

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THÔNG THÔNG TIN



BÁO CÁO ĐÒ ÁN MÔN HỌC KHO DỮ LIỆU VÀ OLAP
ĐỀ TÀI:
**PHÂN TÍCH DỮ LIỆU VỀ XU HƯỚNG THỜI TRANG
BỀN VỮNG-THÂN THIỆN VỚI MÔI TRƯỜNG**

Giảng viên hướng dẫn: ThS. Nguyễn Thị Kim Phụng

Lớp: IS217.P12

Nhóm sinh viên thực hiện:

1. Võ Thanh Nhàn MSSV: 22521008

2. Huỳnh Ngọc Diễm Phúc MSSV: 22521113

TP. HỒ CHÍ MINH, NĂM 2024

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

TP.HCM, ngày ... tháng ... năm 2024

Người nhân xét

(Ký tên và ghi rõ họ tên)

LỜI CẢM ƠN

Lời đầu tiên, nhóm thực hiện đề tài xin gửi tới Cô Nguyễn Thị Kim Phụng – Giảng viên môn *Kho dữ liệu và OLAP* và chị Ngô Thùy Yên Nhi lời cảm ơn chân thành và sâu sắc nhất.

Chúng em xin cảm ơn Cô vì những tri thức mà Cô truyền đạt đến chúng em một cách hăng say và đầy nhiệt huyết. Cảm ơn Cô vì những lời khuyên, hướng dẫn và góp ý đã giúp chúng em có được định hướng đúng đắn cho đề tài của nhóm.

Ngoài ra, chúng em cũng xin cảm ơn Chị Yên Nhi vì những sự chia, giúp đỡ, giải đáp thắc mắc trong suốt quá trình chúng em hoàn thiện đồ án.

Trong quá trình thực hiện đồ án, nhóm chúng em đã vận dụng những kiến thức được Cô cung cấp tại lớp, đồng thời chủ động học hỏi và tiếp thu từ nhiều nguồn tài liệu khác nhau để có thể thực hiện đồ án này một cách tốt nhất. Tuy nhiên, do chưa có nhiều kinh nghiệm nên đồ án vẫn không thể tránh khỏi những điểm thiếu sót, bất cập. Vì vậy, chúng em mong rằng Cô sẽ đưa ra nhận xét, đánh giá và góp ý để chúng em có thể cung cấp kiến thức cũng như rút kinh nghiệm để nhóm có thể thực hiện các đề tài trong tương lai tốt hơn và hoàn thiện hơn.

Chúng em xin chân thành cảm ơn!

Nhóm sinh viên thực hiện.

DANH MỤC BẢNG.....	8
DANH MỤC HÌNH ẢNH	9
1. GIỚI THIỆU	32
1.1 Lý do chọn đề tài.....	32
1.2 Giới thiệu về dataset.....	33
1.2.1 Tổng quan dataset.....	33
1.2.2 Mô tả số dòng, cột, thời gian thu thập dữ liệu.....	34
1.2.3 Làm sạch dữ liệu	34
1.2.4 Mô tả chi tiết các cột thuộc tính sử dụng	38
1.3 Thiết kế kho dữ liệu	39
1.3.1 Lược đồ kho dữ liệu (lược đồ hình sao)	39
1.3.2 Bảng Fact.....	40
1.3.3 Bảng Dim_Brand.....	42
1.3.4 Bảng Dim_Material	42
1.3.5 Bảng Dim_Recycling	43
1.3.6 Bảng Dim_Product	44
1.3.7 Bảng Dim_Certification	44
1.3.8 Bảng Dim_Market_Trend	45
1.3.9 Bảng Dim_Sustainability_Rating.....	45
1.3.10 Bảng Dim_Time	46
2. QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)	48
2.1 Chuẩn bị công cụ.....	48
2.1.1 Chuẩn bị cơ sở dữ liệu.....	50
2.1.2 Tạo mới project SSIS	50
2.2 Tạo bảng Dim và Fact.....	52
2.2.1 Bảng Dim_Brand.....	56
2.2.2 Bảng Dim_Material	64
2.2.3 Bảng Dim_Recycling	67
2.2.4 Bảng Dim_Product	70
2.2.5 Bảng Dim_Certification	74
2.2.6 Bảng Dim_Market_Trend	77
2.2.7 Bảng Dim_Sustainability_Rating.....	81
2.2.8 Bảng Dim_Time	84

2.2.9 Bảng Dim_Eco_Friendly_Manufacturing	88
2.2.10 Bảng Dim_Fact_Raw	92
2.2.11 Merge Dim_Fact_Raw và Dim_Brand thành Fact1	97
2.2.12 Merge Fact1 và Dim_Material thành Fact2	104
2.2.13 Merge Fact2 và Dim_Recycling thành Fact3	110
2.2.14 Merge Fact3 và Dim_Product thành Fact4	117
2.2.15 Merge Fact4 và Dim_Certification thành Fact5	124
2.2.16 Merge Fact5 và Dim_Market_Trend thành Fact6	132
2.2.17 Merge Fact6 và Dim_Time thành Fact7	139
2.2.18 Merge Fact7 và Dim_Eco_Friendly_Manufacturing thành Fact8	145
2.2.19 Merge Fact8 và Dim_Sustainability_Rating thành Fact	153
2.2.20 Tạo các ràng buộc các khóa ngoại	161
2.2.21 Các điều kiện của Original Data	164
2.2.22 Các điều kiện của Data_Warehouse	166
2.3 Chạy dự án SSIS	167
2.3.1 Kiểm tra dữ liệu các bảng	172
2.3.2 Lược đồ dữ liệu sau khi hoàn thành	176
3. PHÂN TÍCH DỮ LIỆU TRONG KHO SSAS	178
3.1 Tạo mới project SSAS:	178
3.1.1 Xác định dữ liệu nguồn (Data Sources)	180
3.1.2 Xác định khung nhìn dữ liệu nguồn (Data Source Views):	188
3.1.3 Xây dựng khối (Cube)	193
3.1.4 Xác định các chiều (Dimensions)	198
3.1.5 Chạy dự án SSAS	208
3.1.6 Tạo Hierarchies và định nghĩa Attribute Relationship	210
3.1.7: Process lại Project	217
3.2 Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối CUBE	219
3.2.1 Câu truy vấn 1	219
3.2.2 Câu truy vấn 2	220
3.2.3 Câu truy vấn 3	221
3.2.4 Câu truy vấn 4	221
3.2.5 Câu truy vấn 5	223
3.2.6 Câu truy vấn 6	225
3.2.7 Câu truy vấn 7	227

3.2.8 Câu truy vấn 8	228
3.2.9 Câu truy vấn 9	229
3.2.10 Câu truy vấn 10	230
3.3 Quá trình phân tích dữ liệu bằng công cụ SSAS trên Excel	231
3.3.1 Câu truy vấn 1	231
3.3.2 Câu truy vấn 2	232
3.3.3 Câu truy vấn 3	234
3.3.4 Câu truy vấn 4	235
3.3.5 Câu truy vấn 5	238
3.3.6 Câu truy vấn 6	239
3.3.7 Câu truy vấn 7	241
3.3.8 Câu truy vấn 8	244
3.3.9 Câu truy vấn 9	246
3.3.10 Câu truy vấn 10	247
3.4 Quá trình phân tích dữ liệu bằng ngôn ngữ MDX	248
3.4.1 Câu truy vấn 1:	248
3.4.2 Câu truy vấn 2:	249
3.4.3 Câu truy vấn 3:	250
3.4.4 Câu truy vấn 4:	251
3.4.5 Câu truy vấn 5:	252
3.4.6 Câu truy vấn 6:	253
3.4.7 Câu truy vấn 7:	254
3.4.8 Câu truy vấn 8:	255
3.4.9 Câu truy vấn 9:	255
3.4.10 Câu truy vấn 10:	256
3.5 Quá trình lập báo biểu bằng công cụ Power BI	257
3.5.1 Tạo mới project và kết nối với cơ sở dữ liệu trên BI	257
3.5.2 Report 1	260
3.5.3 Report 2	264
3.5.4 Report 3	268
3.5.5 Report 4	272
3.5.6 Report 5	277
3.5.7 Report 6	281
3.5.8 Report 7	284

3.5.9 Report 8	288
3.5.10 Report 9	293
3.5.11 Report 10	296
4. QUÁ TRÌNH KHAI THÁC DỮ LIỆU (DATA MINING)	300
4.1 Tiền xử lý dữ liệu	300
4.2 Ứng dụng mô hình thuật toán	304
4.2.1 Chuẩn bị dữ liệu đưa vào mô hình	304
4.2.2 Xây dựng mô hình Decision Tree	306
4.2.3 Xây dựng mô hình Random Forest.	310
5. TỔNG KẾT	314
5.1 Kết quả đạt được.....	314
5.2 Thuận lợi.....	315
5.3 Khó khăn	315
BẢNG PHÂN CÔNG CÔNG VIỆC	316

TỪ NGỮ VIẾT TẮT VÀ THUẬT NGỮ

Bảng 1. Bảng từ viết tắt và thuật ngữ.

STT	Từ viết tắt / thuật ngữ	Ý nghĩa
1	SSIS	SQL Server Integration Services
2	SSAS	SQL Server Analysis Services

DANH MỤC BẢNG

Bảng 1. Bảng từ viết tắt và thuật ngữ	11
Bảng 1. 1. Bảng mô tả chi tiết các cột thuộc tính.	33
Bảng 1. 2. Bảng mô tả Fact_Fatilities.	35
Bảng 1. 3. Bảng mô tả Dim_Person.	36
Bảng 1. 4. Bảng mô tả Dim_Location.	37
Bảng 1. 5. Bảng mô tả Dim_Date.	37
Bảng 1. 6. Bảng mô tả Dim_Injury	38
Bảng 1. 7. Bảng mô tả Dim_Ammunition.	38
Bảng 1. 8. Bảng mô tả Dim_Killed_by	39
Bảng 1. 9. Bảng mô tả Dim_Residence.	39
Bảng 1. 10. Bảng mô tả Dim_Hostilities.	40

DANH MỤC HÌNH ẢNH

Hình 1. 01. Dataset Sustainable Fashion: Eco-Friendly từ Kaggle.	33
Hình 1. 02. Làm sạch dữ liệu (1).	34
Hình 1. 03. Làm sạch dữ liệu (2).	34
Hình 1. 04. Làm sạch dữ liệu (3).	35
Hình 1. 05. Làm sạch dữ liệu (3).	35
Hình 1. 06. Làm sạch dữ liệu (3).	36
Hình 1. 07. Làm sạch dữ liệu (3).	36
Hình 1. 08. Làm sạch dữ liệu (8).	36
Hình 1. 09. Làm sạch dữ liệu (9).	36
Hình 1. 10. Lược đồ kho dữ liệu (hình sao).	38
Hình 2. 01. Giao diện trang chủ Visual Studio 2022	46
Hình 2. 02. Chọn Manage Extensions ở tab Extensions.....	46
Hình 2. 03. Tìm kiếm và cài đặt SQL Server Integration Services Project.	47
Hình 2. 04. Hoàn thành cài đặt SQL Server Integration Services Projects.	47
Hình 2. 05. Kết nối với SQL Server bằng Windows Authentication.	48
Hình 2. 06. Tạo database.....	48
Hình 2. 07. Chọn Create a new project tại trang chủ Visual Studio 2022.....	49
Hình 2. 08. Chọn Integration Services Project làm template.....	49
Hình 2. 09. Đặt tên cho project và chọn nơi lưu.	50
Hình 2. 10. Tạo Data Flow Task.....	50
Hình 2. 11. Tạo Flat File Source lấy dữ liệu từ file .csv.....	50
Hình 2. 12. Flat File Source Editor.	51

Hình 2. 13. Tạo Flat File Connection Manager mới.....	51
Hình 2. 14. Kiểm tra dữ liệu sẽ nạp vào.	52
Hình 2. 15. Các cột thuộc tính sẽ nạp vào Flat File Source.....	52
Hình 2. 16. Tạo Multicast để phân tán dữ liệu.....	53
Hình 2. 17. Tạo Sort đổi tên " Sort_Dim_Brand".....	53
Hình 2. 18. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	54
Hình 2. 19. Tạo OLE DB Destination đổi tên "Dim_Brand".	54
Hình 2. 20. OLE DB Destination Editor.....	55
Hình 2. 21. Thiết lập Data connections.....	55
Hình 2. 22. Thiết lập kết nối mới(1).	56
Hình 2. 23. Thiết lập kết nối mới(2).	57
Hình 2. 24. Kiểm tra kết nối.	57
Hình 2. 25. Data connections hiện có.	57
Hình 2. 26. Tạo bảng Dim_Brand.....	58
Hình 2. 27. Tạo bảng Dim_Brand(2).	58
Hình 2. 28. Tạo bảng Dim_Brand(3).	59
Hình 2. 29. Kiểm tra các cột thuộc tính	59
Hình 2. 30. Tạo Sort đổi tên " Sort_Dim_Material".....	60
Hình 2. 31. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	60
Hình 2. 32. Tạo OLE DB Destination đổi tên "Dim_Material".....	61
Hình 2. 33. Tạo bảng "Dim_Material"(1).....	61
Hình 2. 34. Tạo bảng "Dim_Material"(2).....	62
Hình 2. 35. Tạo bảng "Dim_Material"(3).....	62

Hình 2. 36. Kiểm tra các cột thuộc tính.....	63
Hình 2. 37. Tạo Sort đổi tên " Sort_Dim_Recycling".	63
Hình 2. 38. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	64
Hình 2. 39. Tạo OLE DB Destination đổi tên "Dim_Recycling".....	64
Hình 2. 40. Tạo bảng "Dim_Recycling"(1).	65
Hình 2. 41. Tạo bảng "Dim_Recycling"(2).	65
Hình 2.42. Tạo bảng "Dim_Recycling"(3).	66
Hình 2.43. Kiểm tra các cột thuộc tính.	66
Hình 2. 44. Tạo Sort đổi tên " Sort_Dim_Product".	66
Hình 2.45. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	67
Hình 2. 46. Tạo OLE DB Destination đổi tên "Dim_Product".....	67
Hình 2. 47. Tạo bảng "Dim_Product"(1).	68
Hình 2. 48. Tạo bảng "Dim_Product"(2).	68
Hình 2. 49. Tạo bảng "Dim_Product"(3).	69
Hình 2. 50. Kiểm tra các cột thuộc tính.	69
Hình 2. 51. Tạo Sort đổi tên " Sort_Dim_Certification".	70
Hình 2. 52. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	70
Hình 2. 53. Tạo OLE DB Destination đổi tên "Dim_Certification".....	71
Hình 2. 54. Tạo bảng "Dim_Certification"(1).	71
Hình 2. 55. Tạo bảng "Dim_Certification"(2).	72
Hình 2. 56. Tạo bảng "Dim_Certification"(3).	72
Hình 2. 57. Kiểm tra các cột thuộc tính.	73
Hình 2. 58. Tạo Sort đổi tên "Sort_Dim_Market_Trend".	73

Hình 2. 59. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	74
Hình 2. 60. Tạo OLE DB Destination đổi tên "Dim_Market_Trend".	74
Hình 2. 61. Tạo bảng "Dim_Market_Trend"(1).	74
Hình 2. 62. Tạo bảng "Dim_Market_Trend"(2).	75
Hình 2. 63. Tạo bảng "Dim_Market_Trend"(3).	76
Hình 2. 64. Kiểm tra các cột thuộc tính.	76
Hình 2. 65. Tạo Sort đổi tên "Sort_Dim_Sustainability_Rating"....	76
Hình 2. 66. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	77
Hình 2. 67. Tạo OLE DB Destination đổi tên "Dim_Sustainability_Rating".	77
Hình 2. 68. Tạo bảng "Dim_Sustainability_Rating"(1).....	78
Hình 2. 70. Tạo bảng "Dim_Sustainability_Rating"(3).....	79
Hình 2. 71. Kiểm tra các cột thuộc tính.	79
Hình 2. 72. Tạo Sort đổi tên "Sort_Dim_Time"....	80
Hình 2. 73. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	80
Hình 2. 74. Tạo OLE DB Destination đổi tên "Dim_Time"....	80
Hình 2. 75. Tạo bảng "Dim_Time"(1).....	81
Hình 2. 76. Tạo bảng "Dim_Time"(2).	81
Hình 2. 78. Kiểm tra các cột thuộc tính.	82
Hình 2. 79. Tạo Sort đổi tên "Sort_Dim_Eco_Friendly_Manufacturing"....	83
Hình 2. 80. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.	83
Hình 2. 81. Tạo OLE DB Destination đổi tên "Dim_Eco_Friendly_Manufacturing"....	84
Hình 2. 82. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(1).....	84
Hình 2. 83. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(2).....	85

Hình 2. 84. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(3).....	85
Hình 2. 85. Kiểm tra các cột thuộc tính.....	86
Hình 2. 86. Tạo Sort đổi tên "Sort_Dim_Fact_Raw".....	86
Hình 2. 88. Đổi kiểu dữ liệu.....	88
Hình 2. 89. Tạo OLE DB Destination đổi tên "Dim_Fact_Raw".....	88
Hình 2. 90. Tạo bảng "Dim_Fact_Raw"(1).....	89
Hình 2. 91. Tạo bảng "Dim_Fact_Raw"(2).....	90
Hình 2. 92. Tạo bảng "Dim_Fact_Raw"(3).....	90
Hình 2. 93. Kiểm tra các cột thuộc tính.....	91
Hình 2. 94. Tạo Data Flow Task có tên "Merge Dim_Fact_Raw and Dim_Brand to Fact1".....	92
Hình 2. 95. Các bước thực hiện (1).....	92
Hình 2. 96. Các bước thực hiện (2).....	92
Hình 2. 97. Các bước thực hiện (3).....	93
Hình 2. 98. Các bước thực hiện (4).....	93
Hình 2. 99. Các bước thực hiện (5).....	94
Hình 2. 100. Các bước thực hiện (6).....	94
Hình 2. 101. Các bước thực hiện (7).....	94
Hình 2. 102. Các bước thực hiện (8).....	95
Hình 2. 103. Các bước thực hiện (9).....	95
Hình 2. 104. Các bước thực hiện (10).....	96
Hình 2. 105. Các bước thực hiện (11).....	97
Hình 2. 106. Các bước thực hiện (12).....	97

Hình 2. 107. Các bước thực hiện (13).....	98
Hình 2. 108. Sau khi thực hiện xong.....	98
Hình 2. 109. Tạo Data Flow Task có tên "Merge Fact1 and Dim_Material to Fact2"....	98
Hình 2. 110. Các bước thực hiện (1).....	99
Hình 2. 111. Các bước thực hiện (2).....	99
Hình 2. 112. Các bước thực hiện (3).....	99
Hình 2. 113. Các bước thực hiện (4).....	100
Hình 2. 114. Các bước thực hiện (5).....	100
Hình 2. 115. Các bước thực hiện (6).....	101
Hình 2. 116. Các bước thực hiện (7).....	101
Hình 2. 117. Các bước thực hiện (8).....	101
Hình 2. 118. Các bước thực hiện (9).....	102
Hình 2. 119. Các bước thực hiện (10).....	103
Hình 2. 120. Các bước thực hiện (11).....	104
Hình 2. 121. Các bước thực hiện (12).....	104
Hình 2. 122. Các bước thực hiện (13).....	105
Hình 2. 123. Sau khi thực hiện xong.....	105
Hình 2. 124. Tạo Data Flow Task có tên "Merge Fact2 and Dim_Recycling to Fact3".	105
Hình 2. 125. Các bước thực hiện (1).....	106
Hình 2. 126. Các bước thực hiện (2).....	106
Hình 2. 127. Các bước thực hiện (3).....	107
Hình 2. 128. Các bước thực hiện (4).....	107

Hình 2. 130. Các bước thực hiện (6).....	108
Hình 2. 131. Các bước thực hiện (7).....	108
Hình 2. 132. Các bước thực hiện (8).....	108
Hình 2. 133. Các bước thực hiện (9).....	109
Hình 2. 134. Các bước thực hiện (10).....	109
Hình 2. 135. Các bước thực hiện (11).....	110
Hình 2. 136. Các bước thực hiện (12).....	111
Hình 2. 137. Các bước thực hiện (13).....	111
Hình 2. 138. Sau khi thực hiện xong.....	111
Hình 2. 139. Tạo Data Flow Task có tên "Merge Fact3 and Dim_Product to Fact4" ..	112
Hình 2. 140. Các bước thực hiện (1).....	112
Hình 2. 141. Các bước thực hiện (2).....	113
Hình 2. 142. Các bước thực hiện (3).....	113
Hình 2. 143. Các bước thực hiện (4).....	114
Hình 2. 144. Các bước thực hiện (5).....	114
Hình 2. 145. Các bước thực hiện (6).....	114
Hình 2. 146. Các bước thực hiện (7).....	115
Hình 2. 147. Các bước thực hiện (8).....	115
Hình 2. 148. Các bước thực hiện (9).....	115
Hình 2. 149. Các bước thực hiện (10).....	116
Hình 2. 150. Các bước thực hiện (11).....	117
Hình 2. 151. Các bước thực hiện (12).....	117
Hình 2. 152. Các bước thực hiện (13).....	118

Hình 2. 153. Sau khi thực hiện xong.....	118
Hình 2. 154. Tạo Data Flow Task có tên "Merge Fact4 and Dim_Certificationto Fact5".	118
Hình 2. 155. Các bước thực hiện (1).....	119
Hình 2. 156. Các bước thực hiện (2).....	119
Hình 2. 157. Các bước thực hiện (3).....	120
Hình 2. 158. Các bước thực hiện (4).....	120
Hình 2. 159. Các bước thực hiện (5).....	121
Hình 2. 160. Các bước thực hiện (6).....	121
Hình 2. 161. Các bước thực hiện (7).....	122
Hình 2. 162. Các bước thực hiện (8).....	122
Hình 2. 163. Các bước thực hiện (9).....	122
Hình 2. 164. Các bước thực hiện (10).....	123
Hình 2. 165. Các bước thực hiện (11).....	124
Hình 2. 166. Các bước thực hiện (12).....	124
Hình 2. 167. Các bước thực hiện (13).....	125
Hình 2. 168. Sau khi thực hiện xong.....	125
Hình 2. 169. Tạo Data Flow Task có tên "Merge Fact5 and Dim_Market_Trend to Fact6".....	125
Hình 2. 170. Các bước thực hiện (1).....	126
Hình 2. 171. Các bước thực hiện (2).....	126
Hình 2. 172. Các bước thực hiện (3).....	127
Hình 2. 173. Các bước thực hiện (4).....	127

Hình 2. 174. Các bước thực hiện (5).....	127
Hình 2. 175. Các bước thực hiện (6).....	128
Hình 2. 176. Các bước thực hiện (7).....	128
Hình 2. 177. Các bước thực hiện (8).....	128
Hình 2. 178. Các bước thực hiện (9).....	129
Hình 2. 179. Các bước thực hiện (10).....	129
Hình 2. 180. Các bước thực hiện (11).....	130
Hình 2. 181. Các bước thực hiện (12).....	131
Hình 2. 182. Các bước thực hiện (13).....	131
Hình 2. 182. Sau khi thực hiện xong.....	132
Hình 2. 183. Tạo Data Flow Task có tên "Merge Fact6 and Dim_Dim_Time to Fact7".	132
Hình 2. 184. Các bước thực hiện (1).....	132
Hình 2. 185. Các bước thực hiện (2).....	133
Hình 2. 186. Các bước thực hiện (3).....	133
Hình 2. 187. Các bước thực hiện (4).....	134
Hình 2. 188. Các bước thực hiện (5).....	134
Hình 2. 189. Các bước thực hiện (6).....	134
Hình 2. 190. Các bước thực hiện (7).....	135
Hình 2. 191. Các bước thực hiện (8).....	135
Hình 2. 192. Các bước thực hiện (9).....	135
Hình 2. 193. Các bước thực hiện (10).....	136
Hình 2. 194. Các bước thực hiện (11).....	137

Hình 2. 195. Các bước thực hiện (12).....	137
Hình 2. 196. Các bước thực hiện (13).....	138
Hình 2. 197. Sau khi thực hiện xong.....	138
Hình 2. 198. Tạo Data Flow Task có tên "Merge Fact7 and Dim_Eco_Friendly_Manufacturing to Fact8".....	138
Hình 2. 199. Các bước thực hiện (1).....	139
Hình 2. 200. Các bước thực hiện (2).....	139
Hình 2. 201. Các bước thực hiện (3).....	140
Hình 2. 202. Các bước thực hiện (4).....	140
Hình 2. 203. Các bước thực hiện (5).....	141
Hình 2. 204. Các bước thực hiện (6).....	141
Hình 2. 205. Các bước thực hiện (7).....	142
Hình 2. 206. Các bước thực hiện (8).....	142
Hình 2. 207. Các bước thực hiện (9).....	143
Hình 2. 208. Các bước thực hiện (10).....	143
Hình 2. 209. Các bước thực hiện (11).....	144
Hình 2. 210. Các bước thực hiện (12).....	145
Hình 2. 211. Các bước thực hiện (13).....	145
Hình 2. 212. Sau khi thực hiện xong.....	146
Hình 2. 213. Tạo Data Flow Task có tên "Merge Fact8 and Dim_Sustainability_Rating to Fact".....	146
Hình 2. 214. Các bước thực hiện (1).....	146
Hình 2. 215. Các bước thực hiện (2).....	147

Hình 2. 216. Các bước thực hiện (3).....	148
Hình 2. 217. Các bước thực hiện (4).....	148
Hình 2. 218. Các bước thực hiện (5).....	149
Hình 2. 219. Các bước thực hiện (6).....	149
Hình 2. 220. Các bước thực hiện (7).....	150
Hình 2. 221. Các bước thực hiện (8).....	150
Hình 2. 222. Các bước thực hiện (9).....	151
Hình 2. 223. Các bước thực hiện (10).....	151
Hình 2. 224. Các bước thực hiện (11).....	152
Hình 2. 225. Các bước thực hiện (12).....	153
Hình 2. 226. Các bước thực hiện (13).....	153
Hình 2. 226. Các bước thực hiện (14).....	154
Hình 2. 228. Câu hình phần SQL Statement trong Execute SQL Task Editor.....	156
Hình 2. 228. Sau khi thực hiện xong.....	157
Hình 2. 229. Tạo “Execute SQL Task” đổi tên thành “Các điều kiện của Original Data”.....	157
Hình 2. 230. Câu hình phần SQL Statement trong Execute SQL Task Editor.....	158
Hình 2. 231. Sau khi thực hiện xong.....	158
Hình 2. 232. Tạo “Excute SQL Task” đổi tên thành “Các điều kiện của Data_Warehouse”.....	159
Hình 2. 233. Câu hình phần SQL Statement trong Execute SQL Task Editor.....	159
Hình 2. 234. Sau khi thực hiện xong.....	160
Hình 2. 235. Kết quả chạy dự án SSIS.	160

Hình 2. 236. Cấu hình phần SQL Statement trong Execute SQL Task Editor.....	162
Hình 2. 237. Kết quả (1).	162
Hình 2. 238. Kết quả (2).	163
Hình 2. 239. Kết quả (3).	163
Hình 2. 240. Kết quả (4).	163
Hình 2. 241. Kết quả (5).	163
Hình 2. 242. Kết quả (6).	164
Hình 2. 243. Kết quả (7).	164
Hình 2. 244. Kết quả (8).	164
Hình 2. 245. Kết quả (9).	164
Hình 2. 246. Kết quả (10).	165
Hình 2. 247. Kết quả (11).	165
Hình 2. 248. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Fact.	165
Hình 2. 249. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Brand.	166
Hình 2. 250. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Material.....	166
Hình 2. 251. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Recycling.	167
Hình 2. 252. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Product.....	167
Hình 2. 253. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Certification....	168
Hình 2. 254. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Market_Trend. .	168
Hình 2. 255. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Sustainability_Rating.....	168
Hình 2. 257. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Eco_Friendly_Manufacturing.....	169

Hình 2. 258. Lược đồ sau khi hoàn thành.....	170
Hình 3. 1. Chọn Create a new project từ màn hình bắt đầu của Visual Studio 2022 ..	171
Hình 3. 2. Chọn project template Analysis Services Multidimensional Project.	171
Hình 3. 3. Đặt tên và chọn nơi lưu project.....	172
Hình 3. 4. Chọn Properties của project.....	173
Hình 3.5 Đổi tên server để deploy	173
Hình 3. 6. Cửa sổ Data Source Wizard hiện lên.....	174
Hình 3. 7. Cửa sổ Data Source Wizard hiện lên.....	174
Hình 3. 8. Tạo kết nối với cơ sở dữ liệu được tạo từ quá trình SSIS.....	175
Hình 3. 9. Chọn Provider	176
Hình 3. 10. Chọn server và Database.....	177
Hình 3. 11. Kết nối thành công	177
Hình 3. 12. Chọn Data Connection đã tạo.	178
Hình 3. 13. Chọn Use the service account.....	178
Hình 3. 14. Hoàn tất quá trình xác định Data Source.	179
Hình 3. 15. Hoàn tất quá trình xác định Data Source.	179
Hình 3. 16. Chọn New Data Source View từ Data Source Views.....	180
Hình 3. 17. Cửa sổ Data Source View Wizard hiện lên.	180
Hình 3. 18. Chọn Data Source đã tạo.....	181
Hình 3. 19. Thêm bảng Fact vào Data Source View.	181
Hình 3. 20. Thêm bảng Dim vào Data Source View.	182
Hình 3. 21. Hoàn tất quá trình xác định Data Source View.	182
Hình 3. 22. Hoàn tất quá trình xác định Data Source View.	183

Hình 3. 23. Kết quả quá trình xác định Data Source View.....	184
Hình 3. 24. Chọn New Cube từ Cubes.....	185
Hình 3. 25. Cửa sổ Cube Wizard hiện lên	185
Hình 3. 26. Chọn Use existing tables.....	186
Hình 3. 27. Chọn bảng Fact để phân chia measures.	186
Hình 3. 28. Chọn độ đo để xuất.	187
Hình 3. 29. Chọn các bảng Dim.....	187
Hình 3. 30. Kết quả quá trình xây dựng cube.	188
Hình 3. 31. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Market-Trend.	188
Hình 3.32 Cấu hình cho Dim Market Trend	189
Hình 3. 33. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Recycling.	189
Hình 3.34 Cấu hình cho Dim Recycling	190
Hình 3. 35. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Sustainability Rating.	190
Hình 3. 36. Cấu hình cho Dim Sustainability Rating.	191
Hình 3. 37. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Brand..	192
Hình 3. 38. Cấu hình cho Dim Brand.	192
Hình 3. 39. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Certification	193
Hình 3. 40. Cấu hình cho Dim Certification	193
Hình 3. 41. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Time... ..	194
Hình 3. 42. Cấu hình cho Dim Time.....	194

Hình 3. 43. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Material.	195
Hình 3. 44. Cấu hình cho Dim Material.....	195
Hình 3. 45. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Product.	196
Hình 3. 46. Cấu hình cho Dim Product.....	196
Hình 3. 47. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Eco_Friendly_Manufacturing	197
Hình 3. 48. Cấu hình cho Dim Eco_Friendly_Manufacturing	197
Hình 3. 49. Deploy dự án.....	198
Hình 3. 50. Deploy Project thành công.....	198
Hình 3. 51 Process dữ liệu	199
Hình 3. 52. Phân cấp cho Dim Brand.	199
Hình 3. 53. Đổi loại quan hệ trong Dim Brand.....	200
Hình 3. 54. Chọn Edit Attribute Relationship.....	200
Hình 3. 55. Đổi Source Attribute	201
Hình 3. 56. Kết quả sau khi đổi source attribute.....	201
Hình 3. 57. Kết quả sau khi điều chỉnh Attribute Relationship	202
Hình 3. 58. Chọn properties để cấu hình Dim Brand	202
Hình 3. 59. Cấu hình cột Country trong Dim Brand.....	203
Hình 3. 60. Cấu hình NameColumn cho thuộc tính Country.	203
Hình 3. 61. Cấu hình cho KeyColumn chp thuộc tính Brand Name	204
Hình 3. 62. Cấu hình NameColumn cho thuộc tính Brand Name	204
Hình 3. 63. Phân cấp cho Dim Brand thành công.....	205

Hình 3. 64 Deploy Project thành công.....	205
Hình 3. 65. Deploy Project thành công.....	206
Hình 3. 66. Deploy Project thành công.....	206
Hình 3. 67. Kết quả câu truy vấn 1 dùng khối CUBE	207
Hình 3. 68. Kết quả câu truy vấn 2 dùng CUBE.....	208
Hình 3. 69. Kết quả thực thi câu truy vấn 3 dùng CUBE	209
Hình 3. 70. Thao tác trên thuộc tính Sustainability Rating.....	210
Hình 3. 71. Thao tác trên thuộc tính Year.....	210
Hình 3. 72. Kết quả thực thi câu truy vấn 4 dùng CUBE	211
Hình 3. 73. Thao tác trên thuộc tính Country	212
Hình 3. 74. Thao tác trên thuộc tính Certifications	212
Hình 3. 75. Kết quả thực thi câu truy vấn 5 dùng CUBE	213
Hình 3. 76. Tạo nameset Top_10_Brand.....	214
Hình 3. 77. Kết quả thực thi câu truy vấn 6 bằng CUBE	214
Hình 3. 78. Thao tác trên thuộc tính Market Trend và kết quả thực thi câu truy vấn 7 dùng CUBE	215
Hình 3. 79. Kết quả thực thi câu truy vấn 8 bằng CUBE	216
Hình 3. 80. Kết quả thực thi câu truy vấn 9 bằng CUBE	217
Hình 3. 81. Kết quả thực thi câu truy vấn 10 dùng CUBE	218
Hình 3. 82. Thao tác câu truy vấn 1 bằng Excel.....	219
Hình 3. 83. Kết quả thực thi câu truy vấn 1 bằng Excel.....	219
Hình 3. 84. Biểu đồ kết quả câu truy vấn 1 bằng Excel.....	219
Hình 3. 85. Thao tác câu truy vấn 2 bằng Excel.....	220

Hình 3. 86. Kết quả thực thi câu truy vấn 2 bằng Excel	220
Hình 3. 87. Biểu đồ kết quả câu truy vấn 2 bằng Excel.....	220
Hình 3. 88. Thao tác câu truy vấn 3 bằng Excel.....	221
Hình 3. 89. Kết quả thực thi câu truy vấn 3 bằng Excel	221
Hình 3. 90. Biểu đồ kết quả câu truy vấn 3 bằng Excel.....	222
Hình 3. 91. Thao tác câu truy vấn 4 bằng Excel.....	222
Hình 3. 92. Lọc thuộc tính Year	223
Hình 3. 93. Lọc thuộc tính Sustainability Rating.....	223
Hình 3. 94. Kết quả thực thi câu truy vấn 4 bằng Excel	224
Hình 3. 95. Biểu đồ kết quả câu truy vấn 4 bằng Excel.....	225
Hình 3. 96. Thao tác câu truy vấn 5 bằng Excel	225
Hình 3. 97. Kết quả câu truy vấn 5 bằng Excel	226
Hình 3. 98. Biểu đồ kết quả câu truy vấn 5 bằng Excel.....	226
Hình 3. 99. Thao tác câu truy vấn 6 bằng Excel.....	227
Hình 3. 110. Thao tác trên thuộc tính Brand Name	227
Hình 3. 111. Lấy ra top 10 thương hiệu.....	227
Hình 3. 112. Điều chỉnh giá trị để lấy ra 10 thương hiệu có giá trị trung bình cao nhất	227
Hình 3. 113. Kết quả thực thi câu truy vấn 6 bằng Excel	228
Hình 3. 114. Biểu đồ thực thi câu truy vấn 6 bằng Excel	228
Hình 3. 115. Thao tác câu truy vấn 7 bằng Excel	229
Hình 3. 116. Lọc thuộc tính Market Trend	229
Hình 3. 117. Kết quả thực thi câu truy vấn 7 bằng Excel	229

Hình 3. 118. Thao tác tạo Slicer cho thuộc tính Material Type.....	230
Hình 3. 119. Slider cho thuộc tính Material Type	230
Hình 3. 120. Slider cho thuộc tính Eco Friendly Manufacturing	230
Hình 3. 121. Biểu đồ kết quả thực thi câu truy vấn 7 bằng Excel	231
Hình 3. 122. Thao tác câu truy vấn 8 bằng Excel.....	231
Hình 3. 123. Kết quả thực thi câu truy vấn 8 bằng Excel.....	232
Hình 3. 124. Slicer Sustainability Rating.....	232
Hình 3. 125. Slicer Certifications	232
Hình 3. 126. Biểu đồ kết quả thực thi câu truy vấn 8 bằng Excel	233
Hình 3. 127. Thao tác câu truy vấn 9 bằng Excel.....	233
Hình 3. 128. Lọc thuộc tính Recycling Programs và Eco Friendly Manufacturing	233
Hình 3. 129. Kết quả thực thi câu truy vấn 9 bằng Excel	234
Hình 3. 130. Biểu đồ kết quả thực thi câu truy vấn 9 bằng Excel	234
Hình 3. 131. Thao tác câu truy vấn 10 bằng Excel	234
Hình 3. 132. Kết quả thực thi câu truy vấn 10 bằng Excel	235
Hình 3. 133. Biểu đồ kết quả câu truy vấn 10 bằng Excel.....	235
.....	
Hình 3. 135. Kết quả truy vấn câu 2 bằng MDX	236
Hình 3. 136. Kết quả truy vấn câu 3 bằng MDX	237
Hình 3. 137. Kết quả truy vấn câu 4 bằng MDX	238
Hình 3. 138. Kết quả truy vấn câu 5 bằng MDX	239
Hình 3. 139. Kết quả truy vấn câu 6 bằng MDX	239
Hình 3. 140. Kết quả truy vấn câu 7 bằng MDX	240

Hình 3. 141. Kết quả truy vấn câu 8 bằng MDX	241
Hình 3. 142. Kết quả truy vấn câu 9 bằng MDX	241
Hình 3. 143. Kết quả truy vấn câu 10 bằng MDX	242
Hình 3. 144. Khởi chạy phần mềm Power BI desktop và chọn Get Data.	242
Hình 3. 145. Lựa chọn Get data và SQL Server Analysis Services database.....	243
Hình 3. 146. Điện tên Server và Database	243
Hình 3. 147. Lựa chọn database.....	244
Hình 3. 148. Màn hình BI sau khi tạo kết nối thành công.	244
Hình 3. 149. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	245
Hình 3. 150. Chọn những thông tin cần thiết trong hộp thoại Data.....	245
Hình 3. 151. Chính sửa hiển thị Rows, Values trong hộp thoại Visualizations	246
Hình 3. 152. Biểu đồ Matrix của Report.....	246
Hình 3. 153. Biểu đồ Line Chart của Report1	247
Hình 3. 154. Biểu đồ Donut Chart của Report 1 thứ 1	247
Hình 3. 155. Biểu đồ Donut Chart của Report 1 thứ 2	248
Hình 3. 156. Biểu đồ Donut Chart của Report 1 thứ 3	248
Hình 3. 157. Kết quả Report 1	249
Hình 3. 158. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	249
Hình 3. 159. Chọn những thông tin cần thiết trong hộp thoại Data.....	250
Hình 3. 160. Chính sửa hiển thị Rows, Values trong hộp thoại Visualizations	250
Hình 3. 161. Biểu đồ Matrix của Report 2.....	251
Hình 3. 162. Biểu đồ Stacked Coloum Chart của Report 2	251
Hình 3. 163. Biểu đồ Pie Chart của Report 2.....	252

Hình 3. 164. Kết quả Report 2	252
Hình 3. 165. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	253
Hình 3. 166. Chọn những thông tin cần thiết trong hộp thoại Data.....	253
Hình 3. 167. Chính sửa hiển thị X-ais, Y- ais trong hộp thoại Visualizations	254
Hình 3. 168. Biểu đồ Line and Stacked Column Chart của Report 3	254
Hình 3. 169. Biểu đồ Matrix của Report 2.....	255
Hình 3. 170. Biểu đồ Pie Chart của Report 3.....	255
Hình 3. 171. Slicer điều chỉnh thuộc tính Material_Type của Report 3	255
Hình 3. 172. Kết quả Report 3	256
Hình 3. 173. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	256
Hình 3. 174. Chọn những thông tin cần thiết trong hộp thoại Data.....	257
Hình 3. 175. Chính sửa hiển thị X-ais, Y- ais trong hộp thoại Visualizations	257
Hình 3. 176. Điều kiện lọc trong Filters	258
Hình 3. 177. Biểu đồ Stacked Column Chart của Report 4	259
Hình 3. 178. Biểu đồ Matrix của Report 4.....	260
Hình 3. 179. Biểu đồ Pie Chart của Report 4.....	261
Hình 3. 180. Kết quả Report 4	261
Hình 3. 181. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	261
Hình 3. 182. Chọn những thông tin cần thiết trong hộp thoại Data.....	262
Hình 3. 183. Chính sửa hiển thị Columns trong hộp thoại Visualizations	262
Hình 3. 184. Điều kiện lọc trong Filters	263
Hình 3. 185. Biểu đồ Table của Report 5	263
Hình 3. 186. Biểu đồ Matrix của Report 5.....	263

Hình 3. 187. Biểu đồ Clustered Column Chart của Report 5	264
Hình 3. 188. Biểu đồ Chart của Report 5.....	264
Hình 3. 189. Kết quả Report 5	265
Hình 3. 190. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	265
Hình 3. 191. Chọn những thông tin cần thiết trong hộp thoại Data.....	266
Hình 3. 192. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations.....	266
Hình 3. 193. Điều kiện lọc trong Filters	266
Hình 3. 194. Biểu đồ Matrix của Report 6.....	267
Hình 3. 195. Biểu đồ Clustered Column Chart của Report 6	267
Hình 3. 196. Kết quả Report 6	268
Hình 3. 198. Chọn những thông tin cần thiết trong hộp thoại Data.....	269
Hình 3. 199. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations.....	269
Hình 3. 200. Điều kiện lọc trong Filters	270
Hình 3. 201. Biểu đồ Matrix của Report 7.....	270
Hình 3. 202. Biểu đồ Stacked Column Chart của Report 7	270
Hình 3. 204. Biểu đồ Pie Chart của Report 7.....	271
Hình 3. 204. Kết quả Report 7	271
Hình 3. 205. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations	272
Hình 3. 206. Chọn những thông tin cần thiết trong hộp thoại Data.....	273
Hình 3. 207. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations.....	273
Hình 3. 208. Biểu đồ Matrix của Report 8.....	274

Hình 3. 209. Biểu đồ Clustered Column Chart của Report 8	275
Hình 3. 210. Slicer điều chỉnh thuộc tính của Report 8.....	275
Hình 3. 211. Kết quả Report 8	276
Hình 3. 213. Chọn những thông tin cần thiết trong hộp thoại Data.....	276
Hình 3. 214. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations.....	277
Hình 3. 215. Biểu đồ Matrix của Report 9.....	277
Hình 3. 216. Biểu đồ Clustered Column Chart của Report 9	277
Hình 3. 217. Slicer điều chỉnh thuộc tính của Report 9.....	278
Hình 3. 218. Kết quả Report 9	278
Hình 3. 220. Chọn những thông tin cần thiết trong hộp thoại Data.....	279
Hình 3. 221. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations.....	279
Hình 3. 222. Biểu đồ Matrix của Report 10.....	280
Hình 3. 223. Biểu đồ Clustered Column Chart của Report 10	280
Hình 3. 224. Slicer điều chỉnh thuộc tính của Report 10.....	280
Hình 3. 225. Kết quả Report 10	281
Hình 4. 01. Import thư viện và đọc file csv chứa dữ liệu gốc.	282
Hình 4. 02. Dữ liệu ban đầu	282
Hình 4. 03. Số giá trị NULL trong tập dữ liệu ban đầu.	282
Hình 4. 04. Số giá trị NULL sau khi xử lý.	283
Hình 4. 05. Dữ liệu sau khi xử lý giá trị NULL.....	283
Hình 4. 06. Xóa bỏ cột Brand_ID, Brand_Name và Year.	283

Hình 4. 07. Mã hóa giá trị của các thuộc tính.....	284
Hình 4. 08. Xóa các thuộc tính trước khi mã hóa.....	284
Hình 4. 09. Dữ liệu sau khi được mã hóa.....	285
Hình 4. 10. Chuẩn hóa giá trị của các thuộc tính.....	285
Hình 4. 11. Dữ liệu sau khi được chuẩn hóa.....	285
Hình 4. 12. Lưu dữ liệu đã được xử lý.....	285
Hình 4. 13. Đọc dữ liệu đã lưu trước đó.....	285
Hình 4. 14. Ma trận tương quan giữa các thuộc tính	286
Hình 4. 15. Tách dữ liệu thành hai phần.....	286
Hình 4. 16. Tính độ quan trọng và tìm ra 5 thuộc tính quan trọng nhất.....	287
Hình 4. 17. Lấy dữ liệu top 5 thuộc tính quan trọng và chia tập dữ liệu.....	287
Hình 4. 18. Import thêm thư viện.....	287
Hình 4. 20. Huấn luyện mô hình Decision Tree	288
Hình 4. 21. Độ chính xác của mô hình Decision Tree.....	288
Hình 4. 22. Ma trận nhầm lẫn của mô hình Decision Tree	289
Hình 4. 23. Cây CART.....	290
Hình 4. 24. Tập luật từ cây CART	291
Hình 4. 25. Import thư viện và huấn luyện mô hình Random forest	292
Hình 4. 26. Độ chính xác của mô hình Random forest.....	292
Hình 4. 27. Ma trận nhầm lẫn của mô hình Random forest.....	293
Hình 4. 28. Cây quyết định đầu tiên của mô hình Random forest.....	293
Hình 4. 29. Tập luật từ cây quyết định đầu tiên của Random Forest	294

1. GIỚI THIỆU

1.1 Lý do chọn đề tài

Ngành công nghiệp thời trang là một trong những ngành gây ô nhiễm nhất thế giới, với lượng lớn chất thải và khí thải carbon. Theo trang điện tử Waste Managed, trên toàn cầu, ngành công nghiệp thời trang thải ra khoảng 92 triệu tấn chất thải dệt may mỗi năm, góp phần đáng kể vào tình trạng ô nhiễm môi trường khi ngành này chịu trách nhiệm cho 10% lượng khí thải carbon toàn cầu. Dự đoán đến năm 2030, con số này sẽ tăng lên 102 triệu tấn. Không chỉ dừng lại ở đó, ngành thời trang còn tiêu tốn lượng lớn tài nguyên thiên nhiên. Để sản xuất một chiếc áo thun thông thường, cần khoảng 2.700 lít nước – tương đương lượng nước một người uống trong 2,5 năm. Việc trồng bông, nguyên liệu chính trong ngành dệt may, đòi hỏi lượng lớn thuốc trừ sâu và phân bón, gây ô nhiễm nguồn nước và đất.

Bên cạnh đó, xu hướng thời trang nhanh (fast fashion) đang thúc đẩy nhu cầu tiêu dùng không bền vững, với hàng triệu sản phẩm bị bỏ đi chỉ sau vài lần sử dụng. Điều này tạo áp lực lớn lên hệ thống xử lý rác thải, khi phần lớn các loại vải tổng hợp mất hàng trăm năm để phân hủy. Do đó, việc chuyển đổi sang các mô hình sản xuất và tiêu dùng bền vững là một yêu cầu cấp bách để giảm thiểu tác động tiêu cực của ngành công nghiệp thời trang lên môi trường.

Hiểu được điều đó, nhóm chúng em quyết định chọn đề tài nghiên cứu này nhằm tìm hiểu sâu hơn về những tác động tiêu cực của ngành công nghiệp thời trang đến môi trường. Đồng thời, chúng em mong muốn đề xuất các giải pháp thực tế và khả thi để giảm thiểu lượng chất thải dệt may, hạn chế khí thải carbon, và khuyến khích việc sử dụng các nguyên liệu bền vững. Mục tiêu của nghiên cứu là nâng cao nhận thức cộng đồng về vấn đề này, đồng thời góp phần thúc đẩy xu hướng tiêu dùng có trách nhiệm và bền vững trong xã hội hiện đại.

1.2 Giới thiệu về dataset

1.2.1 Tổng quan dataset

Tên bộ dữ liệu: **Sustainable Fashion: Eco-Friendly Trends**.

Bộ dữ liệu này cung cấp cái nhìn sâu sắc độc đáo về xu hướng thời trang bền vững đang nổi lên trên toàn cầu. Khi thế giới ngày càng ý thức hơn về tác động của môi trường, bộ dữ liệu này làm nổi bật các lĩnh vực chính như vật liệu, quy trình sản xuất và các thương hiệu phô biến thúc đẩy thời trang thân thiện với môi trường. Không những thế dữ liệu còn cung cấp bối cảnh có giá trị để hiểu tác động của tính bền vững đối với các lựa chọn thời trang vào năm 2024.

1.2.2 Mô tả số dòng, cột, thời gian thu thập dữ liệu

Bộ dữ liệu gốc gồm 5000 dòng, 15 cột, được thu thập từ năm 2010 đến năm 2024. Bộ dữ liệu sau khi làm sạch gồm 5000 dòng, 15 cột.

Nguồn tải dataset: <https://www.kaggle.com/datasets/waqi786/sustainable-fashion-eco-friendly-trends>

The screenshot shows the Kaggle website interface. On the left, there's a sidebar with navigation links: Create, Home, Competitions, Datasets (which is highlighted), Models, Code, Discussions, Learn, and More. The main content area displays a dataset titled "Sustainable Fashion: Eco-Friendly Trends" by Waqar Ali, updated 3 months ago. The title has a green leaf icon. Below the title is a subtitle: "Tracking Global Shifts Towards Sustainable Fashion in 2024". There are tabs for Data Card, Code (13), Discussion (0), and Suggestions (0). The Data Card tab is selected. To the right of the title is a thumbnail image of two people wearing large hats and sunglasses. Below the title, there's a brief description: "This dataset offers a unique insight into the rising trend of sustainable fashion across the globe 🌎. With the world becoming increasingly conscious of environmental impact, this dataset highlights key areas such as materials, production processes, and popular brands promoting eco-friendly fashion. Whether you're researching trends, consumer behavior, or industry shifts, this data provides valuable context for understanding the impact of sustainability on fashion choices in 2024 🌱." To the right of the description are sections for Usability (10.00), License (Apache 2.0), and Expected update frequency (Never). At the top right of the main content area are buttons for Sign In, Register, New Notebook, Download, and more options.

