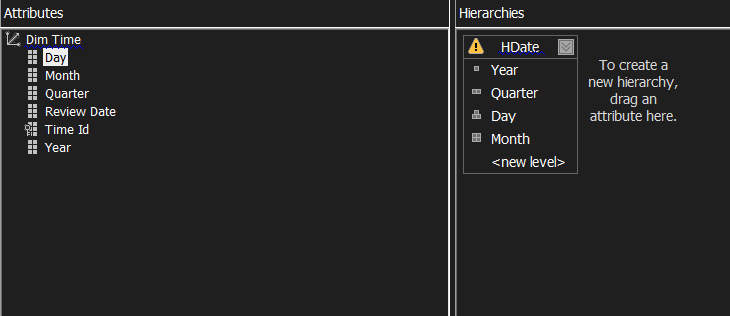
|  |  |
| --- | --- |
| Độ đo | Mô tả |
| Rating | Đánh giá của từng bài review đối với sản phẩm |
| Price | Giá từng sản phẩm |
| Sentiment Score | Điểm cảm xúc của bài đánh giá |
| Review Length | Độ dài của bài đánh giá |
| Fact Review Count | Số bài đánh giá |
| Avg Rating | Trung bình đánh giá của một thương hiệu |
| Avg Sentiment | Trung bình điểm cảm xúc của một thương hiệu |
| Total Reviews | Tổng bài đánh giá của một thương hiệu |
| Positive Count | Số đánh giá tích cực |
| Negative Count | Số đánh giá trung lập |
| Neutral Count | Số đánh giá tiêu cực |
| Fact Brand Summary Count | Số bài đánh giá theo thương |

## 3.5 Phân cấp trong bảng chiều (Dimensions)

**3.5.1 Phân cấp bảng Dim\_Time**

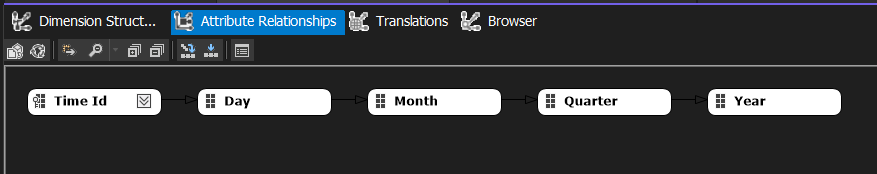
**Bước 1:** Kéo những thuộc tính cần phân cấp qua cửa sổ Hierarchies

**Bước 2:** Sắp xếp lại các thuộc tính phân cấp theo thứ tự: Year ⇒Quarter ⇒ Month ⇒ đổi tên thành HDate.



Hình 3. 72 Attributes

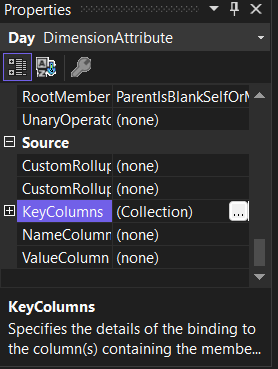
**Bước 3:** Tại panel Attribute Relationships, tạo mối quan hệ như sau



Hình 3. 73 Attribute Relationships

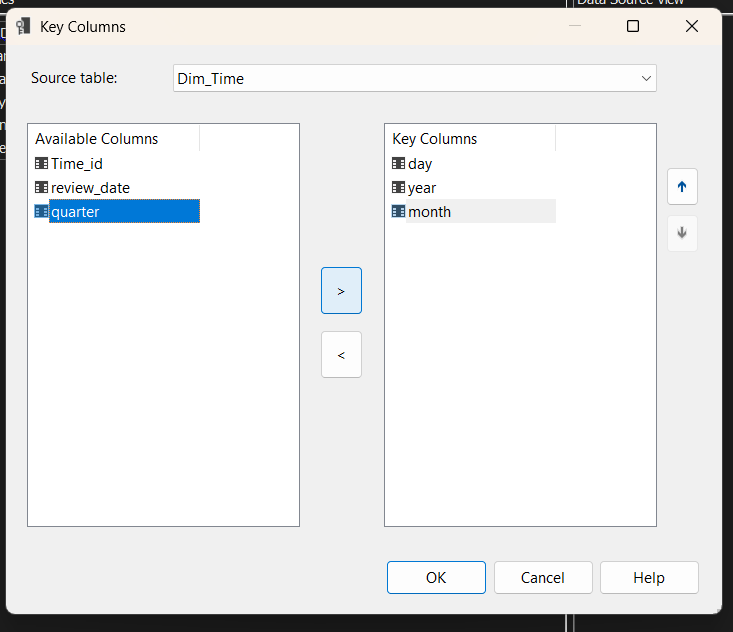
**Bước 4:** Chỉnh khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Day. Vì thuộc tính Day là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

“Một ngày (Day) thuộc về một tháng (Month), tháng thuộc quý (Quarter), quý thuộc năm (Year)” ⇒ ngày (Day) nằm trong tất cả các các phần còn lại của hệ HDate. Thì các thuộc tính con phải có khóa (KeyColumns) chứa toàn bộ các cấp cha.



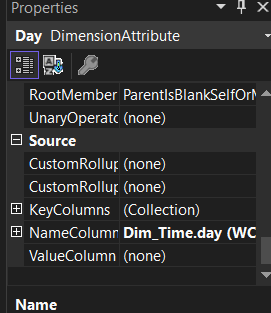
Hình 3. 74 Properties

Thêm các thuộc tính cấp cao hơn (year, quarter, month) vào **KeyColumns**, sau đó chọn OK để hoàn tất.



Hình 3. 75 Bảng Key Columns

Tại thuộc tính Properties của thuộc tính Day, ta chọn Name Column và chọn tên thuộc tính là Day.

****

Hình 3. 76 Chọn Name Columns

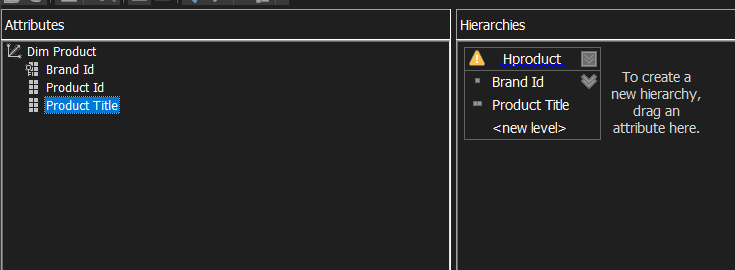
Sau đó thực hiện tương tự các bước với Month, Quarter.

**Bước 5:** Sau đó, Quay lại tab Dimension Structure, “Rebuild” lại dimension → Deploy → Process

**3.5.2 Phân cấp bảng Dim\_Product**

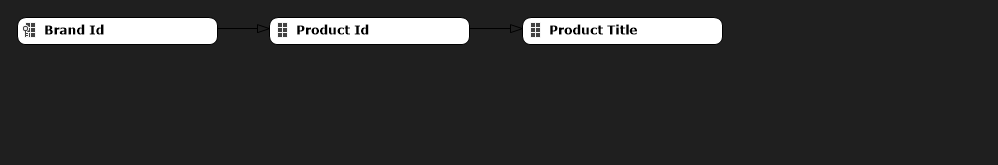
**Bước 1:** Kéo những thuộc tính cần phân cấp qua cửa sổ Hierarchies

**Bước 2:** Sắp xếp lại các thuộc tính phân cấp theo thứ tự: Brand id ⇒ Product Title và  đổi thành Hproduct.



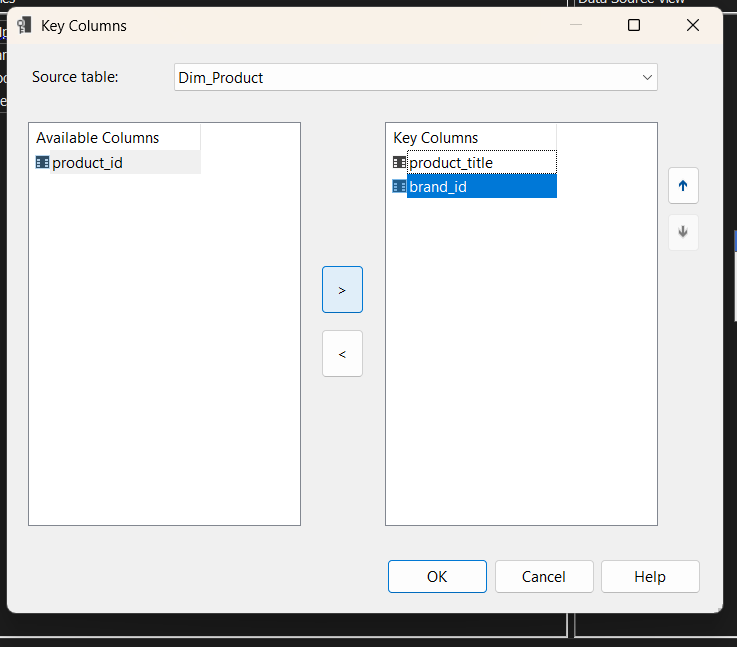
Hình 3. 77 Bảng Attributes

**Bước 3:** Tại panel Attribute Relationships, tạo mối quan hệ như sau



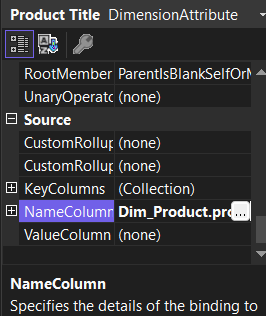
Hình 3. 78 Attribute Relationships

**Bước 4:** Chỉnh khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Product. Vì thuộc tính Product là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn Thêm thuộc tính cấp cao hơn (brand) vào KeyColumns, sau đó chọn OK để hoàn tất.



Hình 3. 79 Bảng Key Columns

Tại thuộc tính Properties của thuộc tính Day, ta chọn Name Column và chọn tên thuộc tính là Day.

****

Hình 3. 80 Product Title

**Bước 5:** Sau đó, Quay lại tab Dimension Structure, “Rebuild” lại dimension → Deploy → Process

**CHƯƠNG 4: POWER BI TRỰC QUAN VÀ RA QUYẾT ĐỊNH**

**4.1 Kết nối dữ liệu với POWER BI**

Power BI Desktop là một ứng dụng máy tính miễn phí của Microsoft, cho phép người dùng kết nối với nhiều nguồn dữ liệu, chuyển đổi, mô hình hóa và trực quan hóa dữ liệu thành các báo cáo tương tác.

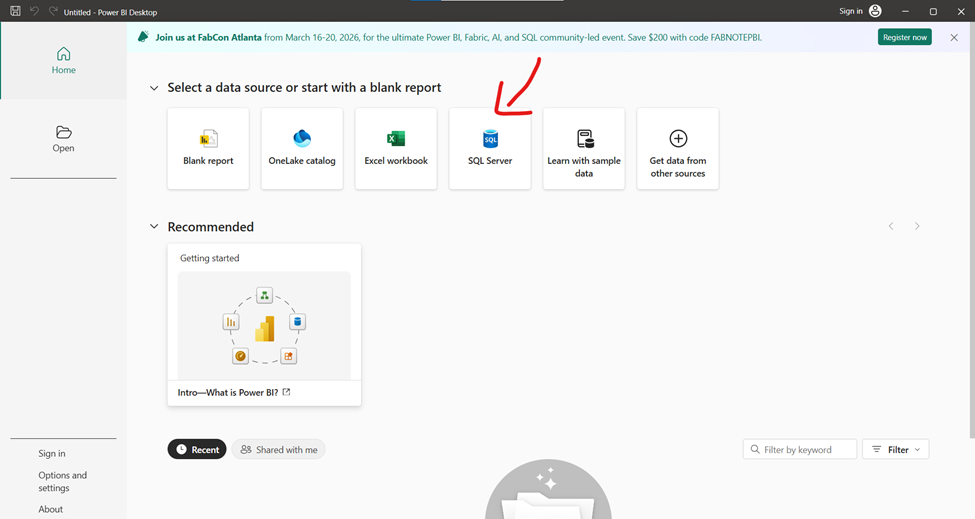
Ứng dụng này là công cụ chính để xây dựng các báo cáo và bảng điều khiển (dashboard) chuyên sâu, sau đó có thể chia sẻ thông qua Dịch vụ Power BI trực tuyến.