Hình 1. 01. Dataset Sustainable Fashion: Eco-Friendly từ Kaggle.

1.2.3 Làm sạch dữ liệu

Nhóm tiến hành sử dụng Python để làm sạch dữ liệu theo các bước sau:

Bước 1: Nhập thư viện và đọc file csv.

IS217 – Kho dữ liệu và OLAP

The screenshot shows a Jupyter Notebook cell with Python code and its output. The code imports necessary libraries and reads a CSV file named 'sustainable_fashion_trends_2024.csv'. It then displays the first five rows of the DataFrame using the head() method.

	0	1	2	3	4
Brand_ID	BRAND-0001	BRAND-0002	BRAND-0003	BRAND-0004	BRAND-0005
Brand_Name	Brand_1	Brand_2	Brand_3	Brand_4	Brand_5
Country	Australia	Japan	USA	Italy	USA
Year	2018	2015	2024	2023	2016
Sustainability_Rating	D	D	A	D	D
Material_Type	Tencel	Vegan Leather	Vegan Leather	Bamboo Fabric	Bamboo Fabric
Eco_Friendly_Manufacturing	No	Yes	No	No	Yes
Carbon_Footprint_MT	1.75	124.39	336.66	152.04	415.63
Water_Usage_Liters	4511152.79	1951566.31	467454.52	899576.9	1809219.9
Waste_Production_KG	97844.11	37267.75	38385.92	32665.45	37295.47
Recycling_Programs	No	No	No	No	Yes
Product_Lines	2	15	2	13	19
Average_Price_USD	38.33	250.07	146.16	165.52	211.63
Market_Trend	Growing	Growing	Growing	Stable	Stable
Certifications	GOTS	GOTS	B Corp	OEKO-TEX	Fair Trade

Hình 1. 02. Làm sạch dữ liệu (1).

Bước 2: Xem kiểu dữ liệu của từng cột thuộc tính.

The screenshot shows the output of the data.info() method. It provides information about the DataFrame, including the number of entries (5000), column names, data types, and memory usage.

#	Column	Non-Null Count	Dtype
0	Brand_ID	5000	object
1	Brand_Name	5000	object
2	Country	5000	object
3	Year	5000	int64
4	Sustainability_Rating	5000	object
5	Material_Type	5000	object
6	Eco_Friendly_Manufacturing	5000	object
7	Carbon_Footprint_MT	5000	float64
8	Water_Usage_Liters	5000	float64
9	Waste_Production_KG	5000	float64
10	Recycling_Programs	5000	object
11	Product_Lines	5000	int64
12	Average_Price_USD	5000	float64
13	Market_Trend	5000	object
14	Certifications	3996	object

Hình 1. 03. Làm sạch dữ liệu (2).

Bước 3: Kiểm tra những giá trị rỗng.

	[] data.isnull().sum()
Brand_ID	0
Brand_Name	0
Country	0
Year	0
Sustainability_Rating	0
Material_Type	0
Eco_Friendly_Manufacturing	0
Carbon_Footprint_MT	0
Water_Usage_Liters	0
Waste_Production_KG	0
Recycling_Programs	0
Product_Lines	0
Average_Price_USD	0
Market_Trend	0
Certifications	1004

Hình 1. 04. Làm sạch dữ liệu (3).

Bước 4: Xử lý các giá trị Null ở cột Certifications bằng cách đổi giá trị của chúng sang None. Sau đó kiểm tra lại một lần nữa và không còn giá trị Null nào trong tập dữ liệu.

```
[ ] data['Certifications'].fillna('None', inplace=True)
print(data.isnull().sum())

Brand_ID          0
Brand_Name        0
Country           0
Year              0
Sustainability_Rating 0
Material_Type     0
Eco_Friendly_Manufacturing 0
Carbon_Footprint_MT 0
Water_Usage_Liters 0
Waste_Production_KG 0
Recycling_Programs 0
Product_Lines     0
Average_Price_USD 0
Market_Trend       0
Certifications    0
dtype: int64
```

Hình 1. 05. Làm sạch dữ liệu (3).

Bước 5: Kiểm tra số bộ dữ liệu trùng nhau. Kết quả là không nên chứng tỏ không có bộ dữ liệu trùng.

```
[ ] data.duplicated().sum()

0
```

Hình 1. 06. Làm sạch dữ liệu (3).

Bước 6: Xóa các ký tự không import được vào SSIS. Thực hiện loại bỏ các ký tự này tại các cột: Brand_ID, Brand_Name, Certifications.

```
[ ] def clean_text(text):
    if pd.isna(text):
        return text
    return re.sub(r'[^\u00e1-\u00f9\u00c1-\u00d9\u00e3-\u00f3\u00c3-\u00d3]+', '', text)

data[['Brand_ID', 'Brand_Name', 'Certifications']] = data[['Brand_ID', 'Brand_Name', 'Certifications']].applymap(clean_text)
```

Hình 1. 07. Làm sạch dữ liệu (3).

Bước 7: Kiểm tra lại dữ liệu đã được xử lý.

	Brand_ID	Brand_Name	Country	Year	Sustainability_Rating	Material_Type	Eco_Friendly_Manufacturing	Carbon_Footprint_Mt	Water_Usage_Liters	Waste_Production_Kg	Recycling_Programs	Product_Lines	Average_Price_USD	Market_Trend	Certifications
0	BRAND0001	Brand1	Australia	2018	D	Tencel	No	1.76	4511152.79	97844.11	No	2	38.33	Growing	GOTS
1	BRAND0002	Brand2	Japan	2015	D	Vegan Leather	Yes	124.39	1951560.31	37267.75	No	15	250.07	Growing	GOTS
2	BRAND0003	Brand3	USA	2024	A	Vegan Leather	No	336.66	467454.52	58385.92	No	2	146.16	Growing	B Corp
3	BRAND0004	Brand4	Italy	2023	D	Bamboo Fabric	No	162.04	899576.90	32665.45	No	13	165.52	Stable	OEKOTEX
4	BRAND0005	Brand5	USA	2016	D	Bamboo Fabric	Yes	415.83	1809219.90	37295.47	Yes	19	211.83	Stable	Fair Trade
...
4995	BRAND4996	Brand4996	Brazil	2010	A	Organic Cotton	Yes	423.73	251758.61	50532.28	No	1	490.95	Stable	None
4996	BRAND4997	Brand4997	Germany	2012	C	Bamboo Fabric	Yes	68.24	2342347.78	84214.57	Yes	13	53.92	Stable	Fair Trade
4997	BRAND4998	Brand4998	Japan	2023	D	Vegan Leather	Yes	321.12	317800.87	81407.21	Yes	12	431.32	Stable	None
4998	BRAND4999	Brand4999	India	2016	A	Tencel	No	124.83	927539.68	34606.22	Yes	20	322.95	Declining	Fair Trade
4999	BRAND5000	Brand5000	Italy	2016	B	Vegan Leather	No	192.11	3402562.70	81958.82	Yes	11	205.62	Growing	None

Hình 1. 08. Làm sạch dữ liệu (8).

Bước 8: Sau khi hoàn thành làm sạch dữ liệu, lưu vào file .csv mới.

```
[ ] data.to_csv('/content/drive/MyDrive/Nam3/OLAP/Do_An_OlAP/dataset_cleaned.csv', index=False)
```

Hình 1. 09. Làm sạch dữ liệu (9).

1.2.4 Mô tả chi tiết các cột thuộc tính sử dụng

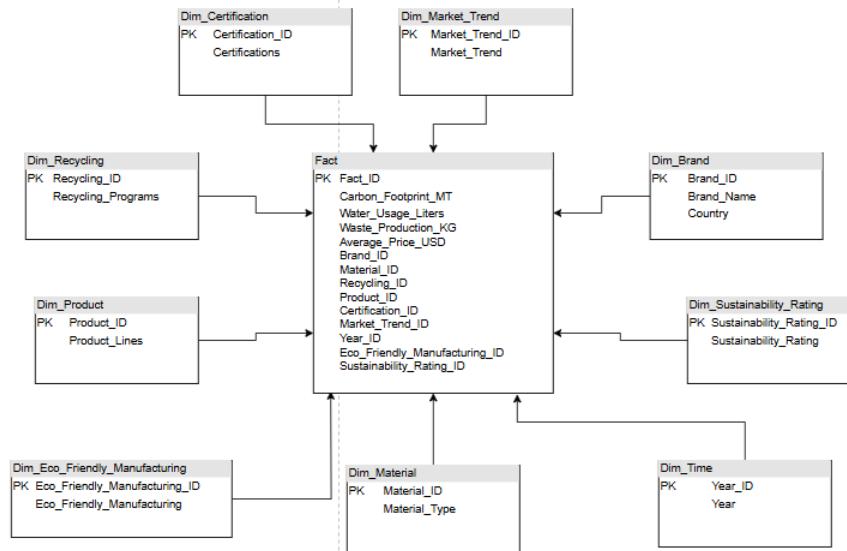
Bảng 1. 1. Bảng mô tả chi tiết các cột thuộc tính.

STT	Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
1	Brand_ID	varchar	Mã thương hiệu
2	Brand_Name	varchar	Tên thương hiệu
3	Country	varchar	Quốc gia nơi thương hiệu đặt trụ sở
4	Year	varchar	Năm dữ liệu được thu thập
5	Sustainability_Rating	varchar	Đánh giá mức độ bền vững của thương hiệu
6	Material_Type	varchar	Loại chất liệu mà thương hiệu sử dụng cho sản phẩm
7	Eco_Friendly_Manufact	varchar	Trạng thái thân thiện với môi trường

	uring		
8	Carbon_Footprint_MT	float	Lượng khí thải carbon tính bằng tấn (MT: Metric Tonnes) mà thương hiệu tạo ra.
9	Water_Usage_Liters	float	Lượng nước sử dụng tính bằng lít trong các quy trình sản xuất.
10	Waste_Production_KG	float	Lượng rác thải sản xuất tính bằng kilogam.
11	Recycling_Program	varchar	Chương trình tái chế của thương hiệu
12	Product_Lines	varchar	Dòng sản phẩm mà thương hiệu cung cấp
13	Average_Price_USD	float	Giá trung bình của sản phẩm thuộc thương hiệu tính bằng USD
14	Market_Trend	varchar	Xu hướng thị trường liên quan đến thương hiệu
15	Certifications	varchar	Chứng nhận về môi trường hoặc bền vững mà thương hiệu đã đạt được

1.3 Thiết kế kho dữ liệu

1.3.1 Lược đồ kho dữ liệu (lược đồ hình sao)



Hình 1. 10. Lược đồ kho dữ liệu (hình sao).

1.3.2 Bảng Fact

Bảng 1. 2. Bảng mô tả Fact.

BẢNG FACT				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Fact_ID	int	Khóa chính	Mã khảo sát
2	Year_ID	int	Khóa ngoại	Mã năm thu thập dữ liệu
3	Material_ID	int	Khóa ngoại	Mã chất liệu
4	Recycling_ID	int	Khóa ngoại	Mã tái chế
5	Product_ID	int	Khóa ngoại	Mã sản phẩm

6	Market_Trend_ID	int	Khóa ngoại	Mã xu hướng thị trường
7	Certification_ID	int	Khóa ngoại	Mã chứng nhận về môi trường hoặc bền vững
8	Brand_ID	int	Khóa ngoại	Mã thương hiệu
9	Eco_Friendly_Manufacturing_ID	int	Khóa ngoại	Mã trạng thái thân thiện với môi trường
10	Sustainability_Rating_ID	int	Khóa ngoại	Mã mức độ bền vững
11	Carbon_Footprint_MT	float		Lượng khí thải carbon tính bằng tấn (MT: Metric Tonnes) mà thương hiệu tạo ra.
12	Water_Usage_Liters	float		Lượng nước sử dụng tính bằng lít trong các quy trình sản xuất.
13	Waste_Production_KG	float		Lượng rác thải sản xuất tính bằng kilogam.

14	Average_Price_US D	float		Giá trung bình của sản phẩm thuộc thương hiệu tính bằng USD
----	-----------------------	-------	--	---

1.3.3 Bảng Dim_Brand

Bảng 1. 3. Bảng mô tả Dim_Brand.

BẢNG DIM_BRAND				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Brand_ID	varchar	Khóa chính	Mã thương hiệu
2	Brand_Name	varchar		Tên thương hiệu
3	Country	varchar		Quốc gia nơi thương hiệu đặt trụ sở

1.3.4 Bảng Dim_Material

Bảng 1. 4. Bảng mô tả Dim_Material.

BẢNG DIM_MATERIAL				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Material_ID	int	Khóa chính	Mã chất liệu
2	Material_Type	varchar		Loại chất liệu mà thương hiệu sử dụng cho sản phẩm

1.3.5 Bảng Dim_Recycling

Bảng 1. 5. Bảng mô tả Dim_Recycling.

BẢNG DIM_RECYCLING				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Recycling_ID	int	Khóa chính	Mã tái chế
2	Recycling_Programs	varchar		Chương trình tái chế của thương hiệu

1.3.6 Bảng Dim_Product

Bảng 1. 6. Bảng mô tả Dim_Product.

BẢNG DIM_PRODUCT				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Product_ID	int	Khóa chính	Mã sản phẩm
2	Product_Lines	varchar		Dòng sản phẩm mà thương hiệu cung cấp

1.3.7 Bảng Dim_Certification*Bảng 1. 7. Bảng mô tả Dim_Certification.*

BẢNG DIM_CERTIFICATION				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Certification_ID	int	Khóa chính	Mã chứng nhận về môi trường hoặc bền vững

2	Certifications	varchar		Chứng nhận về môi trường hoặc bền vững mà thương hiệu đã đạt được
---	----------------	---------	--	---

1.3.8 Bảng Dim_Market_Trend

Bảng 1. 8. Bảng mô tả Dim_Market_Trend.

BẢNG DIM_MARKET_TREND				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Market_Trend_ID	int	Khóa chính	Mã xu hướng thị trường
2	Market_Trend	varchar		Xu hướng thị trường liên quan đến thương hiệu

1.3.9 Bảng Dim_Sustainability_Rating

Bảng 1. 9. Bảng mô tả Dim_Sustainability_Rating.

BẢNG DIM_SUSTAINABILITY_RATING				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Sustainability_Rating_ID	int	Khóa chính	Mã mức độ bền vững
2	Sustainability_Rating	varchar		Mức độ bền vững

1.3.10 Bảng Dim_Time

Bảng 1. 10. Bảng mô tả Dim_Time.

BẢNG DIM_TIME				
STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Year_ID	int	Khóa chính	Mã năm thu thập dữ liệu
2	Year	varchar		Năm dữ liệu được thu thập

1.3.2.10. Bảng Dim_Eco_Friendly_Manufacturing

Bảng 1. 10. Bảng mô tả Dim_Eco_Friendly_Manufacturing.

BẢNG DIM_ECO_FRIENDLY_MANUFACTURING

STT	Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả thuộc tính
1	Eco_Friendly_Manufacturing_ID	int	Khóa chính	Mã trạng thái thân thiện với môi trường
2	Eco_Friendly_Manufacturing	varchar		Trạng thái thân thiện với môi trường

2. QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)

2.1 Chuẩn bị công cụ

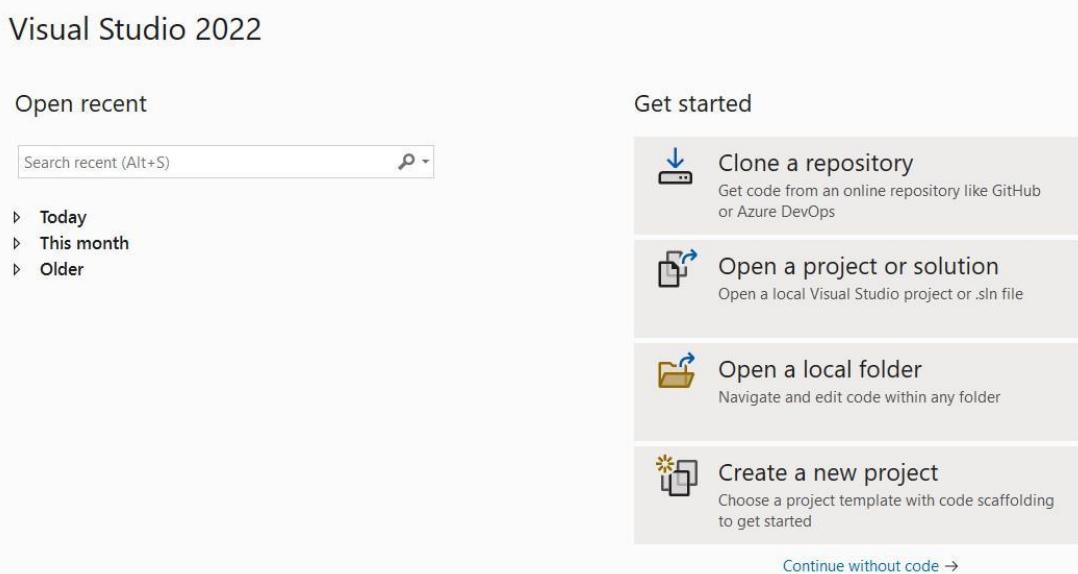
Để thực hiện được quá trình SSIS, cần chuẩn bị các công cụ sau:

- Visual Studio Community 2022.
- SQL Server Integration Services Projects.

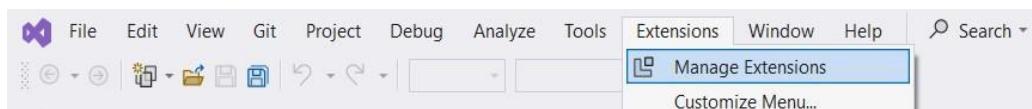
Do trước đó nhóm đã cài đặt sẵn Visual Studio Community 2022 nên hiện tại chỉ cần cài đặt thêm **SQL Server Integration Services Projects**.

Bước 1: Mở phần mềm Visual Studio Community 2022. Chọn *Continue without code*.

Hình 2. 01. Giao diện trang chủ Visual Studio 2022.

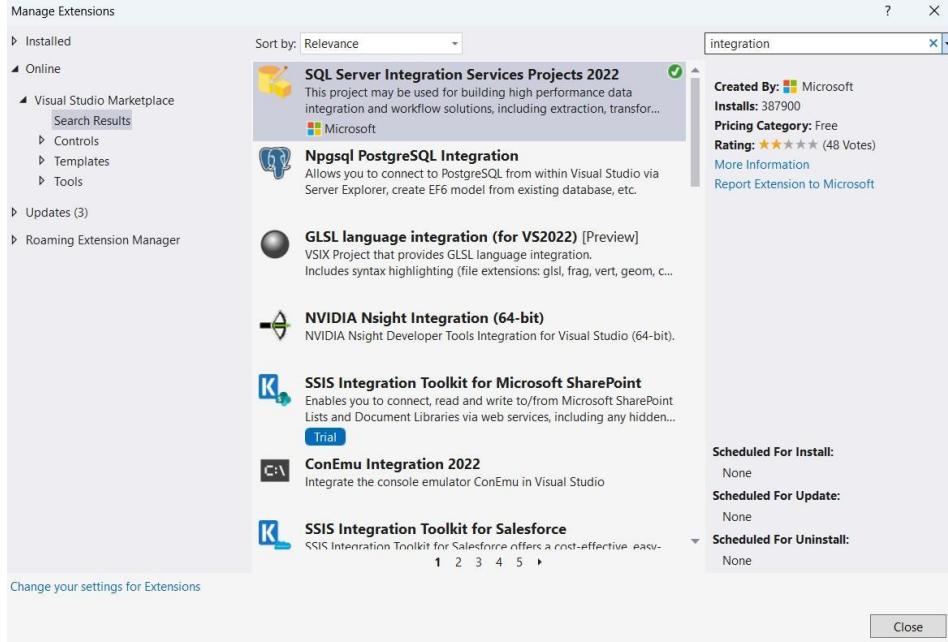


Bước 2: Ở tab Extensions, chọn *Manage Extensions*.



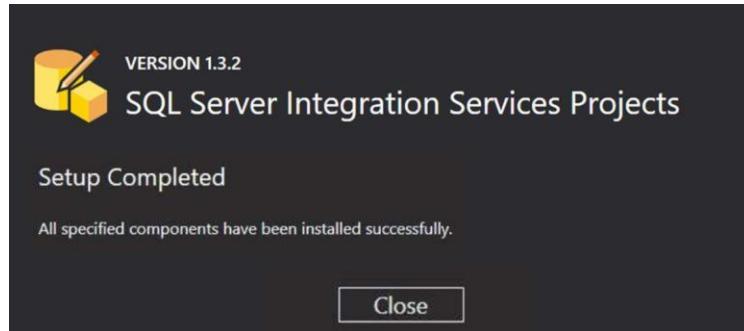
Hình 2. 02. Chọn *Manage Extensions* ở tab Extensions.

Bước 3: Tìm từ khóa “*integration*”, chọn extension có tên như hình dưới và tiến hành cài đặt.



Hình 2. 03. Tìm kiếm và cài đặt SQL Server Integration Services Project.

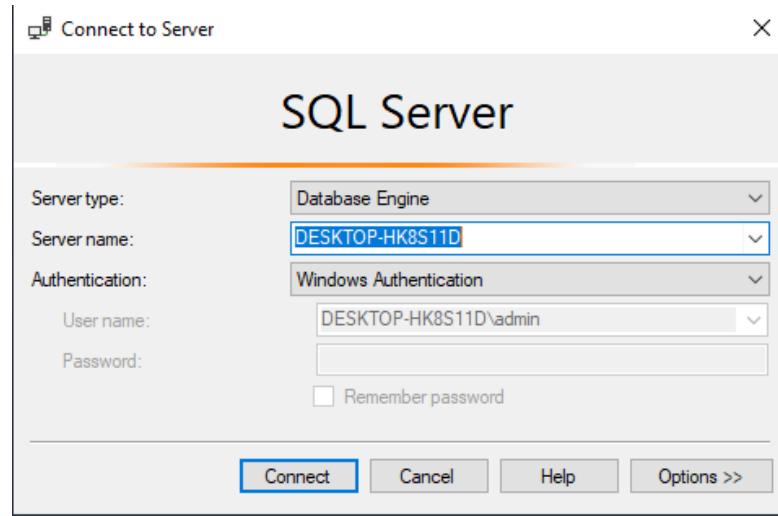
Bước 4: Hoàn thành cài đặt.



Hình 2. 04. Hoàn thành cài đặt SQL Server Integration Services Projects.

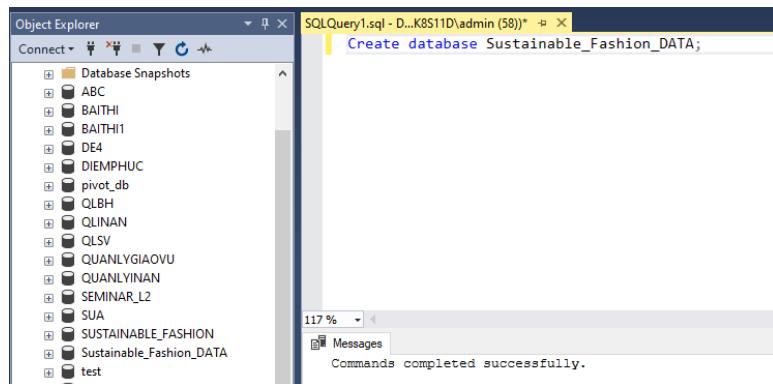
2.1.1 Chuẩn bị cơ sở dữ liệu

Bước 1: Mở phần mềm SQL Server, kết nối đến server bằng **Windows Authentication**.



Hình 2. 05. Kết nối với SQL Server bằng Windows Authentication.

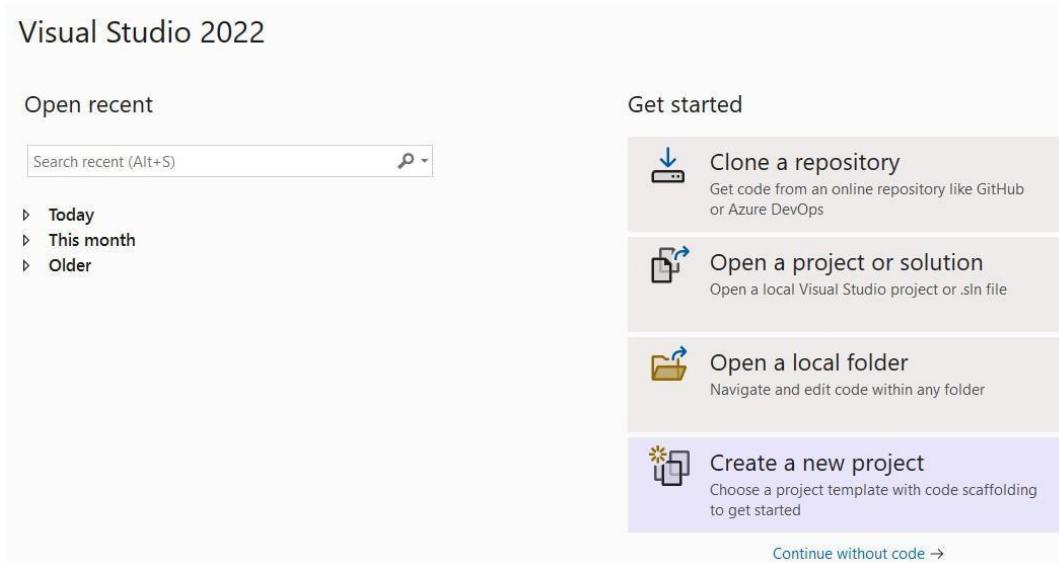
Bước 2: Tạo một database mới có tên Sustainable_Fashion_DATA để lưu các bảng Dim, Fact chứa dữ liệu của bộ dữ liệu.



Hình 2. 06. Tạo database.

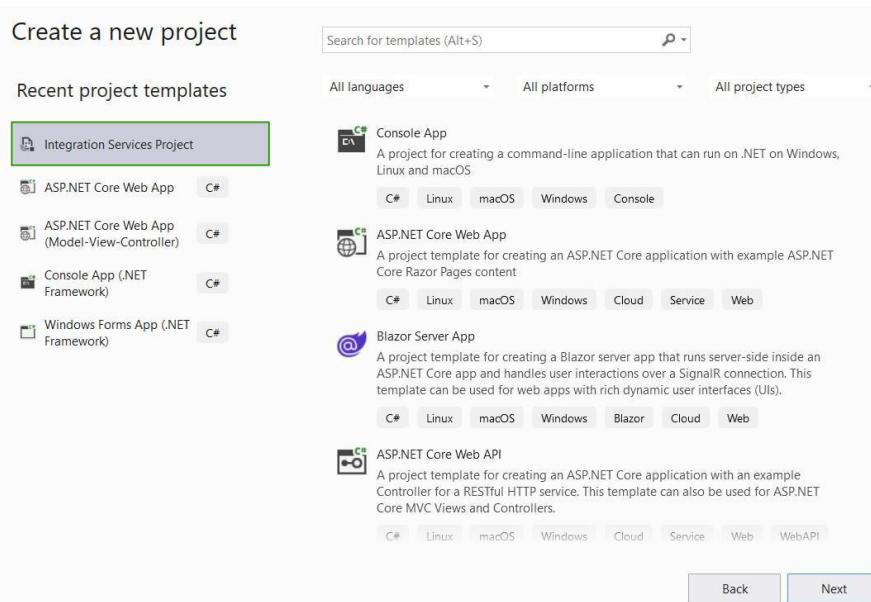
2.1.2 Tạo mới project SSIS

Bước 1: Mở phần mềm Visual Studio Community 2022. Chọn *Create a new project*.



Hình 2. 07. Chọn Create a new project tại trang chủ Visual Studio 2022.

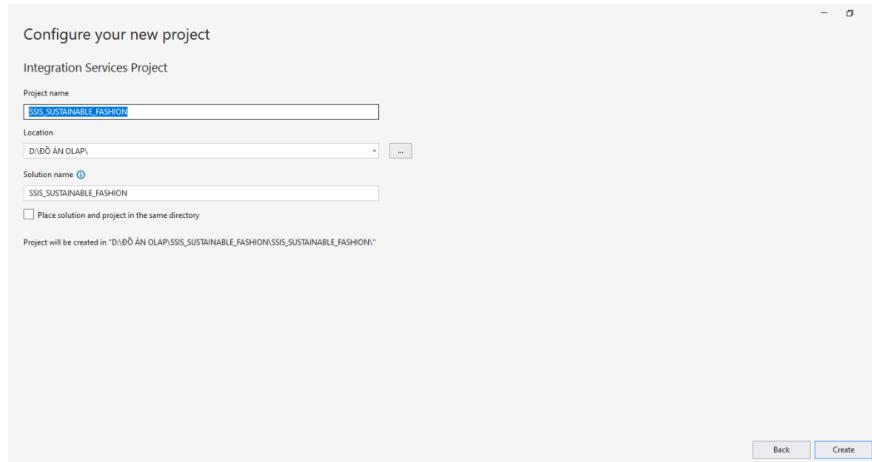
Bước 2: Chọn project template là **Integration Services Project**.



Hình 2. 08. Chọn Integration Services Project làm template.

Bước 3: Đặt tên cho project là **SSIS_SUSTAINABLE_FASHION** và chọn nơi lưu

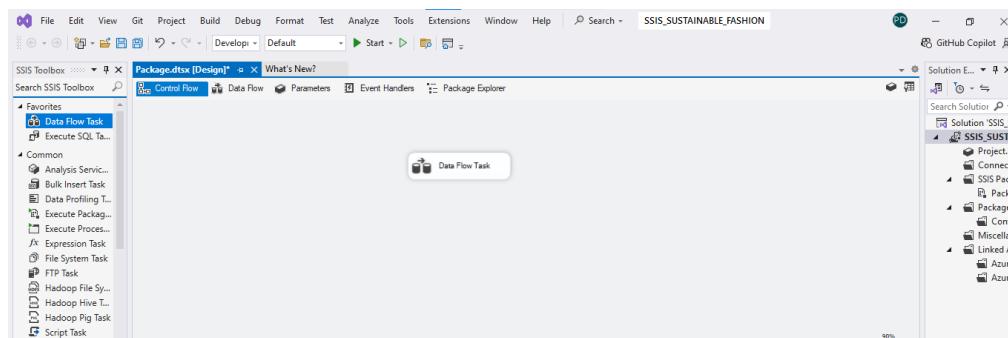
project.



Hình 2. 09. Đặt tên cho project và chọn nơi lưu.

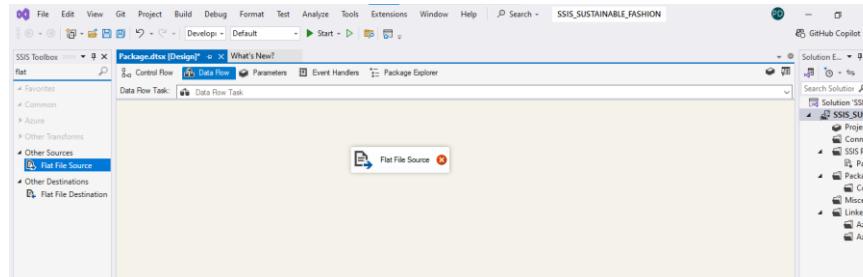
2.2 Tạo bảng Dim và Fact

Bước 1: Tạo một Data Flow Task để import dữ liệu gốc vào.



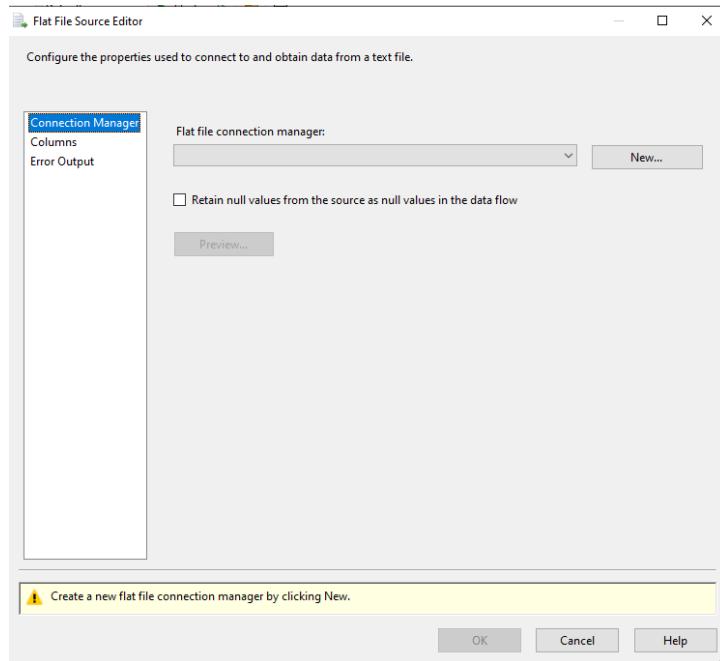
Hình 2. 10. Tạo Data Flow Task

Bước 2: Vào Data Flow đó, tạo một Flat File Source để lấy dữ liệu gốc từ file .csv.



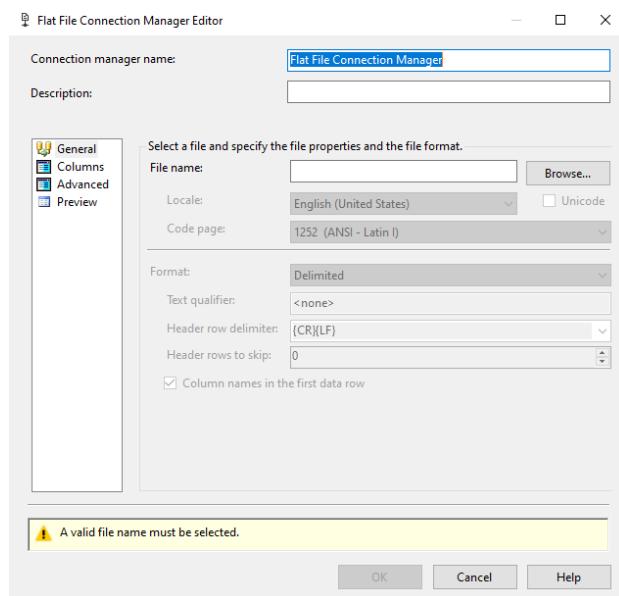
Hình 2. 11. Tạo Flat File Source lấy dữ liệu từ file .csv.

Bước 3: Nhấn đúp vào Flat File Source vừa tạo.



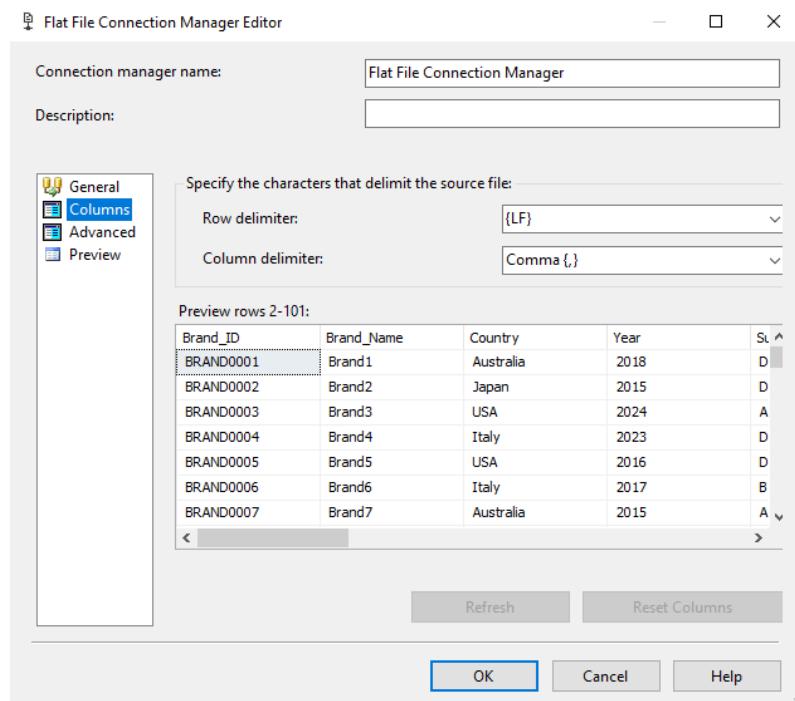
Hình 2. 12. Flat File Source Editor.

Bước 4: Chọn New để tạo Flat file connection manager mới.



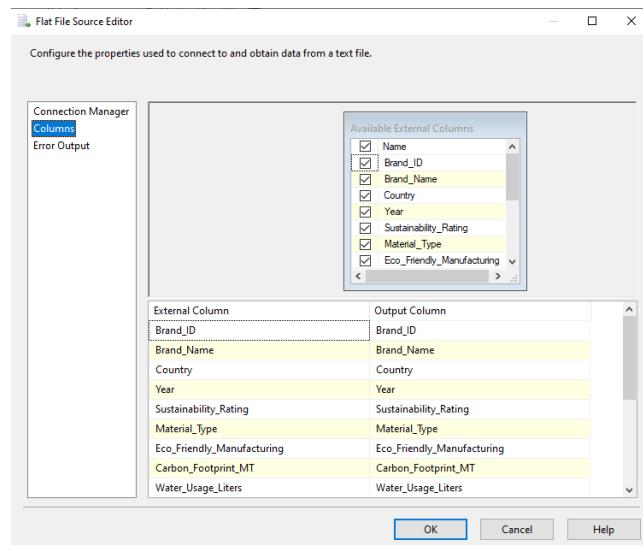
Hình 2. 13. Tạo Flat File Connection Manager mới.

Bước 5: Ở mục *File name*, chọn **Browse** để tải lên file dữ liệu gốc đã lưu trong máy. Sau đó có thể xem trước và kiểm tra các cột thuộc tính cũng như một số dòng dữ liệu sẽ được import. Nhấn chọn **OK** để tiếp tục.



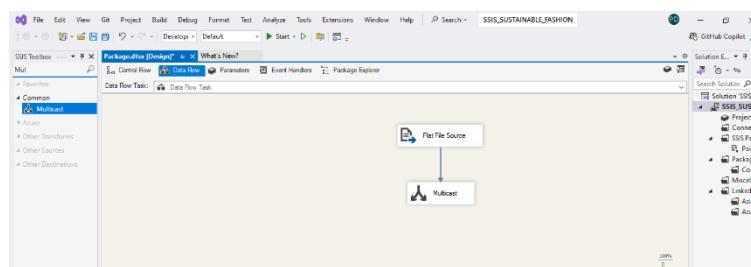
Hình 2. 14. Kiểm tra dữ liệu sẽ nạp vào.

Bước 6: Nhấn **OK** để hoàn thành quá trình nhập dữ liệu gốc vào Flat File Source.



Hình 2. 15. Các cột thuộc tính sẽ nạp vào Flat File Source.

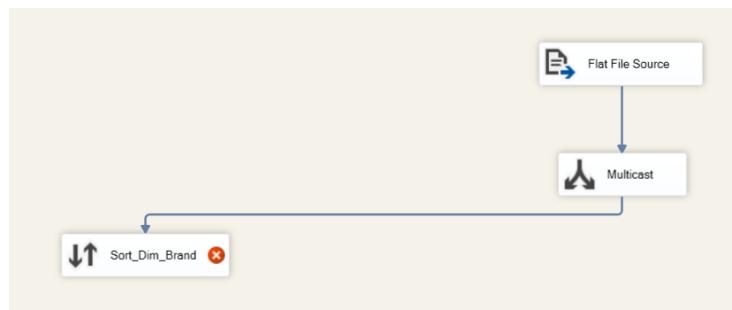
Bước 7 Tạo một “Multicast” để phân tán dữ liệu đến các bảng Dim.



Hình 2. 16. Tạo Multicast để phân tán dữ liệu.

2.2.1 Bảng Dim_Brand

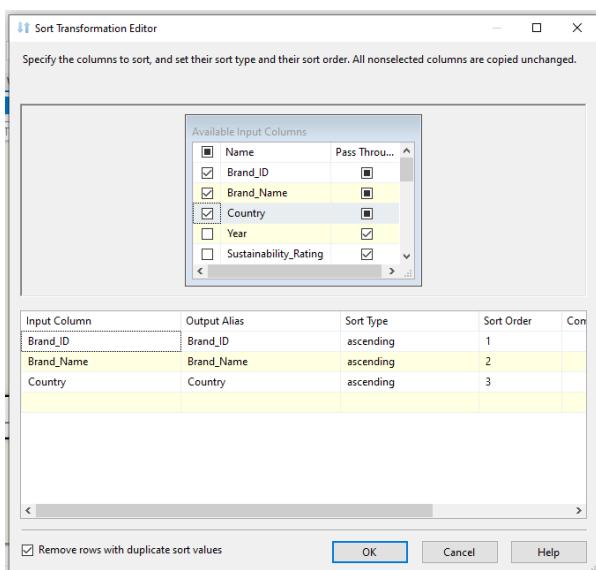
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Brand”. Kết nối từ “Multicast” đến “Sort_Dim_Brand”.



Hình 2. 17. Tạo Sort với tên "Sort_Dim_Brand".

Bước 2:

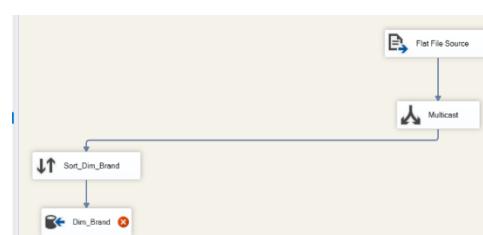
- Nhấn chuột phải vào “Sort_Dim_Brand”, chọn “Edit”.
- Chọn “Brand_ID”, “Brand_Name”, “Country” để làm dữ liệu cho “Sort_Dim_Brand”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn “OK” để hoàn tất.



Hình 2. 18. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

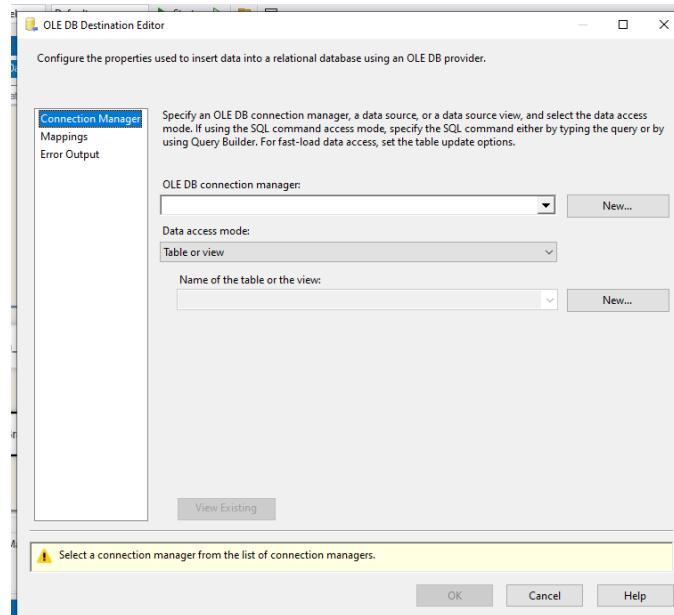
Bước 3: Tiến hành đổ dữ liệu bảng Dim_Brand vào Sustainable_Fashion_DATA.

Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Brand”.



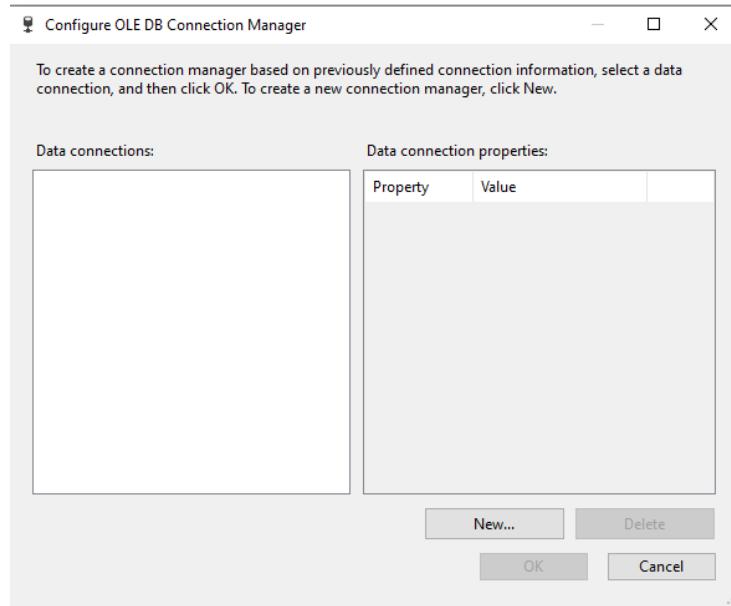
Hình 2. 19. Tạo OLE DB Destination đổi tên "Dim_Brand".

Bước 4: Nhấn chuột phải vào “Dim_Brand”, chọn “Edit”. Chọn “New”



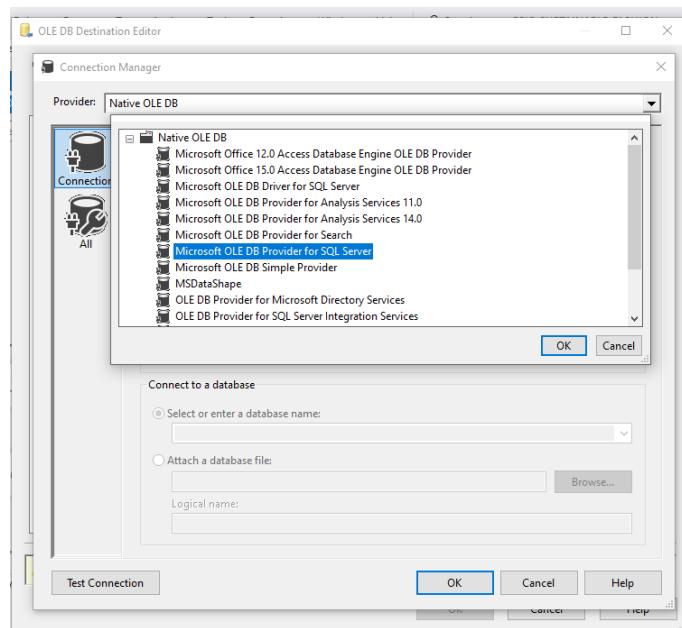
Hình 2. 20. OLE DB Destination Editor.

Chọn tiếp vào “New” để tạo kết nối OLE DB mới.

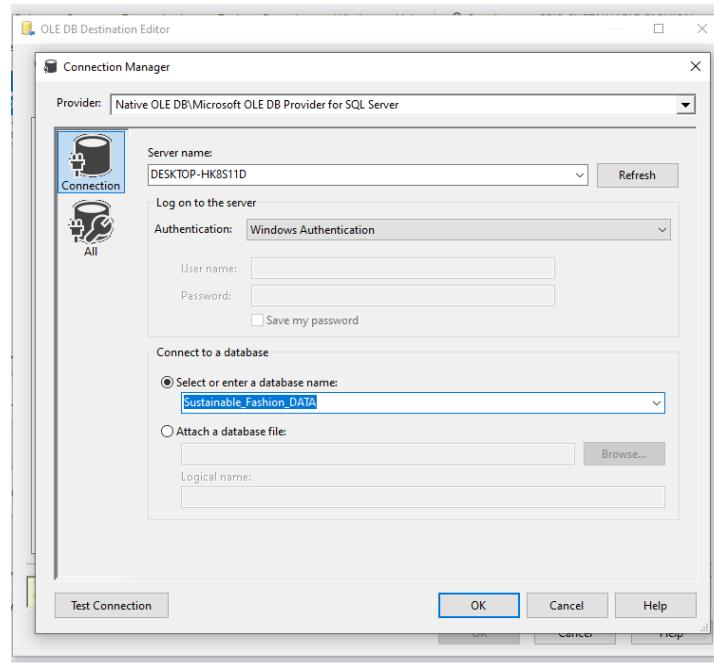


Hình 2. 21. Thiết lập Data connections.

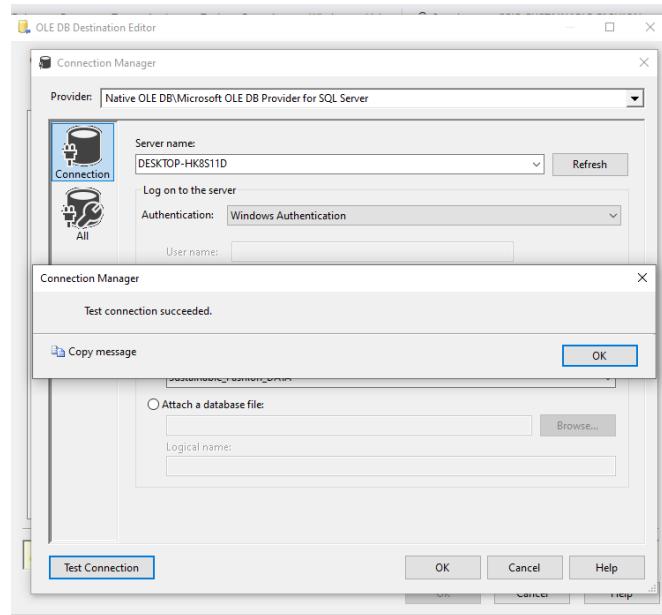
Bước 5: Vào mục Provider để chọn provider thích hợp, sau đó chọn Server name để kết nối. Sau khi đã kết nối thành công đến server, chọn database dùng để chứa bảng Dim đã được tạo trước đó. Nhấn **Test connection** để kiểm tra kết nối. Nếu kết nối thành công, nhấn **OK** để tiếp tục.



Hình 2. 22. Thiết lập kết nối mới(1).