Chức năng chính:

* Kết nối dữ liệu: Kết nối an toàn với hàng trăm nguồn dữ liệu khác nhau, cả trên đám mây và tại chỗ.
* Biến đổi và làm sạch dữ liệu: Sử dụng công cụ Power Query Editor để chuyển đổi, kết hợp và làm sạch dữ liệu từ nhiều nguồn khác nhau.
* Mô hình hóa dữ liệu: Xây dựng mô hình dữ liệu phức tạp bằng cách sử dụng công thức DAX để tạo mối quan hệ giữa các bảng dữ liệu.
* Trực quan hóa dữ liệu: Tạo các biểu đồ và trực quan hóa dữ liệu tương tác từ hơn 100 loại hình ảnh có sẵn, hoặc tạo ra hình ảnh của riêng bạn.
* Phân tích dữ liệu: Khám phá các mẫu và thông tin chi tiết từ dữ liệu để hỗ trợ việc ra quyết định kinh doanh.
* Xây dựng báo cáo: Thiết kế các báo cáo có tính tương tác, bao gồm cả báo cáo di động cho người dùng di chuyển.

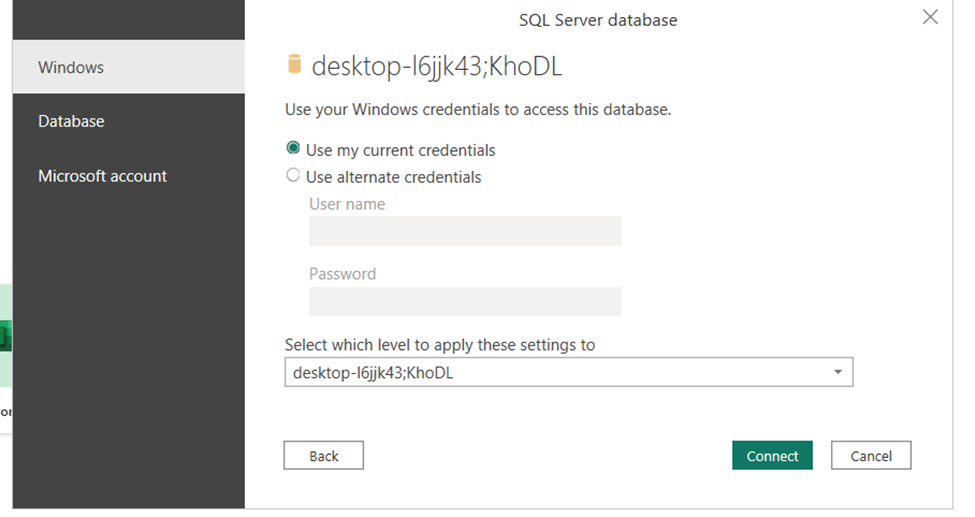
Các bước kết nối:

* Mở Power BI Desktop và chọn "Get Data" hoặc bắt đầu với “SQL Server” để tạo kết nối tới CSDL.



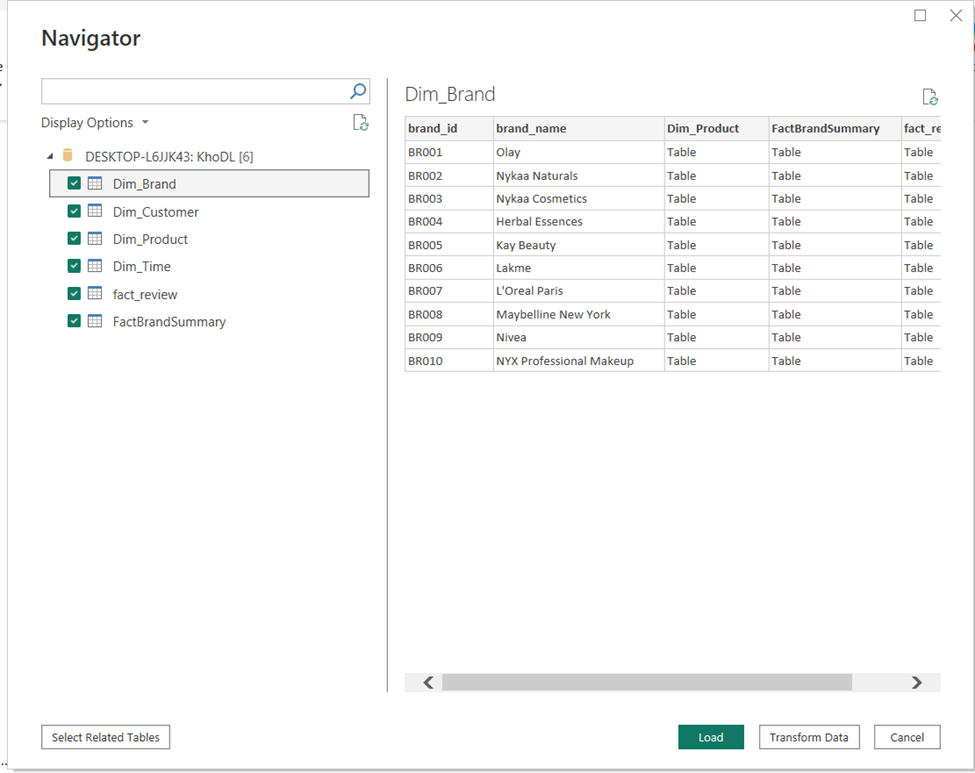
Hình 4. 1 Mở Power BI

* Chọn nguồn dữ liệu là SQL Server.



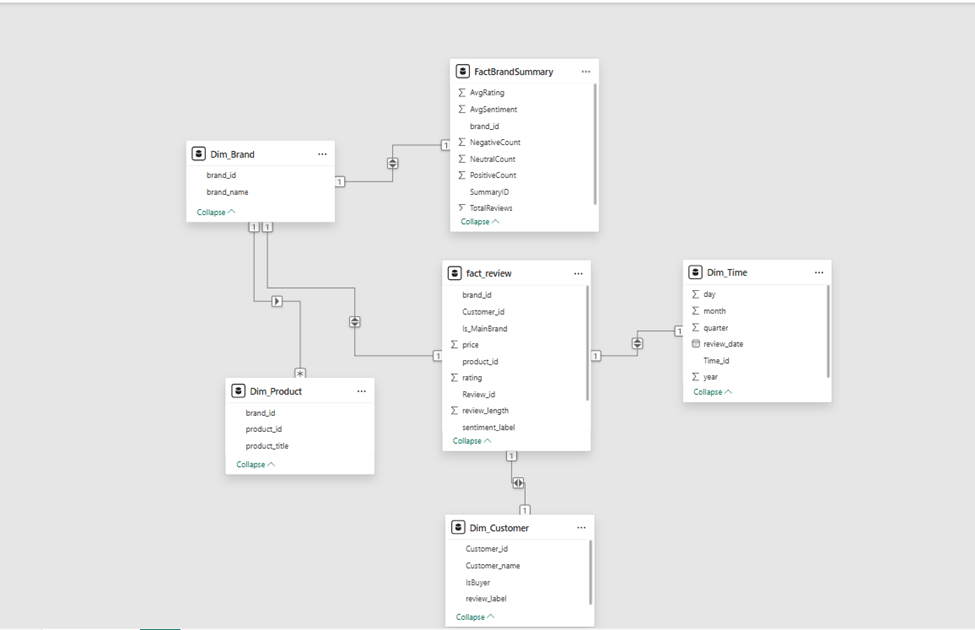
Hình 4. 2 Chọn SQL Server

* Nhập tên server và database đã tạo.
* Chọn các bảng cần sử dụng: *Dim\_Brand, Dim\_Product, Dim\_Customer, Dim\_Time, fact\_Review, factBrandSummary*.



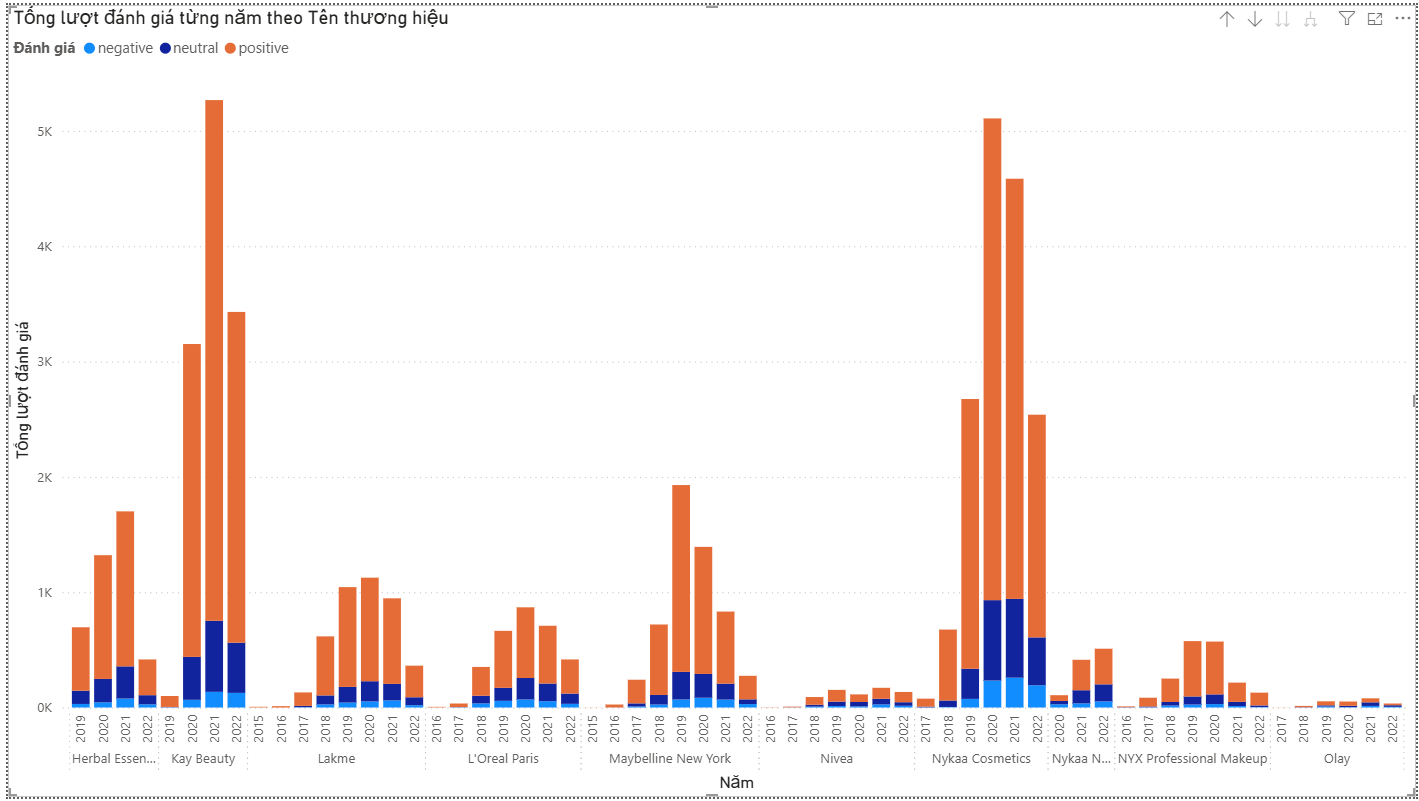
Hình 4. 3 Bảng Navigator

* Mở tab "Model" để thiết lập quan hệ giữa các bảng theo sơ đồ hình sao (FactSale ở giữa, kết nối tới các bảng Dim qua các khóa chính - khóa ngoại).



Hình 4. 4 Thiết lập quan hệ giữa các bảng theo sơ đồ hình sao

**4.1.1 Bảng tổng lượt đánh giá theo từng năm của từng thương hiệu**



Hình 4. 5 Bảng tổng lượt đánh giá theo từng năm của từng thương hiệu

**4.1.1.1 Phân tích**

Xu hướng chung: Biểu đồ cho thấy sự biến động lớn về tổng số lượt đánh giá giữa các thương hiệu và theo thời gian. Phần lớn các đánh giá trên tất cả các thương hiệu đều là tích cực (màu cam), chiếm tỷ trọng cao nhất trong hầu hết các cột.

Thương hiệu nổi bật: "Kay Beauty" và "Nykaa Cosmetics" có sự tăng trưởng đột biến về số lượng đánh giá vào khoảng năm 2020-2021, đạt mức cao nhất trên biểu đồ (trên 5.000 lượt). Tuy nhiên, đi kèm với sự tăng vọt về đánh giá tích cực, khối lượng (số lượng tuyệt đối) của các đánh giá tiêu cực (màu xanh đậm) và trung tính (màu xanh nhạt) cũng tăng lên đáng kể trong giai đoạn này.

Thương hiệu ổn định: Các thương hiệu như "L'Oreal Paris" và "Maybelline New York" có lượng đánh giá ổn định hơn qua các năm, với khối lượng vừa phải. Tỷ lệ đánh giá tiêu cực và trung tính của họ cũng có vẻ nhất quán, tạo thành một phần nhỏ nhưng cố định.

Thương hiệu ít dữ liệu: "Nivea" và "Olay" có số lượng đánh giá rất thấp trên biểu đồ này, khiến việc phân tích tâm trạng trở nên khó khăn hơn do cỡ mẫu nhỏ.

**4.1.1.2 Quyết định**

Ưu tiên xử lý khủng hoảng (cho Kay Beauty & Nykaa Cosmetics): Mặc dù tổng thể tích cực, nhưng khối lượng tuyệt đối của các đánh giá tiêu cực và trung tính trong các năm 2020-2021 là rất lớn. Doanh nghiệp cần khoan sâu (drill-down) vào dữ liệu này ngay lập tức. Cần phân tích nội dung chi tiết của các đánh giá tiêu cực này để xác định các vấn đề cốt lõi (ví dụ: chất lượng sản phẩm ra mắt, vấn đề giao hàng, dịch vụ khách hàng) đã xảy ra trong thời kỳ bùng nổ đó để khắc phục và tránh lặp lại.

Cải tiến liên tục (cho L'Oreal & Maybelline): Với dòng chảy phản hồi ổn định, các thương hiệu này nên thiết lập một hệ thống giám sát thường xuyên. Phân tích định kỳ các chủ đề trong nhóm đánh giá tiêu cực/trung tính để thực hiện các cải tiến nhỏ, liên tục cho sản phẩm và dịch vụ. Đồng thời, xác định các chủ đề trong đánh giá tích cực để khuếch đại trong các chiến dịch marketing.

Tăng cường thu thập dữ liệu (cho Olay & Nivea): Lượng đánh giá quá thấp là một điểm mù. Cần xem xét liệu đây là do thương hiệu ít phổ biến trên thị trường được phân tích, hay do hệ thống kho dữ liệu chưa thu thập đủ nguồn. Nếu là vế sau, cần bổ sung thêm nguồn dữ liệu. Nếu là vế trước, cần triển khai các chiến dịch khuyến khích khách hàng để lại đánh giá để có được cái nhìn rõ ràng hơn về tâm trạng của thị trường.

Quyết định hệ thống (Kho dữ liệu): Biểu đồ này chứng minh sự cần thiết của việc liên kết dữ liệu trực quan (Power BI) với dữ liệu chi tiết. Hệ thống kho dữ liệu phải cho phép người dùng doanh nghiệp (ví dụ: quản lý thương hiệu) nhấp chuột vào một phần (như phần "negative" của Nykaa 2020) và ngay lập tức xem được các từ khóa phổ biến hoặc các đánh giá cụ thể liên quan đến tâm trạng đó.

**4.1.1.3 Cách đọc**

Tiêu đề biểu đồ: Tổng lượt đánh giá từng năm theo Tên thương hiệu.

Loại biểu đồ: Biểu đồ cột chồng (Stacked Bar Chart).

Trục Y (Tung độ): "Tổng lượt đánh giá", biểu thị số lượng tuyệt đối của các lượt đánh giá.

Trục X (Hoành độ): "Năm", được nhóm theo "Tên thương hiệu".

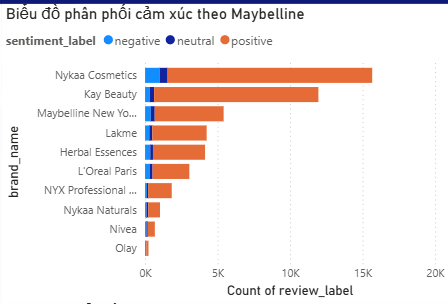
Chú thích (Legend): Phân loại "Đánh giá" thành ba nhóm tâm trạng, được mã hóa bằng màu sắc:

* Positive (Cam): Đánh giá tích cực.
* Negative (Xanh đậm): Đánh giá tiêu cực.
* Neutral (Xanh nhạt): Đánh giá trung tính.

Ý nghĩa: Mỗi cột đại diện cho tổng số lượt đánh giá của một thương hiệu trong một năm cụ thể. Các phân đoạn màu sắc trong mỗi cột cho thấy số lượng và tỷ lệ của từng loại tâm trạng (tích cực, tiêu cực, trung tính) góp phần vào tổng số đó.

**4.2  Xây dựng các biểu đồ trực quan**

**4.2.1 Biểu đồ phân phối cảm xúc theo thương hiệu**



Hình 4. 6 Biểu đồ phân phối cảm xúc theo thương hiệu

**4.2.1.1 Phân tích**

Thương hiệu Maybelline New York đứng thứ ba về tổng số lượt đề cập (hoặc số lượng cảm xúc được phân loại) trong số các thương hiệu được hiển thị. Tổng số lượt đề cập của Maybelline New York nằm ở khoảng 5.000.

Phân phối Cảm xúc:

* Cảm xúc Tích cực (positive): Chiếm phần lớn đáng kể, thể hiện sự hài lòng cao của khách hàng đối với thương hiệu này.
* Cảm xúc Trung lập (neutral): Có một phần nhỏ, nhưng không đáng kể so với tích cực.
* Cảm xúc Tiêu cực (negative): Chiếm tỷ lệ rất nhỏ, hầu như không đáng kể, cho thấy có rất ít phản hồi tiêu cực về thương hiệu này.

**4.2.1.2 Quyết định**

Dựa trên phân tích tích cực này, các quyết định và hành động có thể tập trung vào việc duy trì và củng cố vị thế:

Duy trì Chất lượng Sản phẩm & Dịch vụ (Focus on Strengths):

* Quyết định: Tiếp tục đầu tư vào chất lượng sản phẩm và trải nghiệm khách hàng vì đây là yếu tố chính tạo ra cảm xúc tích cực.
* Hành động: Thực hiện khảo sát chuyên sâu (deep dive) vào các phản hồi tích cực để xác định chính xác những sản phẩm/tính năng nào đang được yêu thích nhất (chẳng hạn như Mascara, son môi, v.v.) để tập trung quảng bá và phát triển phiên bản mới.

Giảm thiểu Phản hồi Tiêu cực (Address Minor Weaknesses):

* Quyết định: Mặc dù tỷ lệ tiêu cực thấp, cần phân tích các phản hồi tiêu cực nhỏ này.
* Hành động: Nhanh chóng xác định nguyên nhân của các phản hồi tiêu cực (ví dụ: vấn đề giao hàng, bao bì, hoặc lô sản phẩm cụ thể) và khắc phục để đưa tỷ lệ tiêu cực về mức gần bằng 0.

**4.2.1.3 Cách đọc biểu đồ**

Biểu đồ là stacked bar chart (biểu đồ thanh ngang xếp chồng), thể hiện:

Trục ngang (x-axis)

* Đại diện cho tổng số lượng review hoặc điểm trung bình review\_id (tùy dữ liệu của bạn).
* Đơn vị “triệu” chỉ giá trị lớn của dữ liệu.
* Trục dọc (y-axis)
* Danh sách các thương hiệu mỹ phẩm.

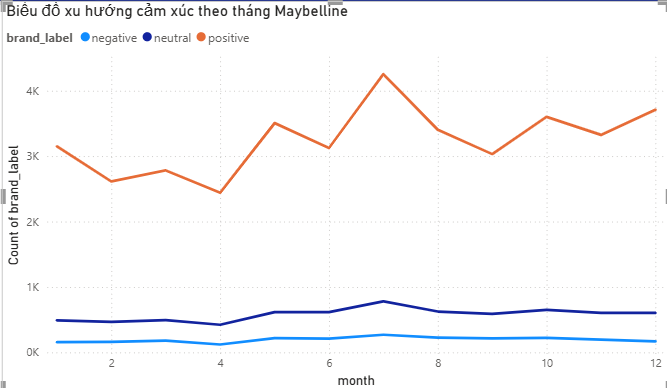
Màu sắc

* Xanh nhạt = Cảm xúc tiêu cực (negative)
* Xanh đậm = Cảm xúc trung tính (neutral)
* Cam = Cảm xúc tích cực (positive)

Ý nghĩa của mỗi thanh

* Chiều dài tổng thanh = tổng lượng review của thương hiệu.
* Tỷ lệ từng màu trong thanh = tỷ lệ cảm xúc của khách hàng dành cho thương hiệu.

**4.2.2 Biểu đồ xu hướng cảm xúc theo tháng**



Hình 4. 7 Biểu đồ xu hướng cảm xúc theo tháng

**4.2.2.1 Phân tích**

Cảm xúc Tích cực (positive - đường màu cam): Chiếm ưu thế áp đảo và là đường có nhiều biến động nhất.

Đạt đỉnh cao nhất vào khoảng Tháng 7 (trên 4k lượt đề cập), sau đó có xu hướng giảm nhẹ và duy trì ở mức cao vào cuối năm (Tháng 10 - Tháng 12).

Mức thấp nhất là khoảng Tháng 4.

Cảm xúc Trung lập (neutral - đường màu xanh đậm): Duy trì tương đối ổn định suốt năm, dao động quanh mức 500 - 1000 lượt đề cập.

Cảm xúc Tiêu cực (negative - đường màu xanh nhạt): Duy trì ở mức thấp và ổn định nhất, dao động quanh mức 200 - 400 lượt đề cập.

#### **4.2.2.2 Quyết định**

Tận dụng Mùa Cao điểm (Tháng 6 - Tháng 8):

Phân tích: Cảm xúc tích cực đạt đỉnh vào Tháng 7 (và tăng mạnh từ Tháng 5). Đây thường là mùa du lịch, mua sắm hè, hoặc các đợt khuyến mãi lớn.

Quyết định: Tăng cường ngân sách tiếp thị và ra mắt sản phẩm mới vào khoảng Tháng 5 và Tháng 6 để tận dụng sự tăng trưởng tự nhiên của sự quan tâm và cảm xúc tích cực.