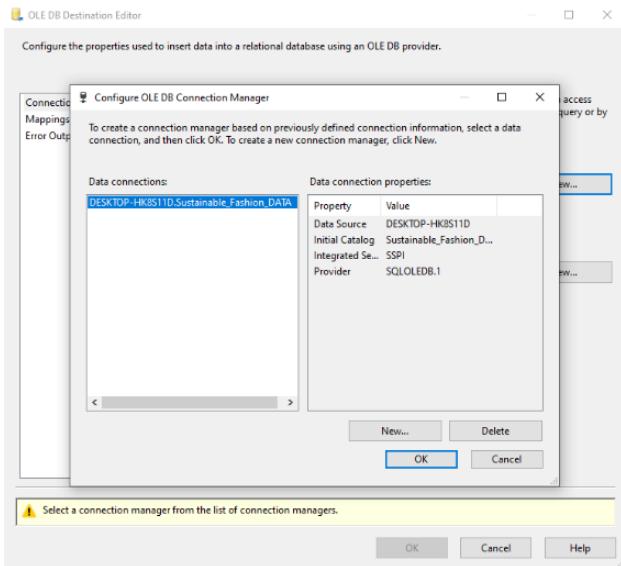


Hình 2. 23. Thiết lập kết nối mới(2).



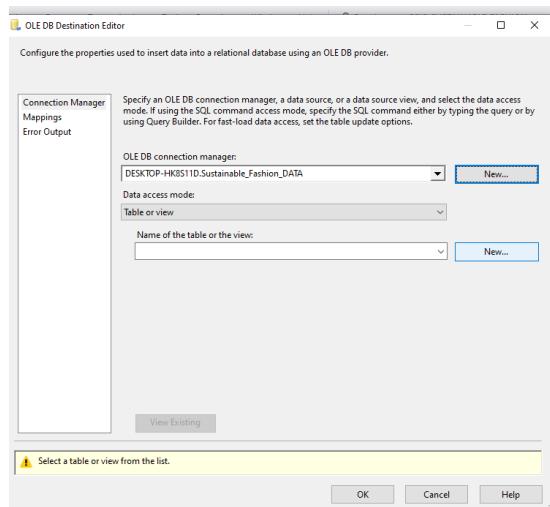
Hình 2. 24. Kiểm tra kết nối.

Bước 6: Chọn vào Connection vừa tạo. Nhấn “OK”



Hình 2. 25. Data connections hiện có.

Bước 7: Chọn “New” ở phần “Name of the table or the view” để tạo bảng Dim_Brand mới trong database.



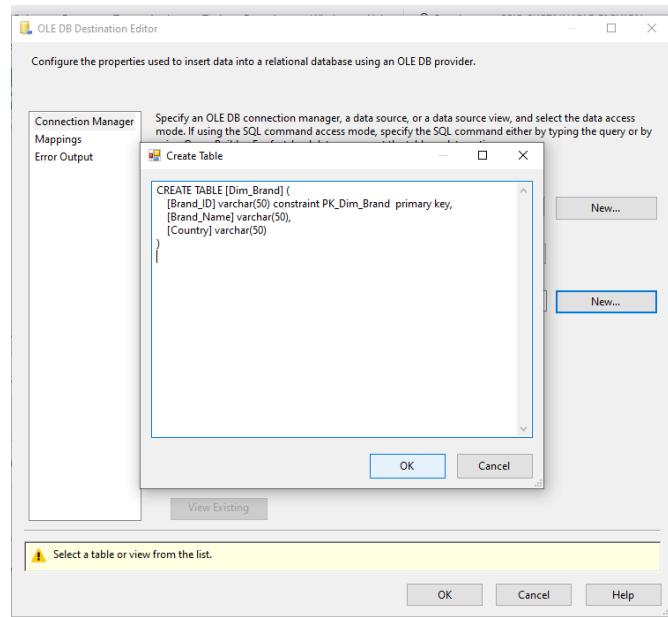
Hình 2. 26. Tạo bảng Dim_Brand.

Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

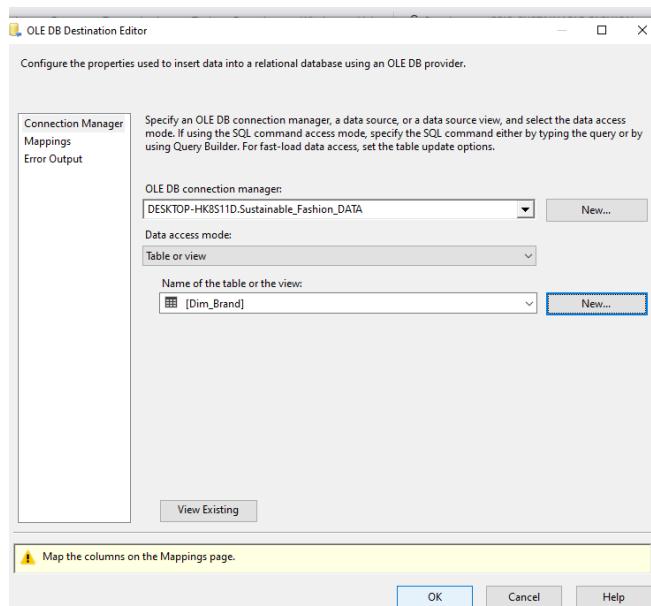
CREATE TABLE [Dim_Brand] (

```
[Brand_ID] varchar(50) constraint PK_Dim_Brand primary key,  
[Brand_Name] varchar(50),  
[Country] varchar(50) )
```

IS217 – Kho dữ liệu và OLAP



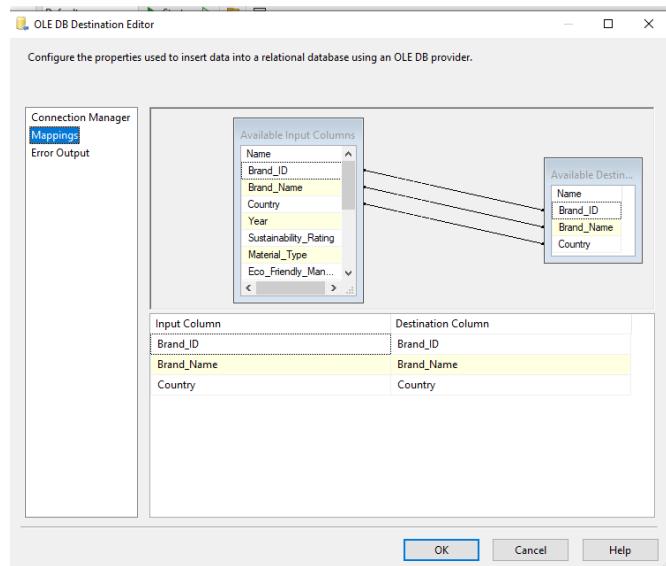
Hình 2. 27. Tạo bảng Dim_Brand(2).



Hình 2. 28. Tạo bảng Dim_Brand(3).

Bước 8: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.

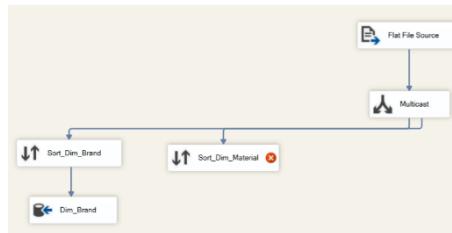
IS217 – Kho dữ liệu và OLAP



Hình 2. 29. Kiểm tra các cột thuộc tính

2.2.2 Bảng Dim_Material

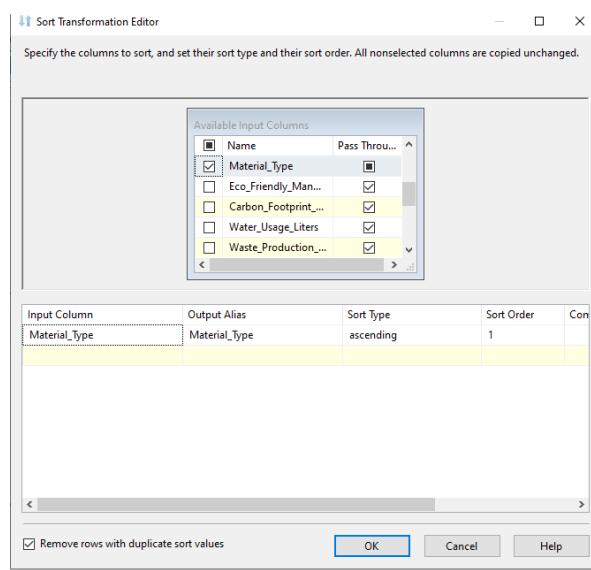
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Material”. Kết nối từ “Multicast” đến “Sort_Dim_Material”.



Hình 2. 30. Tạo Sort đổi tên "Sort_Dim_Material".

Bước 2:

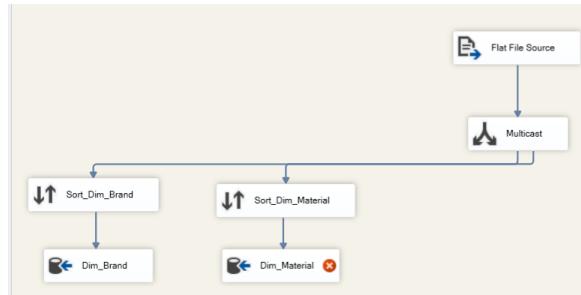
- Nhấn chuột phải vào “Sort_Dim_Material”, chọn “Edit”.
- Chọn “Material_Type” để làm dữ liệu cho “Sort_Dim_Material”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



Hình 2. 31. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

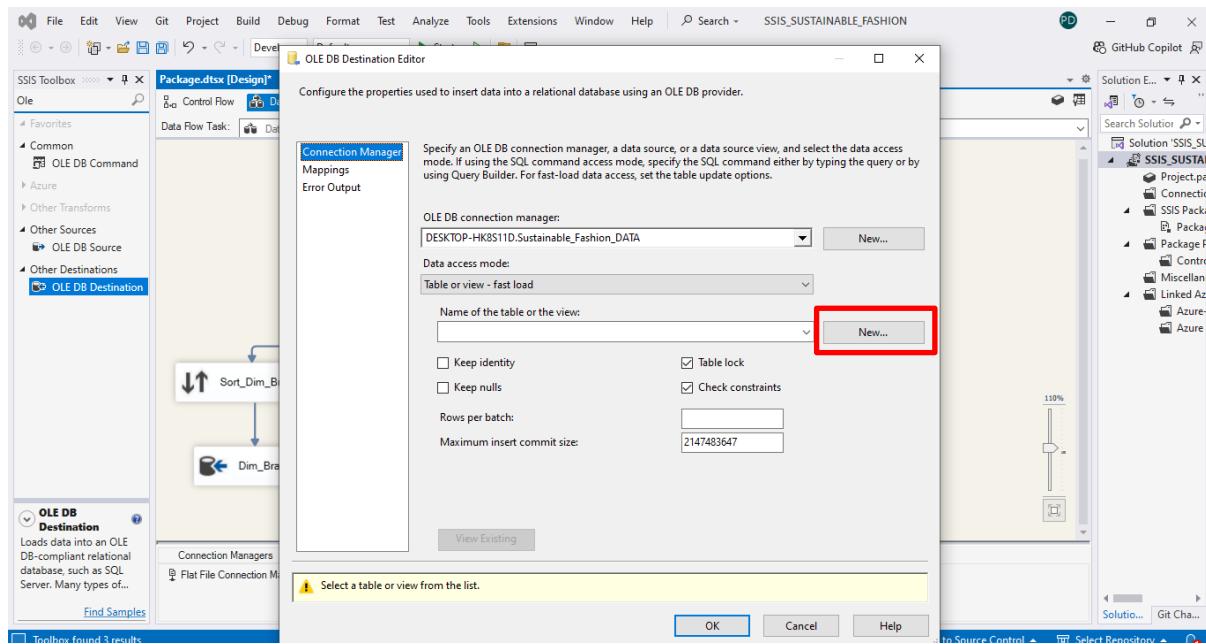
Bước 3: Tiến hành đổ dữ liệu bảng Dim_Material vào

Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Material”.



Hình 2. 32. Tạo OLE DB Destination đổi tên "Dim_Material".

Bước 4: Nhấn chuột phải vào “Dim_Material”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Material mới trong database.



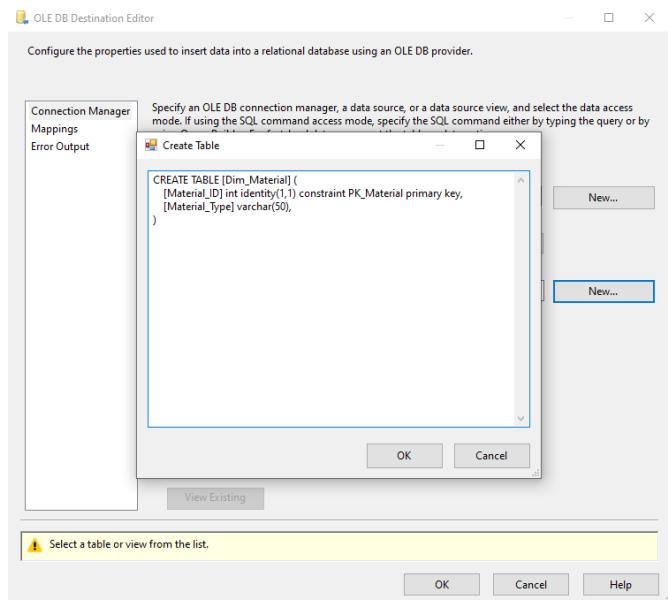
Hình 2. 33. Tạo bảng "Dim_Material"(1).

Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

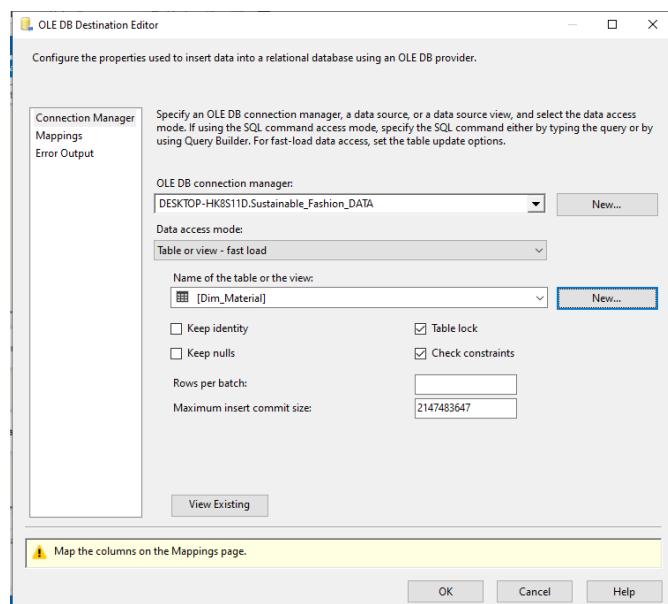
CREATE TABLE [Dim_Material] (

```

[Material_ID] int identity(1,1) constraint PK_Material primary key,
[Material_Type] varchar(50),
  
```

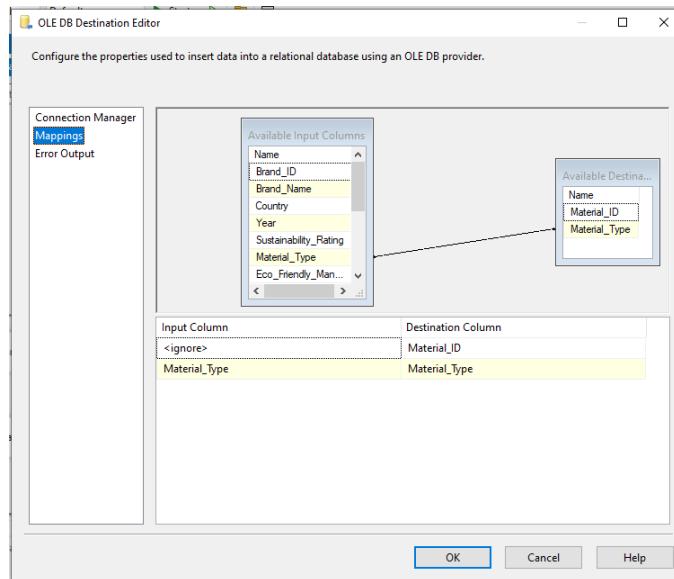


Hình 2. 34. Tạo bảng "Dim_Material"(2).



Hình 2. 35. Tạo bảng "Dim_Material"(3).

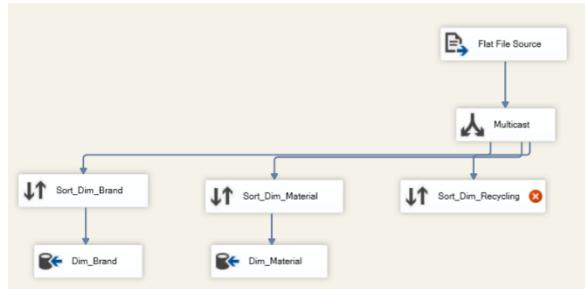
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 36. Kiểm tra các cột thuộc tính.

2.2.3 Bảng Dim_Recycling

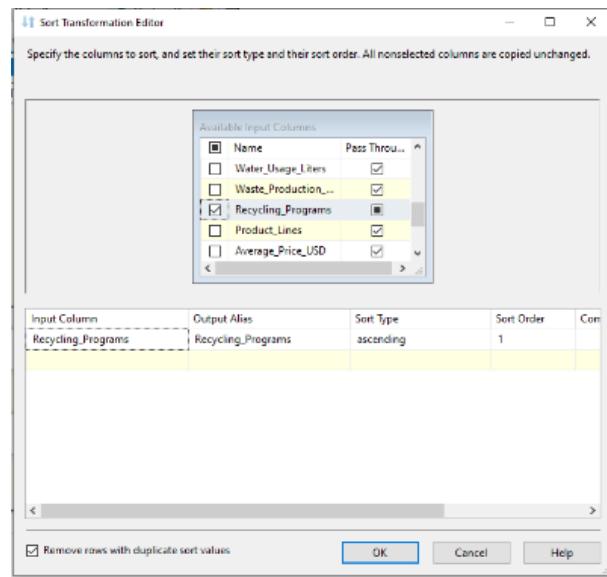
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Recycling”. Kết nối từ “Multicast” đến “Sort_Dim_Recycling”.



Hình 2. 37. Tạo Sort đổi tên "Sort_Dim_Recycling".

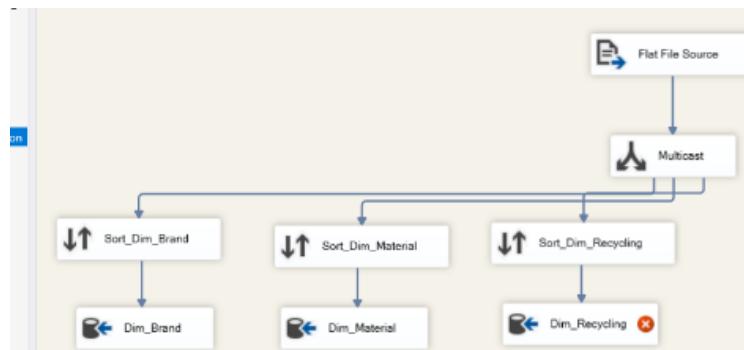
Bước 2:

- Nhấn chuột vào “Sort_Dim_Recycling”, chọn “Edit”.
- Chọn “Recycling_Programs” để làm dữ liệu cho “Sort_Dim_Recycling”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



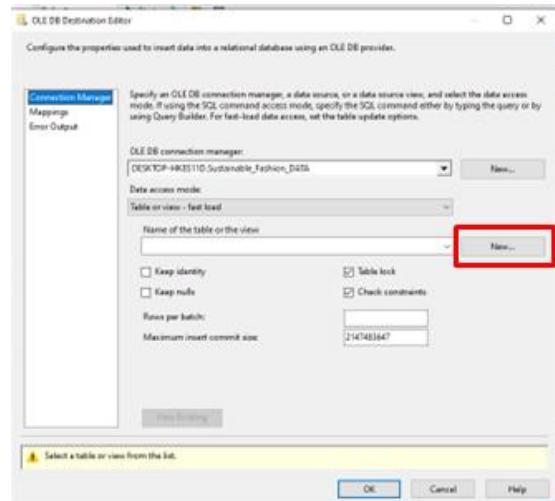
Hình 2. 38. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

Bước 3: Tiến hành đổ dữ liệu bảng Dim_Recycling vào Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Recycling”.



Hình 2. 39. Tạo OLE DB Destination đổi tên "Dim_Recycling".

Bước 4: Nhấn chuột phải vào “Dim_Recycling”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Recycling mới trong database.



Hình 2. 40. Tạo bảng "Dim_Recycling"(1).

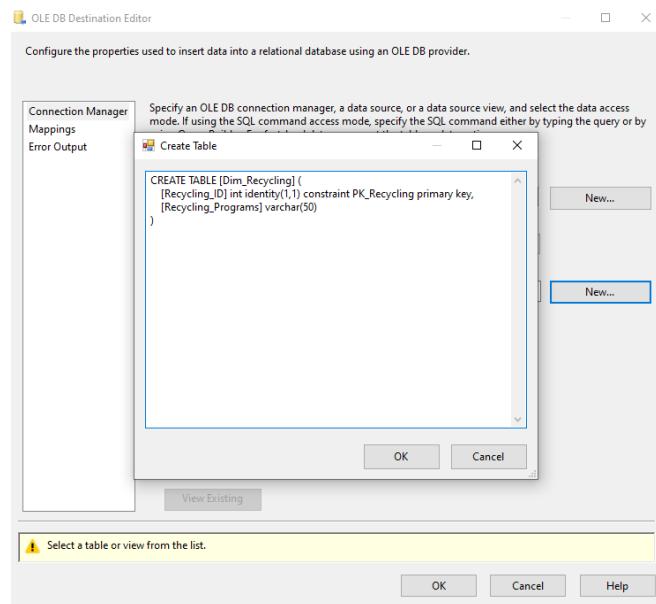
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

CREATE TABLE [Dim_Recycling] (

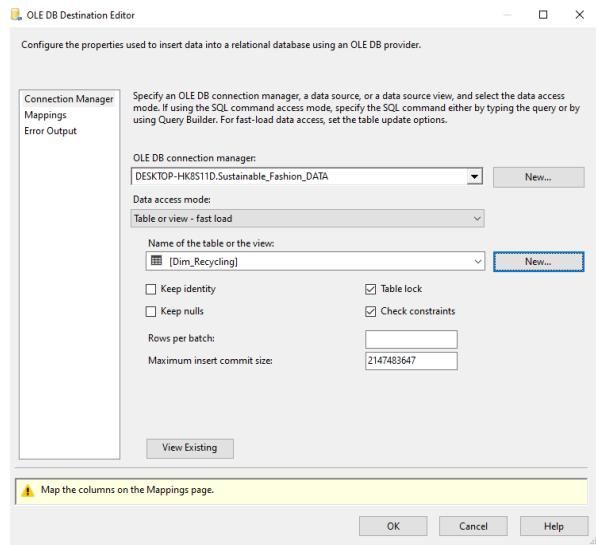
[Recycling_ID] int identity(1,1) constraint PK_Recycling primary key,

[Recycling_Programs] varchar(50)

)

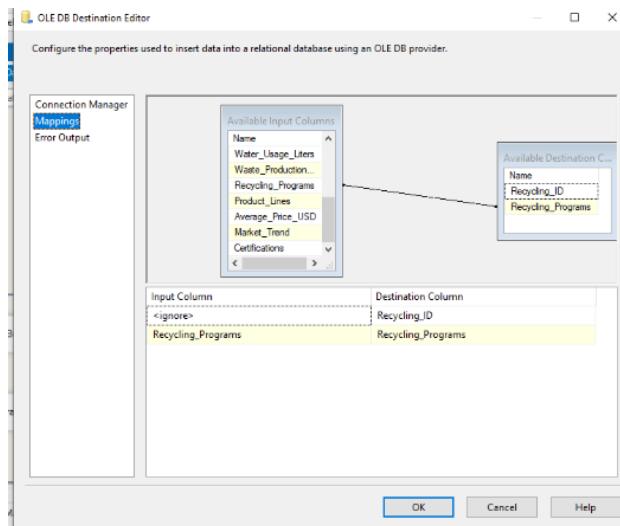


Hình 2. 41. Tạo bảng "Dim_Recycling"(2).



Hình 2.42. Tạo bảng "Dim_Recycling"(3).

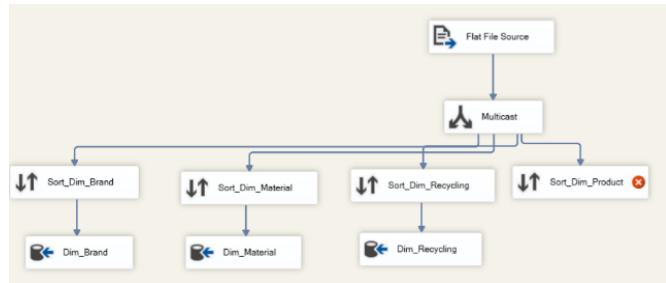
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2.43. Kiểm tra các cột thuộc tính.

2.2.4 Bảng Dim_Product

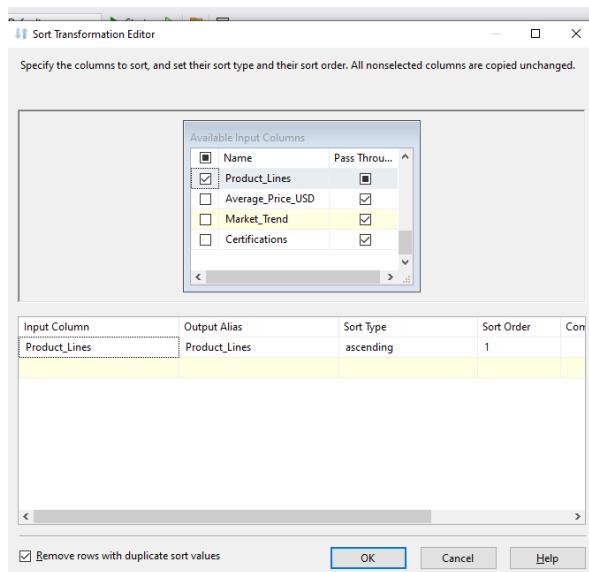
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Product”. Kết nối từ “Multicast” đến “Sort_Dim_Product”.



Hình 2. 44. Tạo Sort đổi tên "Sort_Dim_Product".

Bước 2:

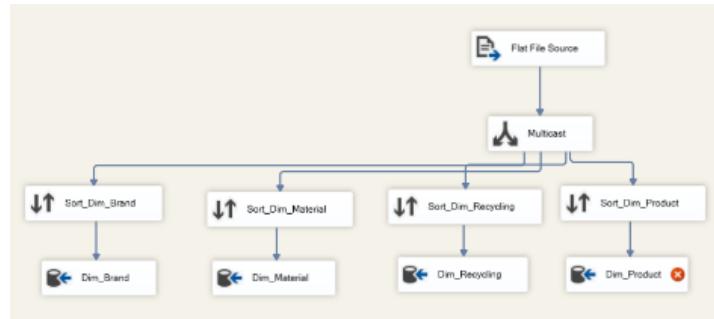
- Nhấn chuột vào “Sort_Dim_Product”, chọn “Edit”.
- Chọn “Product_Lines” để làm dữ liệu cho “Sort_Dim_Product”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



Hình 2.45. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

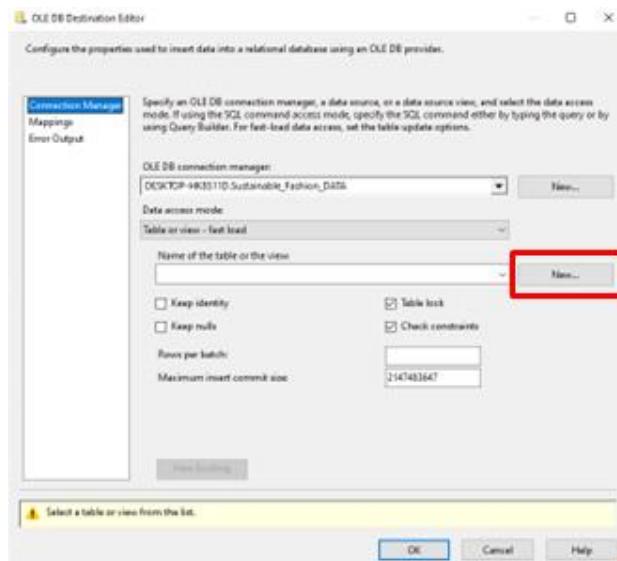
Bước 3: Tiến hành đổ dữ liệu bảng Dim_Product vào

Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Product”.



Hình 2. 46. Tạo OLE DB Destination đổi tên "Dim_Product".

Bước 4: Nhấn chuột phải vào “Dim_Product”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Product mới trong database.

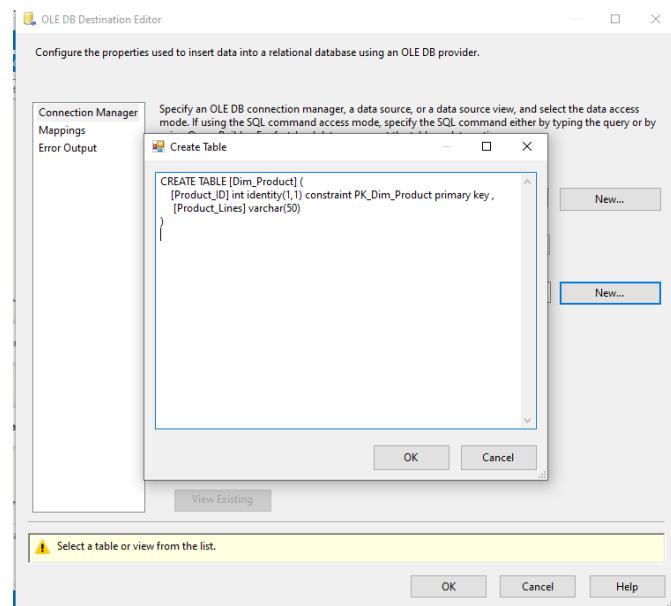


Hình 2. 47. Tạo bảng "Dim_Product"(1).

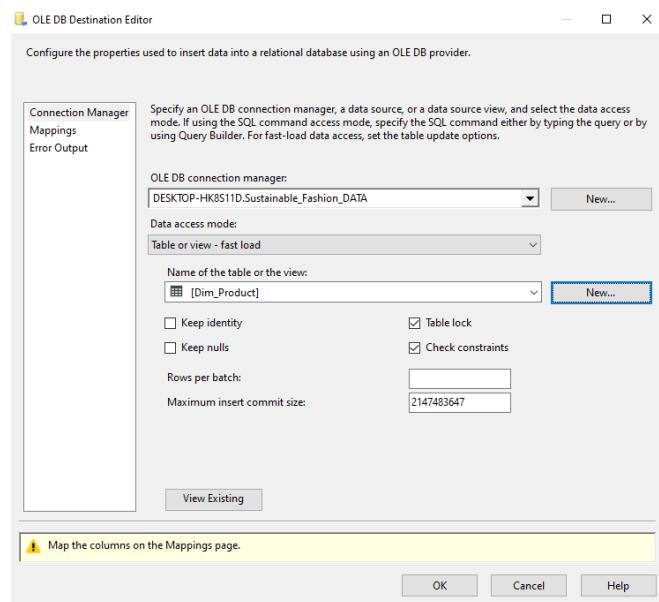
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

```
CREATE TABLE [Dim_Product] (
    [Product_ID] int identity(1,1) constraint PK_Dim_Product primary key ,
    [Product_Lines] varchar(50)
)
```

IS217 – Kho dữ liệu và OLAP

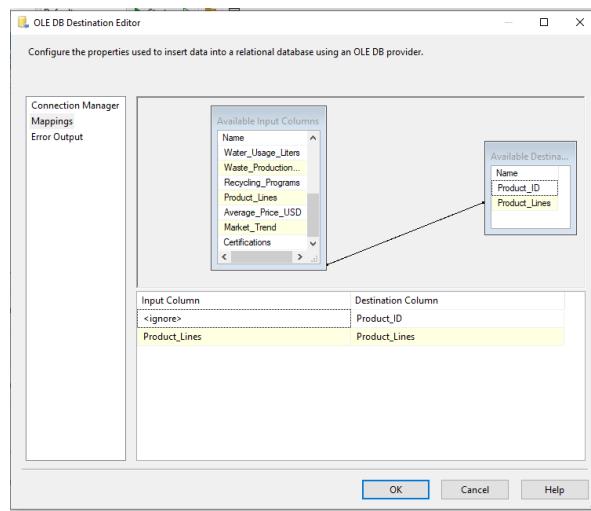


Hình 2. 48. Tạo bảng "Dim_Product"(2).



Hình 2. 49. Tạo bảng "Dim_Product"(3).

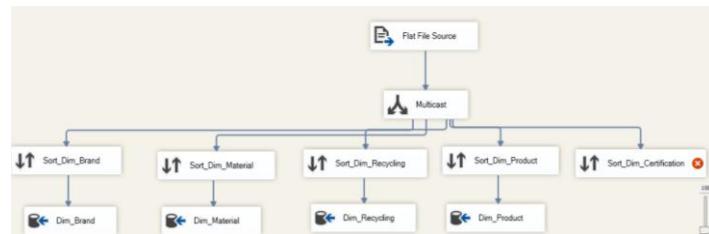
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 50. Kiểm tra các cột thuộc tính.

2.2.5 Bảng Dim_Certification

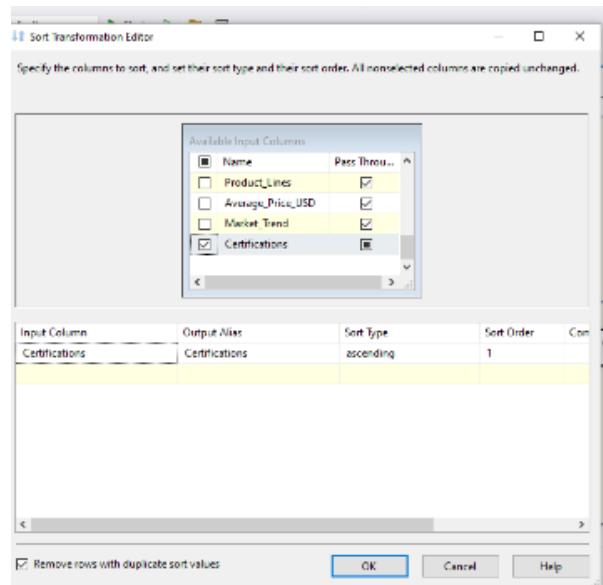
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Certification”. Kết nối từ “Multicast” đến “Sort_Dim_Certification”.



Hình 2. 51. Tạo Sort đổi tên "Sort_Dim_Certification".

Bước 2:

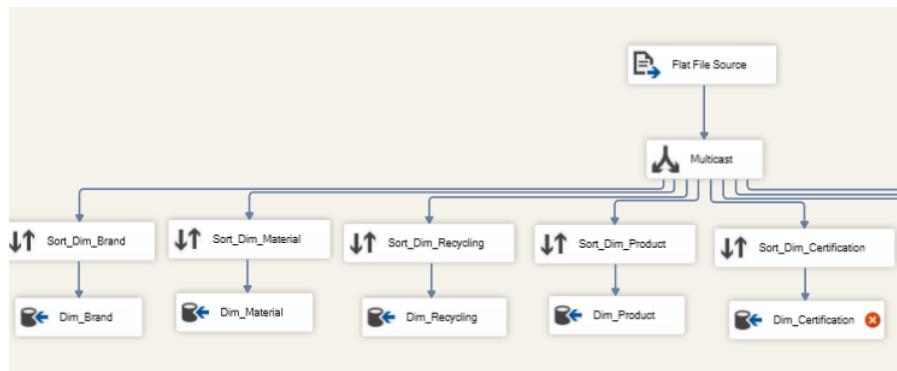
- Nhấn chuột vào “Sort_Dim_Certification”, chọn “Edit”.
- Chọn “Certifications” để làm dữ liệu cho “Sort_Dim_Certification”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



Hình 2. 52. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

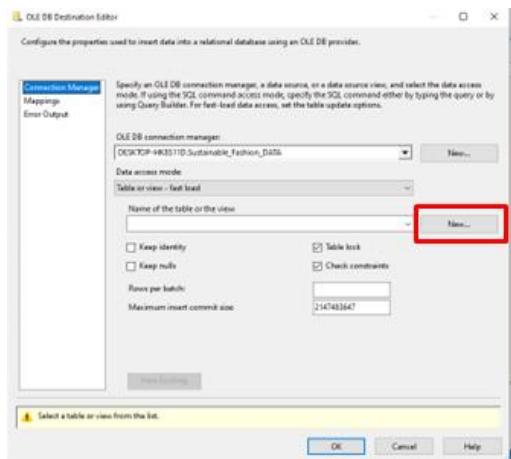
Bước 3: Tiến hành đổ dữ liệu bảng Dim_Certification vào Sustainable_Fashion_DATA.

Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Certification”.



Hình 2. 53. Tạo OLE DB Destination đổi tên "Dim_Certification".

Bước 4: Nhấn chuột phải vào “Dim_Certification”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Certification mới trong database.



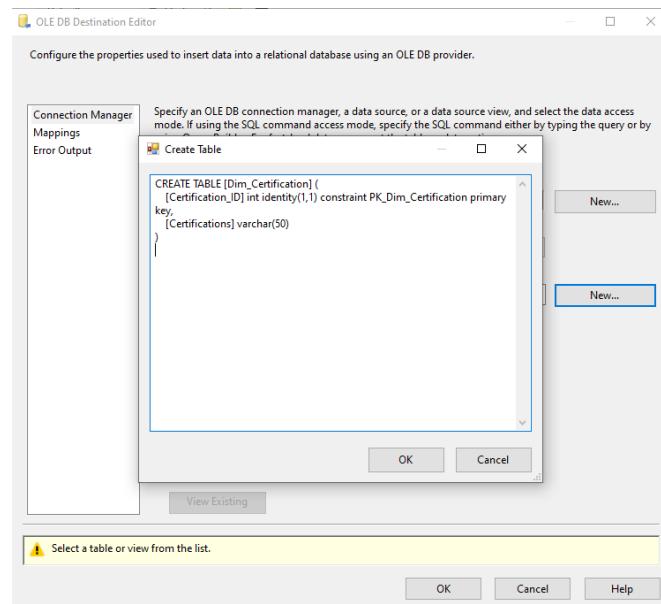
Hình 2. 54. Tạo bảng "Dim_Certification"(1).

Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

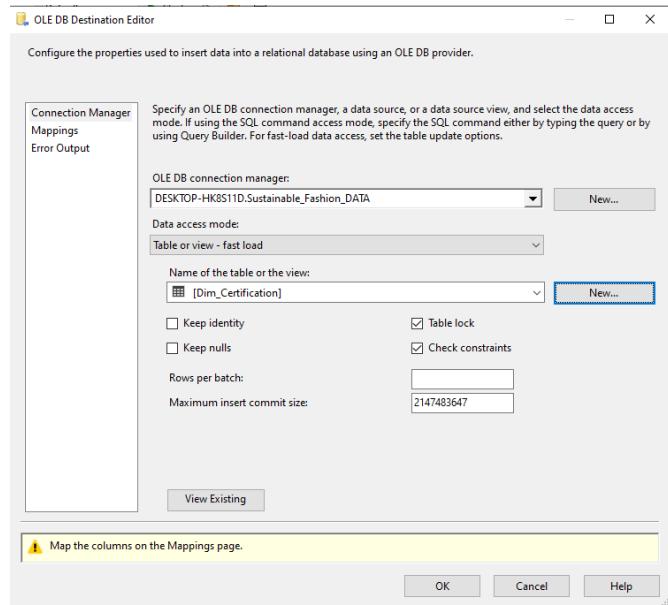
CREATE TABLE [Dim_Certification] (

[Certification_ID] int identity(1,1) constraint PK_Dim_Certification primary key,
[Certifications] varchar(50)

)

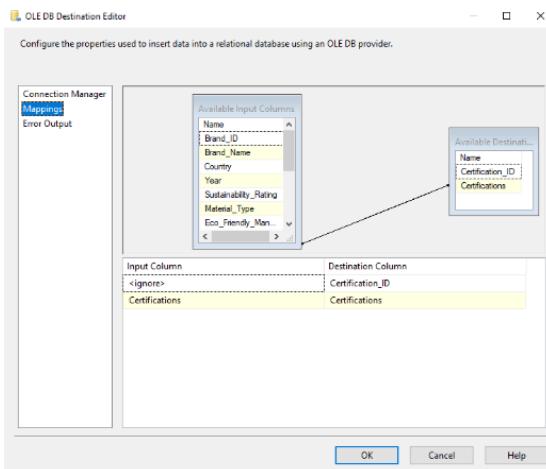


Hình 2. 55. Tạo bảng "Dim_Certification"(2).



Hình 2. 56. Tạo bảng "Dim_Certification"(3).

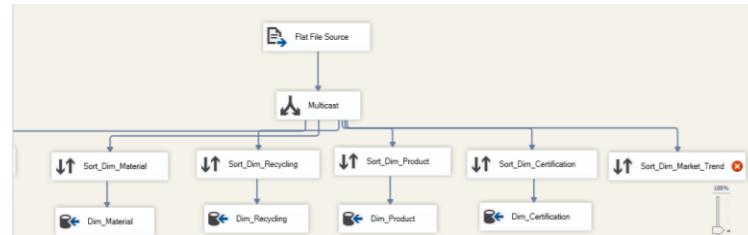
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 57. Kiểm tra các cột thuộc tính.

2.2.6 Bảng Dim_Market_Trend

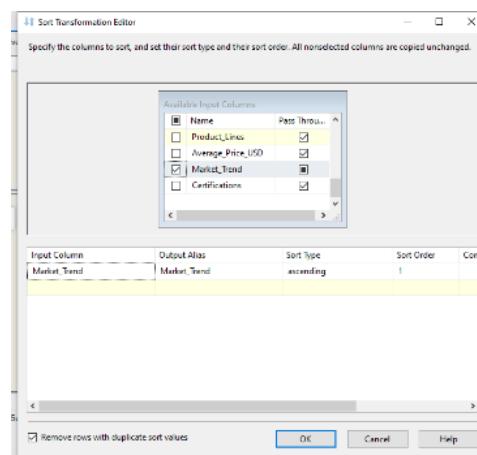
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Market_Trend”. Kết nối từ “Multicast” đến “Sort_Dim_Market_Trend”.



Hình 2. 58. Tạo Sort đổi tên "Sort_Dim_Market_Trend".

Bước 2:

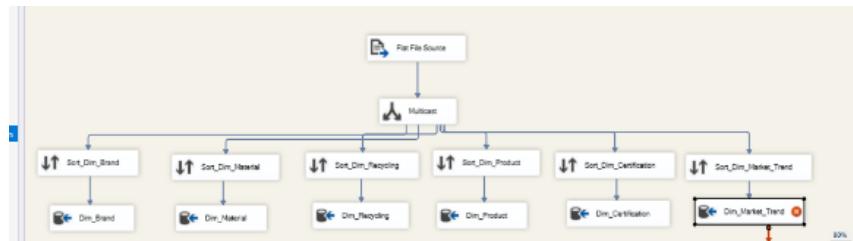
- Nhấn chuột vào “Sort_Dim_Market_Trend”, chọn “Edit”.
- Chọn “Market_” để làm dữ liệu cho “Sort_Dim_Market_Trend”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



Hình 2. 59. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

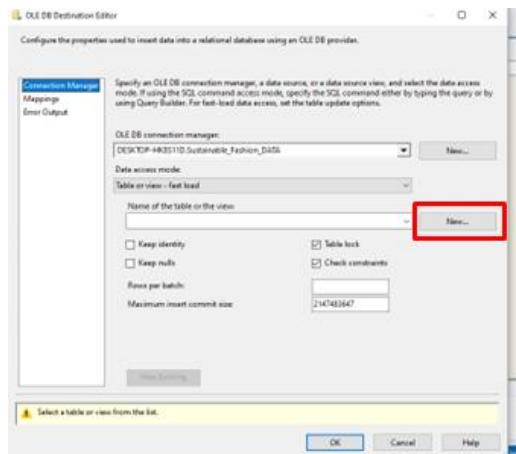
Bước 3: Tiến hành đổ dữ liệu bảng Dim_Market_Trend vào

Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Market_Trend”.



Hình 2. 60. Tạo OLE DB Destination đổi tên "Dim_Market_Trend".

Bước 4: Nhấn chuột phải vào “Dim_Market_Trend”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Market_Trend mới trong database.



Hình 2. 61. Tạo bảng "Dim_Market_Trend"(1).

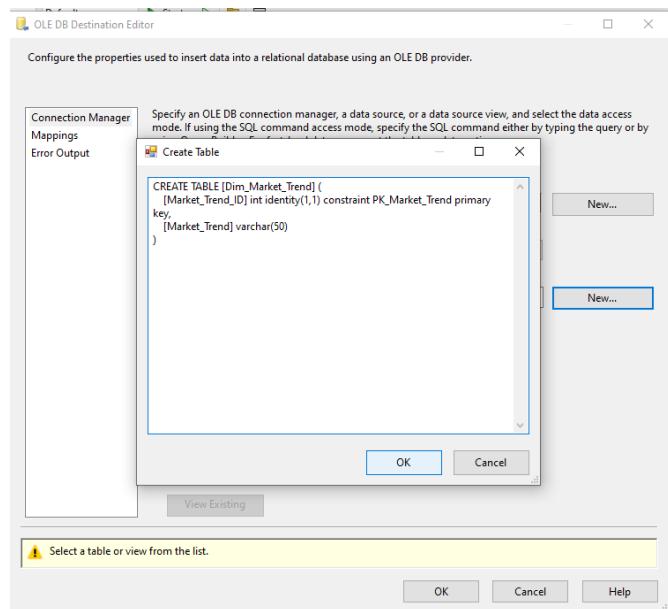
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

CREATE TABLE [Dim_Market_Trend] (

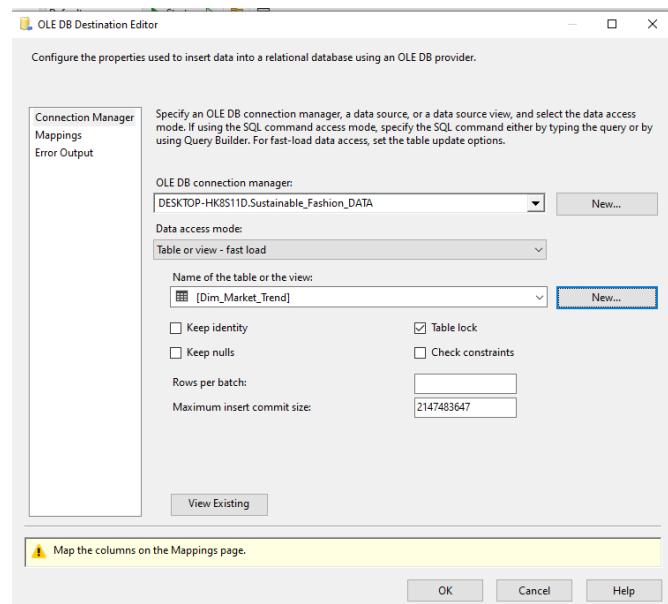
 [Market_Trend_ID] int identity(1,1) constraint PK_Market_Trend primary key,

 [Market_Trend] varchar(50)

)

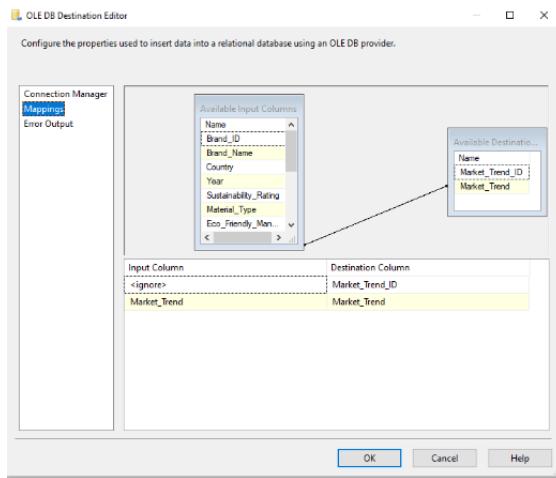


Hình 2. 62. Tạo bảng "Dim_Market_Trend"(2).



Hình 2. 63. Tạo bảng "Dim_Market_Trend"(3).

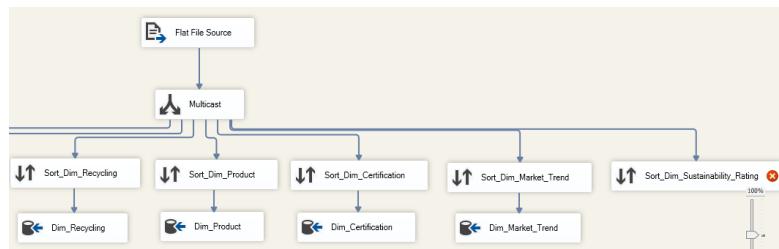
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 64. Kiểm tra các cột thuộc tính.

2.2.7 Bảng Dim_Sustainability_Rating

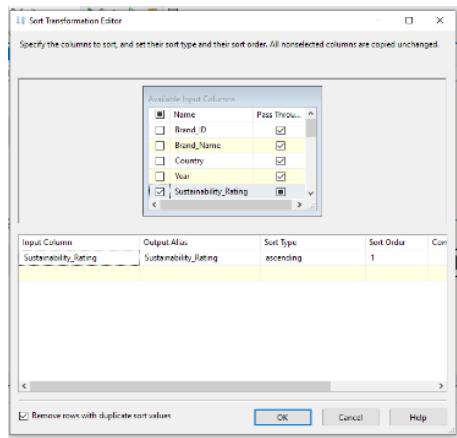
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Sustainability_Rating”. Kết nối từ “Multicast” đến “Sort_Dim_Sustainability_Rating”.



Hình 2. 65. Tạo Sort đổi tên "Sort_Dim_Sustainability_Rating".