Duy trì Kiểm soát Chất lượng (Cảm xúc Tiêu cực):

Phân tích: Cảm xúc tiêu cực rất thấp và ổn định, cho thấy việc xử lý khiếu nại và chất lượng sản phẩm chung là tốt.

Quyết định: Duy trì quy trình kiểm soát chất lượng (QC) và chăm sóc khách hàng hiện tại. Theo dõi sát sao những biến động nhỏ trong đường tiêu cực để kịp thời phát hiện và giải quyết các vấn đề mới xuất hiện (ví dụ: lỗi sản phẩm mới hoặc vấn đề vận chuyển).

**4.2.2.3 Cách đọc biểu đồ**

Đây là biểu đồ xu hướng số lượng review theo tháng, được chia theo nhãn cảm xúc (positive, neutral, negative).

Trục X = Month (1 → 12)

Trục Y = Tổng số review theo từng cảm xúc

Positive (đường màu cam):

Chiếm tỷ trọng lớn nhất, cao hơn hẳn 2 loại còn lại.

Biến động theo tháng, nhưng xu hướng chung là tăng.

Một số đỉnh (peak) xuất hiện ở tháng 7 và tháng 12 → khách hàng đánh giá tích cực mạnh vào các tháng này.

Neutral (đường xanh đậm):

Mức trung bình khoảng 500–900 review / tháng.

Xu hướng tăng nhẹ đến tháng 7 rồi giảm.

Negative (đường xanh nhạt):

Mức thấp nhất, thường chỉ 200–400 review / tháng.

Xu hướng tương đối ổn định, không biến động mạnh.

**4.2.3 Từ khóa phổ biến trong bài đánh giá của Maybelline**



Hình 4. 8 Từ khóa phổ biến trong bài đánh giá của Maybelline

**4.2.3.1 Phân tích**

Các từ khóa lớn nhất thể hiện cảm xúc mạnh mẽ của người dùng:

Amazing (Tuyệt vời): Là một trong những từ lớn nhất, cho thấy một lượng đáng kể khách hàng có trải nghiệm rất tốt.

Good (Tốt) và Love (Yêu thích): Những từ này cũng nổi bật, xác nhận Maybelline vẫn có một nhóm khách hàng trung thành và hài lòng.

Perfect (Hoàn hảo): Gợi ý rằng khi sản phẩm hoạt động tốt, nó được đánh giá rất cao.

Very but: Sự xuất hiện của "Very but" là một dấu hiệu cảnh báo tinh tế. Nó thường đi kèm với các câu như "Rất tốt nhưng..." hoặc "Rất thích nhưng...". Điều này cho thấy khách hàng thường có một trải nghiệm gần như hoàn hảo nhưng bị cản trở bởi một khuyết điểm nhỏ (ví dụ: giá, màu sắc, độ bền...).

Mâu thuẫn Giữa Cảm xúc: Word Cloud giải thích tại sao Maybelline có cả tỷ lệ Tiêu cực cao (từ biểu đồ thanh) và các từ khóa Tích cực lớn ("Amazing," "Love"). Điều này cho thấy thương hiệu có thể tạo ra trải nghiệm tuyệt vời, nhưng sự thiếu sót của sản phẩm ở một số khía cạnh cụ thể đang kích hoạt cảm xúc tiêu cực.

Nguyên nhân Tiêu cực Tiềm ẩn: Các từ khóa như "Color," "Shades," và đặc biệt là cụm từ "Very but" có thể là manh mối cho các khiếu nại. Ví dụ, khách hàng "rất thích" sản phẩm nền nhưng không tìm thấy tông màu phù hợp ("Shades").

**4.2.3.2 Quyết định**

Chiến lược Tông màu (Color Strategy): Vì "Color" và "Shades" rất phổ biến, Maybelline nên đầu tư vào việc mở rộng và chuẩn hóa dải màu (shade range), đặc biệt là các sản phẩm nền. Đảm bảo rằng sự đa dạng màu sắc đáp ứng được nhiều tông da khác nhau để chuyển đổi sự băn khoăn thành sự hài lòng.

Chiến lược Loại bỏ "Nhưng" (Address the "But"): Tiến hành phân tích nội dung chi tiết của các đánh giá có chứa từ "but" để xác định khuyết điểm phổ biến nhất (ví dụ: bao bì, độ bền, mùi hương) và khắc phục chúng. Việc loại bỏ các lỗi nhỏ này có thể giúp chuyển đổi một lượng lớn đánh giá.

Khuếch đại Giá trị: Tận dụng nhận thức thương hiệu là "Affordable" (Giá cả phải chăng) và "Amazing" bằng cách tập trung truyền thông vào các sản phẩm được coi là ”Must Buy” và có đánh giá cao nhất, củng cố vị thế giá trị tốt.

**4.2.3.3 Cách đọc biểu đồ**

Word Cloud là biểu đồ biểu diễn mức độ xuất hiện của các từ trong nội dung đánh giá (review).

Kích thước chữ = Mức độ xuất hiện

Từ càng lớn xuất hiện càng nhiều trong review.

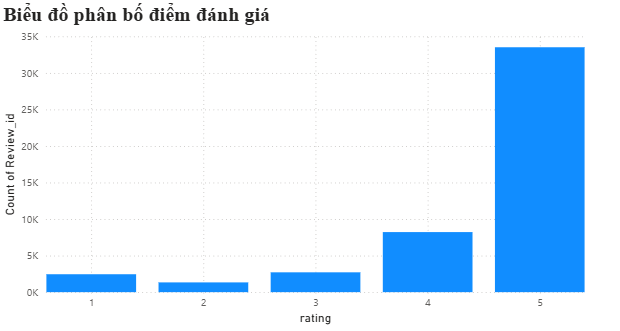
Những từ nổi bật nhất: “product”, “amazing”, “good”, “skin”, “love”, “shade”, “smooth”, “nice”, “perfect”…

Màu sắc chỉ để phân biệt từ

Không thể hiện sentiment (tích cực/tiêu cực).

Chỉ giúp người xem dễ nhìn và phân nhóm từ.

**4.2.4 Biểu đồ phân bố điểm đánh giá**



Hình 4. 9 Biểu đồ phân bố điểm đánh giá

**4.2.4.1 Phân tích**

Đánh giá Tích cực (Rating 4 & 5):

Đánh giá 5 sao có số lượng áp đảo, đạt mức cao nhất ~34K

Đánh giá 4 sao cũng có khối lượng đáng kể ~4K

Đánh giá: Điều này xác nhận rằng Maybelline có một cơ sở khách hàng trung thành lớn và sản phẩm của họ có thể tạo ra trải nghiệm xuất sắc.

Đánh giá Tiêu cực và Trung tính (Rating 1, 2 & 3):

Đánh giá 1 sao (Cực đoan): Khối lượng đáng kể ~3K

Đánh giá 2 sao (tiêu cực): Khối lượng thấp appro ~1.5K

Đánh giá 3 sao tương đương với 1 sao

Đánh giá: Mặc dù đánh giá 1 sao không phải là cột cao nhất, việc có tới 3K đánh giá 1 sao là một dấu hiệu cảnh báo nghiêm trọng về chất lượng.

**4.2.4.2 Quyết định**

Giảm thiểu Rủi ro tiêu cực đánh giá 1 sao

Đây là quyết định cấp bách nhất vì đánh giá 1 sao là nguồn gốc của sự suy giảm BHI.

Hành động: Sử dụng 3K đánh giá đánh giá 1 sao để thực hiện phân tích nguyên nhân gốc rễ (Root Cause Analysis). Xác định chính xác lỗi nào (sản phẩm, bao bì, giao hàng, dịch vụ khách hàng) đang gây ra sự giận dữ tột độ này và cải tiến ngay lập tức (hoặc ngừng sản xuất các SKU lỗi thời).

Chuyển đổi Khách hàng Trung tính (3 sao)

Hành động: Thiết kế các chương trình chăm sóc khách hàng đặc biệt để chuyển 2.5K khách hàng đánh giá 3 sao tthành4 sao hoặc 5 sao.

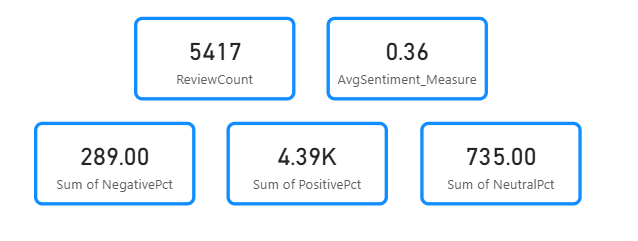
**4.2.4.3 Cách đọc biểu đồ**

Trục Ngang (X-axis): review\_rating (Điểm đánh giá) - Thang điểm từ 1 đến 5 sao.

Trục Dọc (Y-axis): review\_id (Khối lượng đánh giá) - Đo lường số lượng đánh giá trung bình ứng với mỗi điểm sao, tính bằng tỷ (B - Billion).

Chiều cao Cột: Chiều cao của mỗi cột thể hiện tổng số lượng đánh giá mà Maybelline nhận được cho mức điểm sao đó.

**4.2.5 Xây dựng Card KPI**



Hình 4. 10 Xây dựng Card KPI

**4.2.5.1 Phân tích**

Tổng thể thương hiệu Maybelline New York

81% phản hồi nằm trong nhóm Positive → Mức độ hài lòng cao.

AvgSentiment = 0.36, tương đối cao.

Lượng review lớn (5417) cho thấy thương hiệu có độ phủ cao và tệp khách hàng rất lớn

Cơ hội để truyền thông:

Neutral = 735 (13.5%) đánh giá cho thấy:

Người dùng không hài lòng cũng không khen → trải nghiệm chưa tạo “wow”.  
Nội dung review có thể chỉ mô tả sản phẩm, không cảm xúc.

Negative rất thấp (5.3%) – nhưng không được bỏ qua

289 phản hồi tiêu cực là thấp nhưng không được xem nhẹ.

Thông thường, 5% phản hồi tiêu cực có thể ảnh hưởng 15–20% nhận thức thương hiệu,

**4.2.5.2 Quyết định**

Chuyển hóa Neutral thành Positive:

Neutral 735 là “vùng xám”. Tập trung vào nhóm này giúp tăng BHI nhanh ơn tăng số lượng review mới

Chiến lược:Cải thiện mô tả sản phẩm (giá trị, công dụng, phù hợp da nào). Tạo trải nghiệm mở hộp tốt hơn. Gửi email hướng dẫn cách sử dụng tối ưu sản phẩm. Remarketing nhắc người dùng viết review có cảm xúc

Điều tra nhóm Negative

Chỉ 5.3%, nhưng đây là nhóm ảnh hưởng “brand trust”

Dùng Word Cloud + Keyphrase Extraction. Tìm lý do: *khô da*, *không hợp tone*, *vỡ khi ship*, *không bám tốt*?

Gửi bộ phận CSKH xử lý nhanh

**4.2.5.3 Cách đọc biểu đồ**

Biểu đồ Card Chart này tóm tắt các chỉ số hiệu suất chính (KPIs) của từng thương hiệu (Brand) được phân tích trong dữ liệu. Mỗi "thẻ" (Card) đại diện cho một thương hiệu và hiển thị các thông tin sau:

AvgRating (Điểm Đánh Giá Trung Bình): Thang điểm 5.

AvgSentiment (Điểm Tình Cảm Trung Bình): Thang điểm từ -1 (tiêu cực) đến 1 (tích cực).

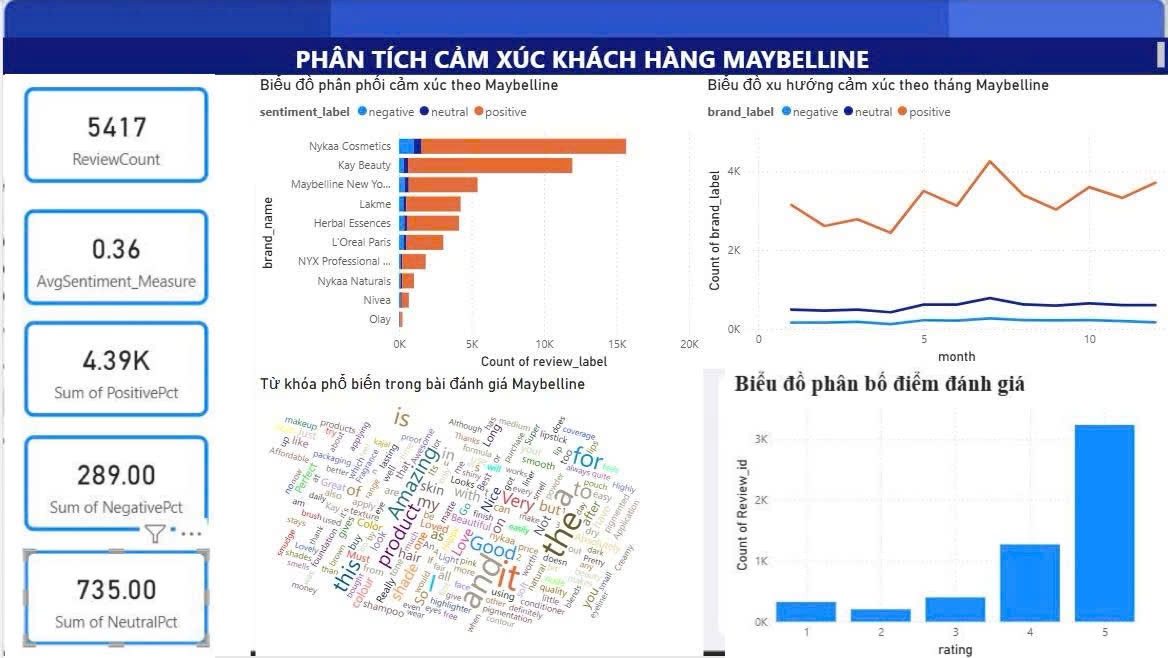
TotalReviews (Tổng Số Đánh Giá): Tổng số lượng đánh giá mà thương hiệu đó nhận được.

PositiveCount (Số Lượng Đánh Giá Tích Cực): Số lượng đánh giá có nhãn 'positive'.

NegativeCount (Số Lượng Đánh Giá Tiêu Cực): Số lượng đánh giá có nhãn 'negative'.

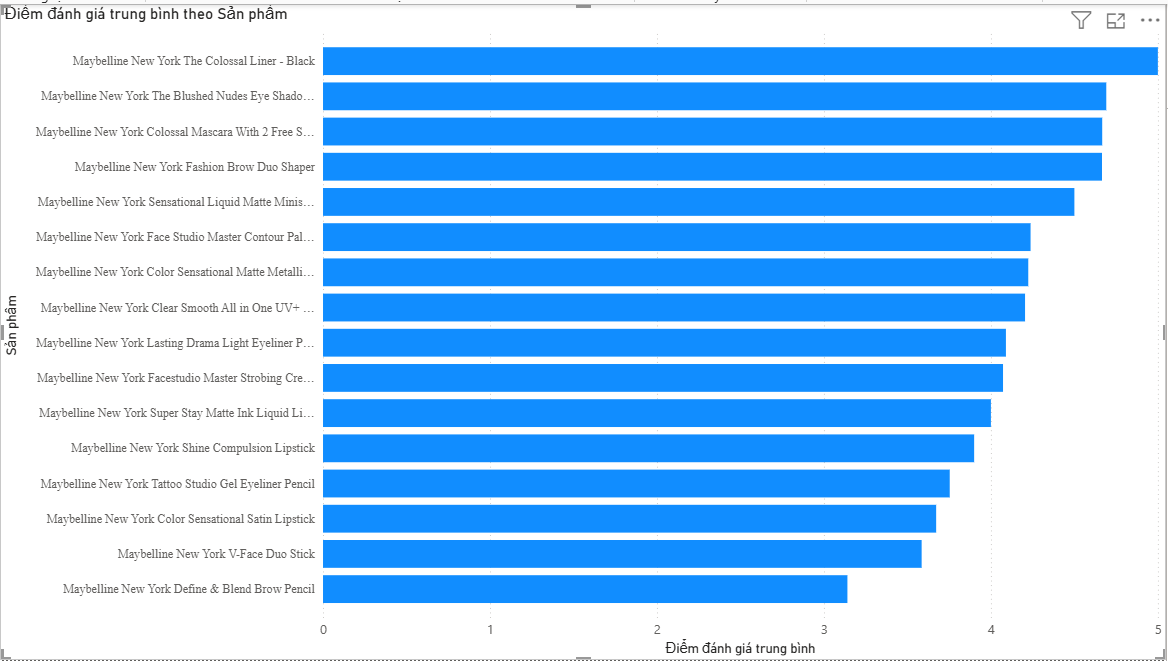
NeutralCount (Số Lượng Đánh Giá Trung Lập): Số lượng đánh giá có nhãn 'neutral'.

**4.2.6 Dashboard cảm xúc khách hàng**



Hình 4. 11 Dashboard Cảm xúc khách hàng

**4.2.7 Biểu đồ thể hiện sản phẩm có điểm đánh giá cao nhất**



Hình 4. 12 Biểu đồ thể hiện sản phẩm có điểm đánh giá cao nhất

**4.2.7.1    Phân tích**

Biểu đồ thể hiện các sản phẩm điểm đánh giá cao nhất  Maybelline, sắp xếp theo thứ tự từ cao xuống thấp. Các sản phẩm có điểm cao nhất gồm:

*The Colossal Liner – Black*

*The Blushed Nudes Eye Shadow Palette*

*Colossal Mascara With 2 Free Samples*

*Fashion Brow Duo Shaper*

Những sản phẩm này có điểm trung bình gần 5, phản ánh mức độ hài lòng rất cao từ người dùng, cho thấy chất lượng ổn định và phù hợp với kỳ vọng của khách hàng. Ngược lại, một số sản phẩm nằm cuối bảng như *V-Face Duo Stick* hoặc *Define & Blend Brow Pencil* có điểm khoảng 3.2–3.4, cho thấy đánh giá thấp hơn, có thể do hiệu quả sử dụng, trải nghiệm người dùng hoặc chất lượng chưa đồng đều.

#### **4.2.7.2  Quyết định**

Tập trung đầu tư mạnh hơn vào các sản phẩm có điểm đánh giá cao, vì đây là nhóm sản phẩm có hiệu suất tốt, nhận được sự tin tưởng và yêu thích từ khách hàng → có thể dùng làm sản phẩm chủ lực để tăng doanh thu.

Đẩy mạnh quảng cáo và truyền thông cho nhóm sản phẩm top đầu nhằm duy trì vị thế cạnh tranh cũng như thu hút thêm khách hàng mới. Thu thập thêm phản hồi chi tiết của khách hàng dành cho nhóm sản phẩm điểm thấp để xác định nguyên nhân và xây dựng kế hoạch cải tiến phù hợp.

**4.2.7.3  Cách đọc biểu đồ**

Tiêu đề biểu đồ: Biểu đồ thể hiện các sản phẩm có điểm đánh giá cao nhất

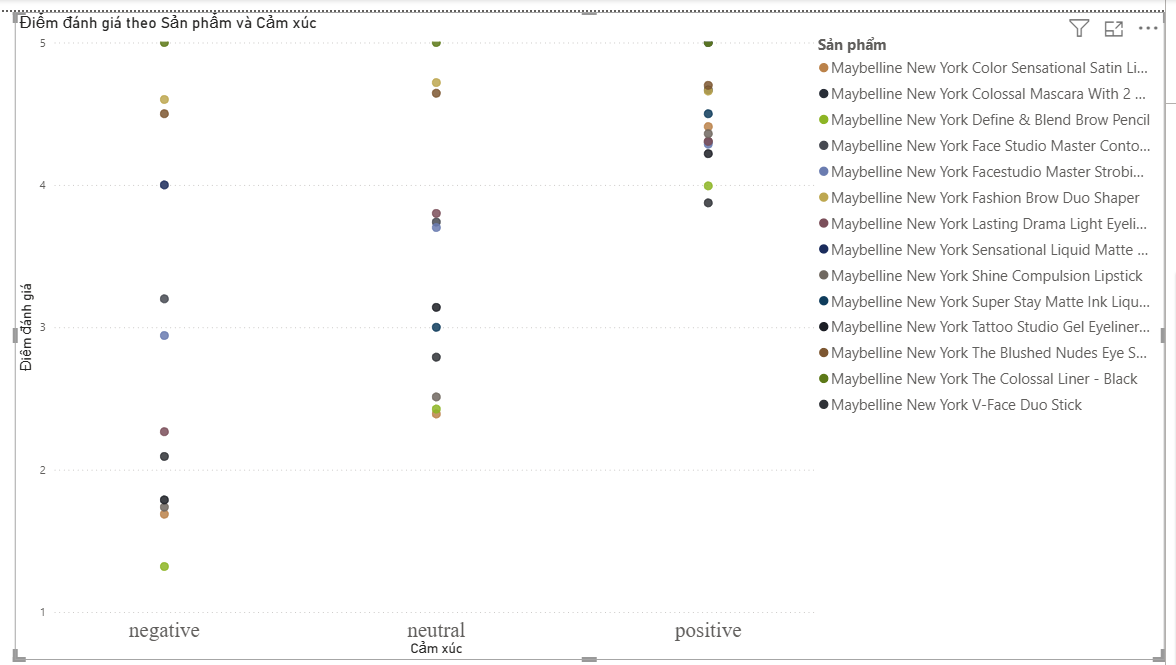
Trục hoành (X): Hiển thị điểm đánh giá trung bình.

Trục tung (Y): Liệt kê tên sản phẩm của Maybelline.

Ý nghĩa màu sắc: Các thanh màu xanh đại diện cho từng sản phẩm, thanh dài hơn thể hiện điểm đánh giá cao hơn.

Cách xem thứ hạng: Sản phẩm được xếp theo chiều từ trên xuống dưới theo mức điểm giảm dần — sản phẩm đứng đầu là sản phẩm được đánh giá tốt nhất.

**4.2.8 Biểu đồ quan hệ giữa cảm xúc và điểm đánh giá**



Hình 4. 13 Biểu đồ quan hệ giữa cảm xúc và điểm đánh giá

**4.2.8.1    Phân tích**

Cảm xúc positive có điểm đánh giá cao nhất: Nhóm đánh giá mang cảm xúc tích cực có điểm trung bình khoảng 4.6 – 4.7, cho thấy khi khách hàng hài lòng, họ sẽ chấm điểm rất cao cho sản phẩm.

Cảm xúc neutral có mức điểm trung bình khá cao (~4.0): Người dùng không quá hài lòng hoặc không quá thất vọng vẫn có xu hướng đánh giá cao, điều này hàm ý sản phẩm vẫn đáp ứng được nhu cầu cơ bản của đa số khách hàng.

Cảm xúc negative có điểm thấp nhất (~2.6): Mặc dù số lượng đánh giá tiêu cực có thể ít, nhưng khi xuất hiện thì điểm đánh giá giảm mạnh. Điều này cho thấy các phản hồi tiêu cực có ảnh hưởng đáng kể đến điểm số trung bình và hình ảnh sản phẩm.

Sự khác biệt giữa sản phẩm: Mặc dù mỗi sản phẩm có một màu sắc đại diện, nhưng do số lượng sản phẩm lớn nên màu trùng và khó phân biệt trên biểu đồ. Tuy nhiên, không có sản phẩm nào có xu hướng lệch quá mạnh so với xu hướng chung.