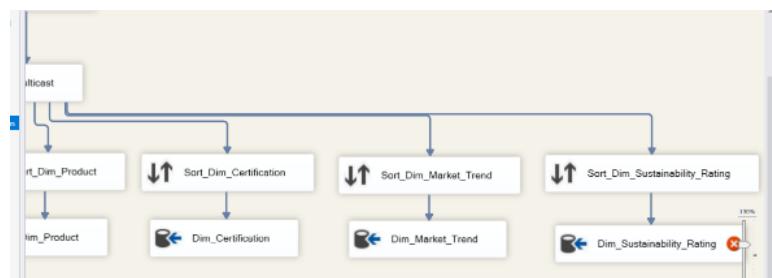
Bước 2:

- Nhấn chuột vào “Sort_Dim_Sustainability_Rating”, chọn “Edit”.
- Chọn “Sustainability_Rating” để làm dữ liệu cho “Sort_Dim_Sustainability_Rating”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



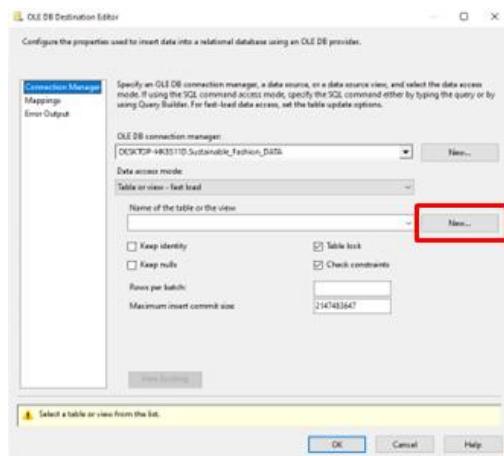
Hình 2. 66. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

Bước 3: Tiến hành đổ dữ liệu bảng Dim_Sustainability_Rating vào Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Sustainability_Rating”.



Hình 2. 67. Tạo OLE DB Destination đổi tên "Dim_Sustainability_Rating".

Bước 4: Nhấn chuột phải vào “Dim_Sustainability_Rating”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Sustainability_Rating mới trong database.



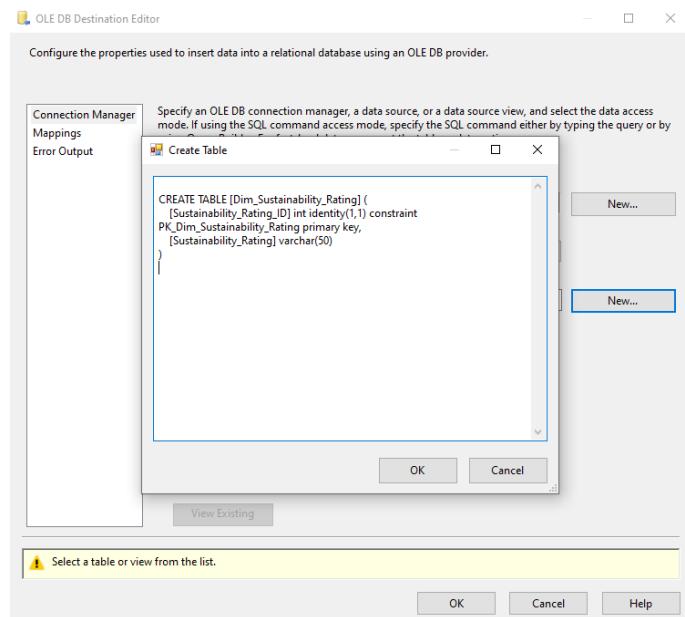
Hình 2. 68. Tạo bảng "Dim_Sustainability_Rating"(1).

Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

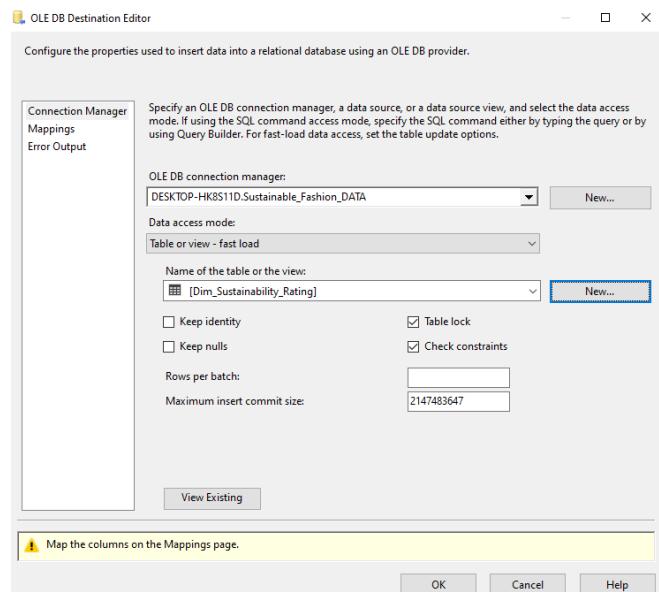
CREATE TABLE [Dim_Sustainability_Rating] (

 [Sustainability_Rating_ID] int identity(1,1) constraint PK_Dim_Sustainability_Rating
 primary key,

 [Sustainability_Rating] varchar(50))

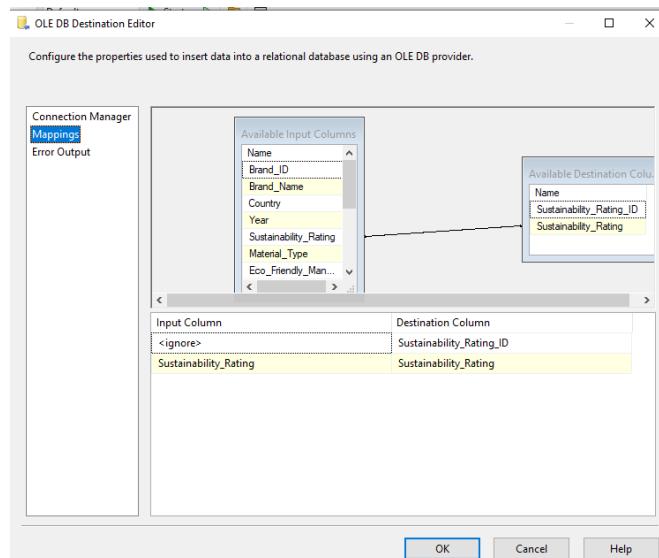


Hình 2. 69. Tạo bảng "Dim_Sustainability_Rating"(2).



Hình 2. 70. Tạo bảng "Dim_Sustainability_Rating"(3).

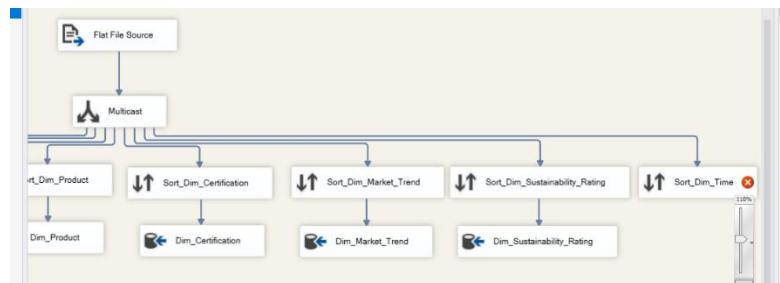
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 71. Kiểm tra các cột thuộc tính.

2.2.8 Bảng Dim_Time

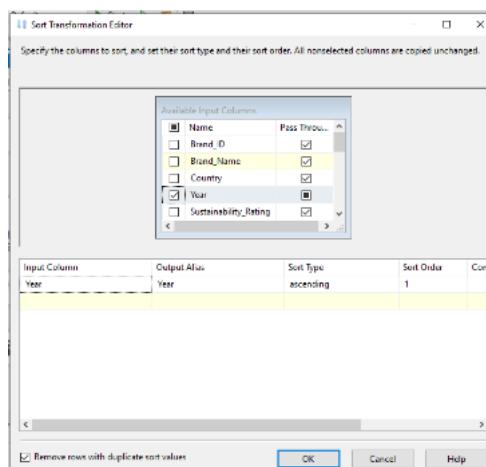
Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Time”. Kết nối từ “Multicast” đến “Sort_Dim_Time”.



Hình 2. 72. Tạo Sort đổi tên "Sort_Dim_Time".

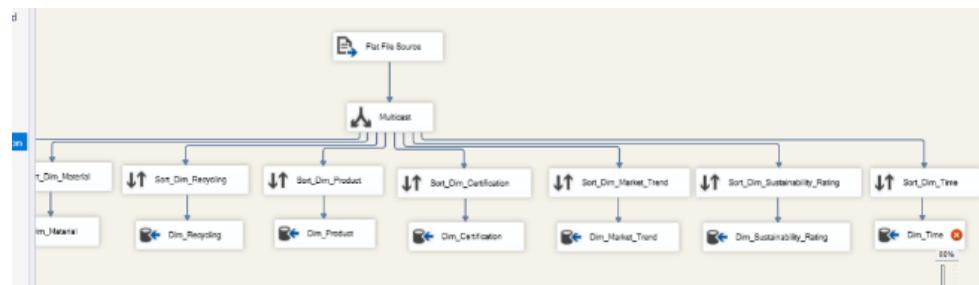
Bước 2:

- Nhấn chuột vào “Sort_Dim_Time”, chọn “Edit”.
- Chọn “Year” để làm dữ liệu cho “Sort_Dim_Time”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



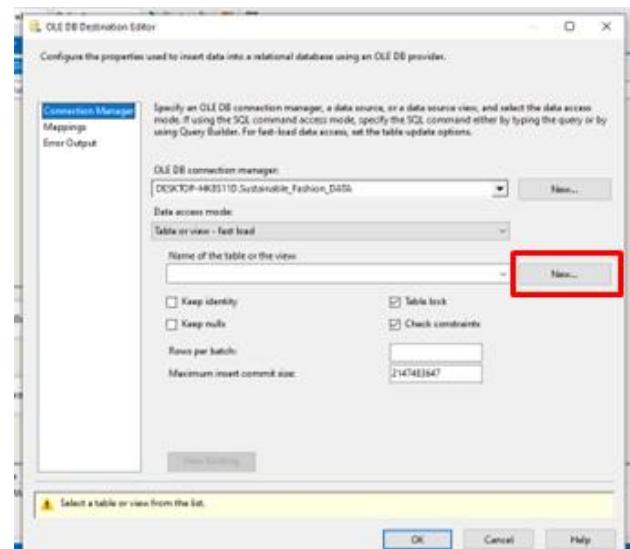
Hình 2. 73. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

Bước 3: Tiến hành đổ dữ liệu bảng Dim_Time vào Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Time”.



Hình 2. 74. Tạo OLE DB Destination đổi tên "Dim_Time".

Bước 4: Nhấn chuột phải vào “Dim_Time”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Time mới trong database.

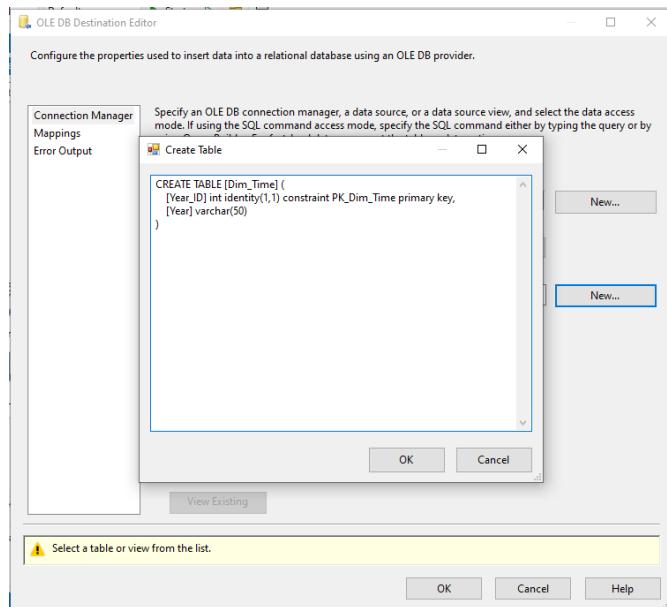


Hình 2. 75. Tạo bảng "Dim_Time"(1).

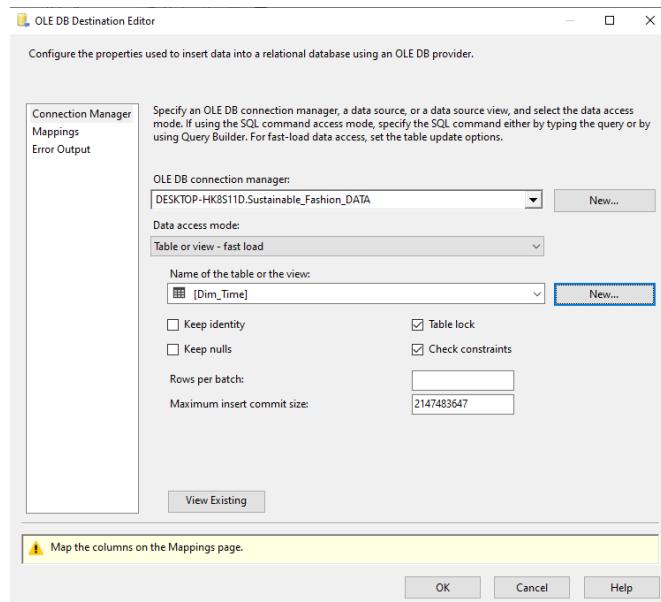
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

```
CREATE TABLE [Dim_Time] (
    [Year_ID] int identity(1,1) constraint PK_Dim_Time primary key,
    [Year] varchar(50)
)
```

IS217 – Kho dữ liệu và OLAP

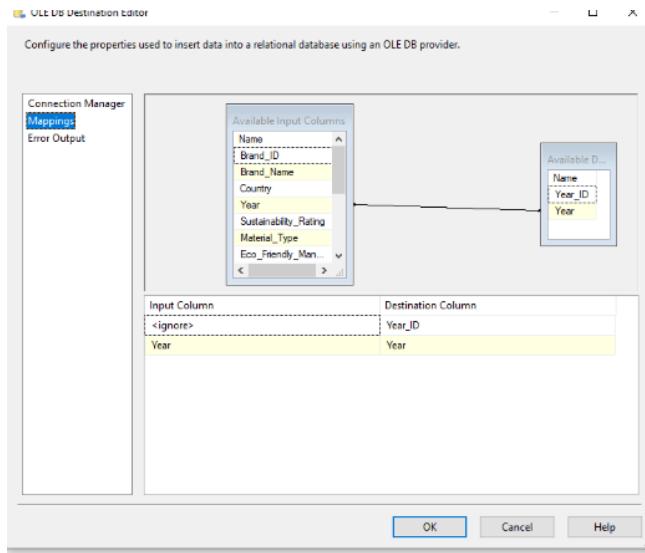


Hình 2. 76. Tạo bảng "Dim_Time"(2).



Hình 2. 77. Tạo bảng "Dim_Time"(3).

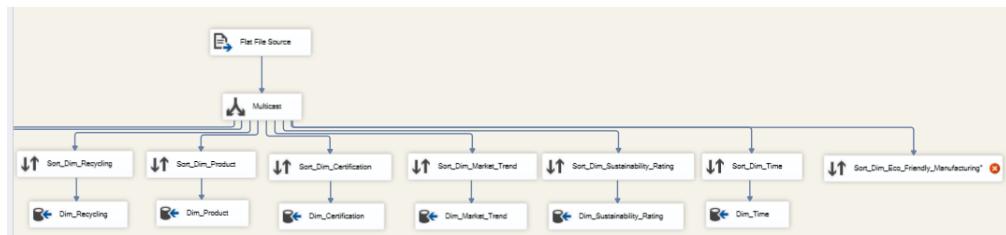
Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 78. Kiểm tra các cột thuộc tính.

2.2.9 Bảng Dim_Eco_Friendly_Manufacturing

Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Eco_Friendly_Manufacturing”. Kết nối từ “Multicast” đến “Sort_Dim_Eco_Friendly_Manufacturing”.

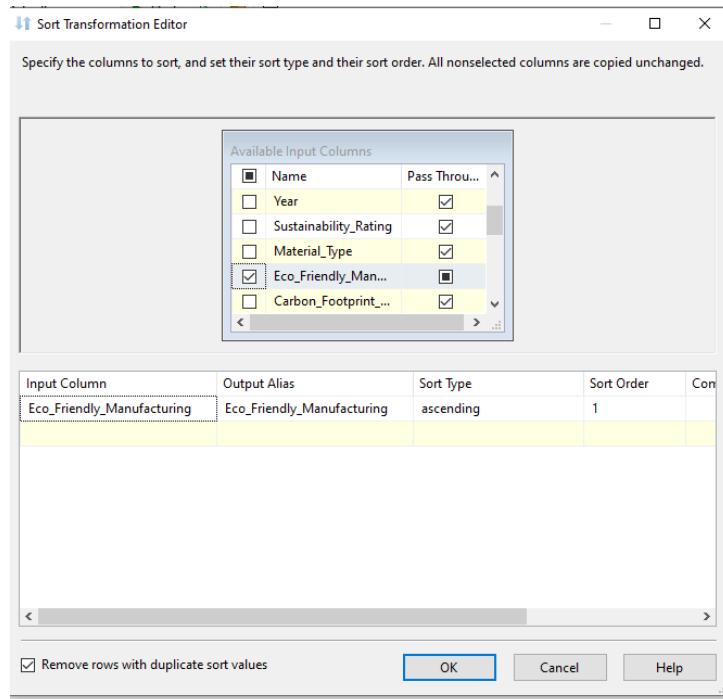


Hình 2. 79. Tạo Sort đổi tên "Sort_Dim_Eco_Friendly_Manufacturing".

Bước 2:

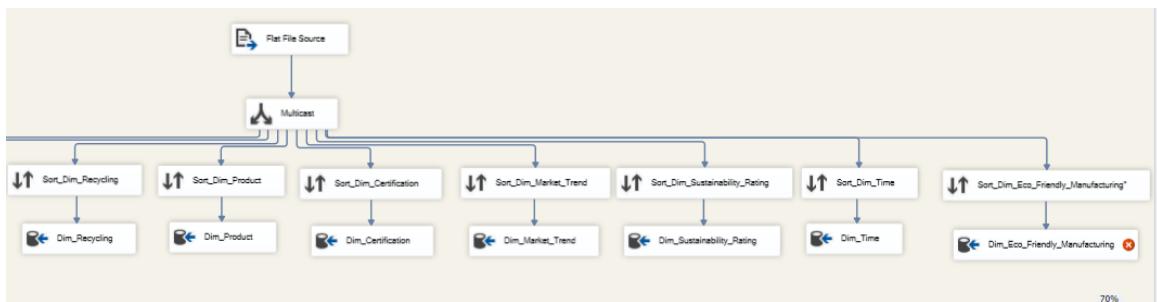
- Nhấn chuột vào “Sort_Dim_Eco_Friendly_Manufacturing”, chọn “Edit”.
- Chọn “Eco_Friendly_Manufacturing” để làm dữ liệu cho “Sort_Dim_Eco_Friendly_Manufacturing”.
- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng

dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.



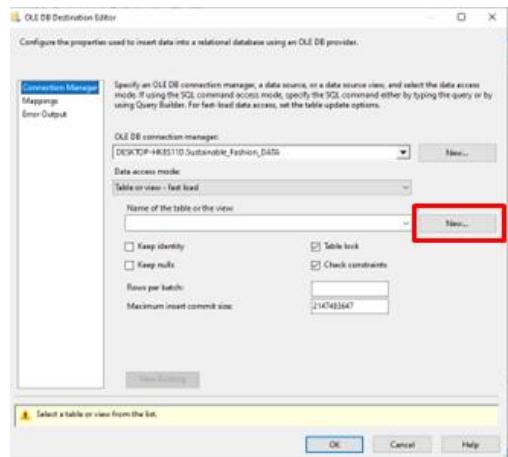
Hình 2. 80. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

Bước 3: Tiến hành đổ dữ liệu bảng Dim_Eco_Friendly_Manufacturing vào Sustainable_Fashion_DATA. Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Eco_Friendly_Manufacturing”.



Hình 2. 81. Tạo OLE DB Destination đổi tên "Dim_Eco_Friendly_Manufacturing".

Bước 4: Nhấn chuột phải vào “Dim_Eco_Friendly_Manufacturing”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Eco_Friendly_Manufacturing mới trong database.

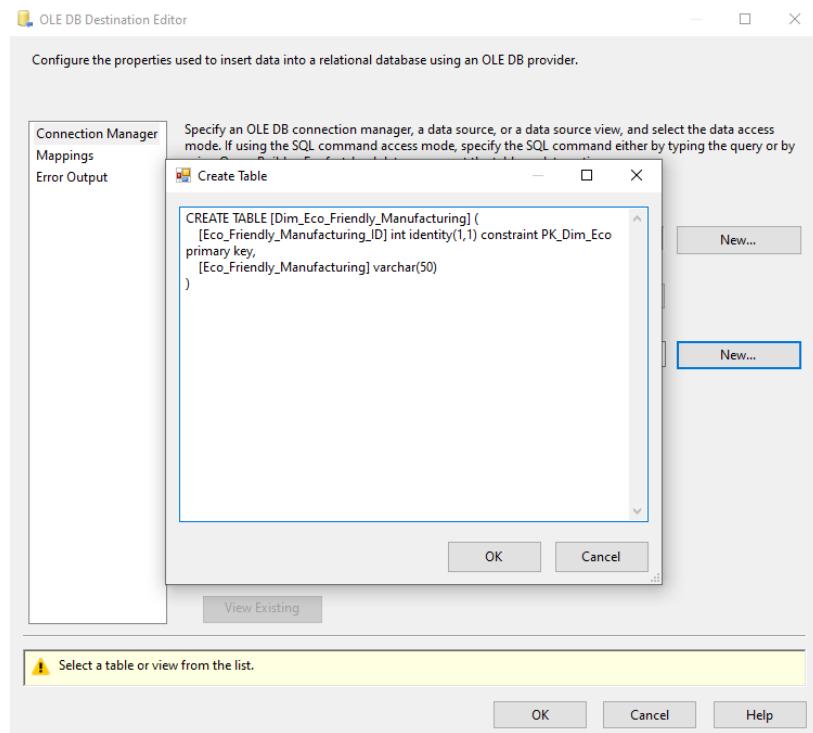


Hình 2. 82. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(1).

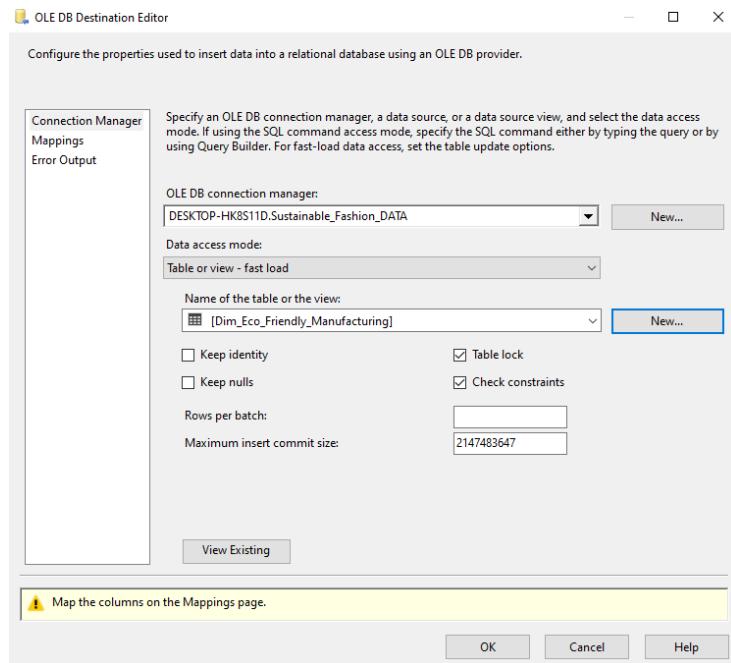
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

```
CREATE TABLE [Dim_Eco_Friendly_Manufacturing] (
    [Eco_Friendly_Manufacturing_ID] int identity(1,1) constraint PK_Dim_Eco primary
key,
    [Eco_Friendly_Manufacturing] varchar(50)
)
```

IS217 – Kho dữ liệu và OLAP



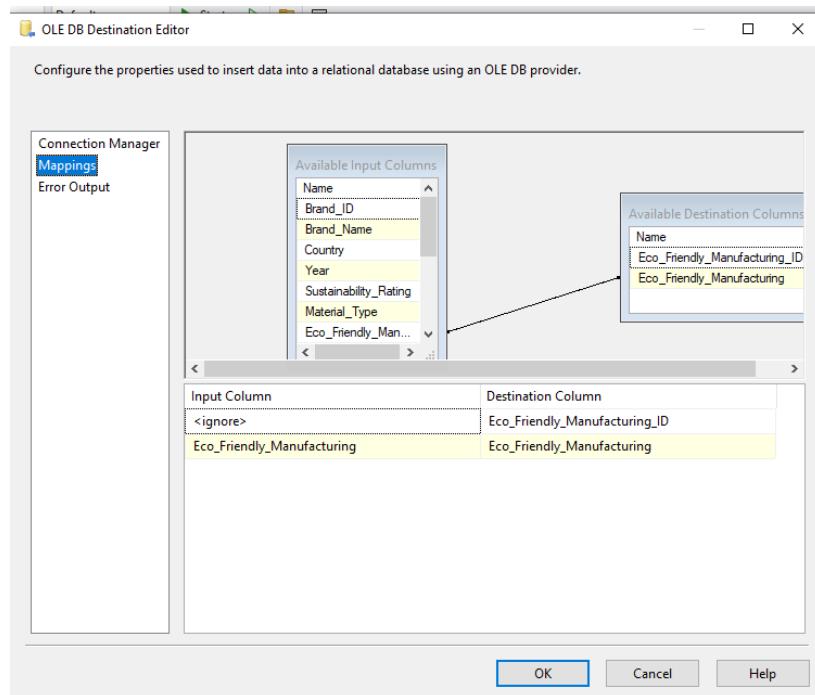
Hình 2. 83. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(2).



Hình 2. 84. Tạo bảng "Dim_Eco_Friendly_Manufacturing"(3).

Bước 5: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để

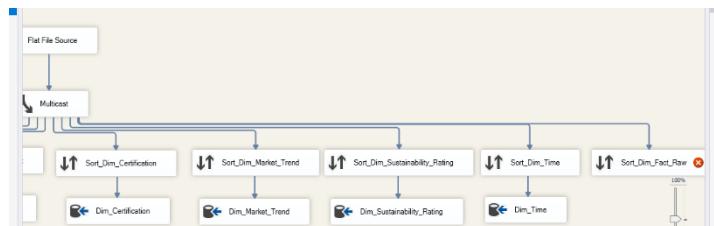
hoàn tất.



Hình 2. 85. Kiểm tra các cột thuộc tính.

2.2.10 Bảng Dim_Fact_Raw

Bước 1: Tạo một “Sort”, đổi tên thành “Sort_Dim_Fact_Raw”. Kết nối từ “Multicast” đến “Sort_Dim_Fact_Raw”.



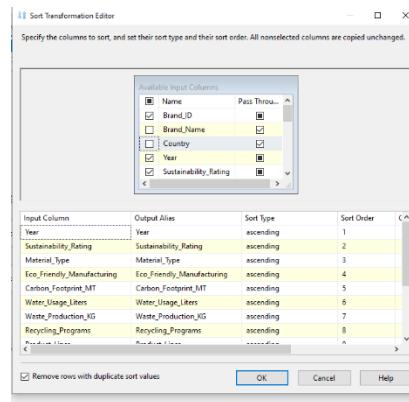
Hình 2. 86. Tạo Sort đổi tên "Sort_Dim_Fact_Raw".

Bước 2:

- Nhấn chuột vào “Sort_Dim_Fact_”, chọn “Edit”.
- Chọn tất cả trừ “Brand_Name”, “Country” để làm dữ liệu cho

“Sort_Dim_Fact_Raw”.

- Chọn vào “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau. Sau đó chọn vào “OK” để hoàn tất.

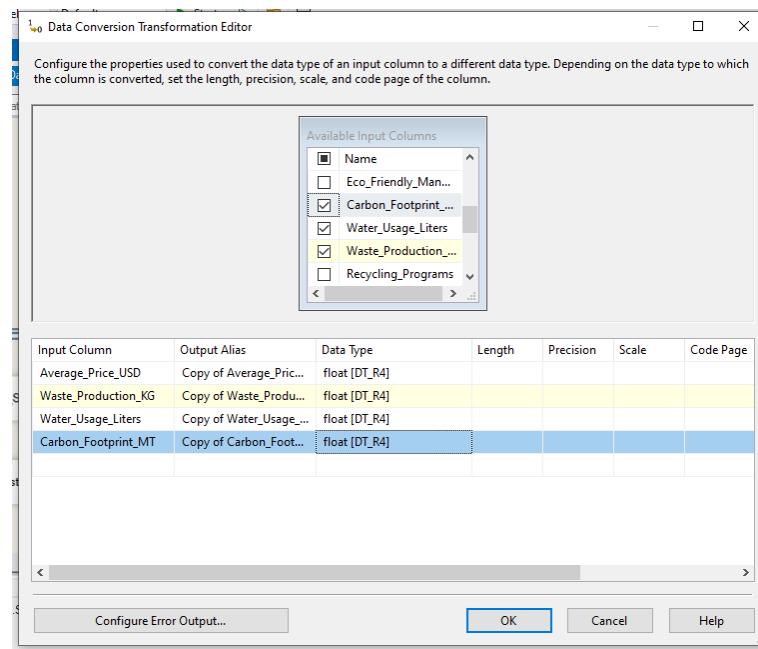


Hình 2. 87. Sắp xếp dữ liệu của các cột thuộc tính bằng Sort.

Bước 3:

Đổi kiểu dữ liệu “Carbon_Footprint_MT”, “Water_Usage_Liters”, “Waste_Production_KG”, “Average_Price_USD”.

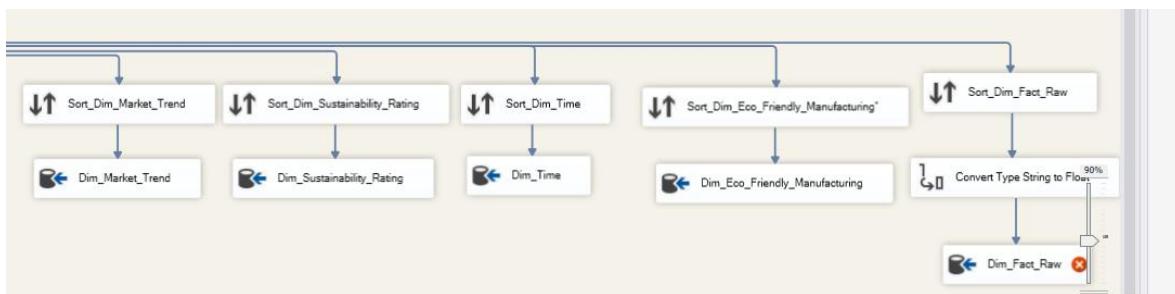
- Tạo một “Data Conversion”, đổi tên thành “Convert type String to Float”
- Trong đây sẽ chuyển “Carbon_Footprint_MT”, “Water_Usage_Liters”, “Waste_Production_KG”, “Average_Price_USD” kiểu String thành kiểu Float.
- Tạo kết nối từ “Sort_Dim_Fact_Raw” đến “Convert type String to Float”.



Hình 2. 88. Đổi kiểu dữ liệu.

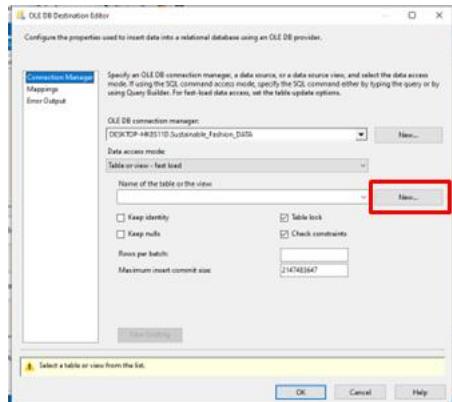
Bước 4: Tiến hành đổi dữ liệu bảng Dim_Fact_Raw vào Sustainable_Fashion_DATA.

Đầu tiên, tạo một OLE DB Destination, đặt tên là “Dim_Fact_Raw”



Hình 2. 89. Tạo OLE DB Destination đổi tên "Dim_Fact_Raw".

Bước 5: Nhấn chuột phải vào “Dim_Fact_Raw”, chọn “Edit”. Chọn “New” để tạo bảng Dim_Fact_Raw mới trong database.

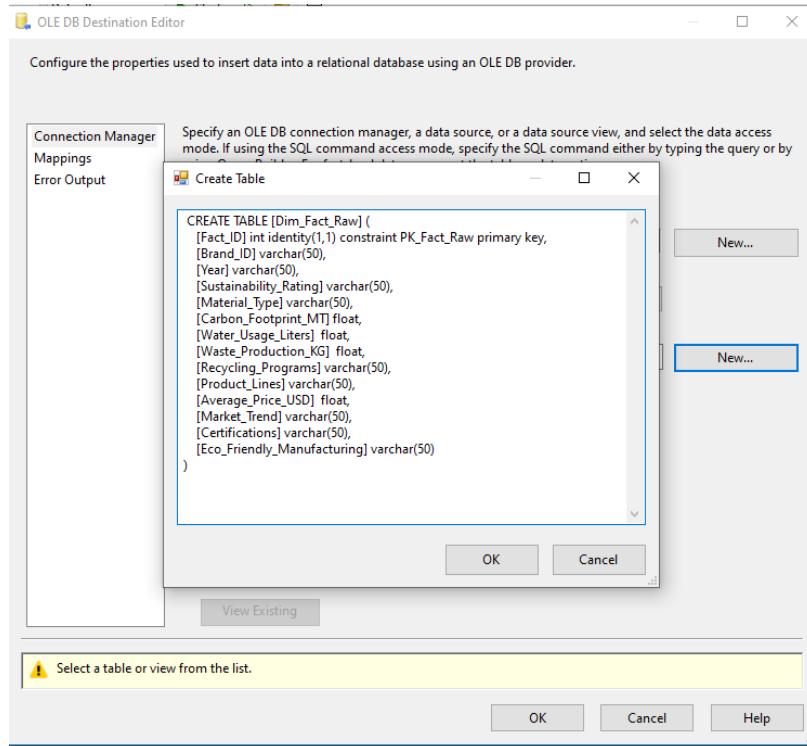


Hình 2. 90. Tạo bảng "Dim_Fact_Raw"(1).

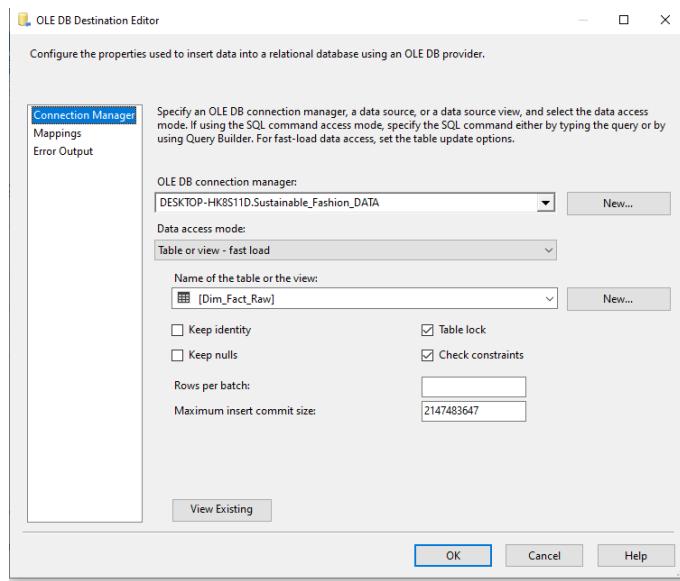
Viết lệnh SQL để tạo bảng, nhấn “OK” để tạo.

```
CREATE TABLE [Dim_Fact_Raw] (
    [Fact_ID] int identity(1,1) constraint PK_Fact_Raw primary key,
    [Brand_ID] varchar(50),
    [Year] varchar(50),
    [Sustainability_Rating] varchar(50),
    [Material_Type] varchar(50),
    [Carbon_Footprint_MT] float,
    [Water_Usage_Liters] float,
    [Waste_Production_KG] float,
    [Recycling_Programs] varchar(50),
    [Product_Lines] varchar(50),
    [Average_Price_USD] float,
    [Market_Trend] varchar(50),
    [Certifications] varchar(50),
    [Eco_Friendly_Manufacturing] varchar(50)
)
```

IS217 – Kho dữ liệu và OLAP

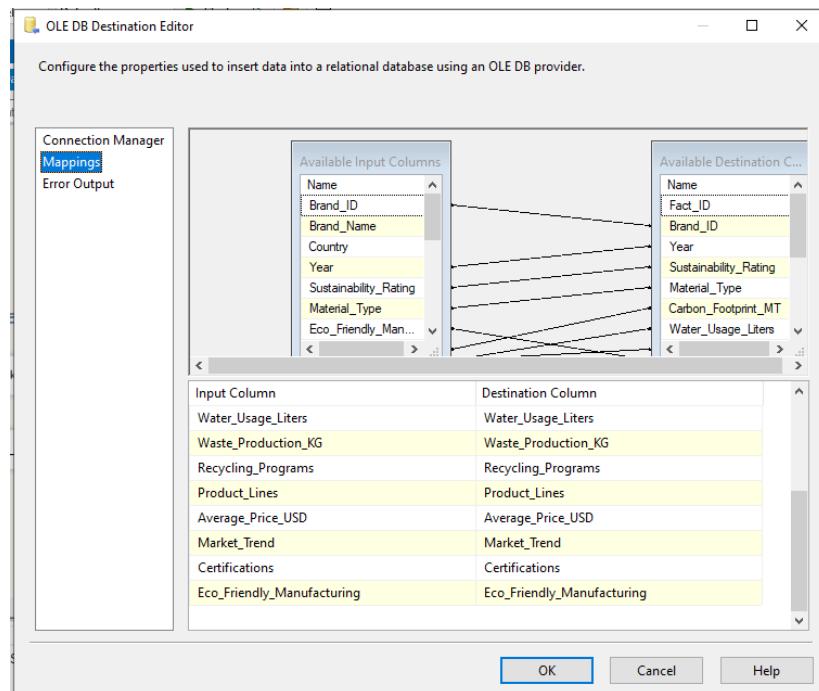


Hình 2. 91. Tạo bảng "Dim_Fact_Raw"(2).



Hình 2. 92. Tạo bảng "Dim_Fact_Raw"(3).

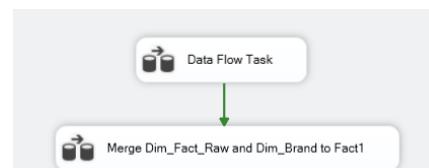
Bước 6: Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 93. Kiểm tra các cột thuộc tính.

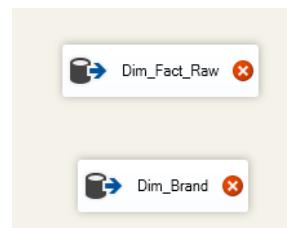
2.2.11 Merge Dim_Fact_Raw và Dim_Brand thành Fact1

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Dim_Fact_Raw và Dim_Brand thành Fact1”.



Hình 2. 94. Tạo Data Flow Task có tên "Merge Dim_Fact_Raw and Dim_Brand to Fact1".

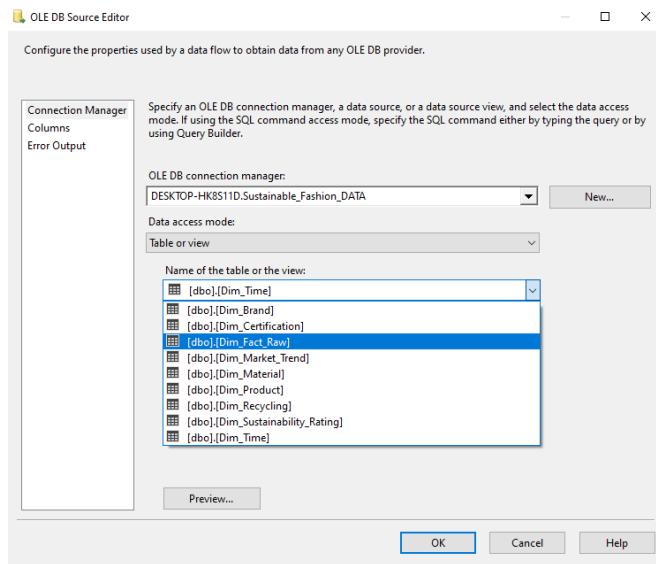
Bước 2: Vào “Merge Dim_Fact_Raw and Dim_Brand to Fact1”, tạo 2 “OLE DB Source” là Dim_Fact_Raw và Dim_Brand



Hình 2. 95. Các bước thực hiện (1).

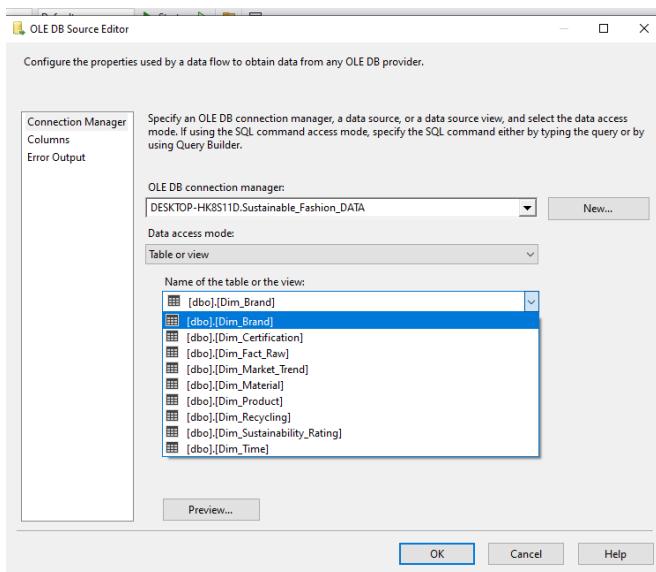
Bước 3: Nhấp chuột phải vào “Dim_Fact_Raw”, chọn “Edit”. Chọn dữ liệu bảng

“Dim_Fact_Raw” đã được chọn trong mục “Name of the table or the view”.



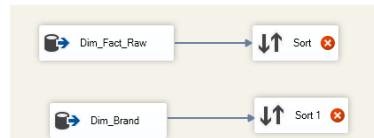
Hình 2. 96. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Brand



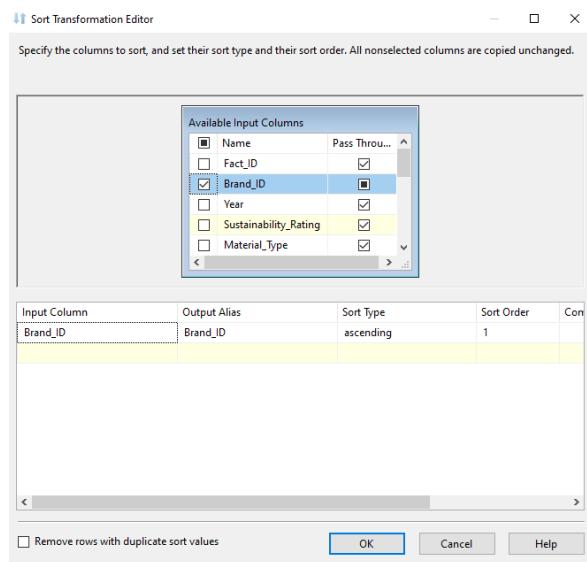
Hình 2. 97. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Dim_Fact_Raw” và “Dim_Brand”, lấy ra thuộc tính để tiến hành kết hai bảng.



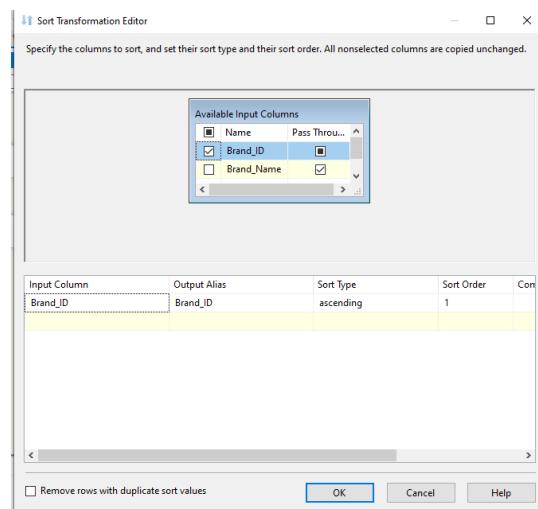
Hình 2. 98. Các bước thực hiện (4).

Với “Sort” kết nối với “Dim_Fact_Raw” sẽ lấy ra thuộc tính “Brand_ID”



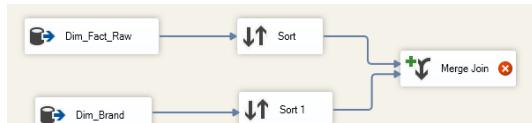
Hình 2. 99. Các bước thực hiện (5).

Với “Sort 1” kết nối với bảng “Dim_Brand” sẽ lấy ra thuộc tính “Brand_ID”



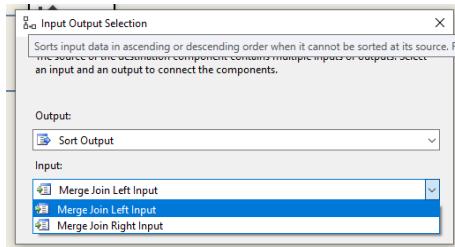
Hình 2. 100. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



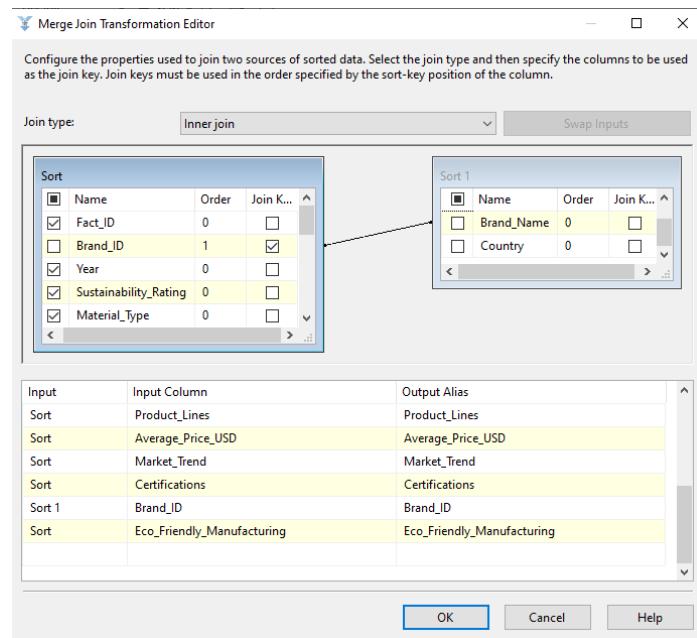
Hình 2. 101. Các bước thực hiện (7).

Kết nối “Sort” của “Dim_Fact_Raw” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Dim_Fact_Raw”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Brand”.



Hình 2. 102. Các bước thực hiện (8).

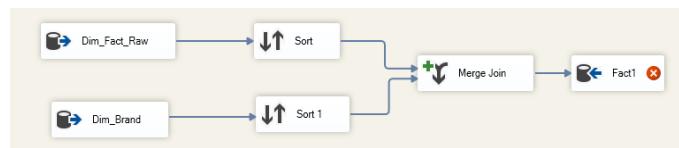
Kết nối “Sort 1” của “Dim_Brand” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 103. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact1 vào Sustainable_Fashion_DATA. Đầu tiên

chọn một “OLE DB Destination”, đặt tên thành “Fact1”



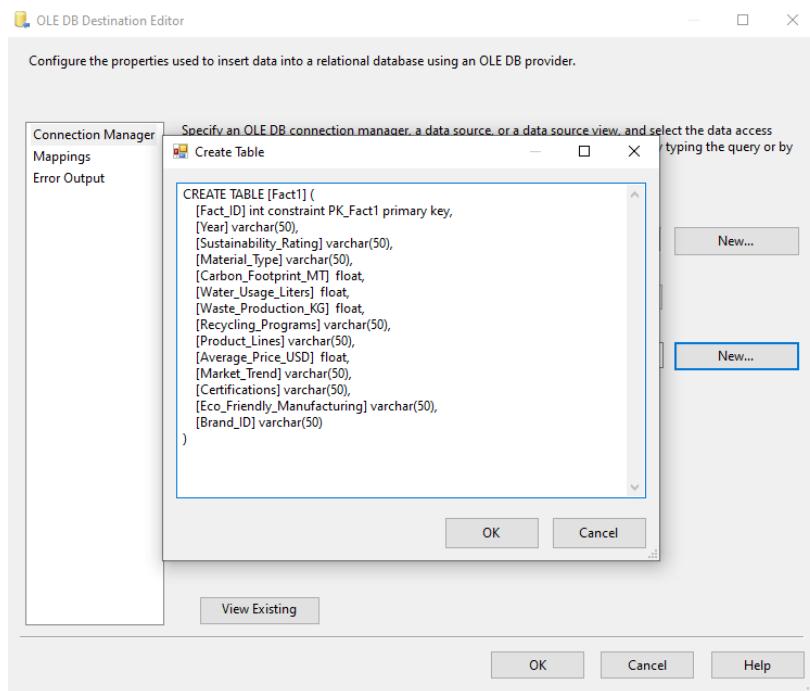
Hình 2. 104. Các bước thực hiện (10).