**4.2.8.2  Quyết định**

Tập trung phân tích nguyên nhân tạo ra cảm xúc negative, nhóm này có điểm rất thấp, cần xem lại: Lỗi sản phẩm, chậm giao hàng, kỳ vọng không trùng khớp với thực tế, trải nghiệm hậu mãi chưa tốt. Việc xử lý các nguyên nhân cốt lõi sẽ giúp giảm đánh giá xấu trong tương lai. Duy trì và phát huy các yếu tố tạo cảm xúc positive, nhóm có điểm cao nhất cần tiếp tục đẩy mạnh: Chất lượng sản phẩm, tính năng nổi bật, trải nghiệm đóng gói – giao hàng, dịch vụ hỗ trợ khách hàng, tối ưu chiến lược truyền thông, có thể tận dụng các đánh giá positive.

**4.2.8.3  Cách đọc biểu đồ**

Trục ngang (X-axis): Thể hiện loại cảm xúc — *negative, neutral, positive*.

Trục dọc (Y-axis): Thể hiện điểm đánh giá từ 1 đến 5.

Cách hiểu các điểm dữ liệu:

Mỗi chấm trên biểu đồ đại diện cho một bài đánh giá của khách hàng.

Vị trí theo chiều ngang → cảm xúc của bài review.

Vị trí theo chiều dọc → điểm số khách hàng chấm.

Màu sắc của từng chấm đại diện cho sản phẩm Maybelline khác nhau.

**4.2.9 Biểu đồ so sánh brand Maybelline và các brand còn lại**



Hình 4. 14 Biểu đồ so sánh Brand Maybelline và các Brand còn lại

**4.2.9.1    Phân tích**

Maybelline New York nằm ở nhóm khá cao (~4.3): Mức điểm này cho thấy sản phẩm của Maybelline được đánh giá tốt, đáp ứng ổn định nhu cầu của phần lớn khách hàng.

Một số thương hiệu có điểm cao hơn Maybelline:

* Kay Beauty (~4.7)
* Nykaa Cosmetics (~4.5)
* NYX Professional Makeup (~4.35–4.4)

Những thương hiệu này đang tạo ra sự hài lòng cao hơn, cho thấy sức cạnh tranh mạnh hơn về chất lượng sản phẩm hoặc trải nghiệm người dùng.

Một số thương hiệu ngang hoặc thấp hơn Maybelline:

* Herbal Essences, Lakme (~4.2–4.3) → tương đương
* L’Oreal Paris (~4.1) → thấp hơn nhẹ
* Nykaa Naturals (~4.05), Nivea (~3.8), Olay (~3.6) → thấp hơn khá nhiều

Điều này cho thấy Maybelline vẫn nằm trong nhóm trung–cao trên thị trường. Khoảng cách điểm giữa các thương hiệu không quá lớn: Mức điểm dao động từ 3.6 đến 4.7 cho thấy chất lượng chung của thị trường khá tốt, nhưng các thương hiệu top đầu lại nổi bật rõ rệt khi vượt ngưỡng 4.5.

**4.2.9.2  Quyết định**

Maybelline New York đạt mức điểm đánh giá trung bình khá cao (khoảng 4.2–4.3), thể hiện rằng thương hiệu này được người dùng đánh giá tích cực về chất lượng sản phẩm. Tuy nhiên, so với các thương hiệu dẫn đầu như Kay Beauty, Nykaa Cosmetics hay NYX Professional Makeup, điểm số của Maybelline vẫn thấp hơn và chưa tạo được sự vượt trội rõ rệt.

Ngược lại, Maybelline lại có điểm đánh giá cao hơn các thương hiệu như L’Oreal Paris, Nykaa Naturals, Nivea và Olay, cho thấy thương hiệu vẫn giữ được vị thế ổn định và được tin tưởng trong nhóm sản phẩm tầm trung.

**4.2.9.3  Cách đọc biểu đồ**

Trục ngang (X-axis): Hiển thị tên các thương hiệu, bao gồm Maybelline và những thương hiệu còn lại.

Trục dọc (Y-axis): Thể hiện điểm đánh giá trung bình (1–5) của từng thương hiệu.

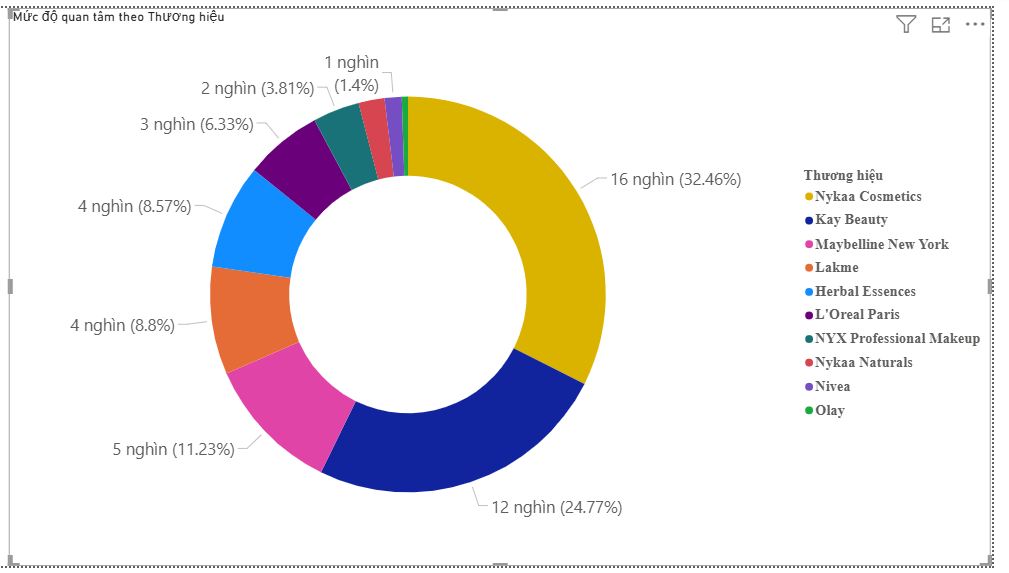
Ý nghĩa độ cao cột:

Cột Maybelline hiển thị vị trí mức điểm so với các thương hiệu khác.

Nếu cột Maybelline thấp hơn → thương hiệu khác đang làm tốt hơn.

Nếu cột Maybelline cao hơn → thương hiệu này đang có lợi thế cạnh tranh.

**4.2.10 Biểu đồ thể hiện mức độ quan tâm đến thương hiệu**



Hình 4. 15 Biểu đồ thể hiện mức độ quan tâm đến thương hiệu

**4.2.10.1    Phân tích**

Maybelline New York: Với mức độ quan tâm đạt 5 nghìn lượt (11.23%), Maybelline nằm trong nhóm thương hiệu trung bình. Dù không nằm trong top đầu, thương hiệu vẫn giữ vị trí khá tốt nhờ mức độ nhận diện cao và sản phẩm đa dạng, phù hợp nhiều phân khúc người dùng.

Maybelline hiện đang giữ vị trí thứ 3 trên thị trường về mức độ quan tâm. Với hơn 10% thị phần. Khoảng cách với Nykaa Cosmetics 32.46% là hơn 21 điểm phần trăm, và với Kay Beauty 24.77% là hơn 13 điểm phần trăm. Điều này cho thấy Maybelline đang thua kém đáng kể trong cuộc đua tạo ra tiếng vang (buzz) và nhận diện thương hiệu.

**4.2.10.2  Quyết định**

Quyết định: Chuyển từ chiến lược duy trì sang Chiến lược Tấn công và Tạo tiếng vang (Aggressive Buzz Generation Strategy).

Triển khai "Siêu Chiến dịch Kỹ thuật số" (Mega Digital Campaign): Tận dụng quy mô toàn cầu của Maybelline để thực hiện các chiến dịch lớn mà đối thủ không thể sao chép (ví dụ: hợp tác với các ngôi sao hạng A toàn cầu, tích hợp công nghệ AR/VR). Mục tiêu là tạo ra sự kiện (Event-driven Marketing) để tăng đột biến mức độ thảo luận.

Đầu tư vào "Sản phẩm Biểu tượng" (Iconic Product Focus): Xác định 2-3 dòng sản phẩm cốt lõi (ví dụ: The Colossal Liner – Black, The Blushed Nudes Eye Shadow Palette) nơi Maybelline có lợi thế chất lượng. Tập trung truyền thông vào các sản phẩm này để chiếm lĩnh thị trường danh mục và dùng chúng làm đầu tàu kéo mức độ quan tâm tổng thể của thương hiệu.

**4.2.10.3  Cách đọc biểu đồ**

Tiêu đề biểu đồ: Biểu đồ thể hiện *mức độ quan tâm theo từng thương hiệu* dựa trên lượng tìm kiếm hoặc tương tác của người tiêu dùng.

Các thương hiệu được thể hiện bằng màu sắc rõ ràng:

* Nykaa Cosmetics: Màu vàng
* Kay Beauty: Màu xanh đậm
* Maybelline New York: Màu hồng
* Lakme: Màu cam
* Herbal Essences: Màu xanh dương
* L’Oreal Paris: Màu tím đậm
* NYX Professional Makeup: Xanh lục đậm
* Nykaa Naturals: Đỏ
* Nivea: Tím nhạt
* Olay: Xanh lá

**4.2.11 Biểu đồ nhóm sản phẩm ảnh hưởng lớn đến thương hiệu**



Hình 4. 16 Biểu đồ nhóm sản phẩm ảnh hưởng lớn đến thương hiệu

**4.2.11.1    Phân tích**

Một số sản phẩm chiếm diện tích lớn hơn rõ rệt: Các ô có kích thước lớn hơn đại diện cho các sản phẩm có số lượng review cao, ví dụ:

*Maybelline New York The Colossal Liner – Black*

*Maybelline New York The Blushed Nudes Eye Shadow Palette*

*Maybelline New York Fashion Brow Duo Shaper*

Những sản phẩm này đóng góp đáng kể vào mức độ nhận diện và mức độ xuất hiện của thương hiệu trên các nền tảng. Các sản phẩm có diện tích nhỏ hơn: Một số sản phẩm nhận được ít review hơn, làm cho mức độ ảnh hưởng đến thương hiệu thấp hơn. Điều này có thể do sản phẩm mới, ít phổ biến, hoặc ít được nhắc đến trong cộng đồng người dùng.

Phân bố review không đồng đều: Một vài sản phẩm chiếm tỷ trọng lớn, trong khi phần còn lại chỉ góp phần nhỏ. Điều này cho thấy thương hiệu đang phụ thuộc vào một số dòng sản phẩm chủ lực.

**4.2.11.2  Quyết định**

Tập trung phát triển các sản phẩm chủ lực: Những sản phẩm có lượng review cao chứng tỏ mức độ quan tâm lớn từ khách hàng. Doanh nghiệp nên tiếp tục đầu tư vào marketing, cải thiện chất lượng và mở rộng phân phối cho các dòng này để duy trì lợi thế cạnh tranh.

Tăng cường truyền thông cho các sản phẩm có ít review: Các sản phẩm có diện tích nhỏ trong biểu đồ có thể cần được thúc đẩy nhận diện — thông qua quảng cáo, KOL/KOC review, ưu đãi hoặc gói combo.

Theo dõi phản hồi khách hàng theo từng nhóm: Những sản phẩm có nhiều review thường mang lại dữ liệu phong phú về cảm xúc và chất lượng. Doanh nghiệp nên ưu tiên phân tích sâu hơn đối với các dòng này để cải thiện trải nghiệm khách hàng.

**4.2.11.3  Cách đọc biểu đồ**

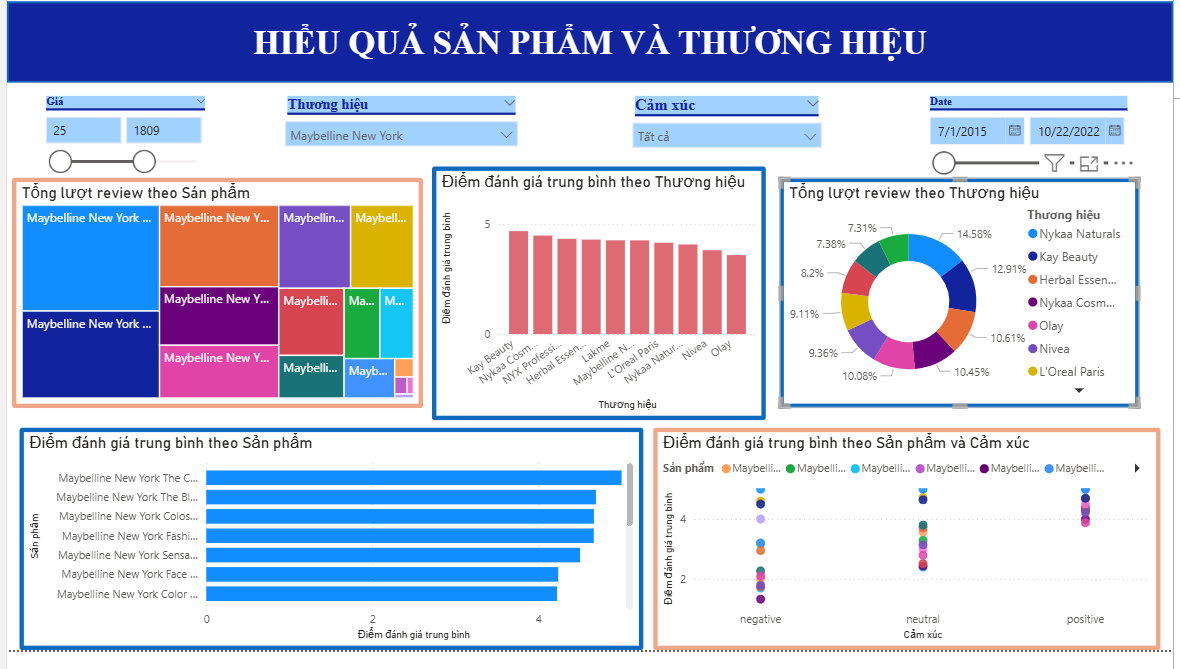
Tiêu đề biểu đồ: Biểu đồ thể hiện nhóm sản phẩm ảnh hưởng lớn đến thương hiệu

Cách hiểu từng ô (trong treemap):

Mỗi ô là một sản phẩm cụ thể thuộc thương hiệu Maybelline New York.

Kích thước ô đại diện cho số lượng review: ô càng lớn → sản phẩm nhận càng nhiều review.

**4.2.12  Dashboard Hiệu quả sản phẩm và thương hiệu**



Hình 4. 17 Dashboard hiệu quả sản phẩm và thương hiệu

**4.2.13 Biểu đồ xu hướng chỉ số BHI theo thời gian**



Hình 4. 18 Biểu đồ xu hướng chỉ số BHI theo thời gian

**4.2.13.1 Phân tích**

BHI – Brand Health Index: chỉ số sức khỏe thương hiệu dùng để đo độ mạnh yếu của một thương hiệu dựa trên hành vi, đánh giá của khách hàng. Được tính bằng công thức:

“BHI = 0.6 \* AvgSentiment + 0.4 \* AvgRating”. 60% của trung bình điểm cảm xúc và 40% của trung bình điểm đánh giá của khách hàng.

2015 - 2017 (Giai đoạn Đỉnh): Bắt đầu ở mức BHI cao (khoảng 2.05 năm 2015) và có xu hướng tăng nhẹ, đạt đỉnh vào khoảng năm 2016 (khoảng 2.15). Đây là giai đoạn sức khỏe thương hiệu của Maybelline rất tốt.

2017 - 2020 (Giai đoạn Suy giảm): Từ 2017, chỉ số BHI của Maybelline bắt đầu giảm rõ rệt từ 2.15 xuống khoảng 1.75 vào năm 2020. Đây là giai đoạn Maybelline mất đi sức khỏe thương hiệu nhanh chóng.

2020 - 2022 (Giai đoạn Ổn định thấp): Chỉ số BHI giữ tương đối ổn định quanh mức 1.75 - 1.8. Mặc dù đã dừng đà giảm mạnh, nhưng BHI hiện tại của Maybelline (khoảng 1.8 năm 2022) nằm ở nhóm trung bình thấp của các thương hiệu.

Tóm lại: Maybelline đã chuyển từ một thương hiệu có sức khỏe tốt nhất nhì năm 2016 sang một thương hiệu có chỉ số trung bình thấp vào năm 2022.

So với Kay Beauty (Màu Xanh Đậm) & Nykaa Naturals (Màu Vàng): Hai thương hiệu này duy trì BHI ở mức cao nhất (trên 2.0) trong hầu hết các năm.

Nykaa Naturals có xu hướng giảm nhẹ nhưng vẫn ở mức BHI cao nhất (trên 1.9 năm 2022). Kay Beauty có chỉ số tăng vọt năm 2019 (đạt đỉnh trên 2.2) và duy trì sự ổn định ở mức cao. Lakme (Màu Đỏ Tươi): Có sự sụt giảm và phục hồi mạnh mẽ. Sụt giảm nghiêm trọng 2017-2020 (xuống 1.15), nhưng phục hồi ấn tượng từ 2020-2022, vượt qua cả Maybelline năm 2021-2022.

Kết luận: Maybelline đã thua xa các thương hiệu dẫn đầu thị trường về sức khỏe thương hiệu kể từ năm 2017.

**4.2.13.2 Quyết định**

Kết hợp với phân tích từ biểu đồ trước (tỷ lệ đánh giá tiêu cực cao), các quyết định chiến lược phải tập trung vào đảo ngược sự suy giảm lịch sử và cải thiện chất lượng thương hiệu.

* Ưu tiên 1: Đảo ngược Xu hướng Giảm Sút (Reverse the Trend)

Quyết định: Đặt mục tiêu rõ ràng là phải tăng BHI vượt qua mốc 1.9 trong vòng 1-2 năm tới.

Hành động: Đầu tư mạnh vào các yếu tố cốt lõi cấu thành BHI: Đổi mới Sản phẩm (để giải quyết đánh giá tiêu cực) và Tăng cường Nhận diện (qua các chiến dịch tiếp thị hiệu quả, đặc biệt là Digital).

* Ưu tiên 2: Phân tích Khoảng cách (Gap Analysis)

Quyết định: Nghiên cứu sâu về các yếu tố giúp Kay Beauty và Nykaa Naturals duy trì BHI cao, và làm thế nào Lakme đạt được sự phục hồi BHI ngoạn mục (từ 2020 đến 2022).

Hành động: Tìm hiểu các chiến lược thành công (ví dụ: mô hình phân phối, định vị phân khúc, chiến lược KOL/KOC của các đối thủ này để tìm ra mô hình học hỏi phù hợp với Maybelline.

**4.2.13.3 Cách đọc biểu đồ**

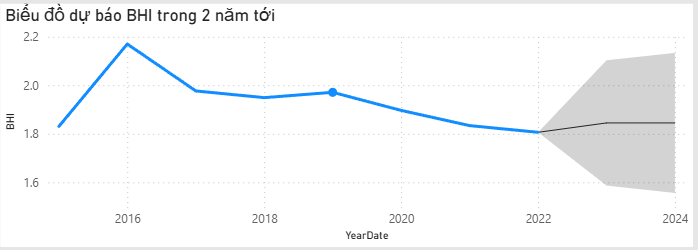
Trục Ngang (X-axis): Year (Năm) - Thể hiện thời gian từ 2015 đến 2022.

Trục Dọc (Y-axis): BHI - Thể hiện giá trị Chỉ số Sức khỏe Thương hiệu (từ 1.0 đến 2.4). Chỉ số BHI càng cao, sức khỏe thương hiệu càng tốt.

Đường Màu: Mỗi đường màu đại diện cho brand\_name (tên thương hiệu).

Maybelline New York (Màu hồng đậm/Magenta): Đường cần tập trung phân tích.

**4.2.14 Biểu đồ dự báo BHI**



Hình 4. 19 Biểu đồ dự báo BHI

**4.2.14.1 Phân tích**

* Phân tích Dữ liệu Lịch sử (2015 - 2022)

Tăng trưởng ban đầu: BHI tăng từ khoảng 1.82 (2015) lên đỉnh 2.18 (2016).

Suy giảm dài hạn: Sau 2016, BHI giảm gần như liên tục qua các năm, từ 2.18 xuống còn khoảng 1.80 vào cuối năm 2022.

Kết luận: Thương hiệu này đã trải qua một sự suy giảm sức khỏe thương hiệu nghiêm trọng và kéo dài trong 6 năm qua, mất đi gần 0.4 điểm BHI. Hiện tại dự báo

* Phân tích Dự báo (2022 - 2024)

Dự báo Trung tâm (Đường Đen): Đường dự báo gần như đi ngang, giữ ổn định ở mức khoảng 1.80 - 1.83 trong 2 năm tới (2023 và 2024). Điều này ngụ ý mô hình dự báo không nhìn thấy sự thay đổi lớn nào trong các yếu tố ảnh hưởng.

Khoảng Tin cậy (Vùng Xám): Vùng tô bóng rất rộng, đặc biệt là vào năm 2024, dao động từ khoảng 1.55 đến 2.15. Hàm ý: Mặc dù dự báo trung bình là ổn định, nhưng sự rộng lớn của khoảng tin cậy cho thấy có rủi ro lớn về khả năng BHI có thể rơi xuống mức rất thấp (1.55), hoặc ngược lại, có cơ hội phục hồi mạnh mẽ (2.15). Thương hiệu này đang ở trong một trạng thái bất ổn cao.

**4.2.14.2 Quyết định**

Cần xem xét các chiến lược đổi mới lớn để tạo ra hiệu ứng tích cực mạnh mẽ, ví dụ: Ra mắt sản phẩm chủ lực (Hero Product Launch): Tung ra một sản phẩm hoàn toàn mới, được nghiên cứu kỹ lưỡng để giải quyết các vấn đề đã gây ra đánh giá tiêu cực (từ biểu đồ trước), tạo ra sự hứng thú và BHI tăng vọt.

Tái định vị Thương hiệu (Rebranding): Cải tổ hình ảnh hoặc thông điệp cốt lõi để kết nối lại với người tiêu dùng, thay đổi nhận thức trì trệ hiện tại. Tập trung vào các biện pháp phòng ngừa và duy trì:

Tối ưu hóa Trải nghiệm Khách hàng (CX): Đảm bảo chất lượng dịch vụ, phân phối, và xử lý phàn nàn là nhất quán để tránh các yếu tố kích hoạt đánh giá tiêu cực diện rộng, vốn là nguyên nhân chính làm giảm BHI.

Giám sát Sentiment liên tục: Thiết lập hệ thống cảnh báo sớm để phát hiện và xử lý ngay lập tức bất kỳ vấn đề nào có thể gây ảnh hưởng xấu đến danh tiếng.

**4.2.14.3 Cách đọc biểu**

Trục Ngang (X-axis): Year Date (Năm) - Thể hiện dữ liệu lịch sử (từ 2015 đến 2022) và khoảng thời gian dự báo (từ 2022 đến 2024).

Trục Dọc (Y-axis): BHI - Chỉ số Sức khỏe Thương hiệu.

Đường Màu Xanh (Dữ liệu Lịch sử): Thể hiện xu hướng BHI thực tế từ 2015 đến 2022.