Bước 7: Nhấn chuột phải vào “Fact1”, chọn “Edit”. Chọn “New” để tạo bảng “Fact1” mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

```

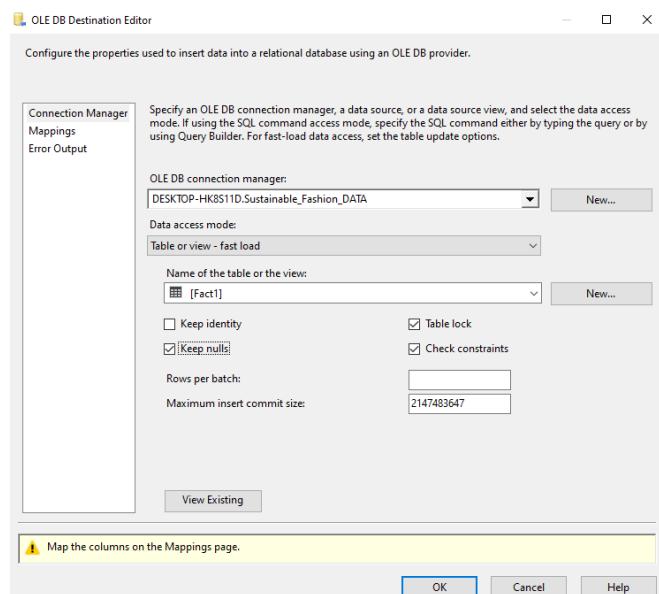
CREATE TABLE [Fact1] (
    [Fact_ID] int constraint PK_Fact1 primary key,
    [Year] varchar(50),
    [Sustainability_Rating] varchar(50),
    [Material_Type] varchar(50),
    [Carbon_Footprint_MT] float,
    [Water_Usage_Liters] float,
    [Waste_Production_KG] float,
    [Recycling_Programs] varchar(50),
    [Product_Lines] varchar(50),
    [Average_Price_USD] float,
    [Market_Trend] varchar(50),
    [Certifications] varchar(50),
    [Eco_Friendly_Manufacturing] varchar(50),
    [Brand_ID] varchar(50)
)
  
```

IS217 – Kho dữ liệu và OLAP



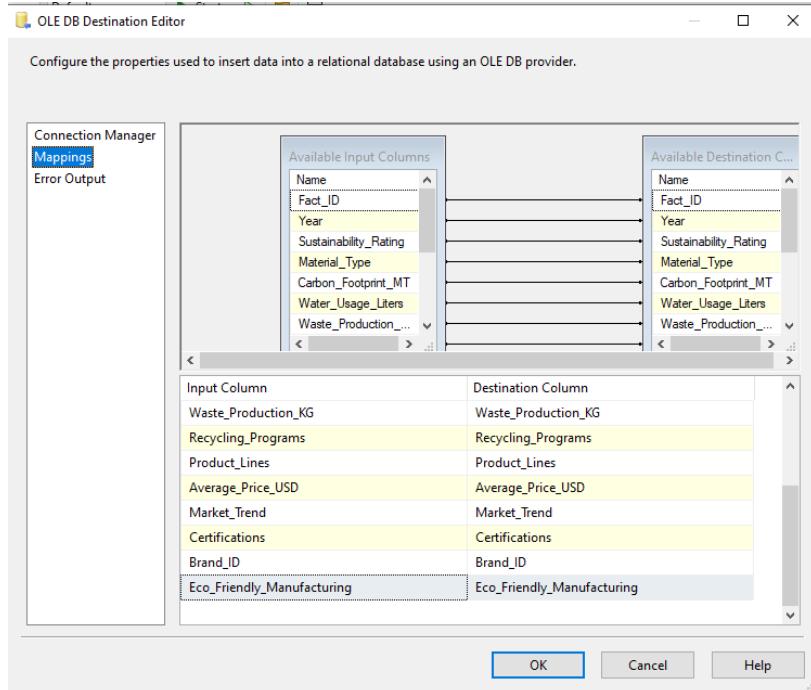
Hình 2. 105. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần.

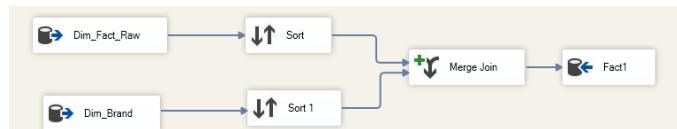


Hình 2. 106. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



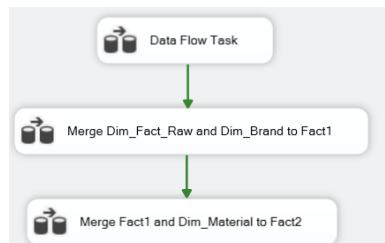
Hình 2. 107. Các bước thực hiện (13).



Hình 2. 108. Sau khi thực hiện xong.

2.2.12 Merge Fact1 và Dim_Material thành Fact2

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact1 và Dim_Material thành Fact2”.



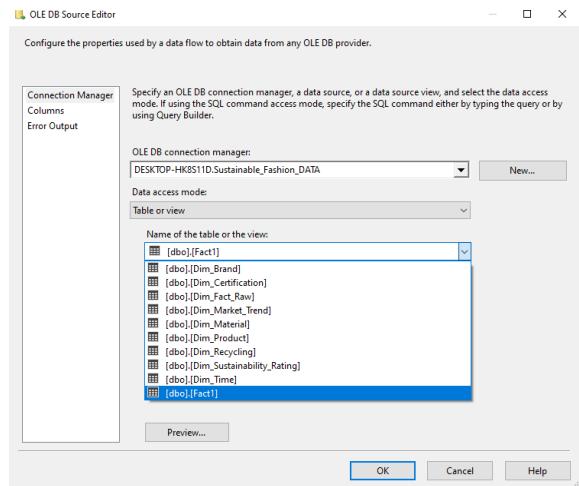
Hình 2. 109. Tạo Data Flow Task có tên "Merge Fact1 and Dim_Material to Fact2".

Bước 2: Vào “Merge Fact1 and Dim_Material to Fact2”, tạo 2 “OLE DB Source” là Fact1 và Dim_Material.



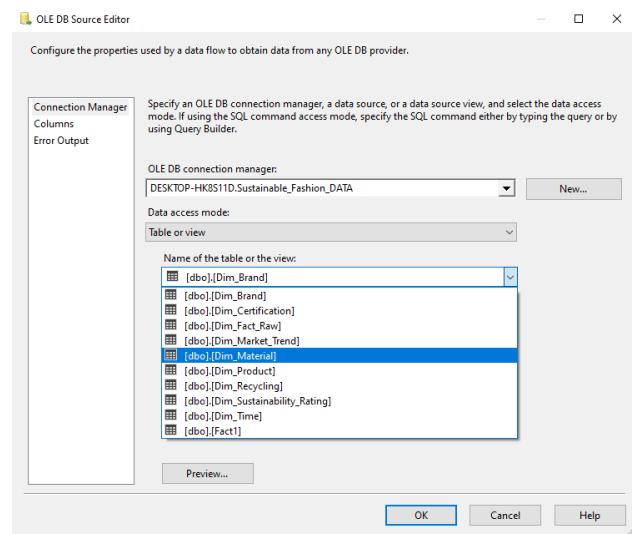
Hình 2. 110. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact1”, chọn “Edit”. Chọn dữ liệu bảng “Fact1” đã đỗ trong mục “Name of the table or the view”.



Hình 2. 111. Các bước thực hiện (2).

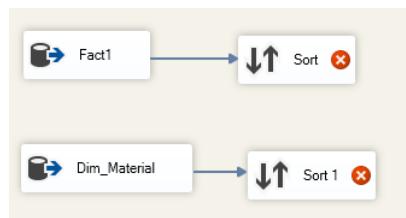
Thực hiện tương tự với bảng Dim_Material



Hình 2. 112. Các bước thực hiện (3).

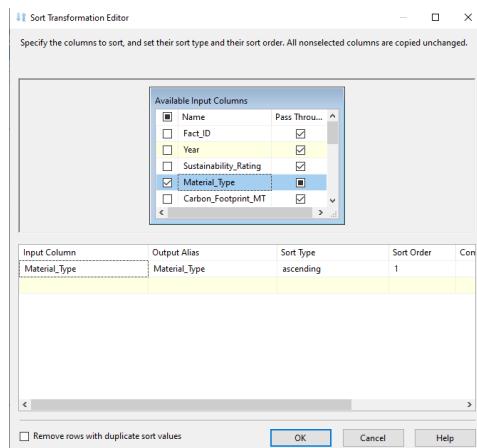
Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact1” và “Dim_Material” , lấy ra thuộc

tính để tiến hành kết hai bảng.



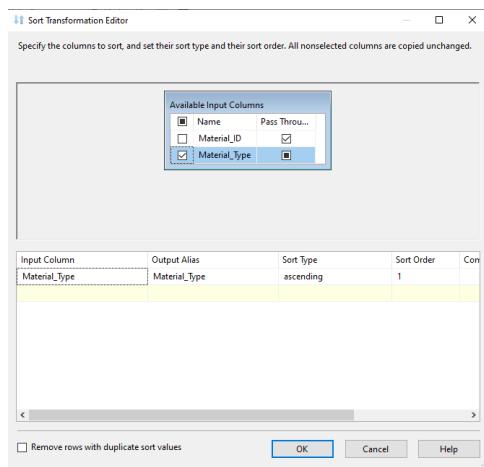
Hình 2. 113. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact1” sẽ lấy ra thuộc tính “Material_Type”.



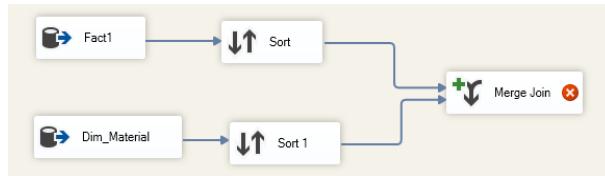
Hình 2. 114. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Material” sẽ lấy ra thuộc tính “Material_Type”.



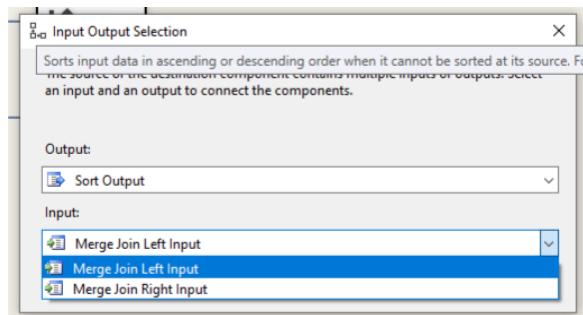
Hình 2. 115. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



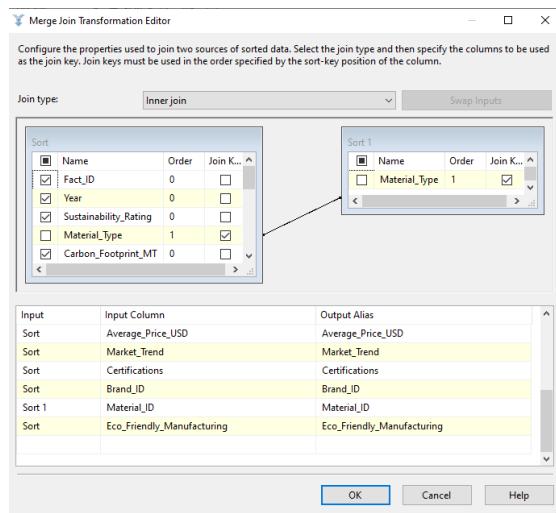
Hình 2. 116. Các bước thực hiện (7).

Kết nối “Sort” của “Fact1” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact1”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Material”.



Hình 2. 117. Các bước thực hiện (8).

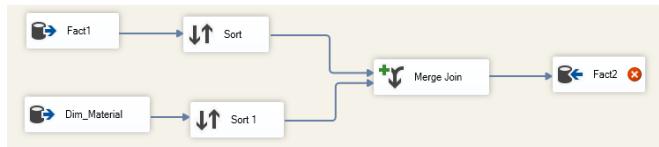
Kết nối “Sort 1” của “Dim_Material” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 118. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact2 vào Sustainable_Fashion_DATA. Đầu tiên

chọn một “OLE DB Destination”, đặt tên thành “Fact2”.



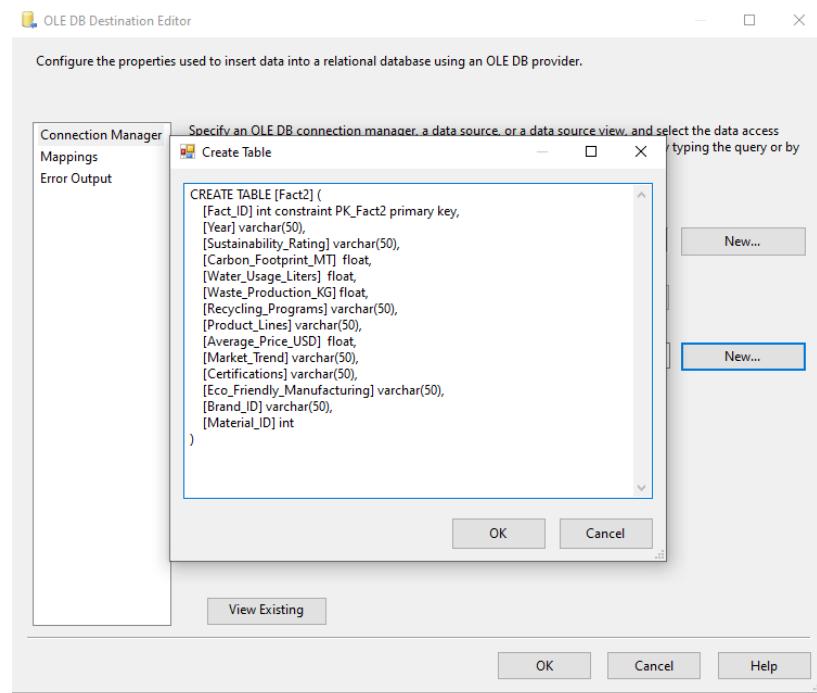
Hình 2. 119. Các bước thực hiện (10).

Bước 7: Nhấn chuột phải vào “Fact2”, chọn “Edit”. Chọn “New” để tạo bảng “Fact” mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

```

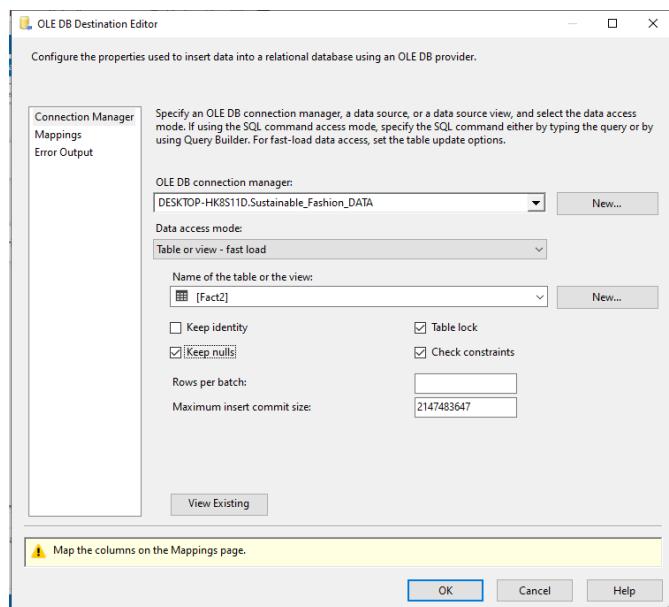
CREATE TABLE [Fact2] (
    [Fact_ID] int constraint PK_Fact2 primary key,
    [Year] varchar(50),
    [Sustainability_Rating] varchar(50),
    [Carbon_Footprint_MT] float,
    [Water_Usage_Liters] float,
    [Waste_Production_KG] float,
    [Recycling_Programs] varchar(50),
    [Product_Lines] varchar(50),
    [Average_Price_USD] float,
    [Market_Trend] varchar(50),
    [Certifications] varchar(50),
    [Eco_Friendly_Manufacturing] varchar(50),
    [Brand_ID] varchar(50),
    [Material_ID] int
)
  
```

IS217 – Kho dữ liệu và OLAP



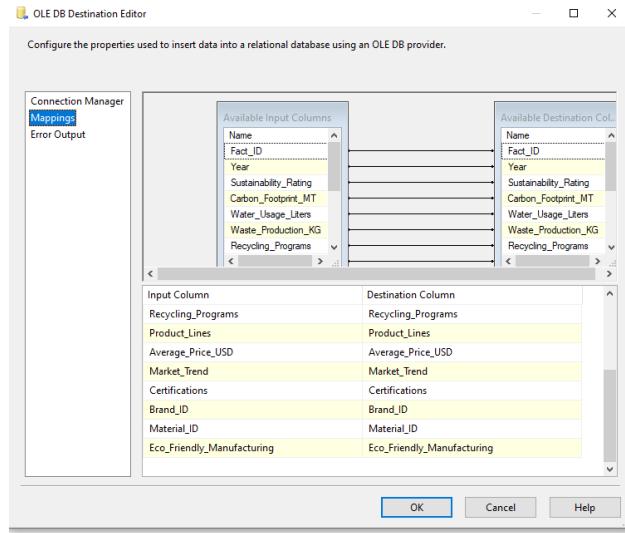
Hình 2. 120. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

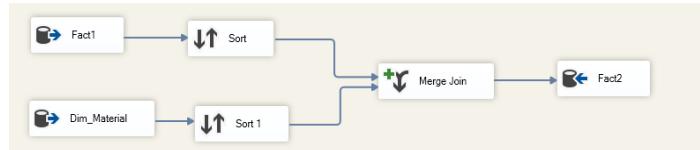


Hình 2. 121. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



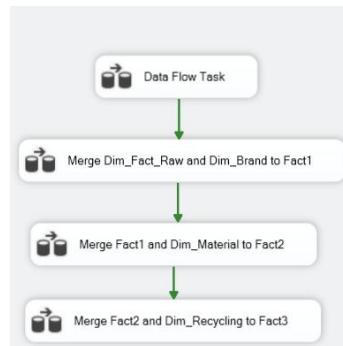
Hình 2. 122. Các bước thực hiện (13).



Hình 2. 123. Sau khi thực hiện xong.

2.2.13 Merge Fact2 và Dim_Recycling thành Fact3

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact2 và Dim_Recycling thành Fact3”.



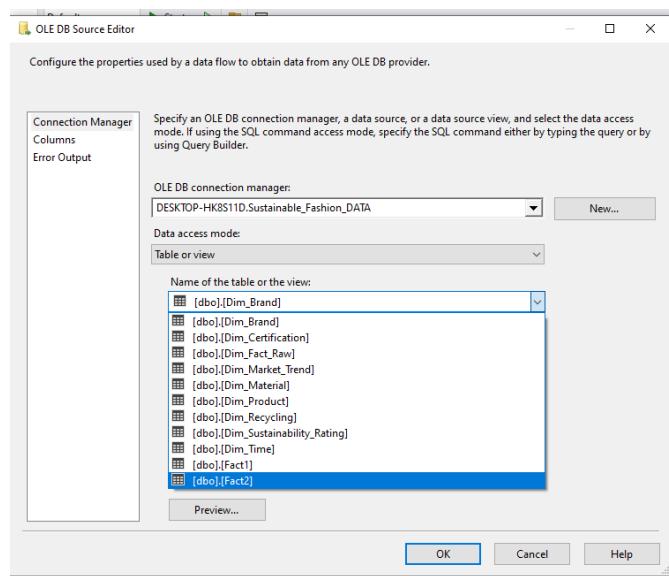
Hình 2. 124. Tạo Data Flow Task có tên "Merge Fact2 and Dim_Recycling to Fact3".

Bước 2: Vào “Merge Fact2 and Dim_Recycling to Fact3”, tạo 2 “OLE DB Source” là Fact2 và Dim_Recycling.



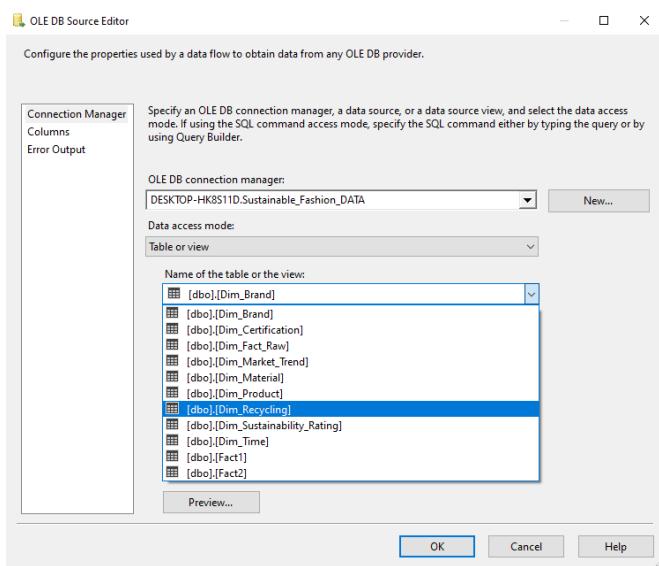
Hình 2. 125. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact2”, chọn “Edit”. Chọn dữ liệu bảng “Fact2” đã đỗ trong mục “Name of the table or the view”.



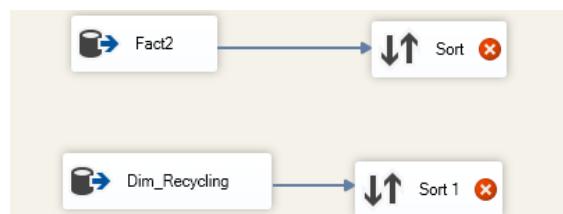
Hình 2. 126. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Recycling.



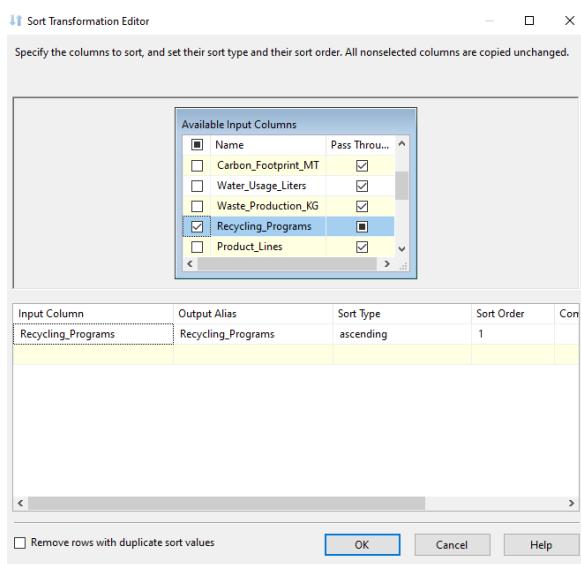
Hình 2. 127. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact2” và “Dim_Recycling, lấy ra thuộc tính để tiến hành kết hai bảng.



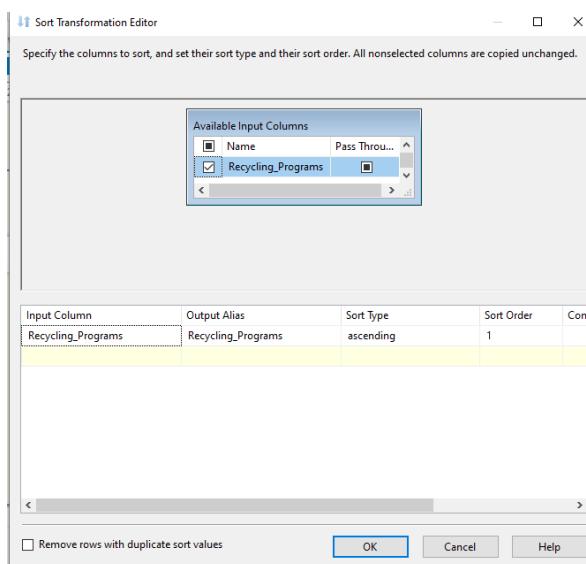
Hình 2. 128. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact2” sẽ lấy ra thuộc tính “Recycling_Programs”.



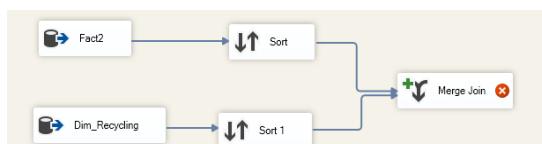
Hình 2. 129. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Recycling” sẽ lấy ra thuộc tính “Recycling_Programs”.



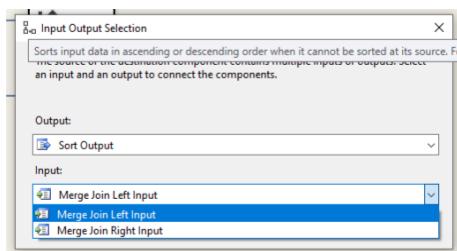
Hình 2. 130. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



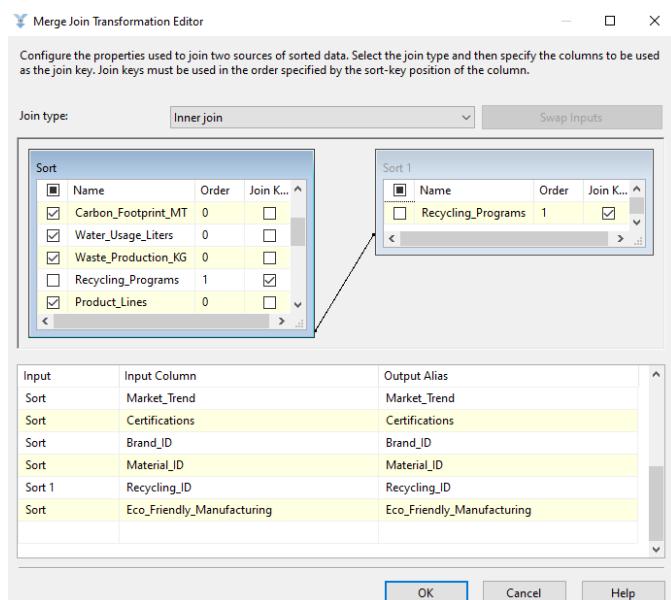
Hình 2. 131. Các bước thực hiện (7).

Kết nối “Sort” của “Fact2” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact1”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Recycling”.



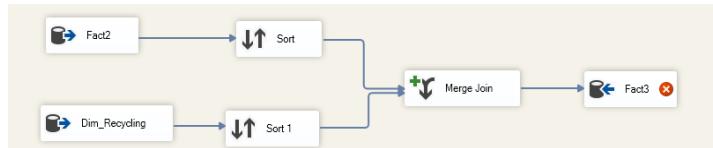
Hình 2. 132. Các bước thực hiện (8).

Kết nối “Sort 1” của “Dim_Recycling” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 133. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact3 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact3”.



Hình 2. 134. Các bước thực hiện (10).

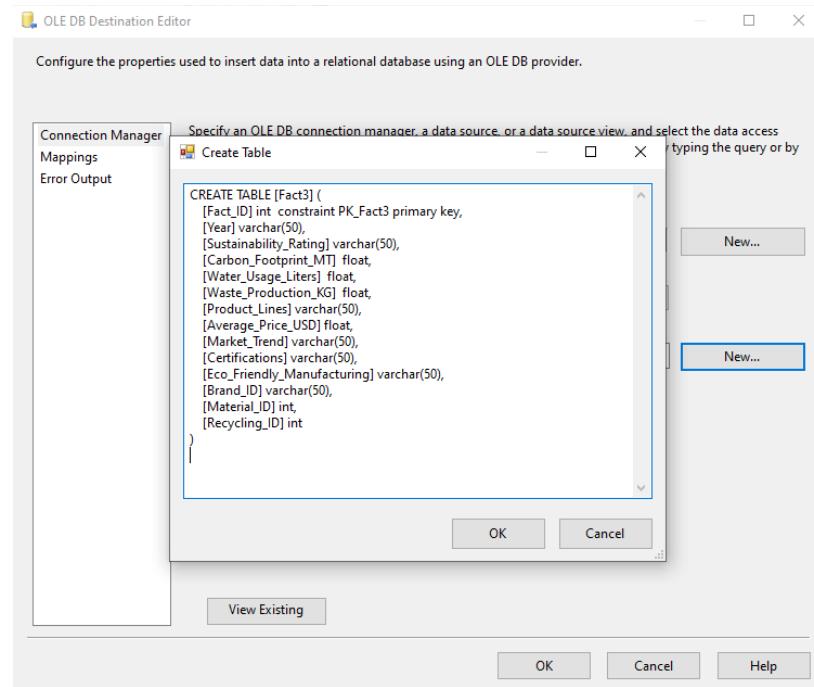
Bước 7: Nhấn chuột phải vào “Fact3”, chọn “Edit”. Chọn “New” để tạo bảng “Fact3” mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

CREATE TABLE [Fact3] (

```
[Fact_ID] int constraint PK_Fact3 primary key,  
[Year] varchar(50),  
[Sustainability_Rating] varchar(50),  
[Carbon_Footprint_MT] float,  
[Water_Usage_Liters] float,  
[Waste_Production_KG] float,  
[Product_Lines] varchar(50),  
[Average_Price_USD] float,  
[Market_Trend] varchar(50),  
[Certifications] varchar(50),  
[Eco_Friendly_Manufacturing] varchar(50),  
[Brand_ID] varchar(50),  
[Material_ID] int,  
[Recycling_ID] int
```

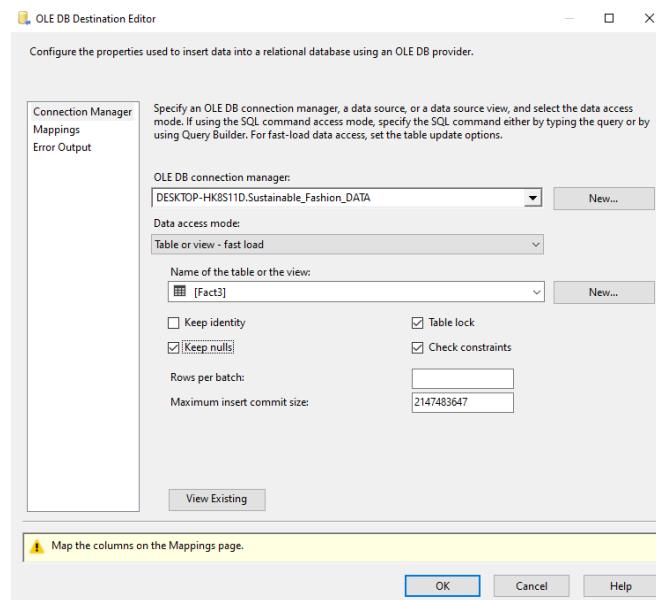
)

IS217 – Kho dữ liệu và OLAP



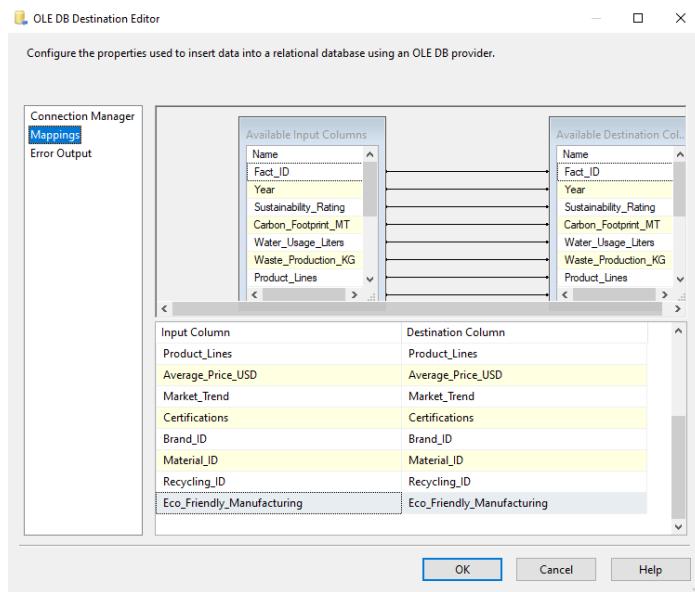
Hình 2. 135. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

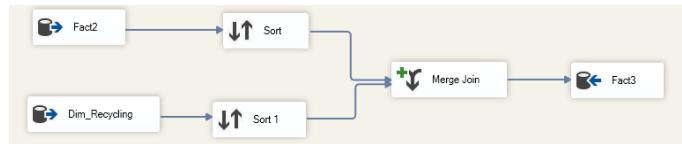


Hình 2. 136. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



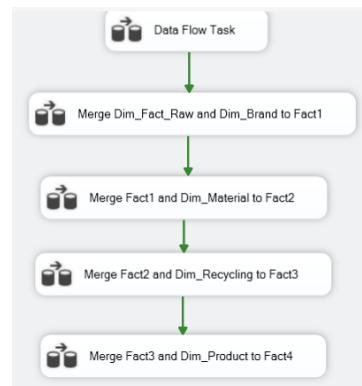
Hình 2. 137. Các bước thực hiện (13).



Hình 2. 138. Sau khi thực hiện xong.

2.2.14 Merge Fact3 và Dim_Product thành Fact4

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact3 và Dim_Product thành Fact4”.



Hình 2. 139. Tạo Data Flow Task có tên "Merge Fact3 and Dim_Product to Fact4".

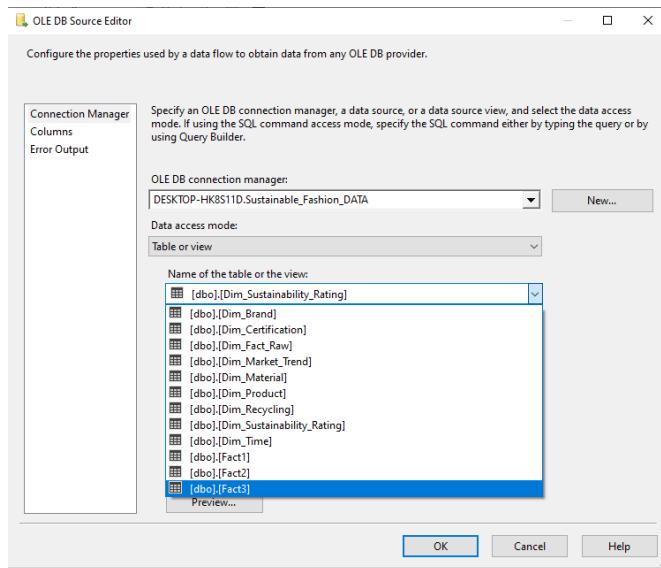
Bước 2: Vào “Merge Fact3 and Dim_Product to Fact4”, tạo 2 “OLE DB Source” là Fact3

và Dim_Product.



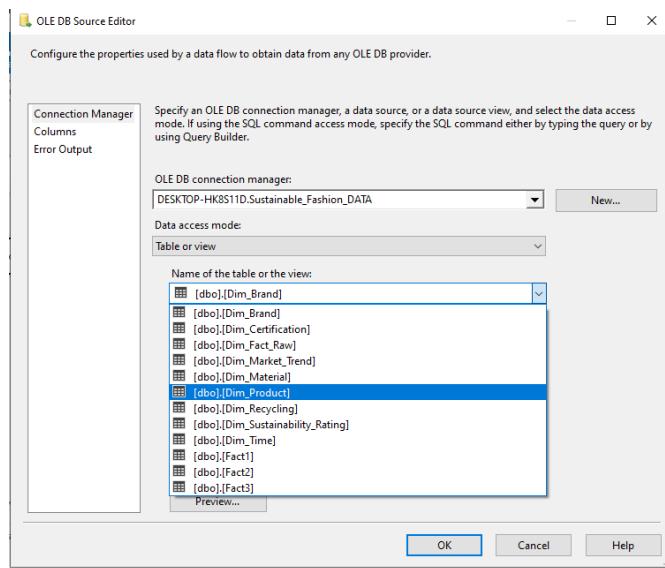
Hình 2. 140. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact3”, chọn “Edit”. Chọn dữ liệu bảng “Fact3” đã đỗ trong mục “Name of the table or the view”.



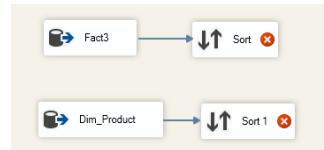
Hình 2. 141. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Product.



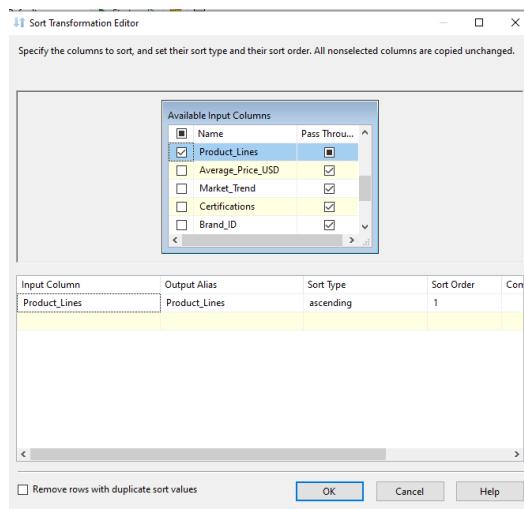
Hình 2. 142. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact3” và “Dim_Product”, lấy ra thuộc tính để tiến hành kết hai bảng.



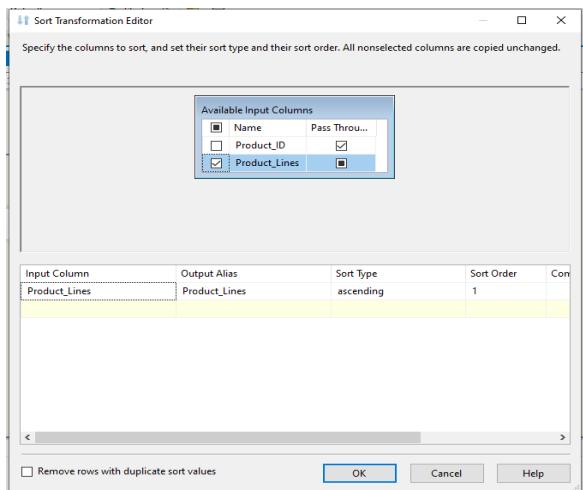
Hình 2. 143. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact3” sẽ lấy ra thuộc tính “Product_Lines”.



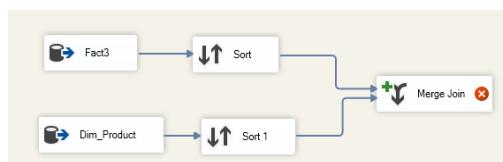
Hình 2. 144. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Product” sẽ lấy ra thuộc tính “Product_Lines”.



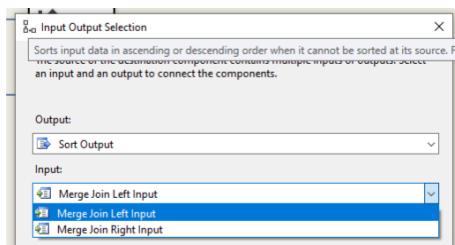
Hình 2. 145. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



Hình 2. 146. Các bước thực hiện (7).

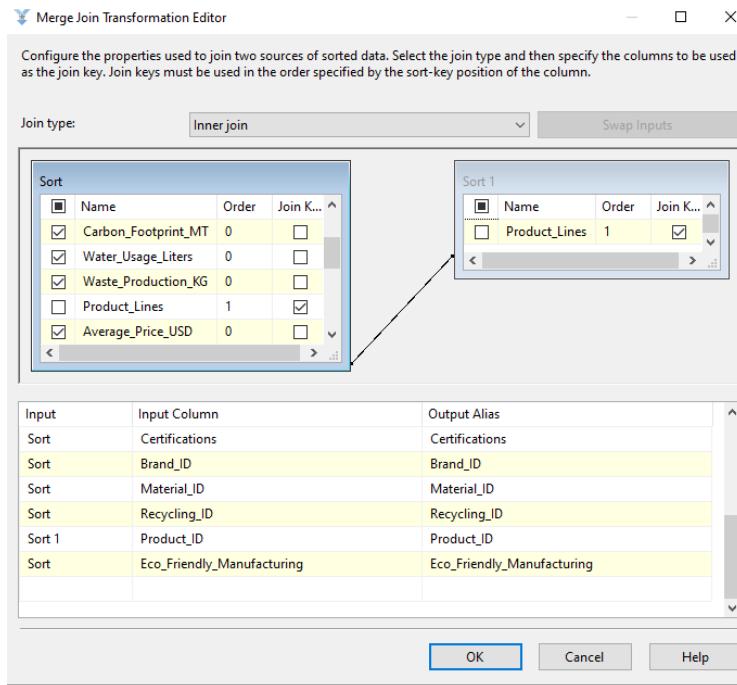
Kết nối “Sort” của “Fact3” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact1”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Recycling”.



Hình 2. 147. Các bước thực hiện (8).

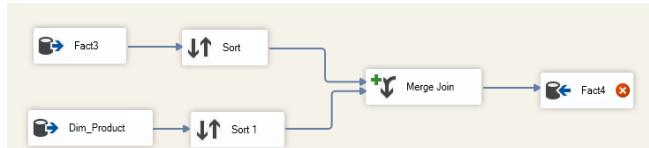
Kết nối “Sort 1” của “Dim_Product” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai

bảng.



Hình 2. 148. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact4 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact4”.



Hình 2. 149. Các bước thực hiện (10).

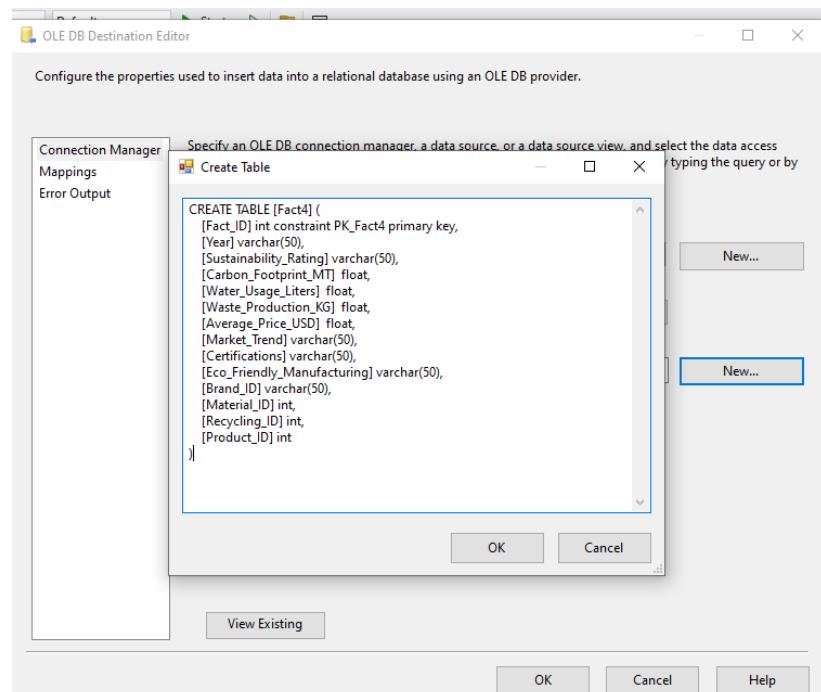
Bước 7: Nhấn chuột phải vào “Fact4”, chọn “Edit”. Chọn “New” để tạo bảng “Fact4” mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

CREATE TABLE [Fact4] (

```

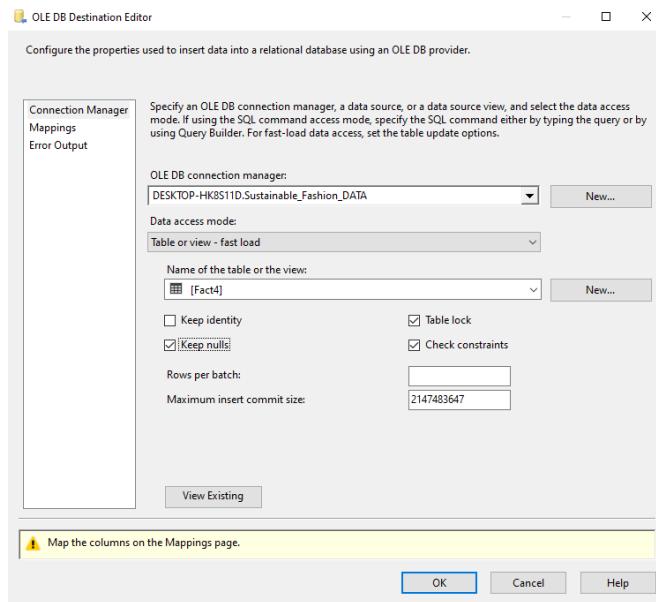
    [Fact_ID] int constraint PK_Fact4 primary key,
    [Year] varchar(50),
    [Sustainability_Rating] varchar(50),
    [Carbon_Footprint_MT] float,
  
```

```
[Water_Usage_Liters] float,  
[Waste_Production_KG] float,  
[Average_Price_USD] float,  
[Market_Trend] varchar(50),  
[Certifications] varchar(50),  
[Eco_Friendly_Manufacturing] varchar(50),  
[Brand_ID] varchar(50),  
[Material_ID] int,  
[Recycling_ID] int,  
[Product_ID] int  
)
```



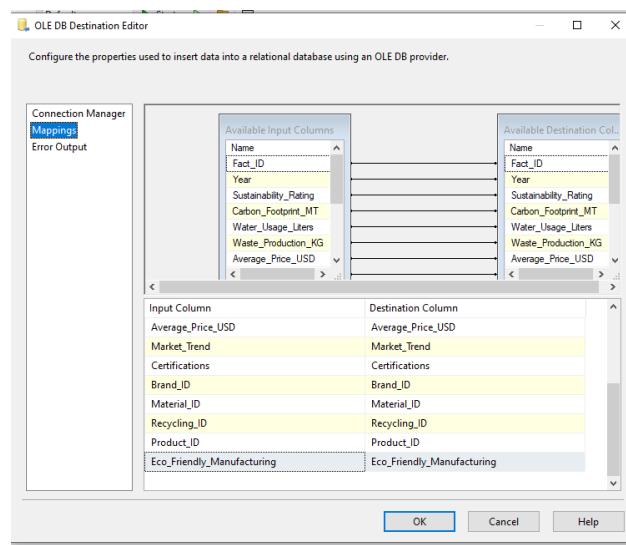
Hình 2. 150. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

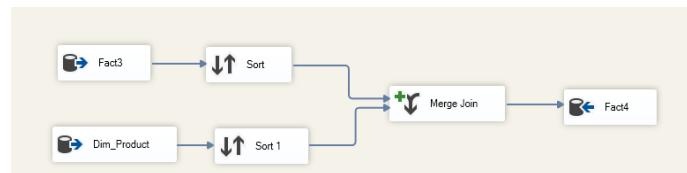


Hình 2. 151. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



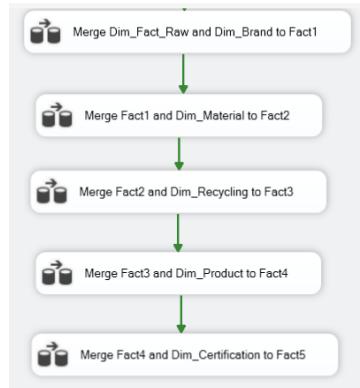
Hình 2. 152. Các bước thực hiện (13).



Hình 2. 153. Sau khi thực hiện xong.

2.2.15 Merge Fact4 và Dim_Certification thành Fact5

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact4 and Dim_Certification to Fact5”.



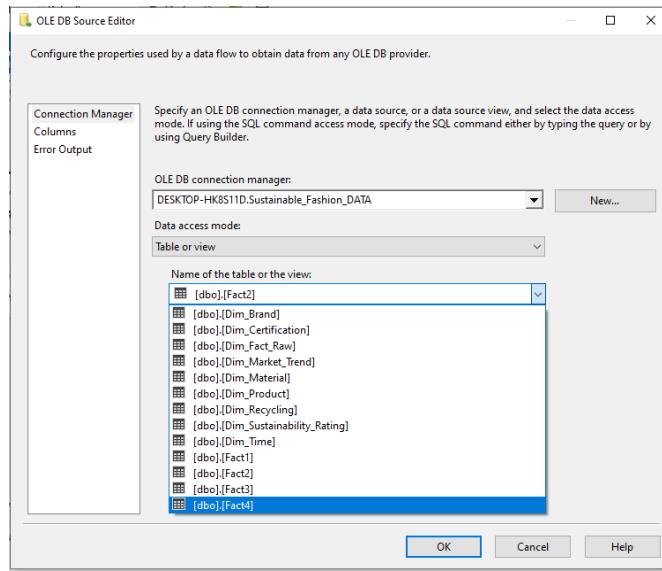
Hình 2. 154. Tạo Data Flow Task có tên "Merge Fact4 and Dim_Certification to Fact5".

Bước 2: Vào “Merge Fact4 and Dim_Certification to Fact5”, tạo 2 “OLE DB Source” là Fact4 và Dim_Certification.



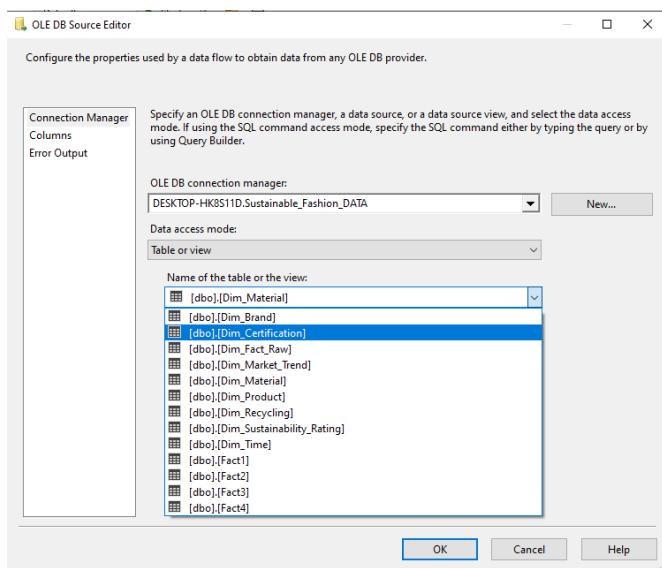
Hình 2. 155. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact4”, chọn “Edit”. Chọn dữ liệu bảng “Fact4” đã đỗ trong mục “Name of the table or the view”.



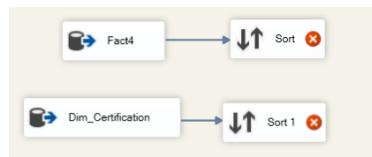
Hình 2. 156. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Certification.



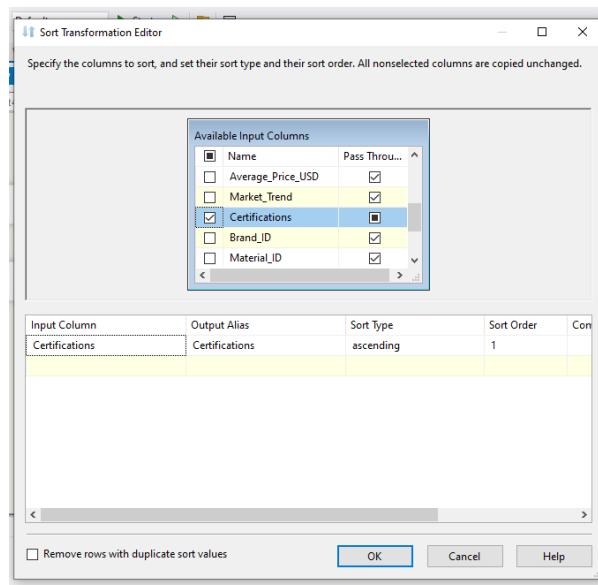
Hình 2. 157. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact4” và “Dim_Certification”, lấy ra thuộc tính để tiến hành kết hai bảng.



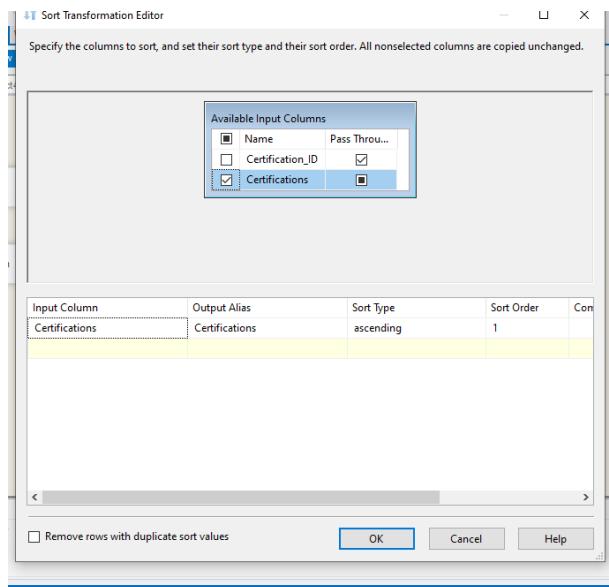
Hình 2. 158. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact4” sẽ lấy ra thuộc tính “Certifications”.



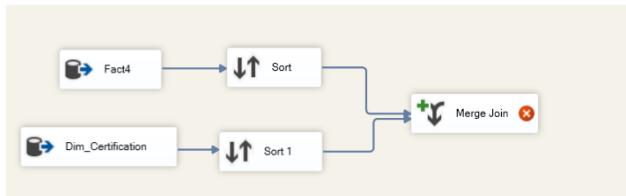
Hình 2. 159. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Certification” sẽ lấy ra thuộc tính “Certifications”.



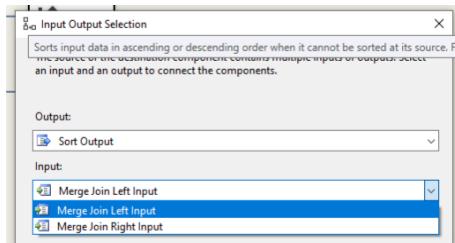
Hình 2. 160. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



Hình 2. 161. Các bước thực hiện (7).

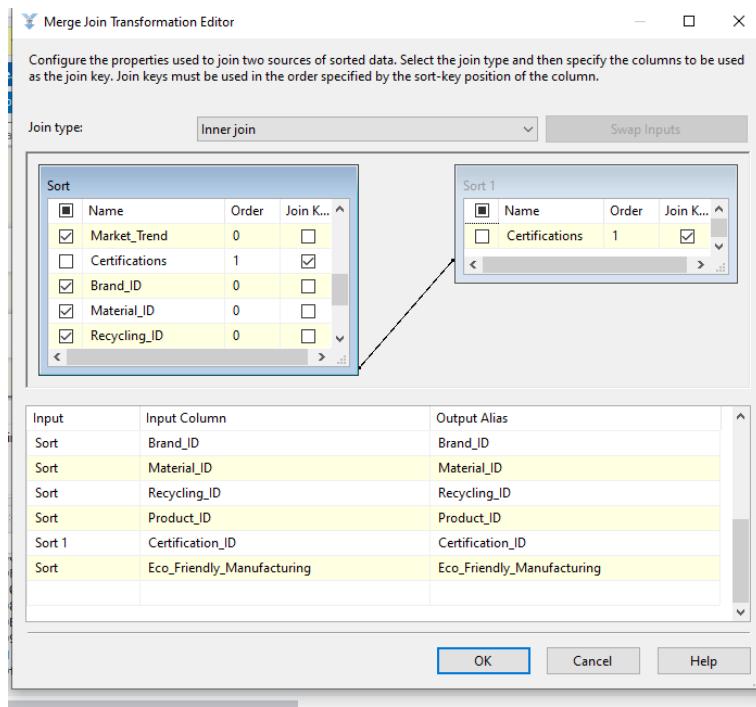
Kết nối “Sort” của “Fact4” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Certification”.



Hình 2. 162. Các bước thực hiện (8).

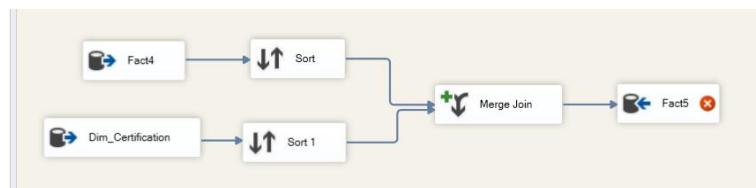
Kết nối “Sort 1” của “Dim_Certification” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết

nối hai bảng.



Hình 2. 163. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact5 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact5”.



Hình 2. 164. Các bước thực hiện (10).

Bước 7: Nhấn chuột phải vào “Fact5”, chọn “Edit”. Chọn “New” để tạo bảng Fact5 mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

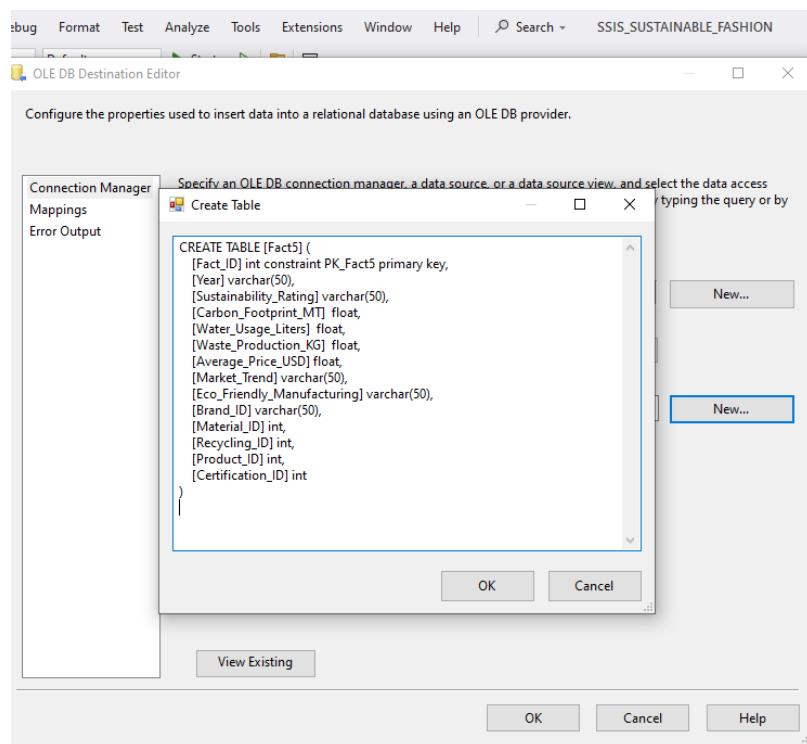
CREATE TABLE [Fact5] (

```

    [Fact_ID] int constraint PK_Fact5 primary key,
    [Year] varchar(50),
  
```

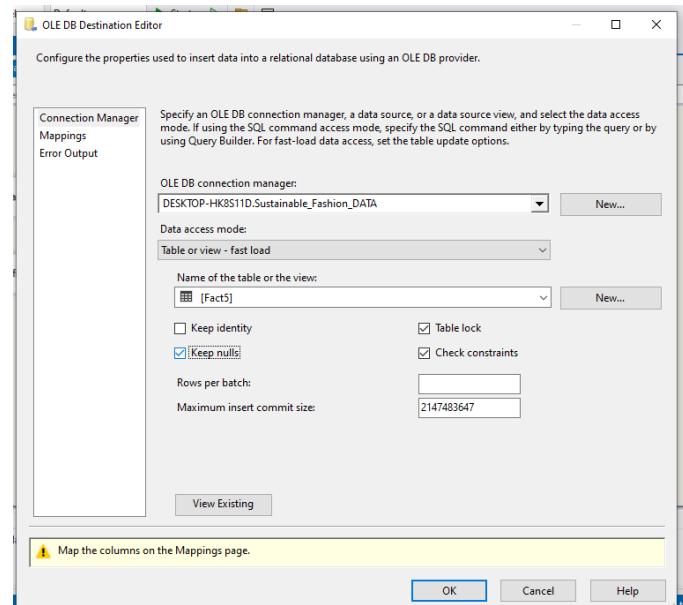
```
[Sustainability_Rating] varchar(50),
[Carbon_Footprint_MT] float,
[Water_Usage_Liters] float,
[Waste_Production_KG] float,
[Average_Price_USD] float,
[Market_Trend] varchar(50),
[Eco_Friendly_Manufacturing] varchar(50),
[Brand_ID] varchar(50),
[Material_ID] int,
[Recycling_ID] int,
[Product_ID] int,
[Certification_ID] int
```

)



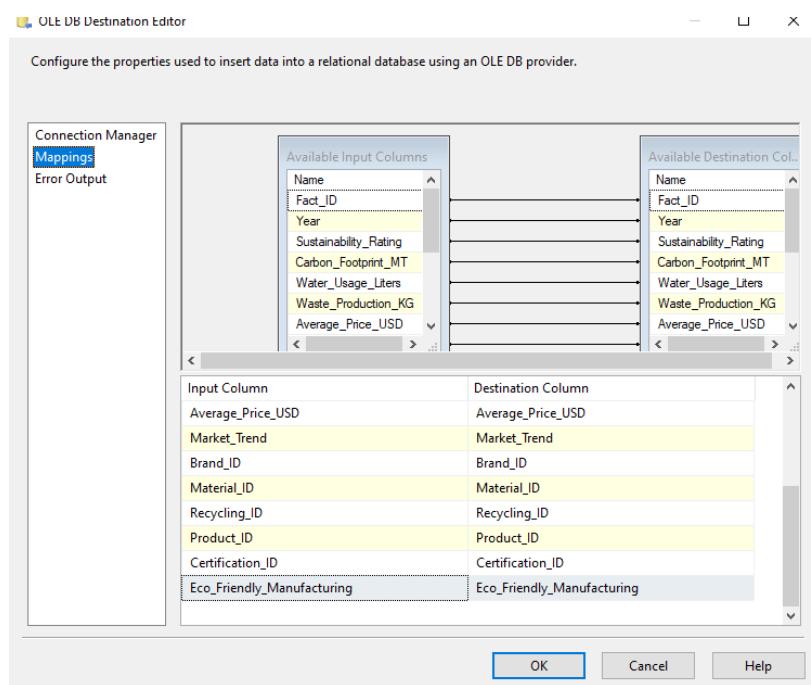
Hình 2. 165. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

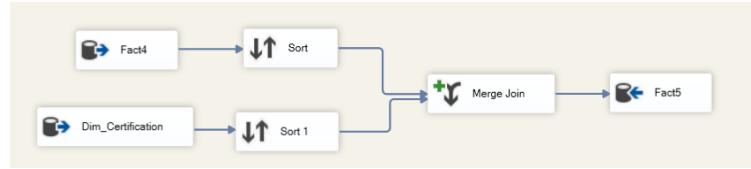


Hình 2. 166. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 167. Các bước thực hiện (13).



Hình 2. 168. Sau khi thực hiện xong.

2.2.16 Merge Fact5 và Dim_Market_Trend thành Fact6

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact5 and Dim_Market_Trend to Fact6”.



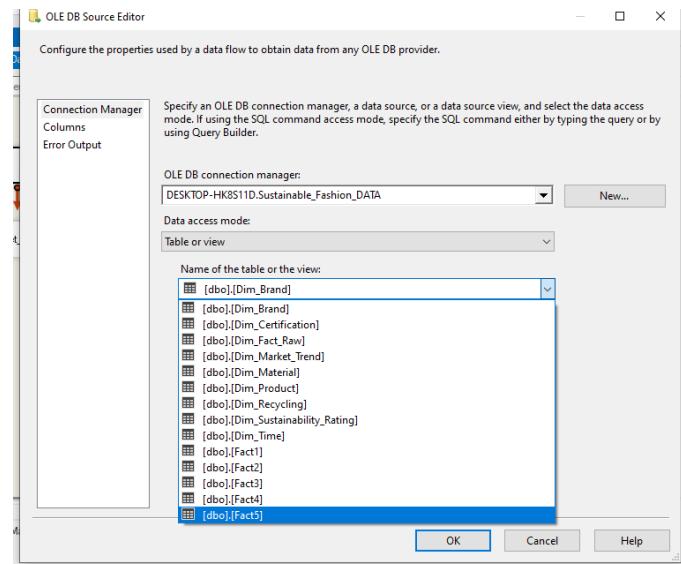
Hình 2. 169. Tạo Data Flow Task có tên "Merge Fact5 and Dim_Market_Trend to Fact6".

Bước 2: Vào “Merge Fact5 and Dim_Market_Trend to Fact6”, tạo 2 “OLE DB Source” là Fact5 và Dim_Market_Trend.



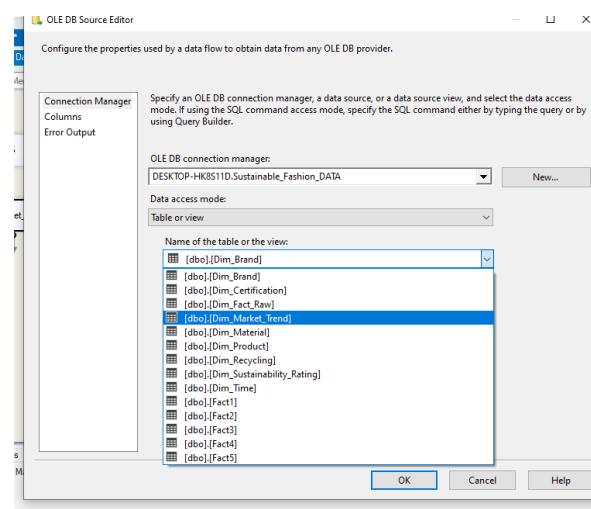
Hình 2. 170. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact5”, chọn “Edit”. Chọn dữ liệu bảng “Fact5” đã đỗ trong mục “Name of the table or the view”.



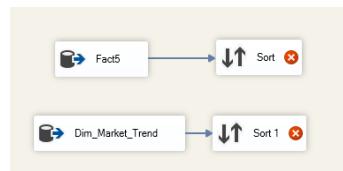
Hình 2. 171. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Market_Trend.



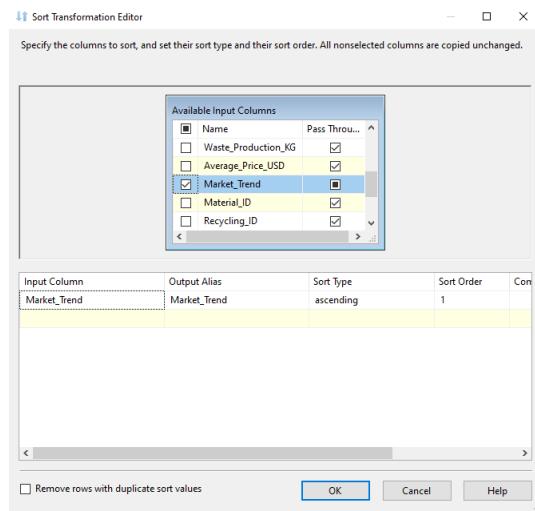
Hình 2. 172. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact5” và “Dim_Market_Trend, lấy ra thuộc tính để tiến hành kết hai bảng.



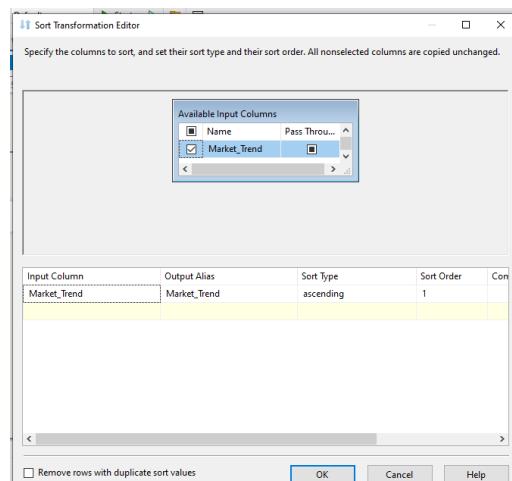
Hình 2. 173. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact5” sẽ lấy ra thuộc tính “Market_Trend”.



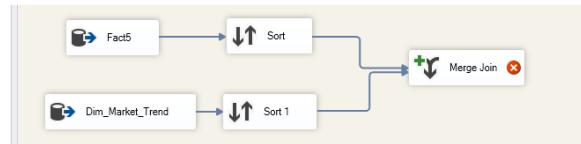
Hình 2. 174. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Market_Trend” sẽ lấy ra thuộc tính “Market_Trend”.



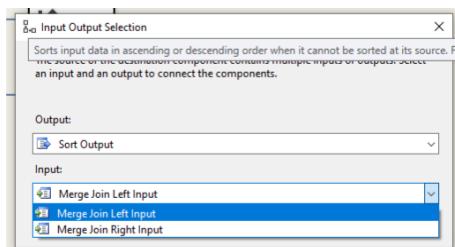
Hình 2. 175. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



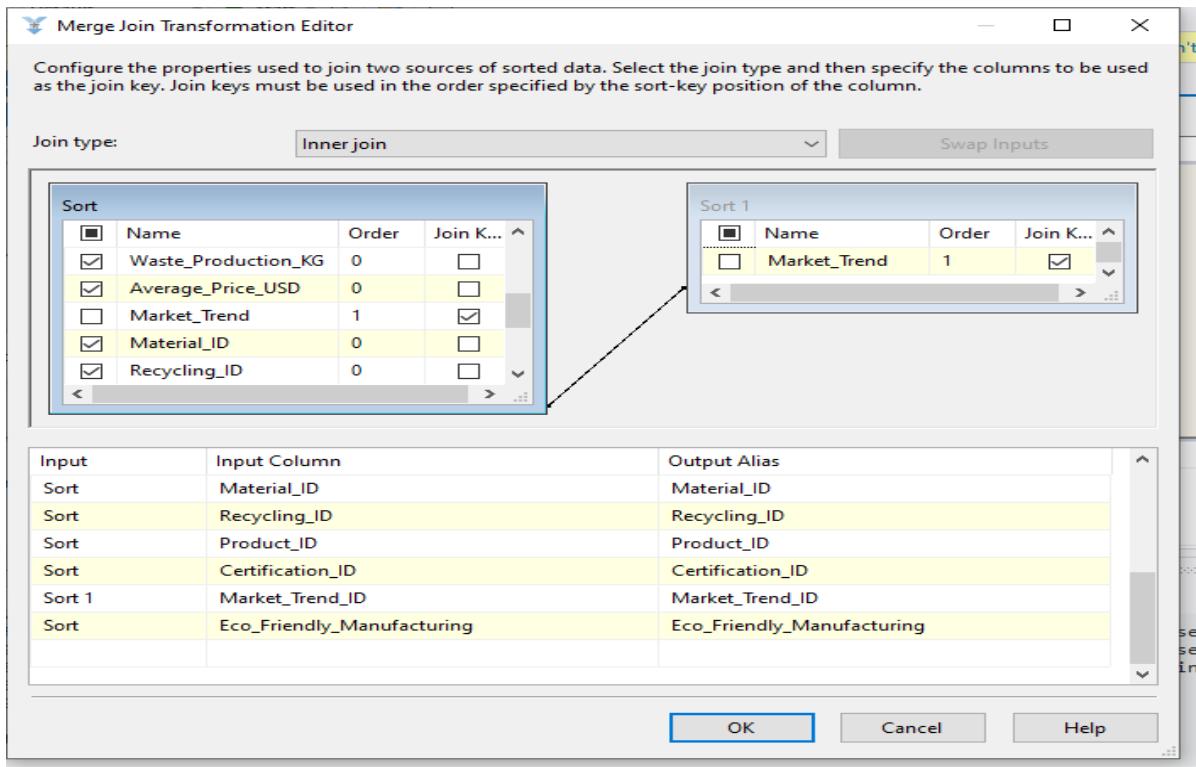
Hình 2. 176. Các bước thực hiện (7).

Kết nối “Sort” của “Fact5” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact5”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Market_Trend”.



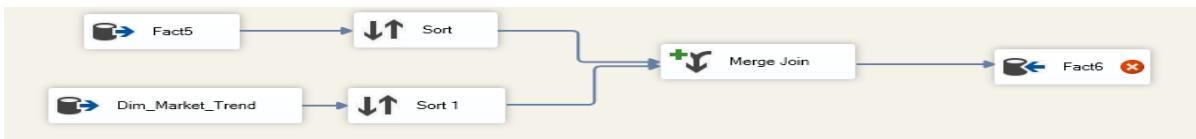
Hình 2. 177. Các bước thực hiện (8).

Kết nối “Sort 1” của “Dim_Market_Trend” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 178. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact6 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact6”.



Hình 2. 179. Các bước thực hiện (10).

Bước 7: Nhấn chuột phải vào “Fact6”, chọn “Edit”. Chọn “New” để tạo bảng Fact6 mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

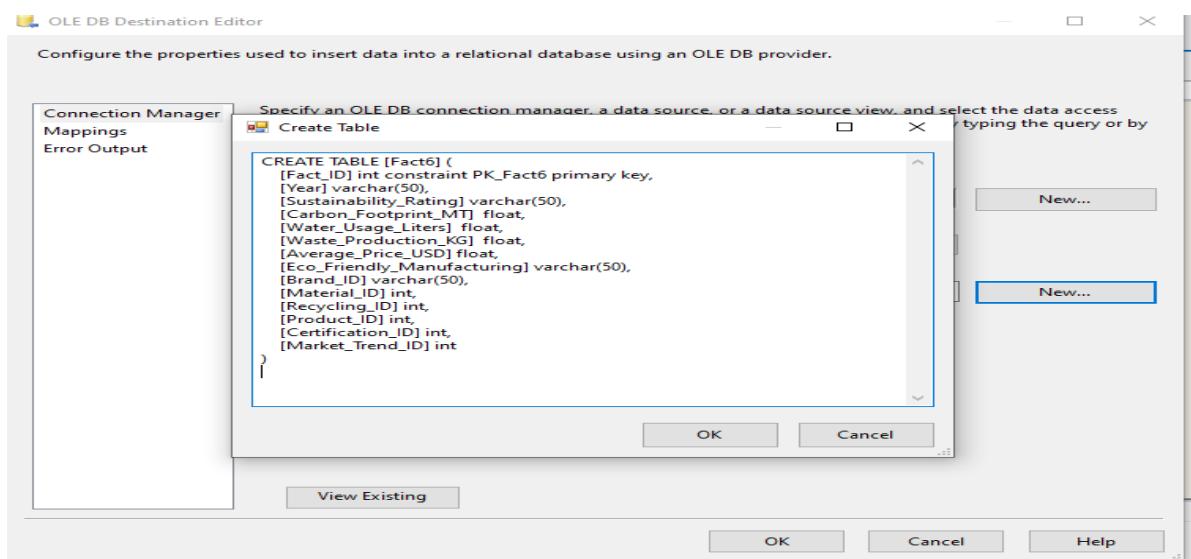
CREATE TABLE [Fact6] (

```

[Fact_ID] int constraint PK_Fact6 primary key,
[Year] varchar(50),
[Sustainability_Rating] varchar(50),
[Carbon_Footprint_MT] float,
[Water_Usage_Liters] float,
  
```

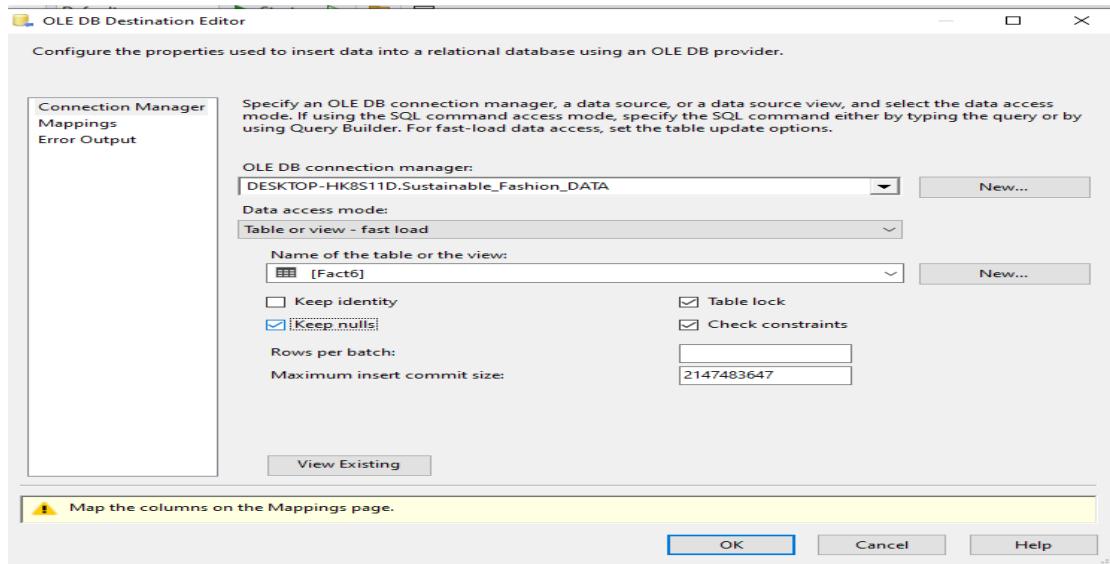
```
[Waste_Production_KG] float,  
[Average_Price_USD] float,  
[Eco_Friendly_Manufacturing] varchar(50),  
[Brand_ID] varchar(50),  
[Material_ID] int,  
[Recycling_ID] int,  
[Product_ID] int,  
[Certification_ID] int,  
[Market_Trend_ID] int
```

)



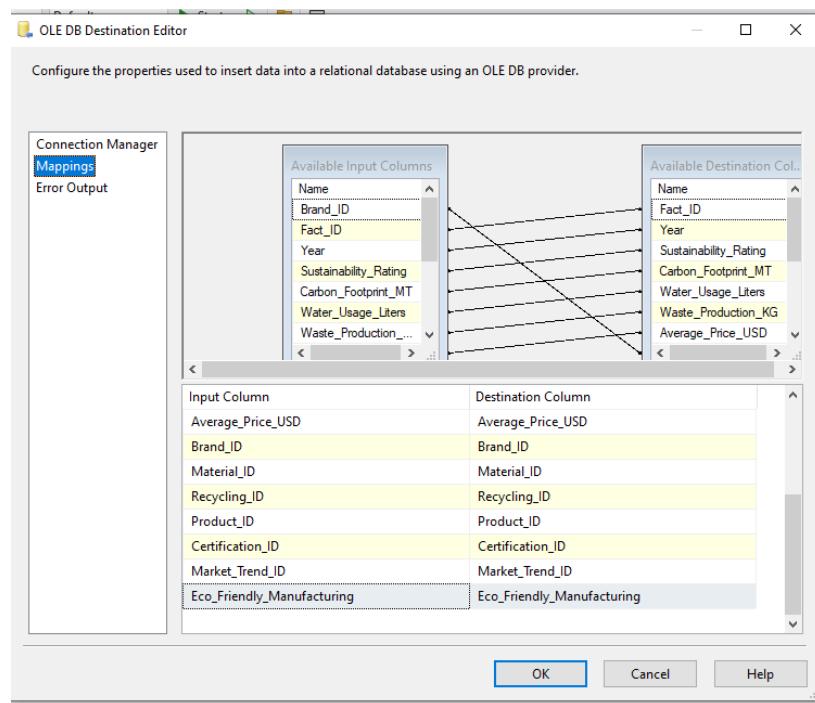
Hình 2. 180. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

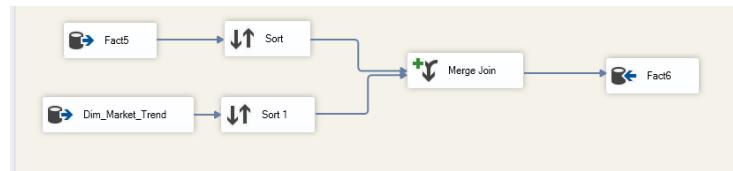


Hình 2. 181. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



Hình 2. 182. Các bước thực hiện (13).



Hình 2. 182. Sau khi thực hiện xong.

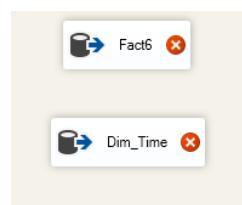
2.2.17 Merge Fact6 và Dim_Time thành Fact7

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact6 and Dim_Time to Fact7”.



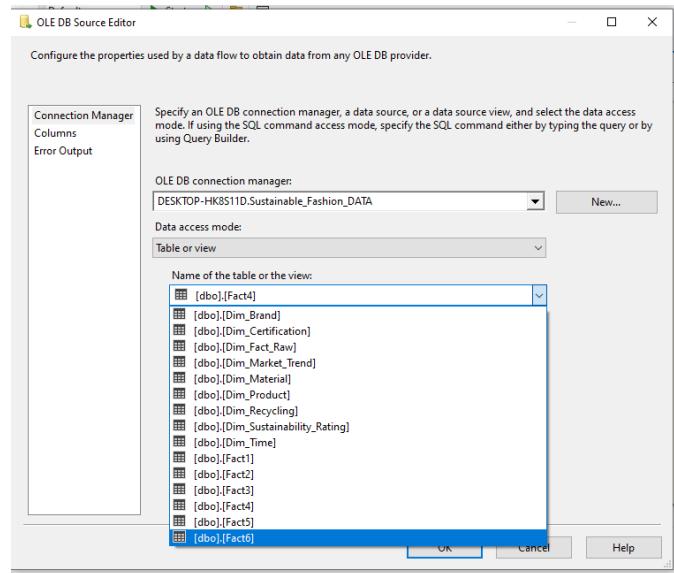
Hình 2. 183. Tạo Data Flow Task có tên "Merge Fact6 and Dim_Dim_Time to Fact7".

Bước 2: Vào “Merge Fact6 and Dim_Time to Fact7”, tạo 2 “OLE DB Source” là Fact6 và Dim_Time.



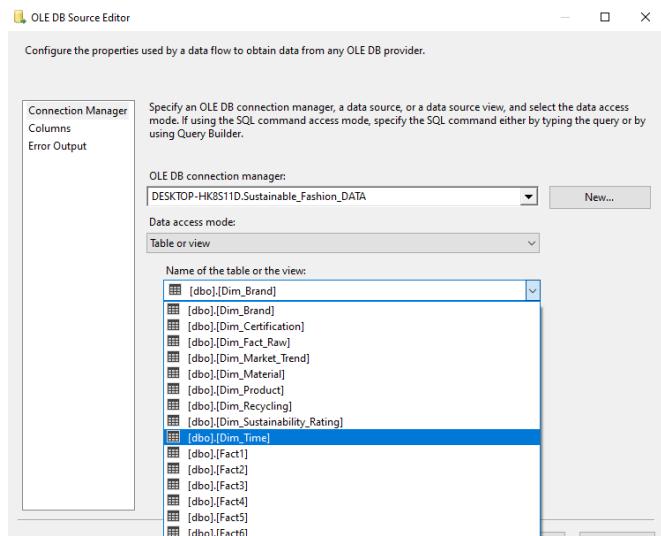
Hình 2. 184. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact6”, chọn “Edit”. Chọn dữ liệu bảng “Fact6” đã đỗ trong mục “Name of the table or the view”.



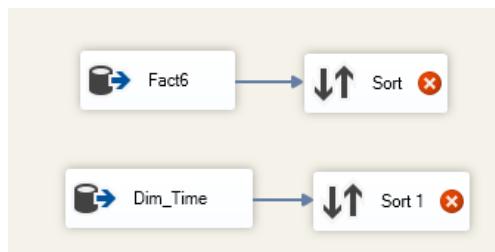
Hình 2. 185. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Time.



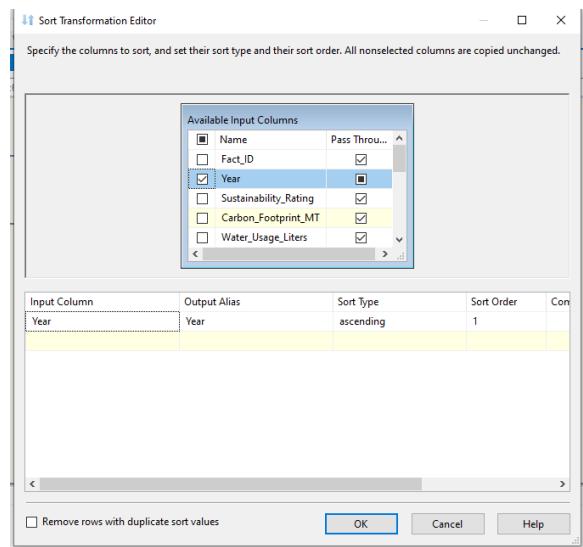
Hình 2. 186. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact6” và “Dim_Time lấy ra thuộc tính để tiến hành kết hai bảng.



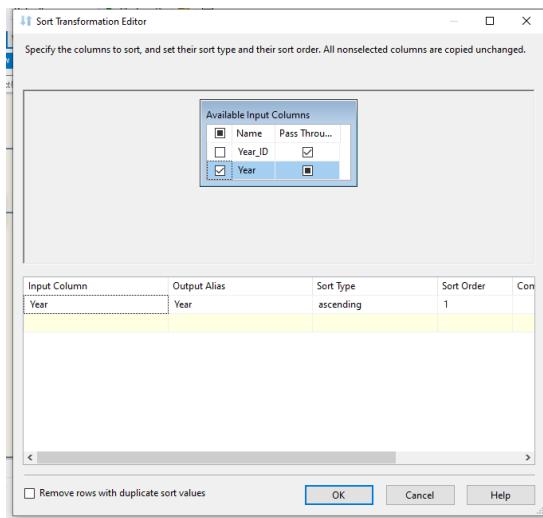
Hình 2. 187. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact6” sẽ lấy ra thuộc tính “Year”.



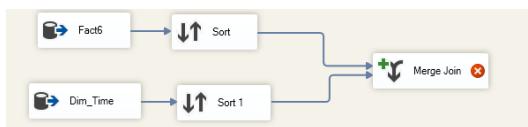
Hình 2. 188. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Time” sẽ lấy ra thuộc tính “Year”.



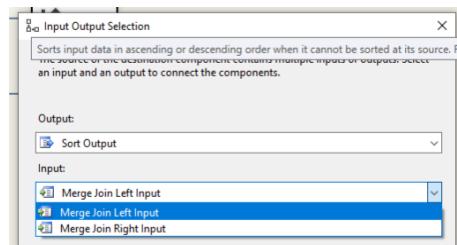
Hình 2. 189. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



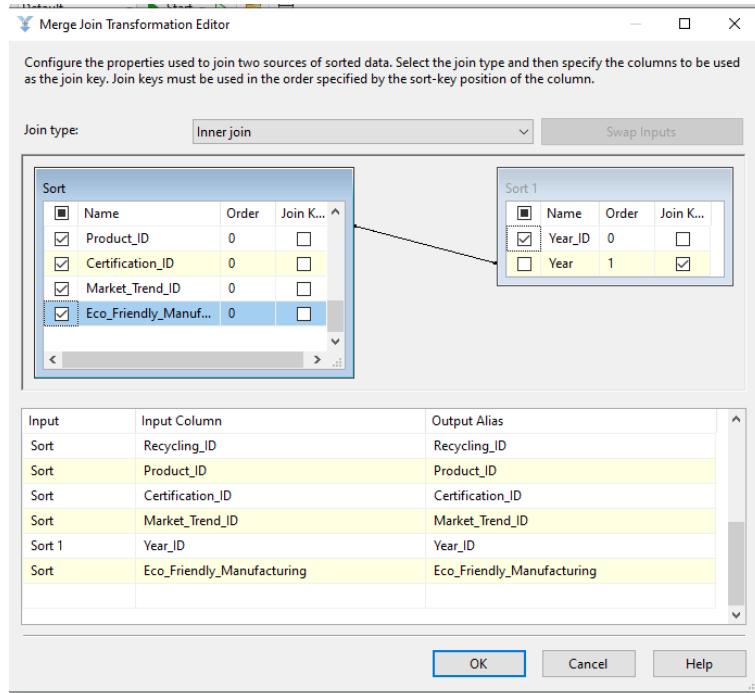
Hình 2. 190. Các bước thực hiện (7).

Kết nối “Sort” của “Fact6” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact5”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Time”.



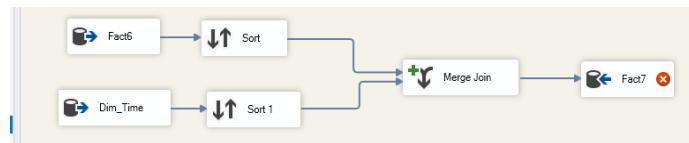
Hình 2. 191. Các bước thực hiện (8).

Kết nối “Sort 1” của “Dim_Time” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 192. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact7 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact7”.



Hình 2. 193. Các bước thực hiện (10).

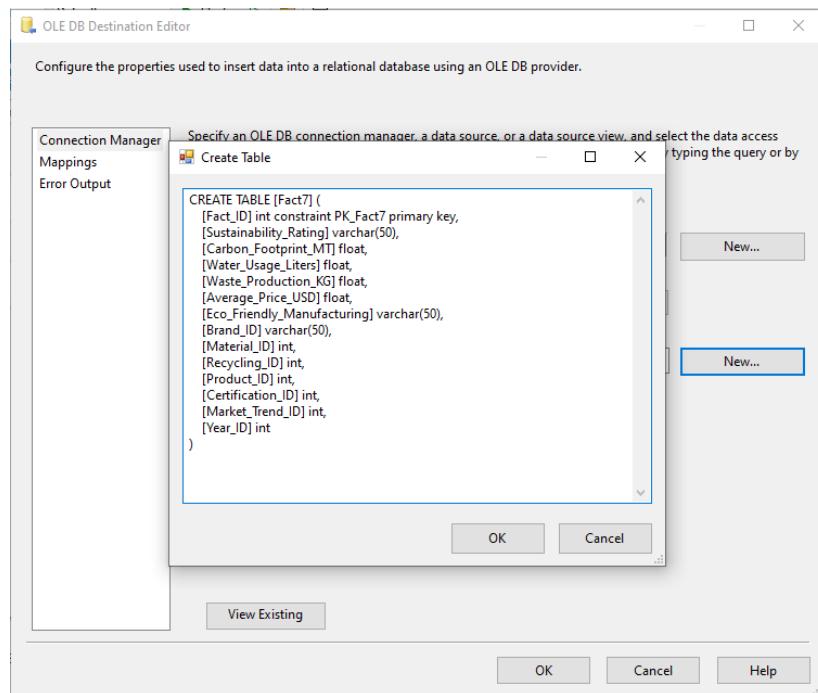
Bước 7: Nhấn chuột phải vào “Fact7”, chọn “Edit”. Chọn “New” để tạo bảng Fact” mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

CREATE TABLE [Fact7] (

```

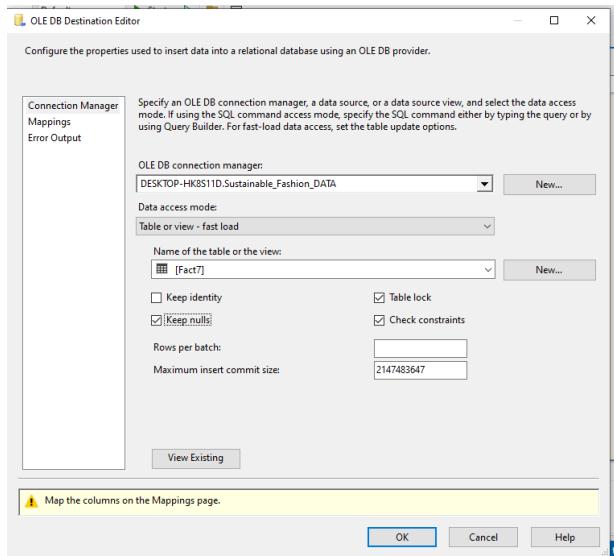
[Fact_ID] int constraint PK_Fact7 primary key,
[Sustainability_Rating] varchar(50),
[Carbon_Footprint_MT] float,
[Water_Usage_Liters] float,
[Waste_Production_KG] float,
[Average_Price_USD] float,
  
```

```
[Eco_Friendly_Manufacturing] varchar(50),  
[Brand_ID] varchar(50),  
[Material_ID] int,  
[Recycling_ID] int,  
[Product_ID] int,  
[Certification_ID] int,  
[Market_Trend_ID] int,  
[Year_ID] int  
)
```



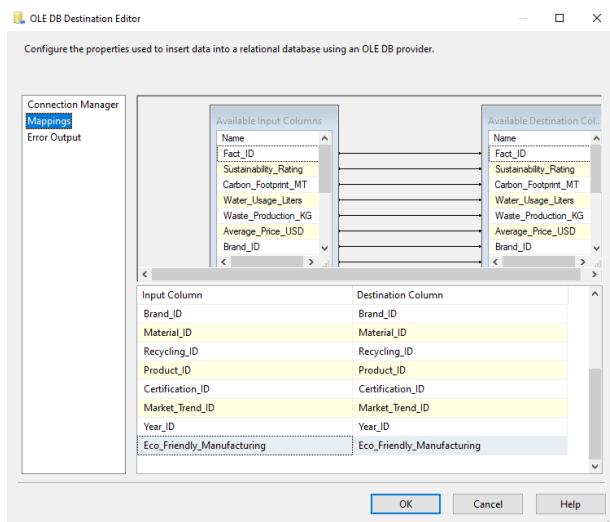
Hình 2. 194. Các bước thực hiện (11).

Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

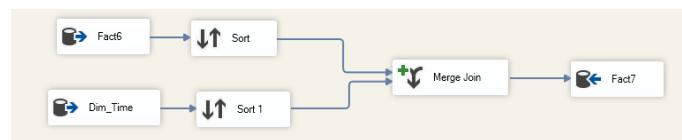


Hình 2. 195. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



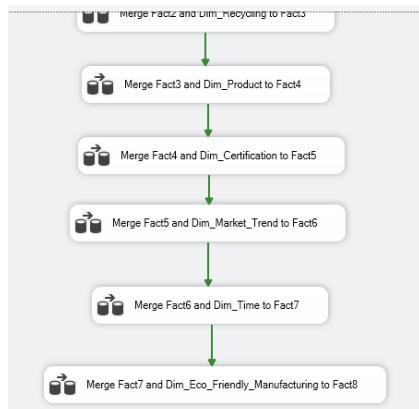
Hình 2. 196. Các bước thực hiện (13).



Hình 2. 197. Sau khi thực hiện xong.

2.2.18 Merge Fact7 và Dim_Eco_Friendly_Manufacturing thành Fact8

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact7 and Dim_Eco_Friendly_Manufacturing to Fact8”.



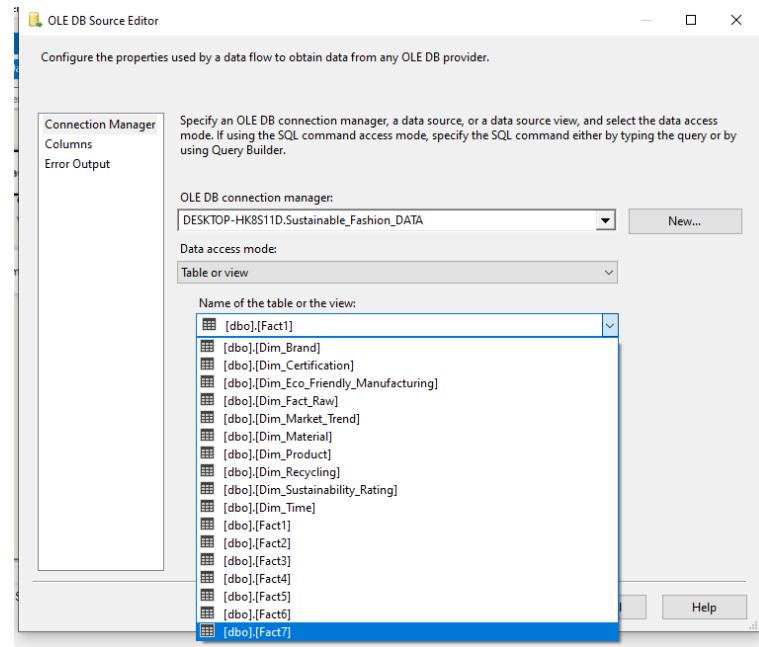
Hình 2. 198. Tạo Data Flow Task có tên "Merge Fact7 and Dim_Eco_Friendly_Manufacturing to Fact8".

Bước 2: Vào “Merge Fact7 and Dim_Eco_Friendly_Manufacturing to Fact8”, tạo 2 “OLE DB Source” là Fact7 và Dim_Eco_Friendly_Manufacturing.



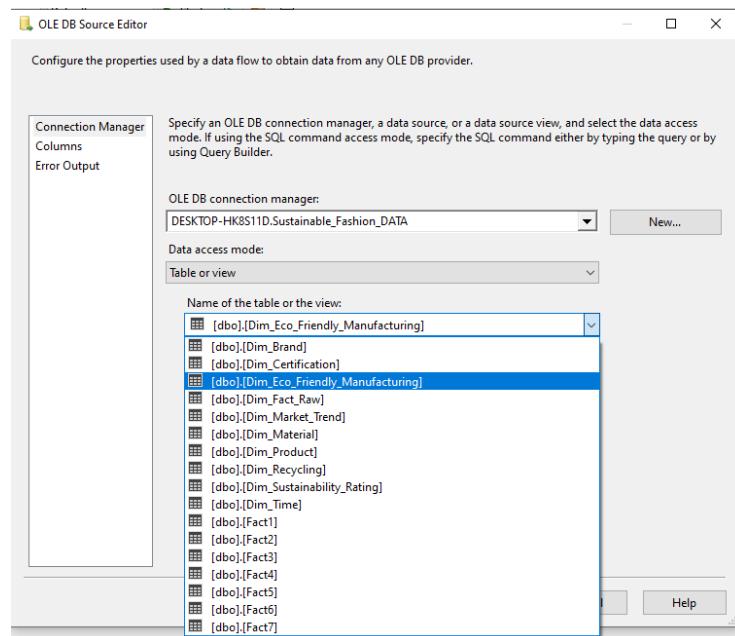
Hình 2. 199. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact7”, chọn “Edit”. Chọn dữ liệu bảng “Fact7” đã đỗ trong mục “Name of the table or the view”.



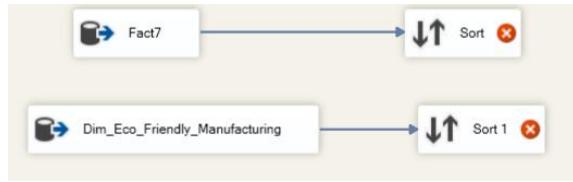
Hình 2. 200. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Eco_Friendly_Manufacturing.



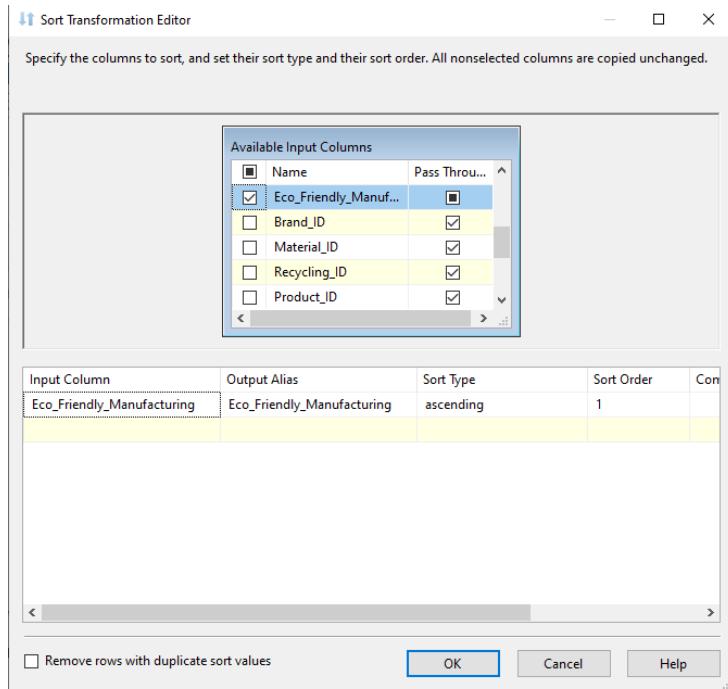
Hình 2. 201. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact7” và “Dim_Eco_Friendly_Manufacturing” áy ra thuộc tính để tiến hành kết hai bảng.