Đường Màu Đen (Dự báo Trung tâm): Thể hiện giá trị BHI dự báo trung bình/khả năng cao nhất trong 2 năm tới (2022-2024).

Vùng Tô Bóng Màu Xám (Khoảng tin cậy): Thể hiện phạm vi khả năng ( khoảng 95% tin cậy) mà giá trị BHI thực tế có thể rơi vào.

**4.2.15 Biểu đồ phân rã nguyên nhân gây sự biến động BHI**



Hình 4. 20 Biểu đồ phân rã nguyên nhân gây sự biến động BHI

**4.2.15.1 Phân tích**

Biểu đồ mở rộng cho thấy các năm có BHI trung bình cao nhất và thấp nhất đối với Maybelline:

* Tác động Tích cực (BHI Cao): 2016 (2.17) và 2017 (1.98). Điều này khớp với biểu đồ xu hướng trước đó, nơi 2016 là năm đỉnh điểm BHI của Maybelline.
* Tác động Tiêu cực (BHI Thấp): 2021 (1.83) và 2015 (1.83) (mặc dù 2015 là năm bắt đầu cao). 2018 (1.95) dường như là một năm có BHI ở mức trung bình của Maybelline.

Biểu đồ mở rộng đường dẫn đến tầng tiếp theo là year và month, ví dụ: 2018 và tháng 5, sau đó phân rã theo cảm xúc:

* Yếu tố Tác động mạnh nhất (Tích cực):

Sentiment Label = Positive (Tích cực): BHI trung bình là 2.02. Rating = 5: BHI trung bình là 2.32. (Đây là giá trị BHI/Rating cao nhất trong đường dẫn). Rating = 2: BHI trung bình là 1.04. (Đây là giá trị BHI/Rating thấp nhất trong đường dẫn).

* Yếu tố Tác động yếu nhất (Tiêu cực):

Sentiment Label = Negative (Tiêu cực): BHI trung bình là 1.50.

Kết luận tổng quan: Biểu đồ này xác nhận một cách định lượng rằng cảm xúc đánh giá (Sentiment Label) và xếp hạng sao (Rating) là các yếu tố ảnh hưởng mạnh nhất đến sự biến động BHI của Maybelline. Đặc biệt, việc nhận được rating 5 có thể đẩy BHI lên rất cao, trong khi rating 2 kéo BHI xuống rất thấp (1.04). Có thể thay đổi theo nhu cầu năm và tháng muốn xem hoặc phân tích.

**4.2.15.2 Quyết định**

Quyết định: Cân nhắc Tác động của Thời gian

Mục tiêu: Nghiên cứu sâu hơn các hoạt động xảy ra trong các năm BHI cao (2016) và năm BHI thấp (2021) để tìm ra mô hình thành công.

Hành động: Phân tích các chiến dịch ra mắt sản phẩm, mức chi tiêu tiếp thị, hoặc các sự kiện cạnh tranh xảy ra vào năm 2016 (đỉnh cao 2.17) và cố gắng tái tạo lại yếu tố thành công đó. Ngược lại, điều tra điều gì đã làm giảm BHI vào năm 2021 (1.83) để tránh lặp lại.

**4.2.15.3 Cách đọc biểu đồ**

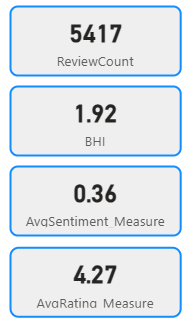
Chỉ số Mục tiêu (Root): BHI (1.92). Đây là giá trị BHI trung bình của dữ liệu được chọn (trong trường hợp này, dữ liệu đã được lọc sẵn cho Maybelline New York).

Các Cấp độ Phân tích (Nodes): Các cột thể hiện các chiều phân tích dữ liệu: brand\_name, product\_title, year, month, sentiment\_label, rating.

Các Thanh Ngang (Bars): Độ dài của thanh ngang thể hiện giá trị BHI trung bình tương ứng với giá trị đó (ví dụ: BHI trung bình của Maybelline New York là 1.92).

Các Mũi Tên (Arrows): Các mũi tên chỉ lên/xuống trên tiêu đề cột cho thấy biểu đồ đang được cấu hình để tìm kiếm các đường dẫn (paths) dẫn đến giá trị BHI cao nhất hoặc giá trị BHI thấp nhất. Trong biểu đồ này, các mũi tên có thể được dùng để mở rộng theo các yếu tố tác động mạnh nhất.

**4.2.15 Xây dựng KPI Card tổng hợp của Maybelline New York**



Hình 4. 21 Xây dựng KPI Card tổng hợp của Maybelline New York

**4.2.15.1 Phân tích**

ReviewCount = 5,417:  mức review rất lớn. Thương hiệu được quan tâm cao.

Mẫu dữ liệu đủ mạnh để đánh giá sức khỏe. Maybelline có mức độ nhận diện & tham gia (engagement) tốt trên các kênh đánh giá

Chỉ số BHI: BHI ~1.92 phản ánh sức khỏe thương hiệu ổn định, tích cực.  
Tương ứng với mức cảm xúc và đánh giá hiện tại → Maybelline có Brand Health khá mạnh trong phân khúc mỹ phẩm bình dân – trung cấp.

Trung bình điểm cảm xúc: 0.36 > 0.05 → Sentiment tích cực (Positive), Người tiêu dùng có xu hướng viết review mang tính khen, hài lòng. Trung bình điểm đánh giá: Mức rating khá tốt (trên 4.0).

#### **4.2.15.2 Quyết định**

Quyết định: Đặt mục tiêu BHI phục hồi lên 2.0 và tăng Khối lượng Đánh giá lên 15% trong vòng 2 năm.

Hành động Cốt lõi: Chiến dịch "Khuếch đại Đánh giá Tích cực" (Positive Amplification).

Tăng cường Tiếp cận: Tăng cường hoạt động Digital Marketing tập trung vào việc tạo ra nội dung hấp dẫn và tương tác cao để tăng khối lượng thảo luận tổng thể.

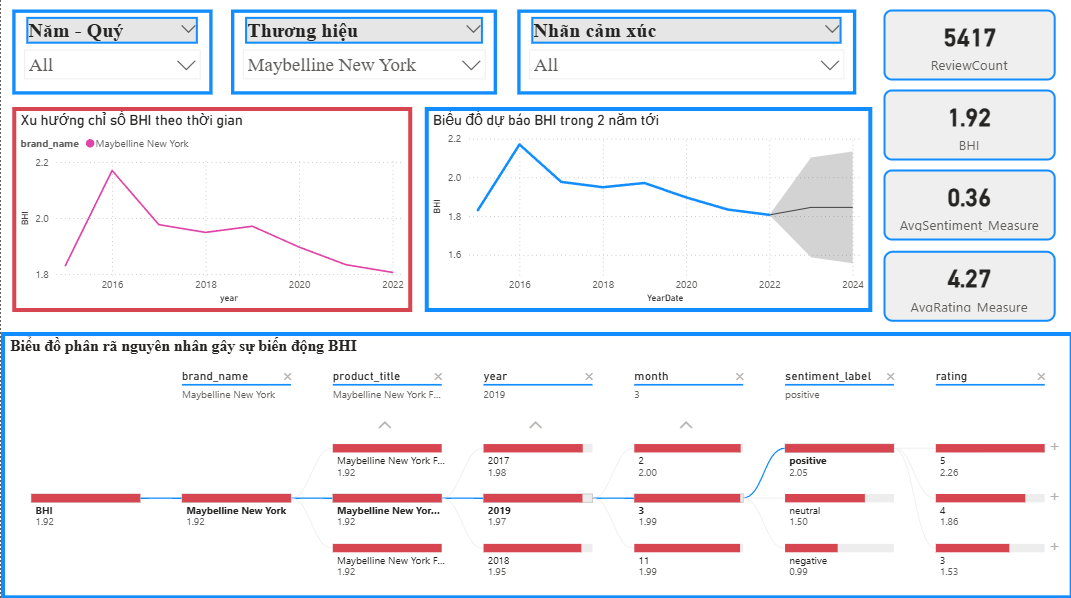
Chuyển đổi Tích cực: Sau khi chất lượng sản phẩm được đảm bảo, triển khai các chương trình thưởng/kích thích đánh giá 5 sao (Rating 5) để chuyển đổi khách hàng hài lòng thành bằng chứng xã hội.

**4.2.15.3 Cách đọc biểu đồ**

KPI Chính (Ví dụ: BHI): Giá trị sức khỏe thương hiệu tổng thể: 1.92

Đánh giá Trung bình (Average Rating): Giá trị xếp hạng sao trung bình của các sản phẩm: 4.27. Phân bổ Cảm xúc (Sentiment Split): Điểm cảm xúc trung bình của thương hiệu: 0.36 (Positvie). Tổng số bài đánh giá: số lượng bài đánh giá dành cho thương hiệu trong khoảng thời gian.

### 4.2.16 Dashboard sức khỏe thương hiệu và chiến lược



Hình 4. 22 Dashboard sức khỏe thương hiệu và chiến lược

**4.3 Đánh giá kết quả**

Kết quả trực quan hóa cho thấy Maybelline là thương hiệu có lượng review cao, thể hiện mức độ quan tâm lớn từ khách hàng. Tỷ lệ cảm xúc tích cực chiếm ưu thế, phản ánh sự hài lòng về chất lượng và độ phổ biến của các dòng son, kem nền.

Điểm đánh giá (rating) của Maybelline chủ yếu ở mức cao, tuy nhiên vẫn có một số phản hồi tiêu cực liên quan đến độ bám màu và giá ở một số dòng sản phẩm. Xu hướng cảm xúc theo thời gian khá ổn định, ít biến động mạnh và thường tăng trong các giai đoạn ra mắt sản phẩm mới.

So với các thương hiệu khác, Maybelline duy trì hình ảnh tốt, có chỉ số KPI (BHI, Rating trung bình, Sentiment Score) cao và ổn định.

**4.4 Hướng phát triển trong tương lai**

Mở rộng phạm vi dữ liệu thu thập: Trong tương lai, dữ liệu cảm xúc có thể được thu thập từ nhiều nền tảng thương mại và mạng xã hội như Shopee, TikTok Shop, Facebook, Instagram… nhằm mang lại cái nhìn toàn diện hơn về hành vi người tiêu dùng.

Tích hợp mô hình AI phân tích cảm xúc nâng cao: Ứng dụng các mô hình học sâu như BERT, GPT hoặc mô hình phân tích chủ đề (Topic Modeling) để phân loại cảm xúc chính xác hơn và phát hiện chủ đề được nhắc đến nhiều trong đánh giá.

Dự báo xu hướng thương hiệu: Kết hợp các thuật toán dự báo (Forecasting) để dự đoán chỉ số BHI, số lượng đánh giá, rating trung bình và mức độ quan tâm sản phẩm trong tương lai.

Xây dựng hệ thống cảnh báo tự động: Thiết lập cảnh báo khi lượng cảm xúc tiêu cực tăng đột ngột, giúp doanh nghiệp xử lý vấn đề kịp thời và giảm thiểu rủi ro hình ảnh thương hiệu.

Tăng cường khả năng tương tác dashboard: Cải thiện bộ lọc (slicer), bổ sung phân tích theo phân khúc khách hàng, vùng địa lý và giá sản phẩm nhằm giúp Power BI trở thành công cụ theo dõi chiến lược toàn diện.

Tích hợp với hệ thống CRM/ERP: Việc đồng bộ dữ liệu cảm xúc với hệ thống quản trị khách hàng sẽ tạo ra một chu trình khép kín, hỗ trợ doanh nghiệp trong việc cá nhân hóa trải nghiệm khách hàng và tối ưu dịch vụ.