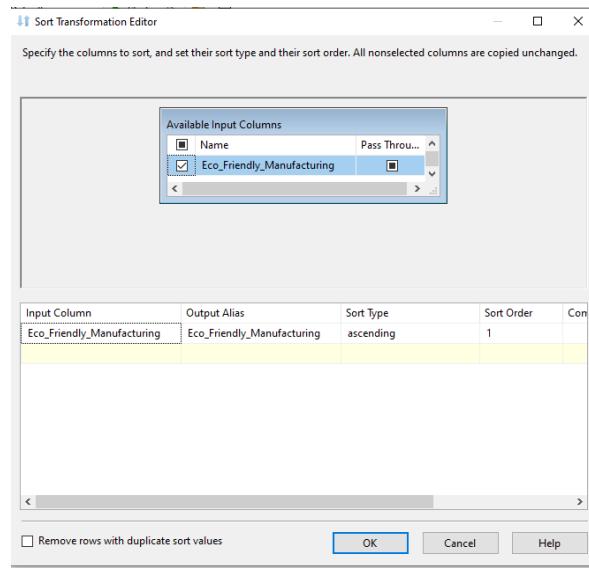
Hình 2. 202. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact7” sẽ lấy ra thuộc tính “Eco_Friendly_Manufacturing”.



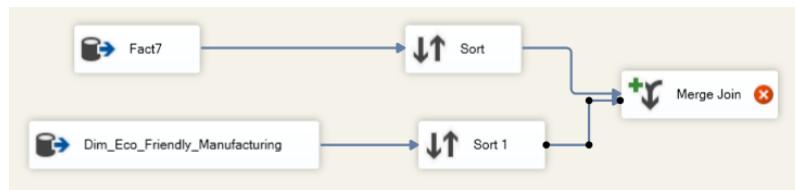
Hình 2. 203. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Eco_Friendly_Manufacturing” sẽ lấy ra thuộc tính “Eco_Friendly_Manufacturing”.



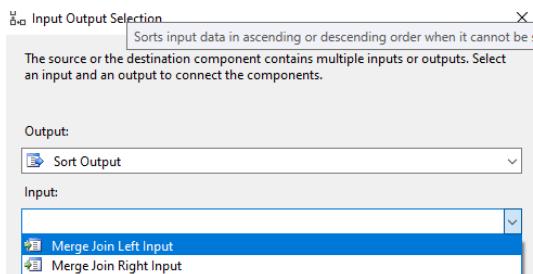
Hình 2. 204. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



Hình 2. 205. Các bước thực hiện (7).

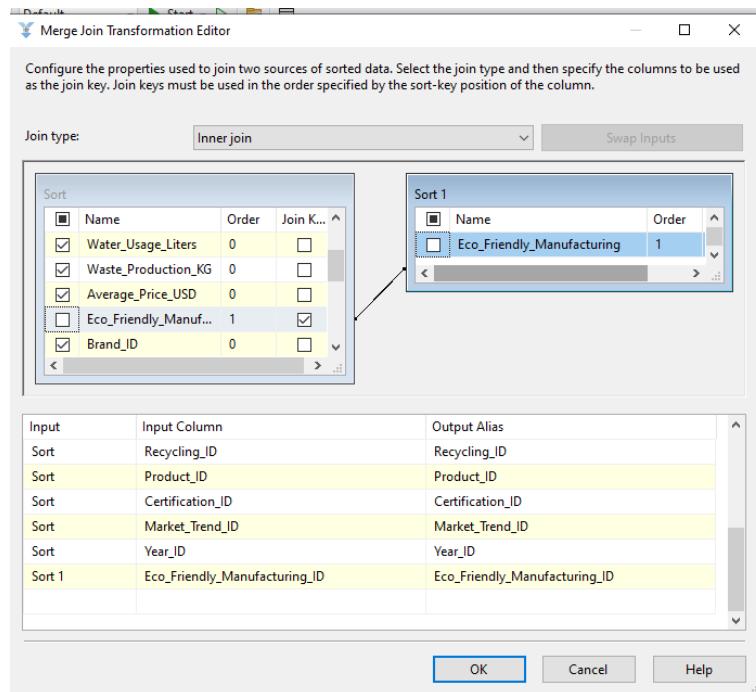
Kết nối “Sort” của “Fact7” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact5”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Eco_Friendly_Manufacturing”.



Hình 2. 206. Các bước thực hiện (8).

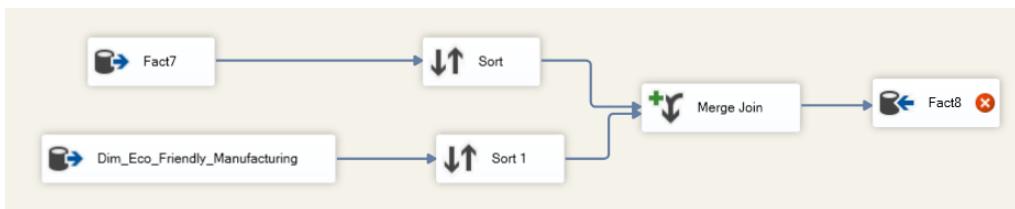
Kết nối “Sort 1” của “Dim_Eco_Friendly_Manufacturing” vào “Merge Join”. Nhấn

chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính kết nối hai bảng.



Hình 2. 207. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact8 vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact8”.



Hình 2. 208. Các bước thực hiện (10).

Bước 7: Nhấn chuột phải vào “Fact8”, chọn “Edit”. Chọn “New” để tạo bảng Fact8 mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

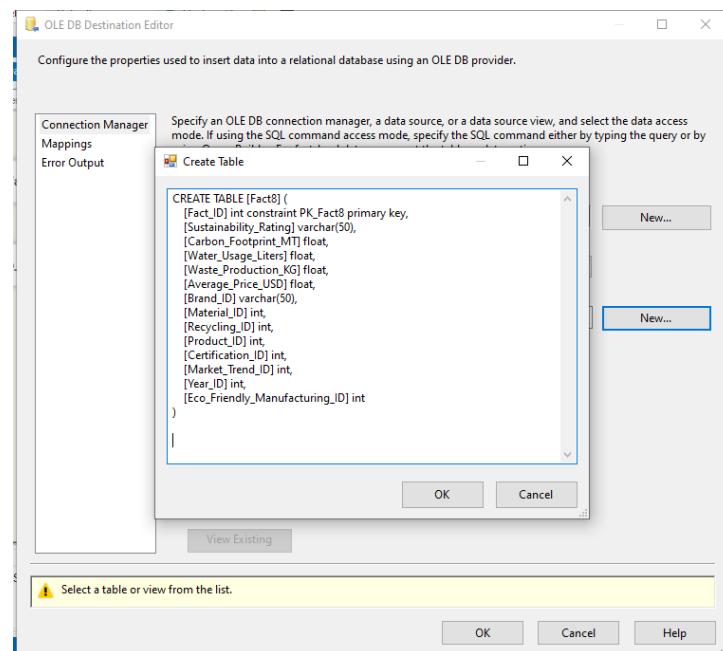
CREATE TABLE [Fact8] (

```

[Fact_ID] int constraint PK_Fact8 primary key,
[Sustainability_Rating] varchar(50),
[Carbon_Footprint_MT] float,
  
```

```
[Water_Usage_Liters] float,
[Waste_Production_KG] float,
[Average_Price_USD] float,
[Brand_ID] varchar(50),
[Material_ID] int,
[Recycling_ID] int,
[Product_ID] int,
[Certification_ID] int,
[Market_Trend_ID] int,
[Year_ID] int,
[Eco_Friendly_Manufacturing_ID] int
```

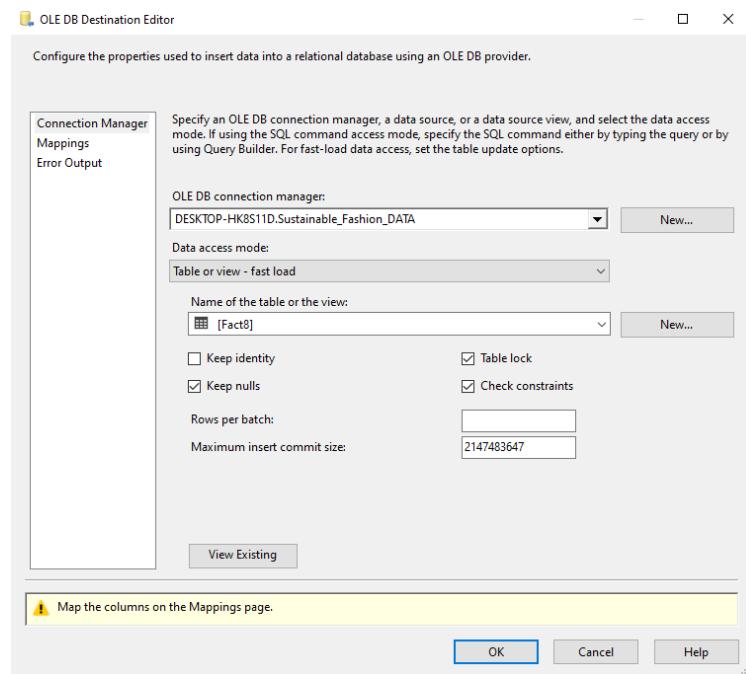
)



Hình 2. 209. Các bước thực hiện (11).

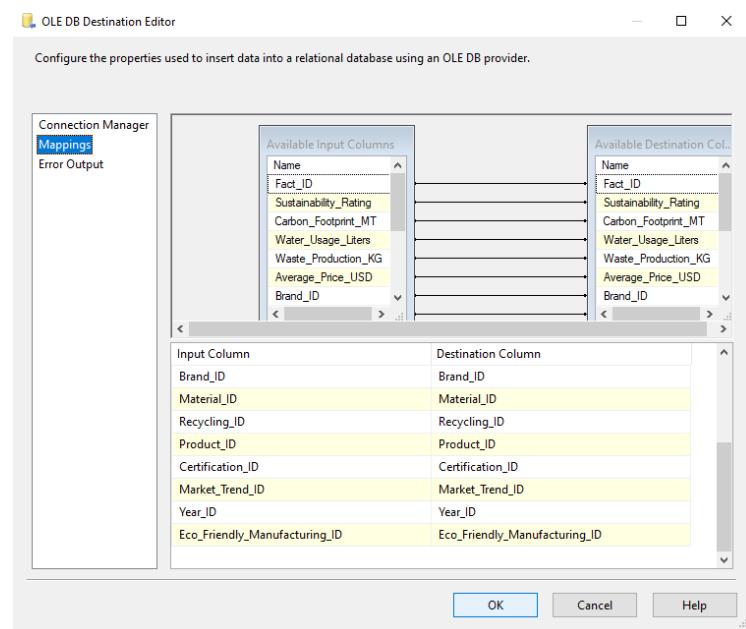
Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

IS217 – Kho dữ liệu và OLAP

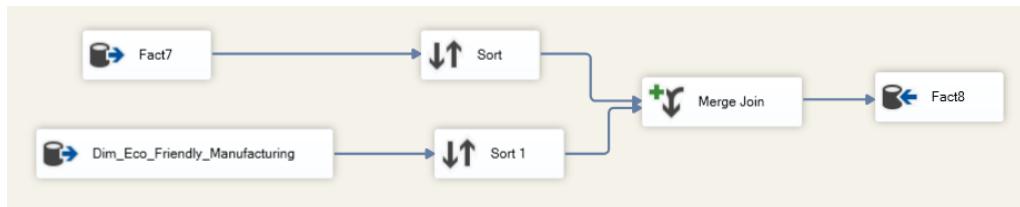


Hình 2. 210. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



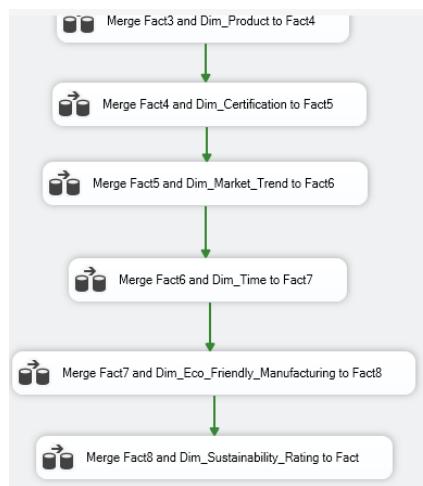
Hình 2. 211. Các bước thực hiện (13).



Hình 2. 212. Sau khi thực hiện xong.

2.2.19 Merge Fact8 và Dim_Sustainability_Rating thành Fact

Bước 1: Tạo một “Data Flow Task” đổi tên thành “Merge Fact8 and Dim_Sustainability_Rating to Fact”.



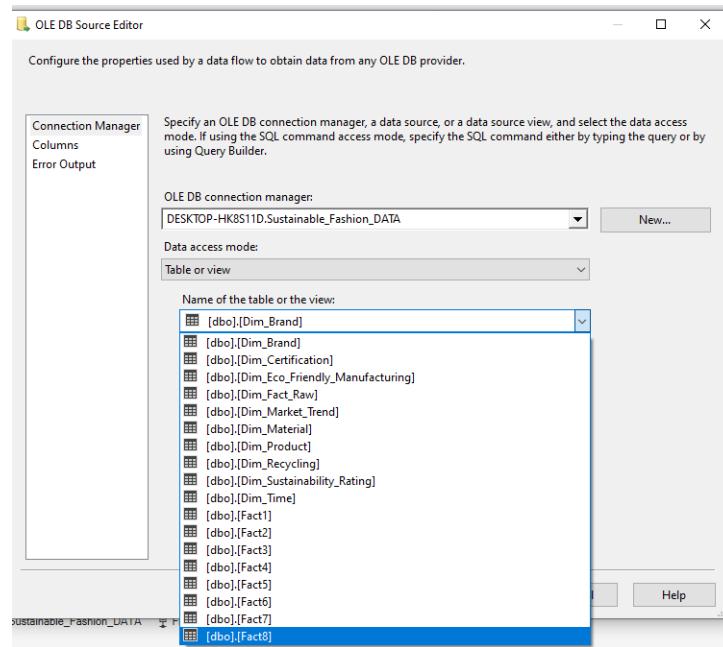
Hình 2. 213. Tạo Data Flow Task có tên "Merge Fact8 and Dim_Sustainability_Rating to Fact"

Bước 2: Vào “Merge Fact8 and Dim_Sustainability_Rating to Fact”, tạo 2 “OLE DB Source” là Fact8 và Dim_Sustainability_Rating.



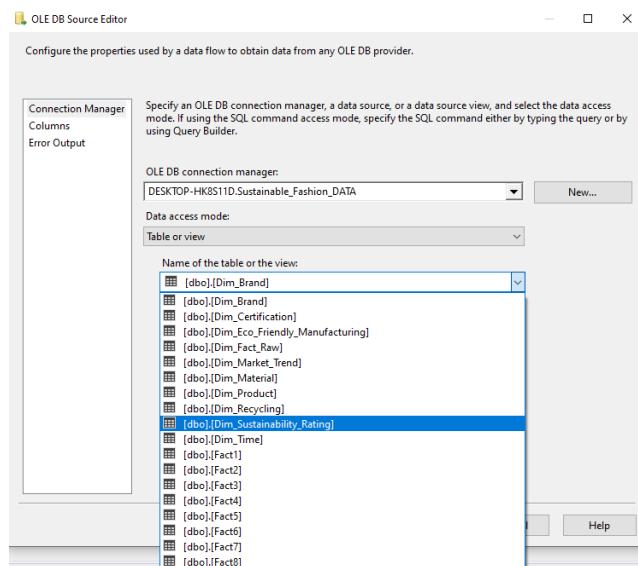
Hình 2. 214. Các bước thực hiện (1).

Bước 3: Nhấp chuột phải vào “Fact8”, chọn “Edit”. Chọn dữ liệu bảng “Fact” đã đỗ trong mục “Name of the table or the view”.



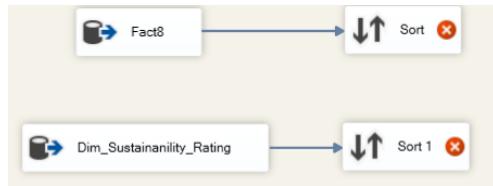
Hình 2. 215. Các bước thực hiện (2).

Thực hiện tương tự với bảng Dim_Sustainability_Rating.



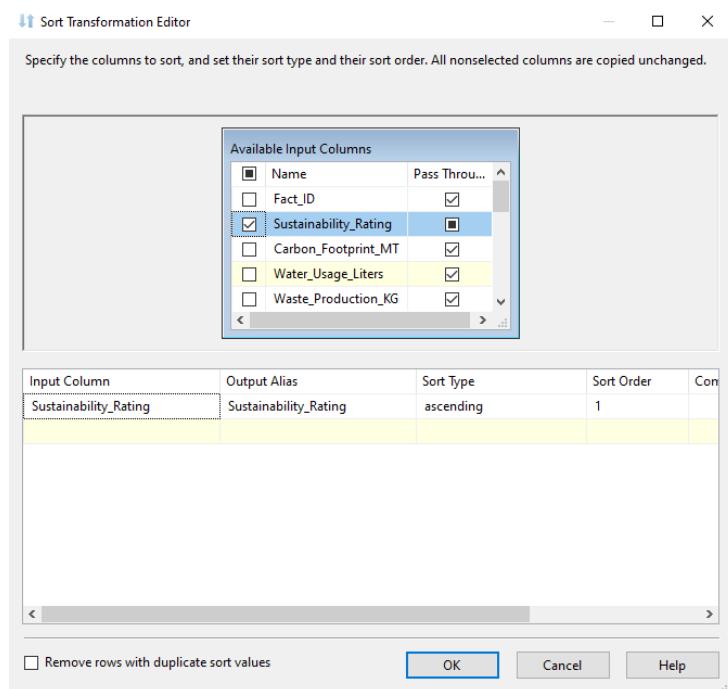
Hình 2. 216. Các bước thực hiện (3).

Bước 4: Tạo 2 “Sort” tương ứng với hai bảng “Fact7” và “Dim_Sustainability_Rating” lấy ra thuộc tính để tiến hành kết hai bảng.



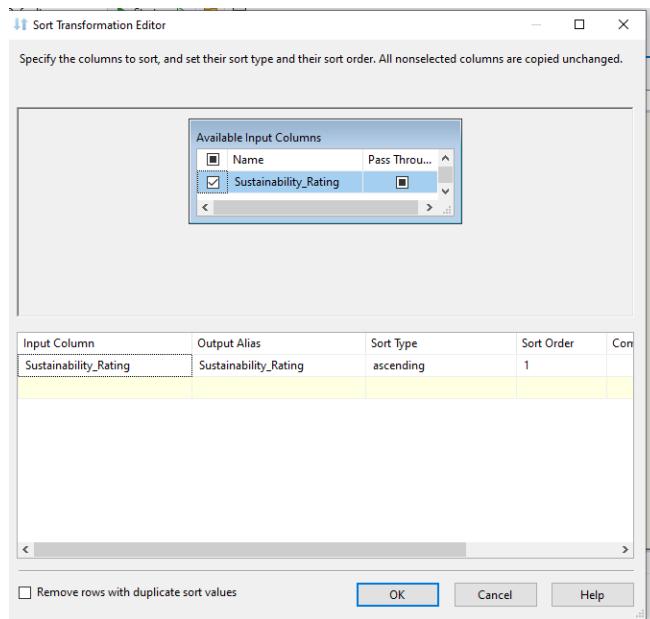
Hình 2. 217. Các bước thực hiện (4).

Với “Sort” kết nối với “Fact7” sẽ lấy ra thuộc tính “Sustainability_Rating”.



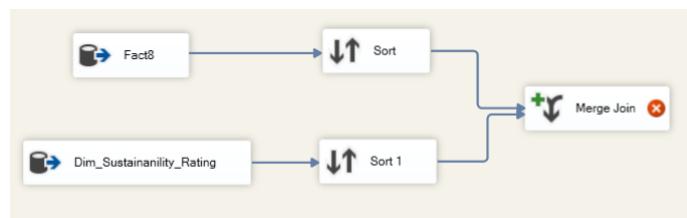
Hình 2. 218. Các bước thực hiện (5).

Với “Sort1” kết nối với “Dim_Sustainability_Rating” sẽ lấy ra thuộc tính “Sustainability_Rating”.



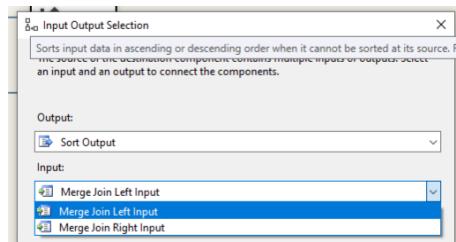
Hình 2. 219. Các bước thực hiện (6).

Bước 5: Tạo một “Merge Join” để tiến hành merge hai bảng.



Hình 2. 220. Các bước thực hiện (7).

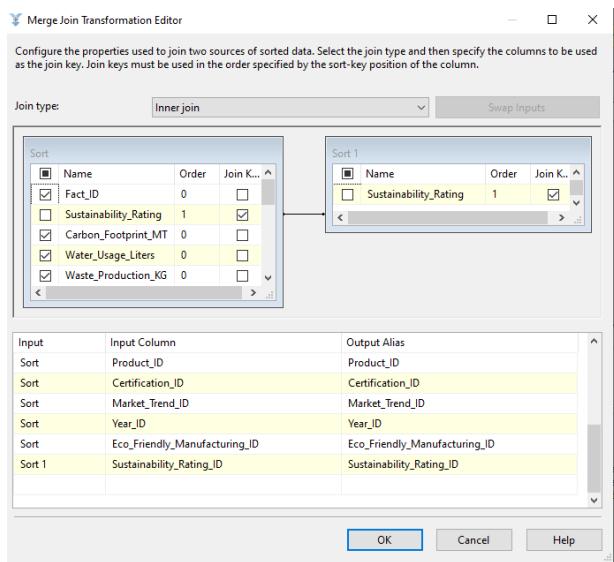
Kết nối “Sort” của “Fact7” vào “Merge Join”. Chọn vào “Merge Join Left Input” để giữ các thuộc tính của bảng “Fact5”, chỉ thay đổi một số thuộc tính dựa trên bảng “Dim_Sustainability_Rating”.



Hình 2. 221. Các bước thực hiện (8)

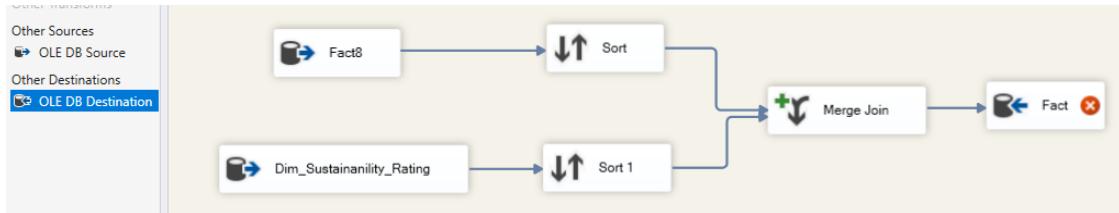
Kết nối “Sort 1” của “Dim_Sustainability_Rating” vào “Merge Join”. Nhấn chuột phải vào “Merge Join”, chọn “Edit”. Ở hai bảng chọn tất cả các thuộc tính trừ các thuộc tính

kết nối hai bảng.



Hình 2. 222. Các bước thực hiện (9).

Bước 6: Tiến hành đổ dữ liệu bảng Fact vào Sustainable_Fashion_DATA. Đầu tiên chọn một “OLE DB Destination”, đặt tên thành “Fact”.



Hình 2. 223. Các bước thực hiện (10).

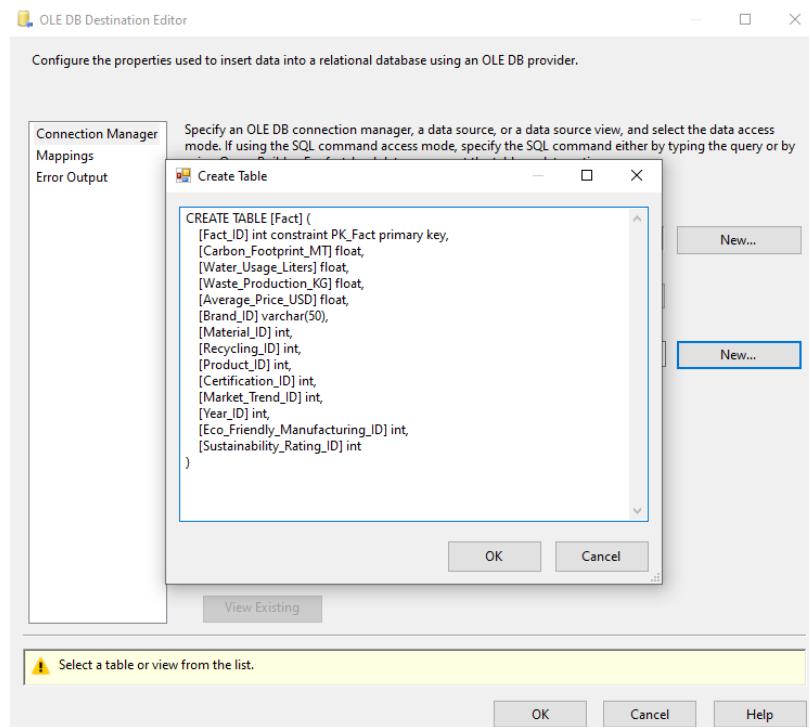
Bước 7: Nhấn chuột phải vào “Fact”, chọn “Edit”. Chọn “New” để tạo bảng Fact mới trong database. Viết câu lệnh sql để tạo bảng, chọn “OK”.

CREATE TABLE [Fact] (

```

[Fact_ID] int constraint PK_Fact primary key,
[Carbon_Footprint_MT] float,
[Water_Usage_Liters] float,
[Waste_Production_KG] float,
[Average_Price_USD] float,
[Brand_ID] varchar(50),
  
```

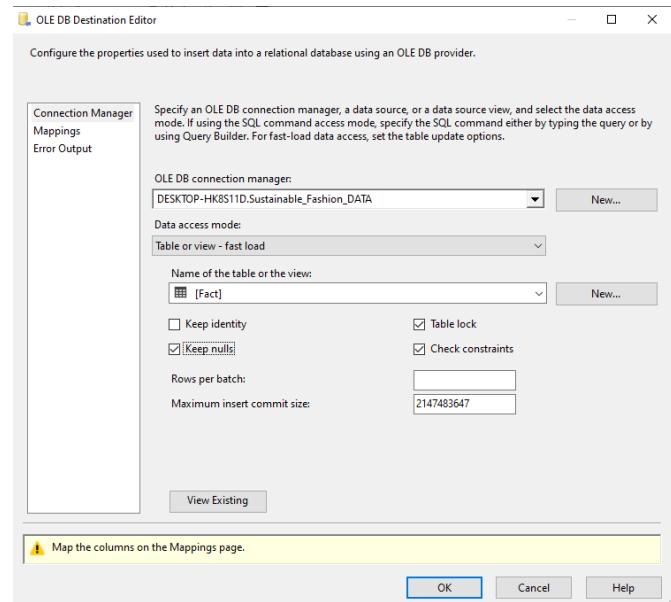
```
[Material_ID] int,  
[Recycling_ID] int,  
[Product_ID] int,  
[Certification_ID] int,  
[Market_Trend_ID] int,  
[Year_ID] int,  
[Eco_Friendly_Manufacturing_ID] int,  
[Sustainability_Rating_ID] int  
)
```



Hình 2. 224. Các bước thực hiện (11).

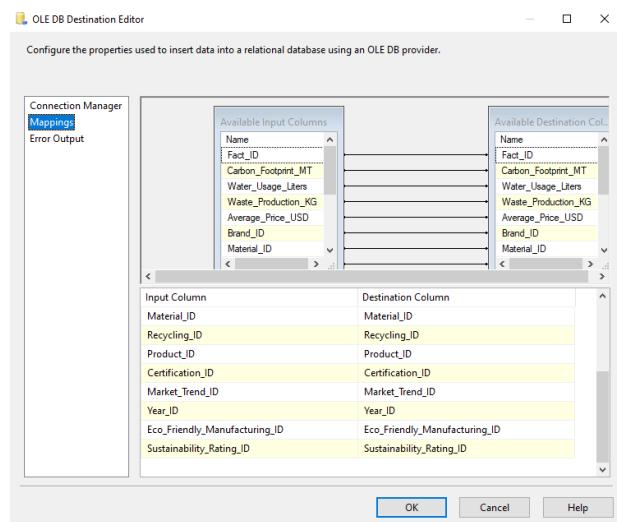
Tích chọn “Keep nulls” để chọn những giá trị null cần thiết.

IS217 – Kho dữ liệu và OLAP

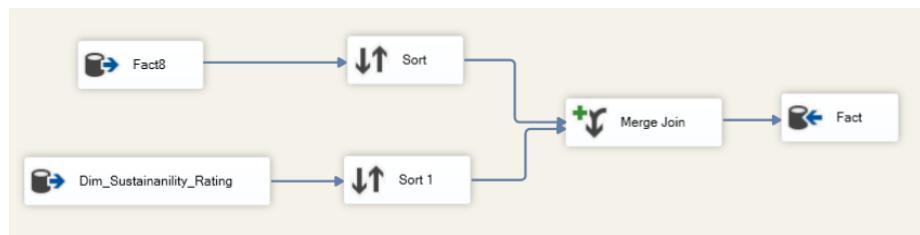


Hình 2. 225. Các bước thực hiện (12).

Vào mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Nhấn “OK” để hoàn tất.



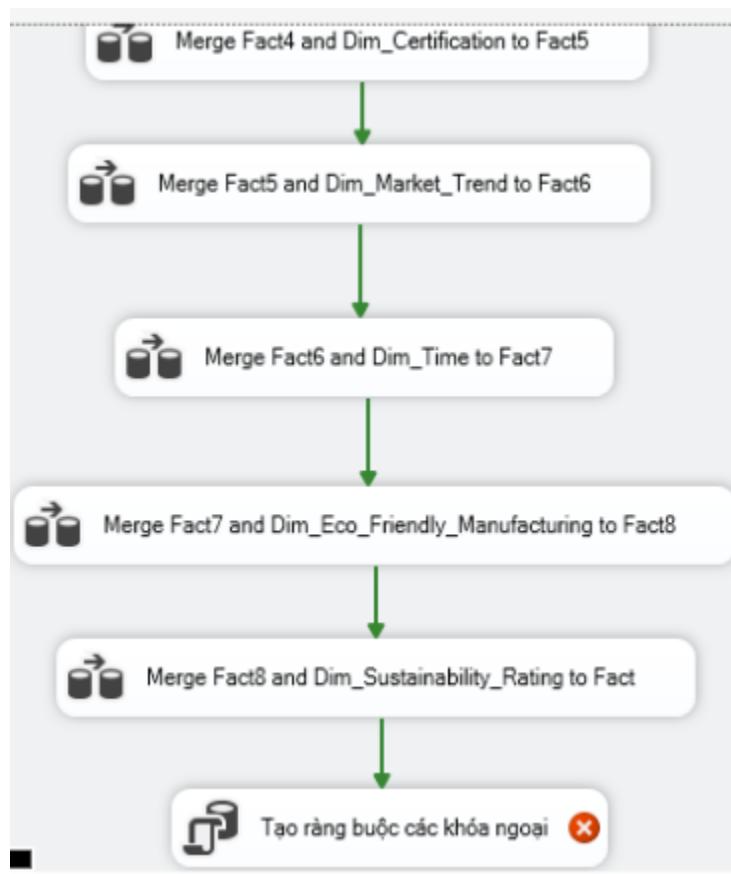
Hình 2. 226. Các bước thực hiện (13).



Hình 2. 226. Các bước thực hiện (14).

2.2.20 Tạo các ràng buộc các khóa ngoại

Bước 1: Tạo một “Excute SQL Task” đổi tên thành “Tạo ràng buộc các khóa ngoại”.



Hình 2. 227. Tạo “Execute SQL Task” đổi tên thành “Tạo ràng buộc các khóa ngoại”.

Bước 2: Nhấn chuột phải, chọn “Edit”.

Ở mục Connection sẽ chọn kết nối đến database trước đó.

Ở mục SQLStatement sẽ viết các lệnh SQL tạo khóa ngoại.

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Brand FOREIGN KEY (Brand_ID) REFERENCES  
Dim_Brand(Brand_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Material FOREIGN KEY (Material_ID) REFERENCES  
Dim_Material(Material_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Recycling FOREIGN KEY (Recycling_ID)  
REFERENCES Dim_Recycling(Recycling_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Product FOREIGN KEY (Product_ID) REFERENCES  
Dim_Product(Product_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Certification FOREIGN KEY (Certification_ID)  
REFERENCES Dim_Certification(Certification_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Market_Trend FOREIGN KEY (Market_Trend_ID)  
REFERENCES Dim_Market_Trend(Market_Trend_ID);
```

ALTER TABLE Fact

```
ADD CONSTRAINT FK_Fact_Sustainability_Rating FOREIGN KEY  
(Sustainability_Rating_ID) REFERENCES Dim_Sustainability_Rating
```

(Sustainability_Rating_ID);

ALTER TABLE Fact

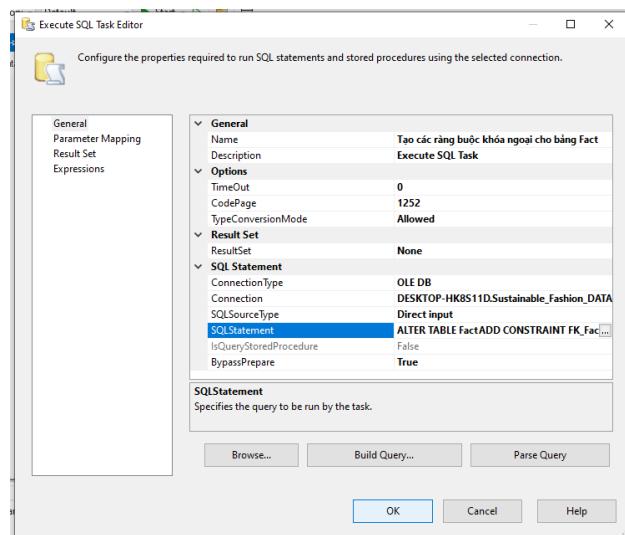
ADD CONSTRAINT FK_Fact_Time FOREIGN KEY (Year_ID) REFERENCES

Dim_Time (Year_ID);

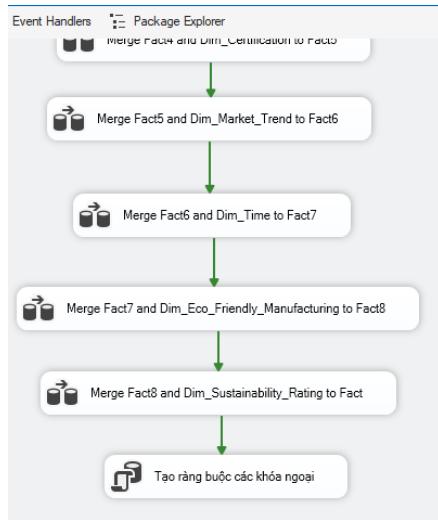
ALTER TABLE Fact

ADD CONSTRAINT FK_Fact_Eco FOREIGN KEY (Eco_Friendly_Manufacturing_ID)

REFERENCES Dim_Eco_Friendly_Manufacturing (Eco_Friendly_Manufacturing_ID);



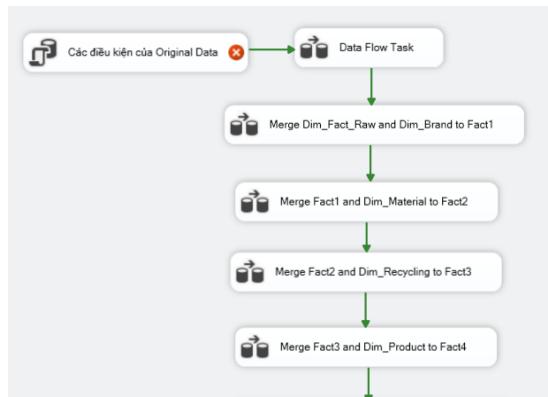
Hình 2. 228. Cấu hình phần SQL Statement trong Execute SQL Task Editor.



Hình 2. 228. Sau khi thực hiện xong.

2.2.21 Các điều kiện của Original Data

Bước 1: Tạo một “Execute SQL Task” đổi tên thành “Các điều kiện của Original Data”.

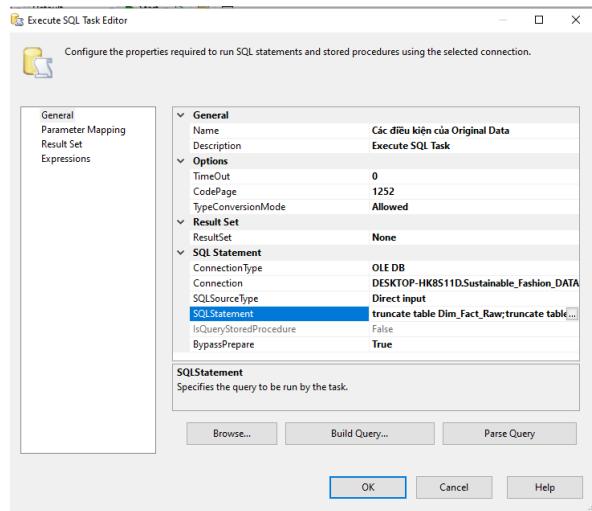


Hình 2. 229. Tạo “Execute SQL Task” đổi tên thành “Các điều kiện của Original Data”.

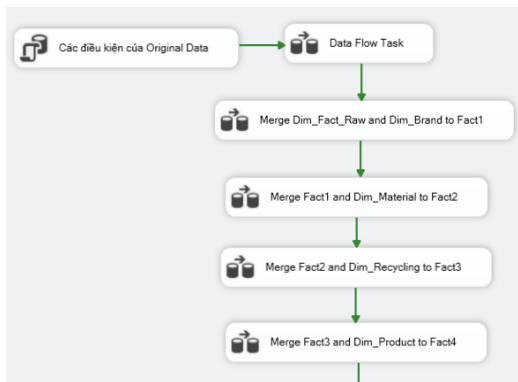
Bước 2: Nhấn chuột phải, chọn “Edit”. Ở mục Connection sẽ chọn kết nối đến database trước đó. Ở mục SQLStatement sẽ viết các lệnh SQL.

```
truncate table Dim_Fact_Raw;  
truncate table Fact1;  
truncate table Fact2;  
truncate table Fact3;  
truncate table Fact4;  
truncate table Fact5;
```

truncate table Fact6;
truncate table Fact7;
truncate table Fact8;



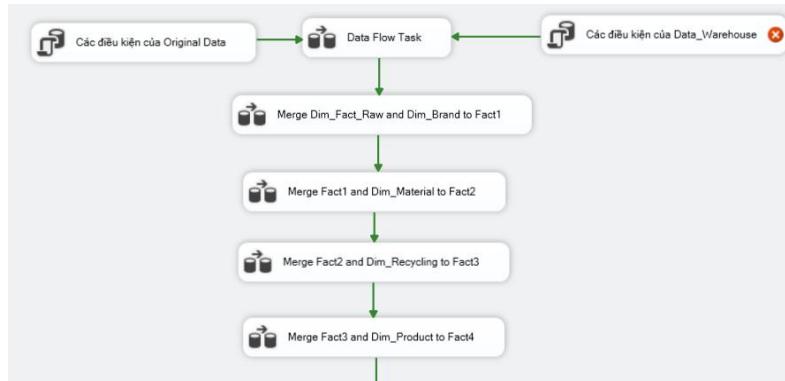
Hình 2. 230. Cấu hình phần SQL Statement trong Execute SQL Task Editor.



Hình 2. 231. Sau khi thực hiện xong.

2.2.22 Các điều kiện của Data_Warehouse

Bước 1: Tạo một “Excute SQL Task” đổi tên thành “Các điều kiện của Data_Warehouse”.



Hình 2. 232. Tạo “Excute SQL Task” đổi tên thành “Các điều kiện của Data_Warehouse”.

Bước 2: Nhấn chuột phải, chọn “Edit”. Ở mục Connection sẽ chọn kết nối đến database trước đó. Ở mục SQLStatement sẽ viết các lệnh SQL.

truncate table Dim_Brand;

truncate table Dim_Product;

truncate table Dim_Market_Trend;

truncate table Dim_Certification;

truncate table Dim_Material;

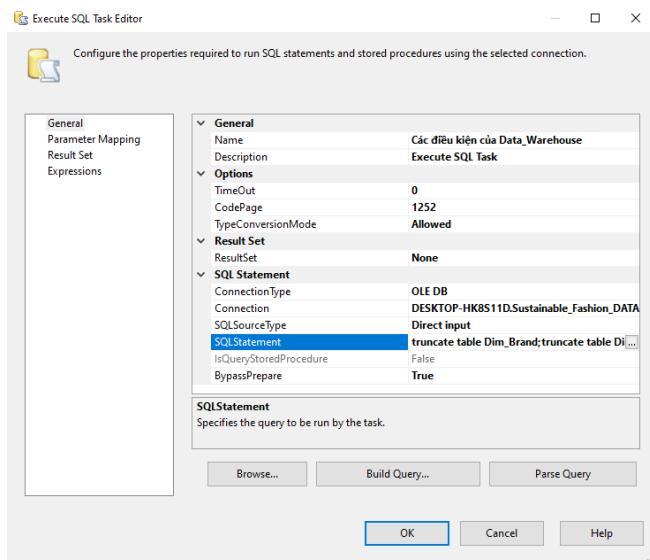
truncate table Dim_Recycling;

truncate table Dim_Sustainability_Rating;

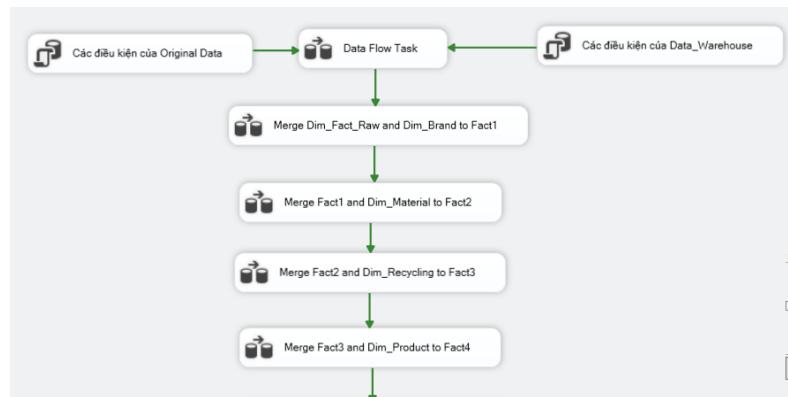
truncate table Dim_Time;

truncate table Dim_Eco_Friendly_Manufacturing;

truncate table Fact;



Hình 2. 233. Cấu hình phần SQL Statement trong Execute SQL Task Editor.



Hình 2. 234. Sau khi thực hiện xong.

2.3 Chạy dự án SSIS

Bước 1: Nhấn “Start” để chạy

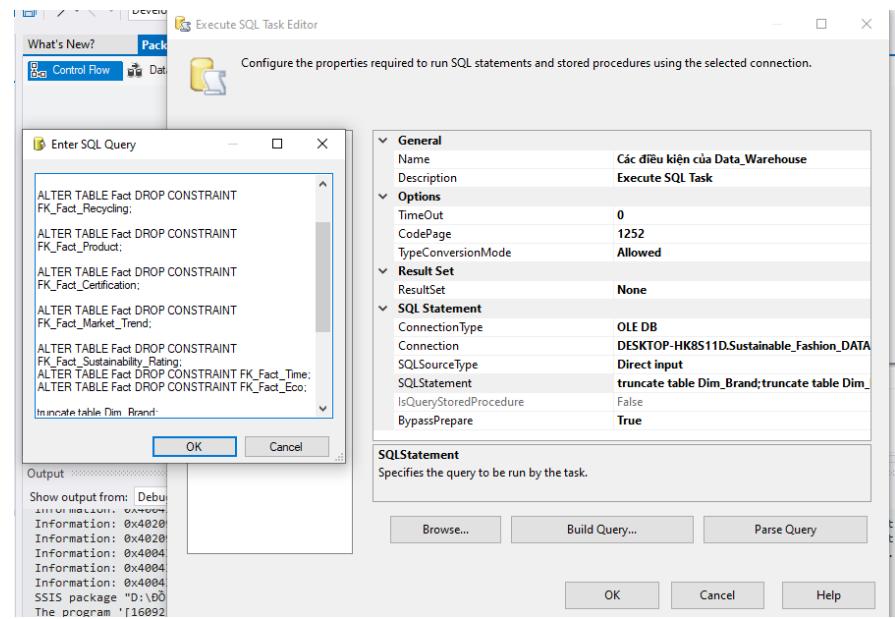
Kết quả sau khi chạy:



Hình 2. 235. Kết quả chạy dự án SSIS.

Sau khi chạy project lần đầu thành công, khóa ngoại của bảng Fact tham chiếu đến các Dimension đã được tạo. Thêm các lệnh SQL để xóa và thêm lại khóa ngoại vào mỗi lần khởi chạy lại project tiếp theo.

```
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Brand;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Material;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Recycling;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Product;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Certification;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Market_Trend;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Sustainability_Rating;  
ALTER TABLE Fact DROP CONSTRAINT FK_Fact_Time;
```



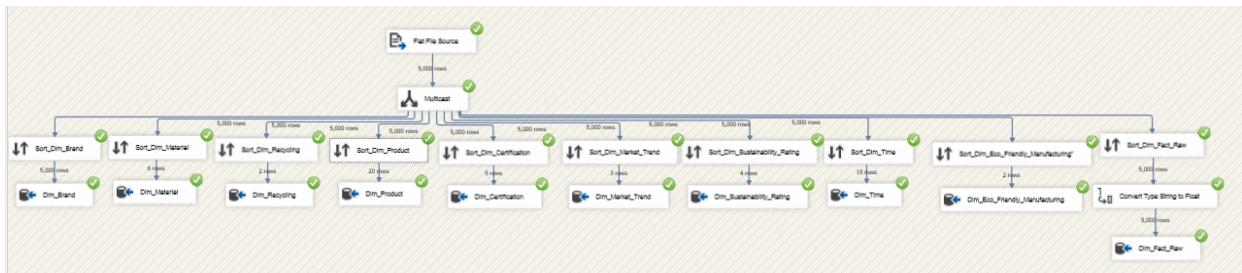
Hình 2. 236. Cấu hình phần SQL Statement trong Execute SQL Task Editor.

Bước 2: Kết quả sau khi chạy lần 2

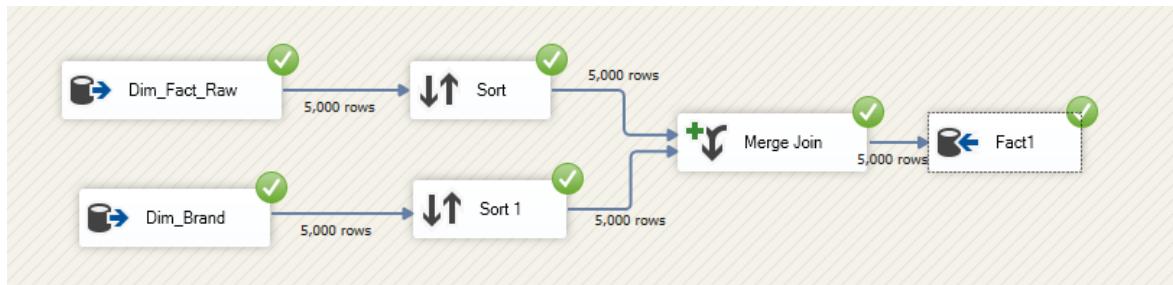


Hình 2. 237. Kết quả (1).

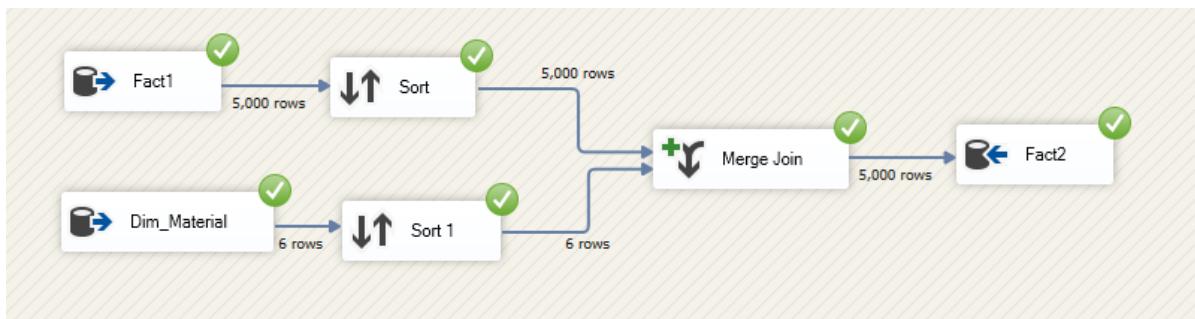
IS217 – Kho dữ liệu và OLAP



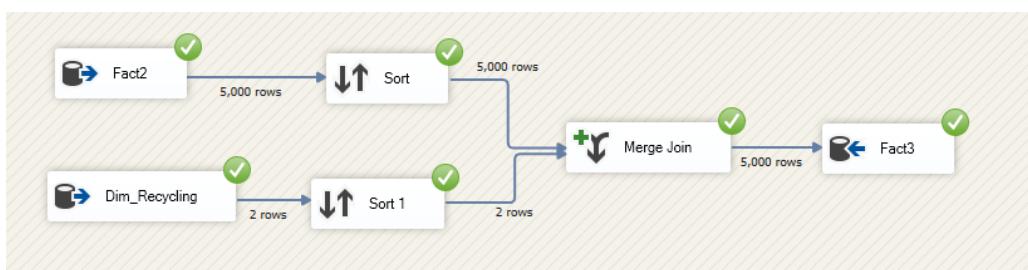
Hình 2. 238. Kết quả (2).



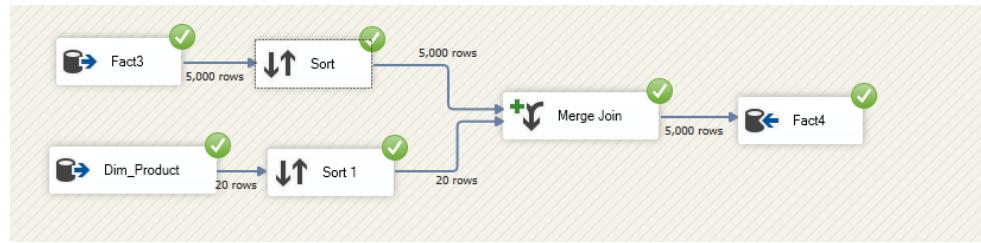
Hình 2. 239. Kết quả (3).



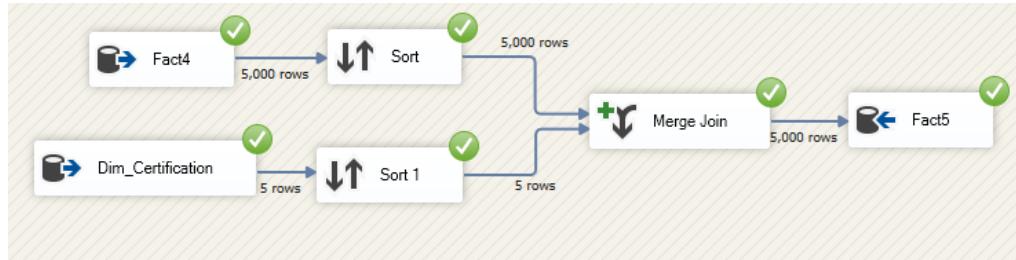
Hình 2. 240. Kết quả (4).



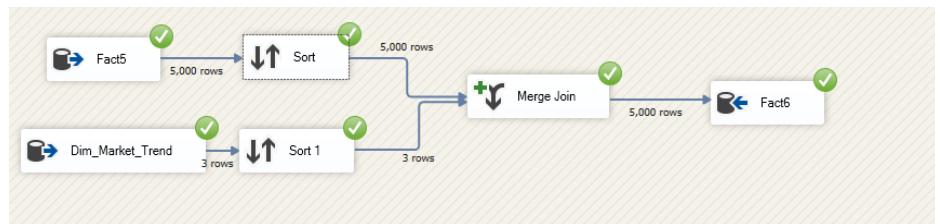
Hình 2. 241. Kết quả (5).



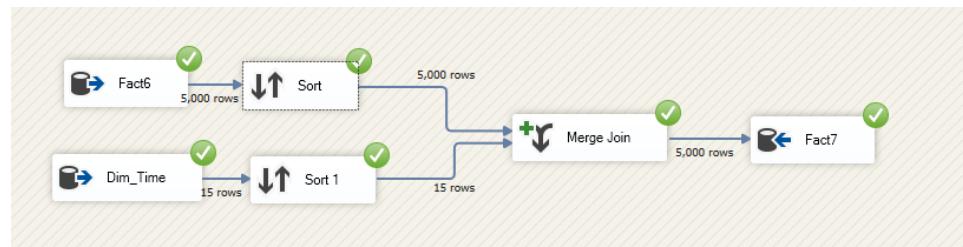
Hình 2. 242. Kết quả (6).



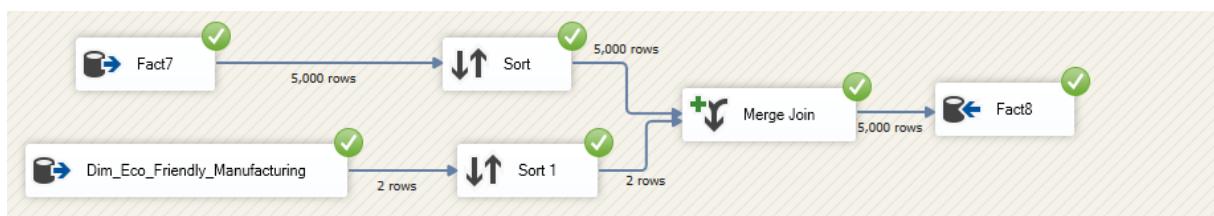
Hình 2. 243. Kết quả (7).



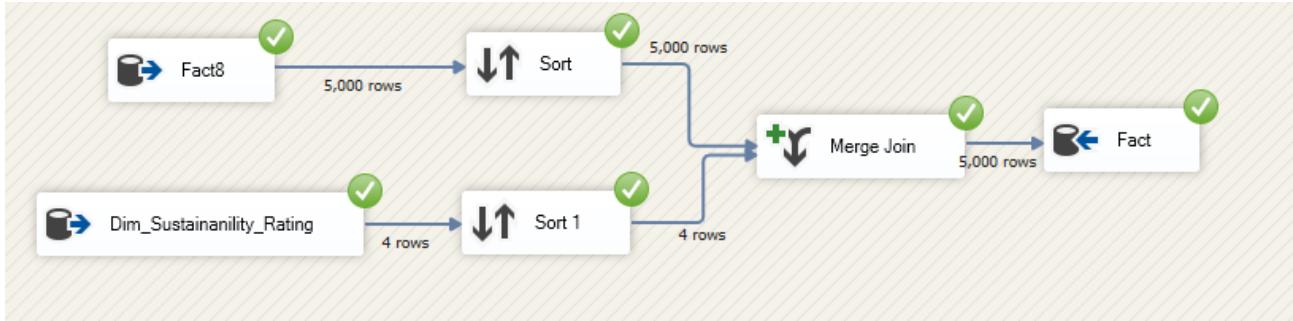
Hình 2. 244. Kết quả (8).



Hình 2. 245. Kết quả (9).



Hình 2. 246. Kết quả (10).



Hình 2. 247. Kết quả (11).

2.3.1 Kiểm tra dữ liệu các bảng

```

Select * from Dim_Recycling;
Select * from Dim_Product;
Select * from Dim_Certification;
Select * from Dim_Market_Trend;
Select * from Dim_Sustainability_Rating;
Select * from Dim_Time;
Select * from Dim_Eco_Friendly_Manufacturing;
Select * from Fact;
  
```

Fact_ID	Carbon_Footprint_MT	Water_Usage_Litres	Waste_Production_KG	Average_Price_USD	Brand_ID	Material_ID	Recycling_ID	Product_ID	Certification_ID	Market_Trend_ID
1	1.75	4511152.79	97844.11	38.33	BRAND0001	5	1	12	3	2
2	124.39	1951566.31	37267.75	250.07	BRAND0002	6	1	7	3	2
3	336.66	467454.52	38385.92	146.16	BRAND0003	6	1	12	1	2
4	152.04	899576.9	32665.45	165.52	BRAND0004	1	1	5	5	3
5	415.63	1809219.9	37295.47	211.63	BRAND0005	1	2	11	2	3
6	447.65	2244115.38	69017.63	196.45	BRAND0006	4	2	2	1	3
7	345.64	1020265.49	18373.35	376.14	BRAND0007	2	1	9	3	3
8	336.7	3847649.16	28717.69	59.34	BRAND0008	5	1	10	5	1
9	386.68	1494891.28	45916.88	51.12	BRAND0009	4	2	3	3	1
10	250.94	3892548.66	58636.15	426.55	BRAND0010	1	1	20	4	3

Hình 2. 248. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Fact.

IS217 – Kho dữ liệu và OLAP

The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a T-SQL script window titled 'Fill_Baocao.sql' with the following code:

```
Create database Sustainable_Fashion_DATA;
Select * from Dim_Brand;
```

The 'Results' tab on the right shows the output of the query:

	Brand_ID	Brand_Name	Country
1	BRAND0001	Brand1	Australia
2	BRAND0002	Brand2	Japan
3	BRAND0003	Brand3	USA
4	BRAND0004	Brand4	Italy
5	BRAND0005	Brand5	USA
6	BRAND0006	Brand6	Italy
7	BRAND0007	Brand7	Australia
8	BRAND0008	Brand8	Brazil
9	BRAND0009	Brand9	France
10	BRAND0010	Brand10	India
11	BRAND0011	Brand11	Italy

At the bottom of the results pane, it says 'Query executed successfully.'

Hình 2. 249. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Brand.

The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a T-SQL script window titled 'Fill_Baocao.sql' with the following code:

```
Create database Sustainable_Fashion_DATA;
Select * from Dim_Brand;
Select * from Dim_Material;
```

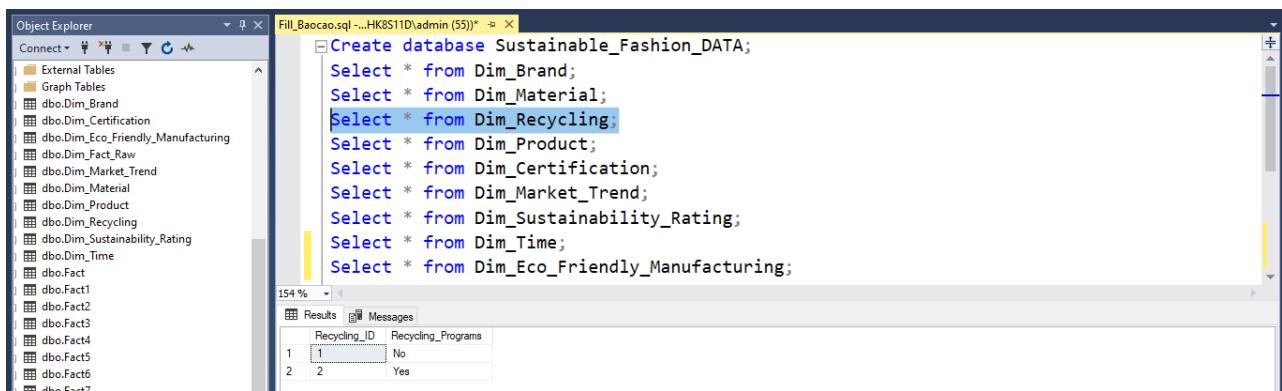
The 'Results' tab on the right shows the output of the query:

	Material_ID	Material_Type
1	1	Bamboo Fabric
2	2	Hemp
3	3	Organic Cotton
4	4	Recycled Polyester
5	5	Tencel
6	6	Vegan Leather

At the bottom of the results pane, it says 'Query executed successfully.'

Hình 2. 250. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Material.

IS217 – Kho dữ liệu và OLAP

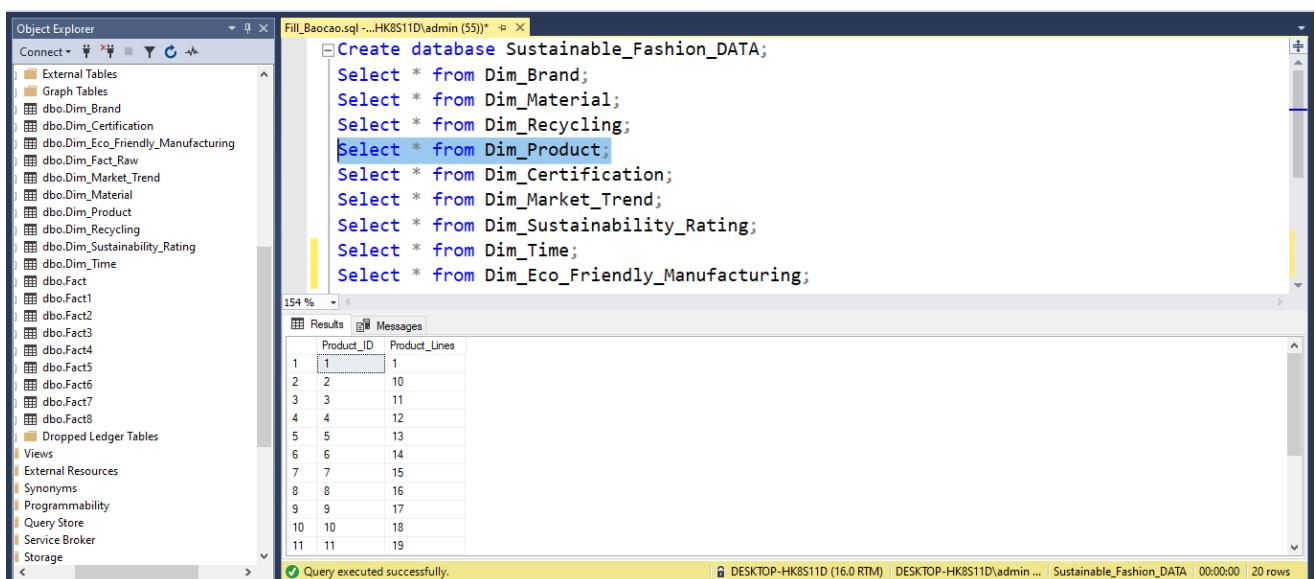


The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a T-SQL script for creating a database and selecting from several dimension tables. The 'Results' tab on the right shows the output of the 'Select * from Dim_Recycling' query.

```
Fill_Baocao.sql -...HK8S11D\admin (55)* - x
Create database Sustainable_Fashion_DATA;
Select * from Dim_Brand;
Select * from Dim_Material;
Select * from Dim_Recycling;
Select * from Dim_Product;
Select * from Dim_Certification;
Select * from Dim_Market_Trend;
Select * from Dim_Sustainability_Rating;
Select * from Dim_Time;
Select * from Dim_Eco_Friendly_Manufacturing;
```

Recycling_ID	Recycling_Programs
1	No
2	Yes

Hình 2. 251. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Recycling.



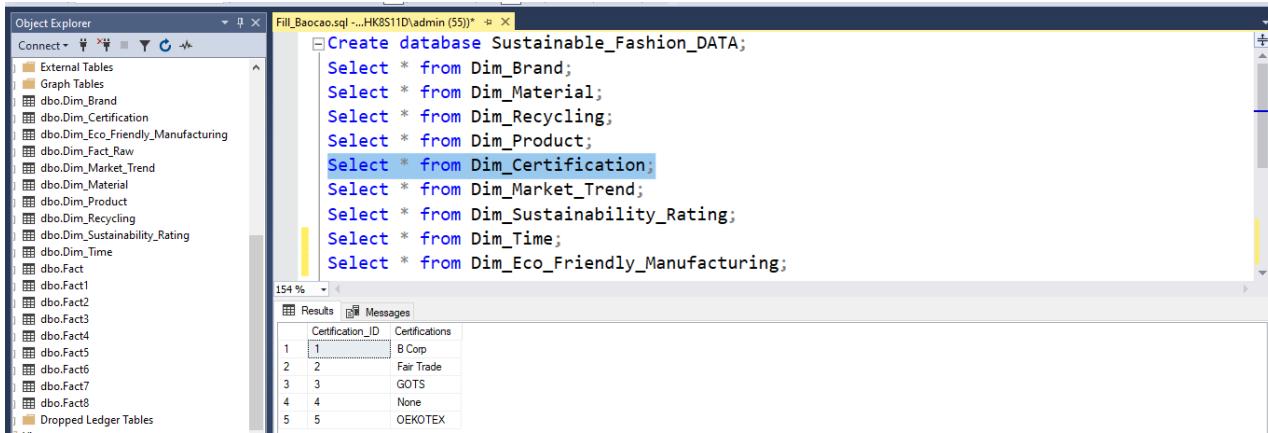
The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a T-SQL script for creating a database and selecting from several dimension tables. The 'Results' tab on the right shows the output of the 'Select * from Dim_Product' query.

```
Fill_Baocao.sql -...HK8S11D\admin (55)* - x
Create database Sustainable_Fashion_DATA;
Select * from Dim_Brand;
Select * from Dim_Material;
Select * from Dim_Recycling;
Select * from Dim_Product;
Select * from Dim_Certification;
Select * from Dim_Market_Trend;
Select * from Dim_Sustainability_Rating;
Select * from Dim_Time;
Select * from Dim_Eco_Friendly_Manufacturing;
```

Product_ID	Product_Lines
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11

Hình 2. 252. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Product.

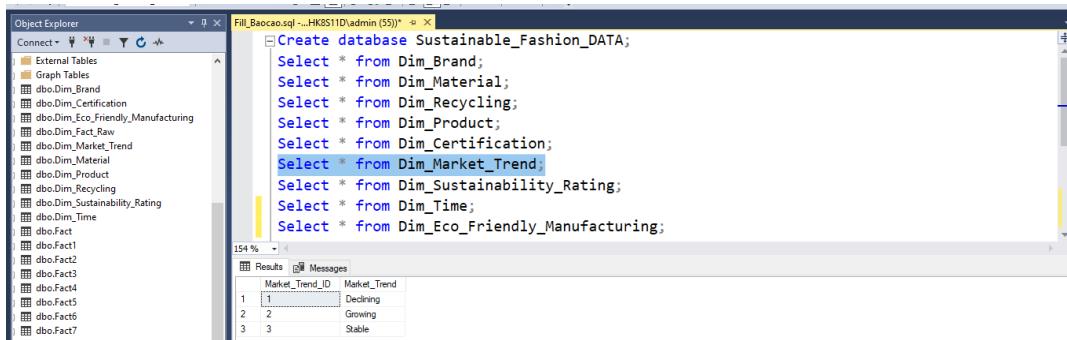
IS217 – Kho dữ liệu và OLAP



The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a SQL script named 'Fill_Baocaosql...'. A specific line of code, 'Select * from Dim_Certification;', is highlighted in blue. Below the script, the results pane shows a table with five rows of data:

Certification_ID	Certifications
1	B Corp
2	Fair Trade
3	GOTS
4	None
5	OEKOTEX

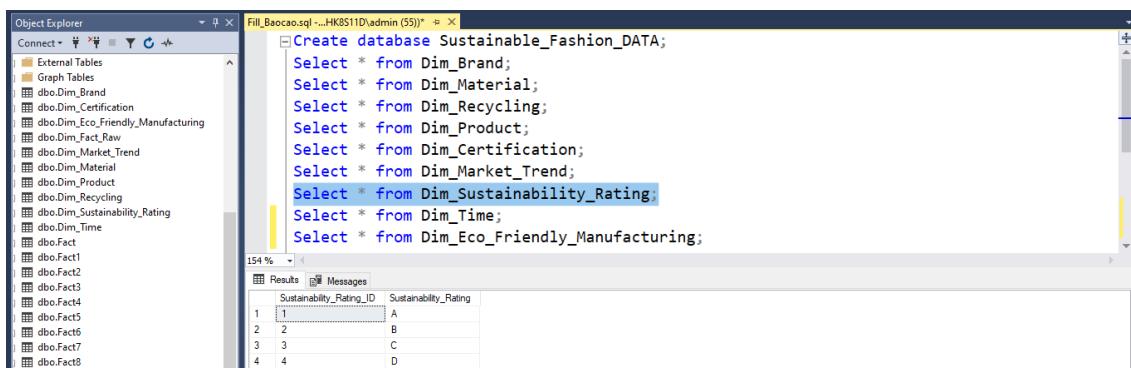
Hình 2. 253. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Certification.



The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a SQL script named 'Fill_Baocaosql...'. A specific line of code, 'Select * from Dim_Market_Trend;', is highlighted in blue. Below the script, the results pane shows a table with three rows of data:

Market_Trend_ID	Market_Trend
1	Declining
2	Growing
3	Stable

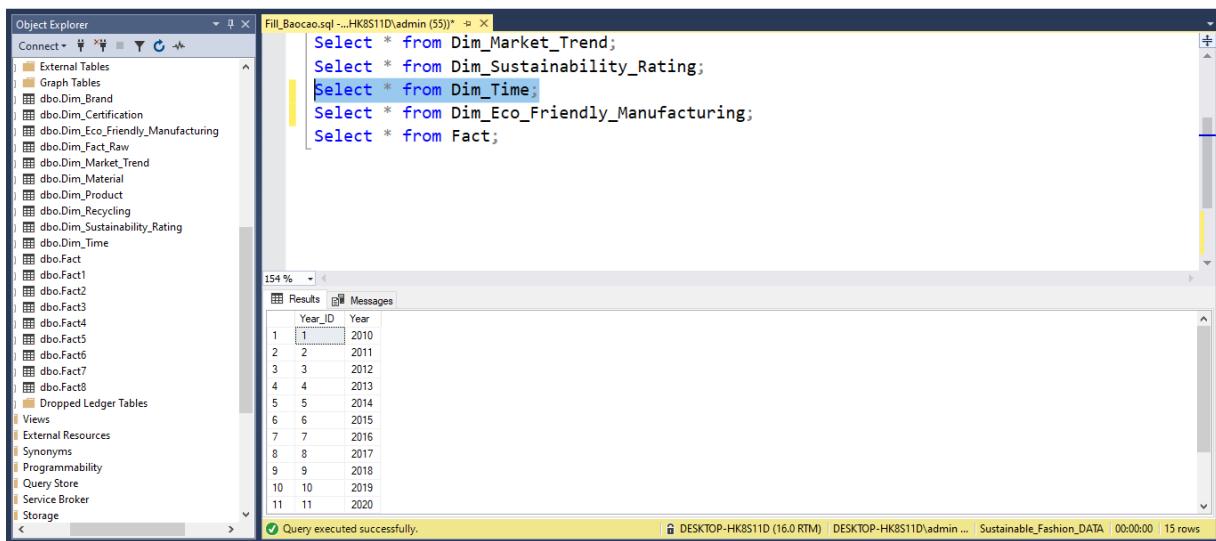
Hình 2. 254. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Market_Trend.



The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a SQL script named 'Fill_Baocaosql...'. A specific line of code, 'Select * from Dim_Sustainability_Rating;', is highlighted in blue. Below the script, the results pane shows a table with four rows of data:

Sustainability_Rating_ID	Sustainability_Rating
1	A
2	B
3	C
4	D

Hình 2. 255. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Sustainability_Rating.



The screenshot shows the Object Explorer on the left with various database objects listed. The central pane displays a T-SQL query:

```

Select * from Dim_Market_Trend;
Select * from Dim_Sustainability_Rating;
Select * from Dim_Time;
Select * from Dim_Eco_Friendly_Manufacturing;
Select * from Fact;

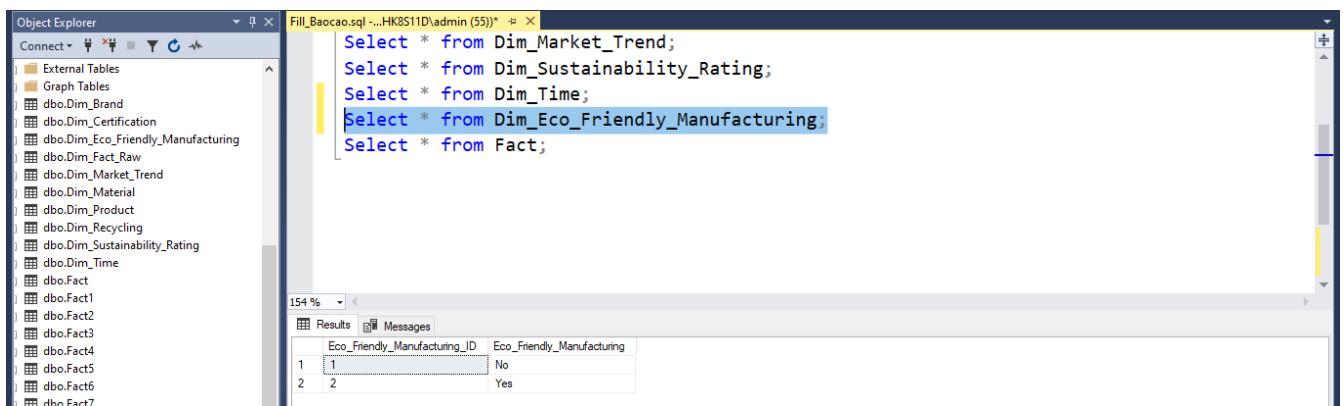
```

The Results tab on the right shows the output of the query, specifically the Dim_Time table:

Year_ID	Year
1	2010
2	2011
3	2012
4	2013
5	2014
6	2015
7	2016
8	2017
9	2018
10	2019
11	2020

A status bar at the bottom indicates "Query executed successfully." and provides connection information.

Hình 2. 256. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Time.



The screenshot shows the Object Explorer on the left. The central pane displays a T-SQL query:

```

Select * from Dim_Market_Trend;
Select * from Dim_Sustainability_Rating;
Select * from Dim_Time;
Select * from Dim_Eco_Friendly_Manufacturing;
Select * from Fact;

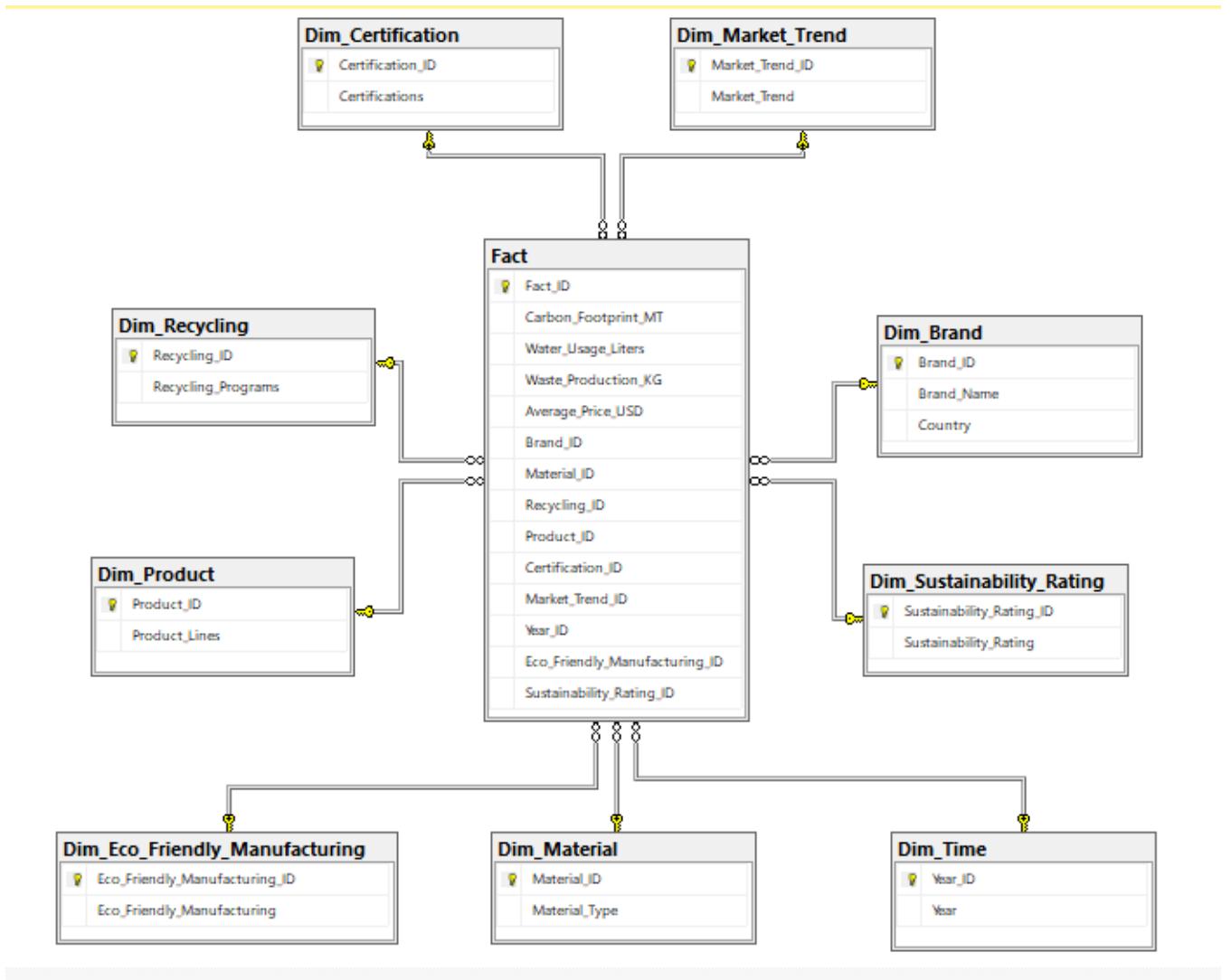
```

The Results tab on the right shows the output of the query, specifically the Dim_Eco_Friendly_Manufacturing table:

Eco_Friendly_Manufacturing_ID	Eco_Friendly_Manufacturing
1	No
2	Yes

Hình 2. 257. Thực hiện lệnh SELECT để in ra dữ liệu từ bảng Dim_Eco_Friendly_Manufacturing.

2.3.2 Lược đồ dữ liệu sau khi hoàn thành

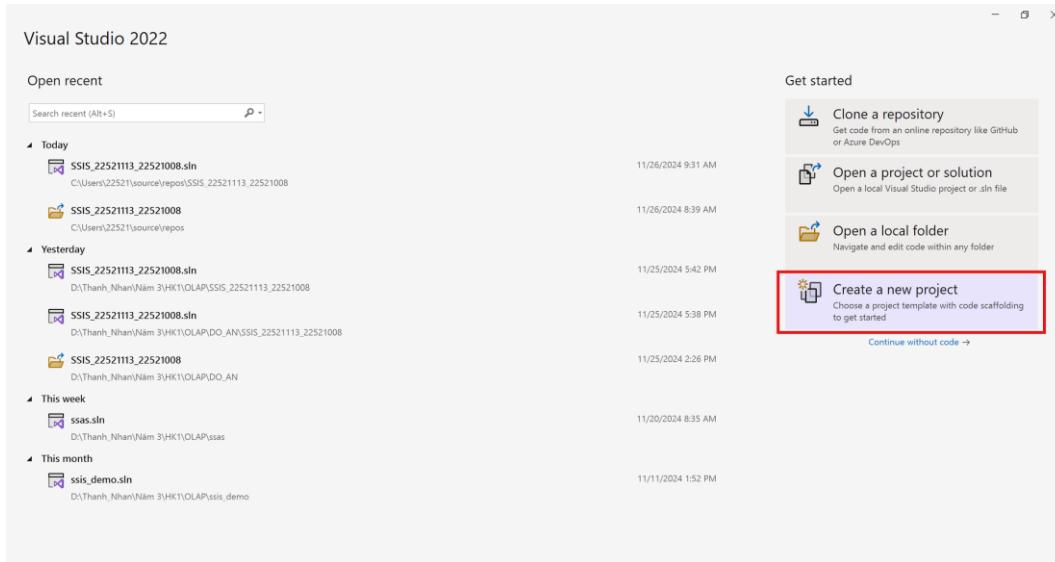


Hình 2. 258. Lược đồ sau khi hoàn thành.

3. PHÂN TÍCH DỮ LIỆU TRONG KHO SSAS

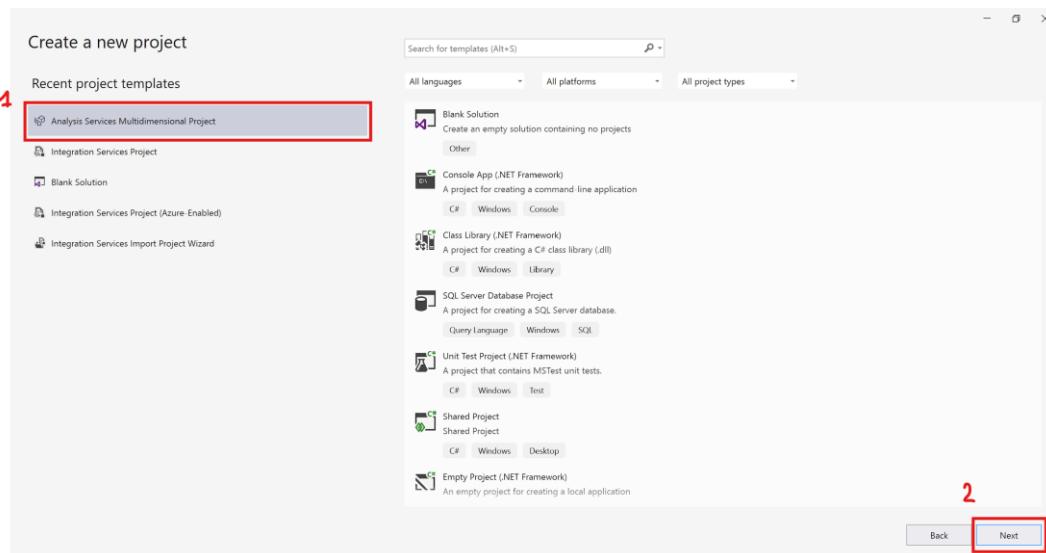
3.1 Tạo mới project SSAS:

Bước 1: Mở Visual Studio 2022, chọn Create a new project.



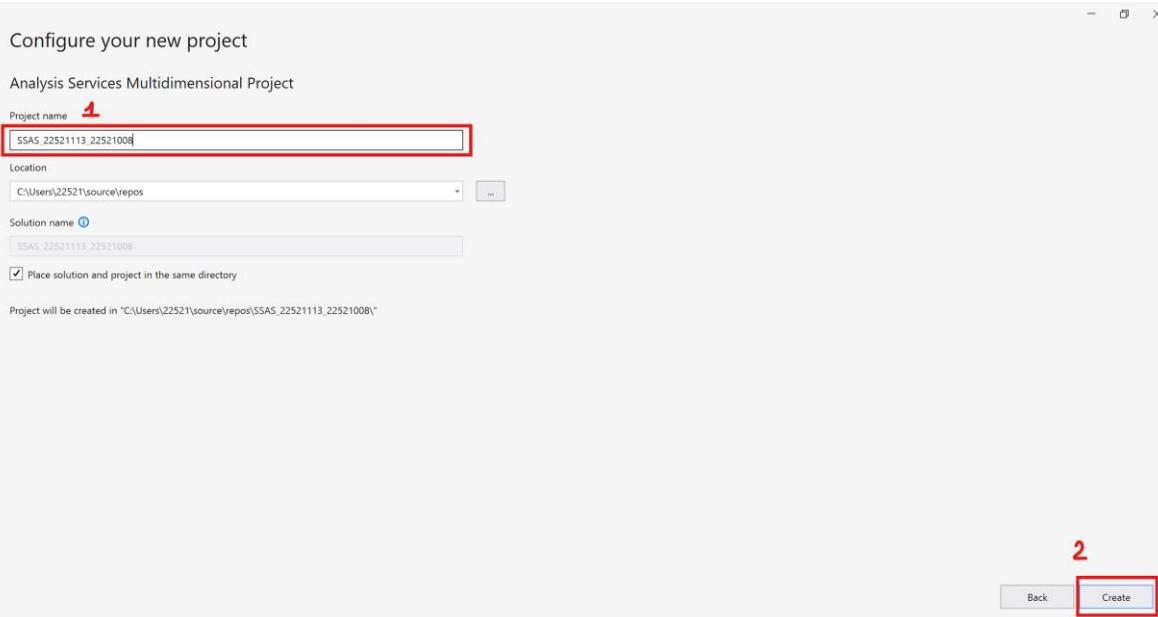
Hình 3. 1. Chọn Create a new project từ màn hình bắt đầu của Visual Studio 2022.

Bước 2: Chọn project template là **Analysis Services Multidimensional Project**, sau đó chọn **New**.



Hình 3. 2. Chọn project template Analysis Services Multidimensional Project.

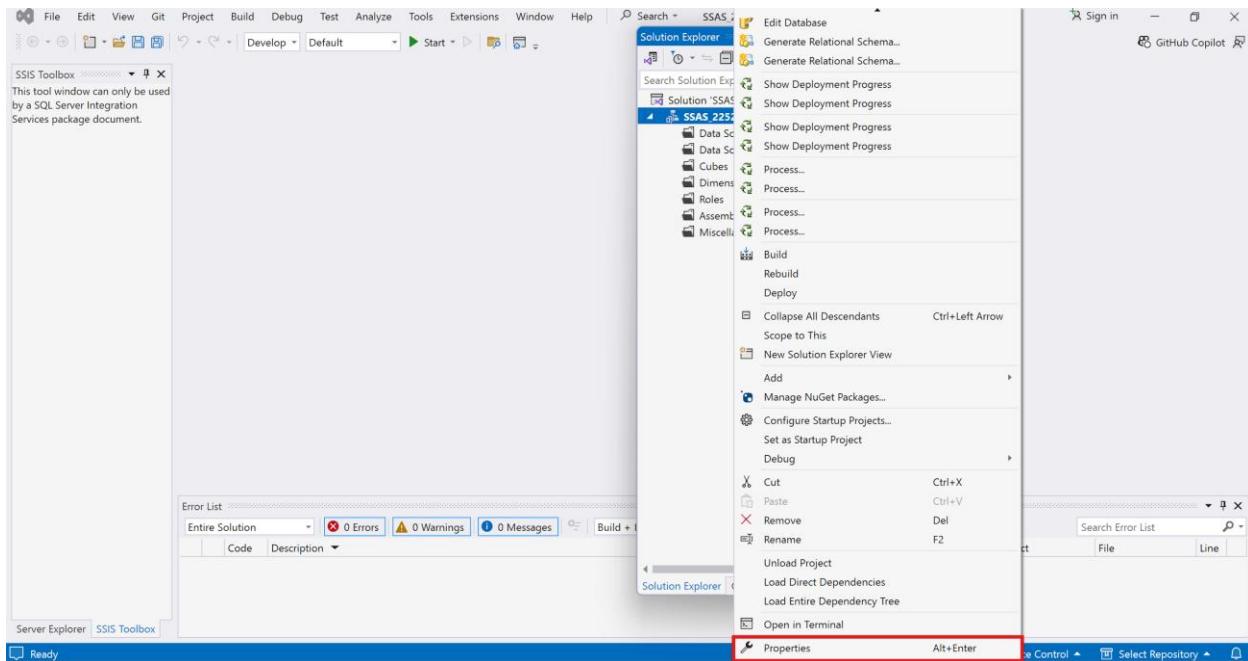
Bước 3: Đặt tên project là **SSAS_22521113_22521008**, chọn nơi lưu sau đó nhấn **Create** để tạo project.



Hình 3. 3. Đặt tên và chọn nơi lưu project.

3.1.1 Xác định dữ liệu nguồn (Data Sources)

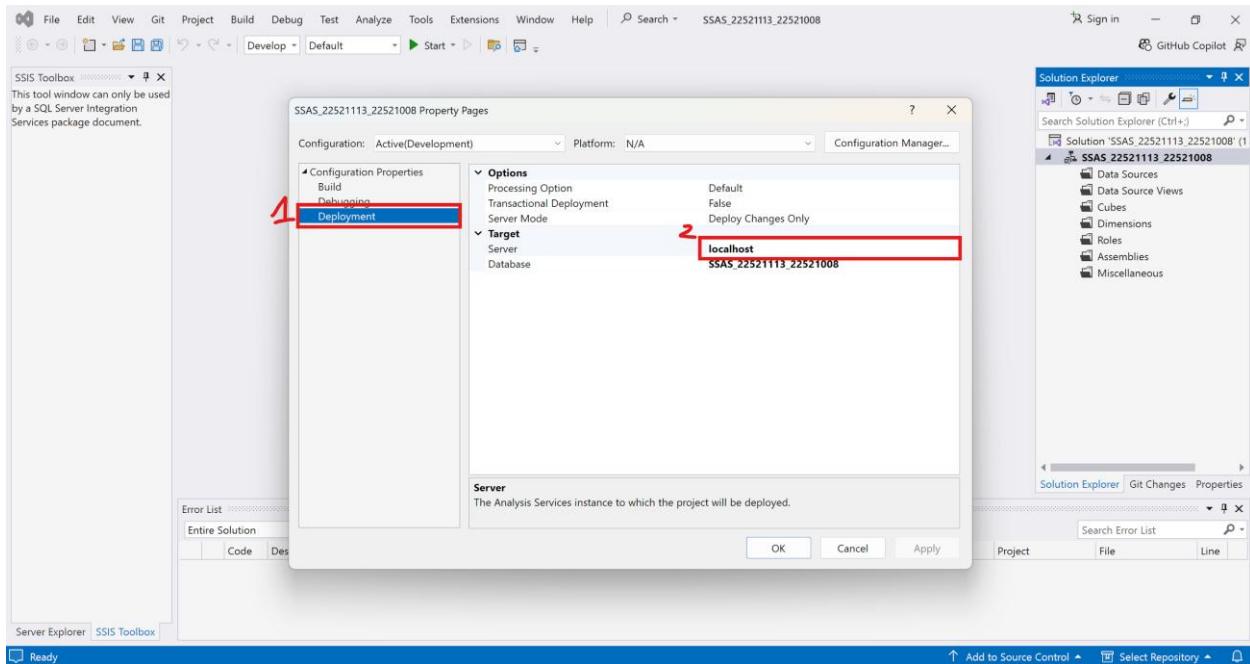
Bước 1: Trong **Solution Explorer**, click chuột phải vào tên project là **SSAS_22521113_22521008**, chọn mũi tên xuống dưới và click vào **Properties**.



Hình 3. 4. Chọn Properties của project.

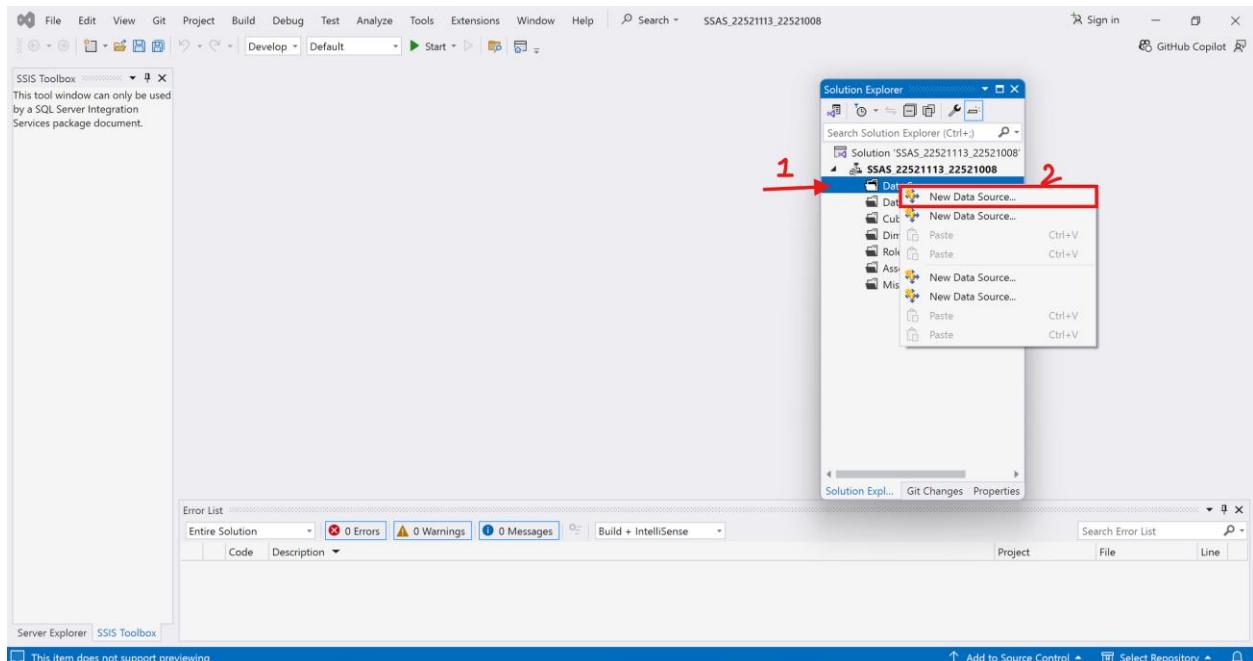
Bước 2: Trong cửa sổ hiện lên. Chọn **Deployment** (bên góc trái). Ở phần **Target**, đổi Server từ **localhost** sang tên server của máy , ở đây là **ASUS_TN**. Nhấn OK để lưu và đóng cửa sổ.

IS217 – Kho dữ liệu và OLAP



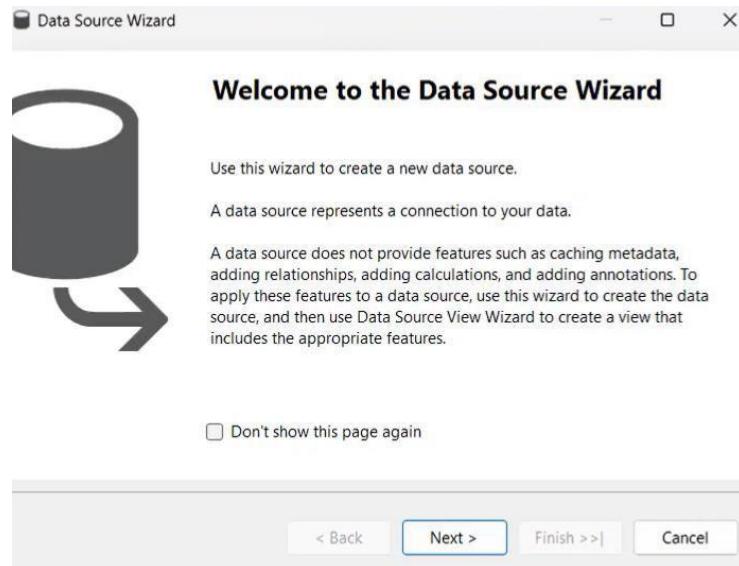
Hình 3.5 Đổi tên server để deploy

Bước 3: Tại Solution Explorer, ta click chuột phải vào thư mục Data Sources và chọn New Data Source.



Hình 3. 6. Cửa sổ Data Source Wizard hiện lên.

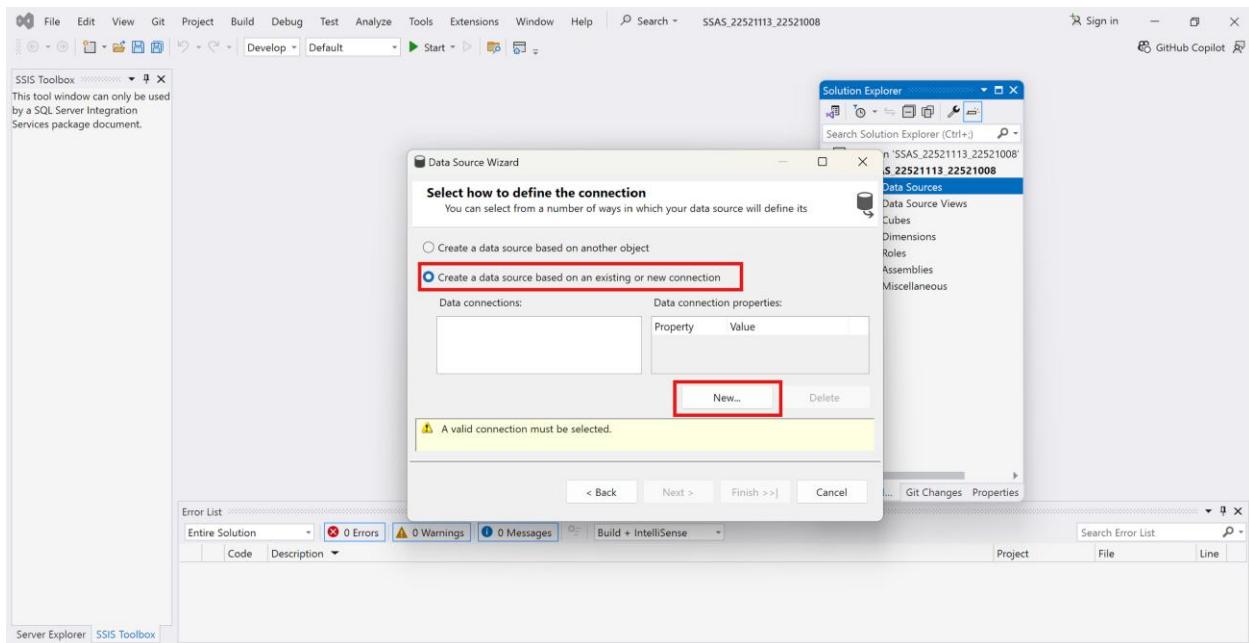
Bước 4: Cửa sổ Data Source Wizard hiện lên. Nhấn Next để tiếp tục.



Hình 3. 7. Cửa sổ Data Source Wizard hiện lên.

Bước 5: Chọn Create a data source based on an existing or new connection.

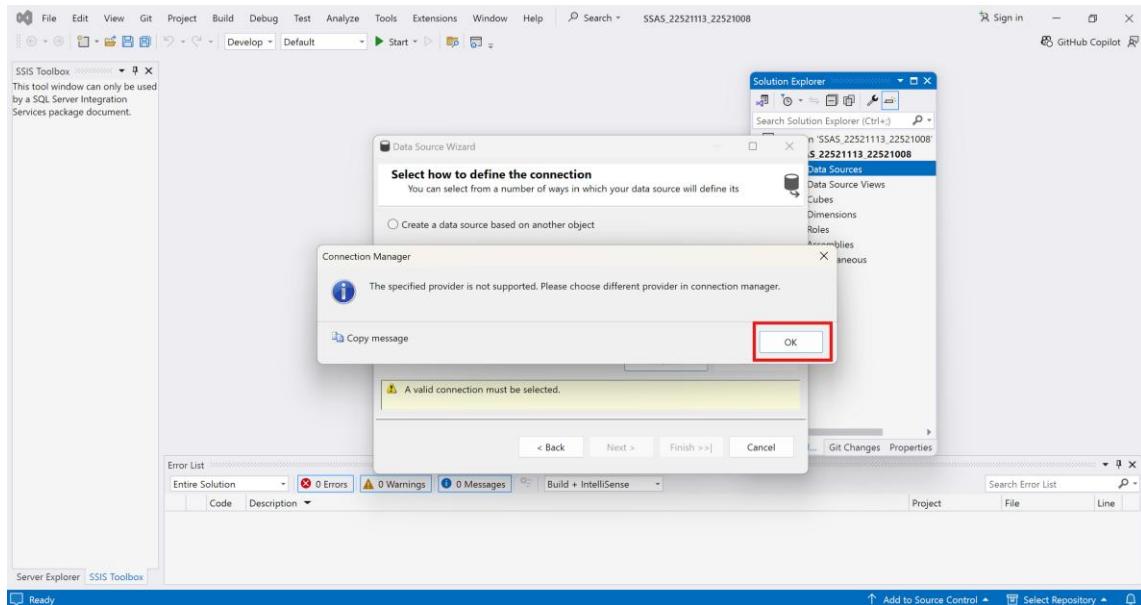
Sau đó chọn New để tạo kết nối với cơ sở dữ liệu đã được tạo từ quá trình SSIS.



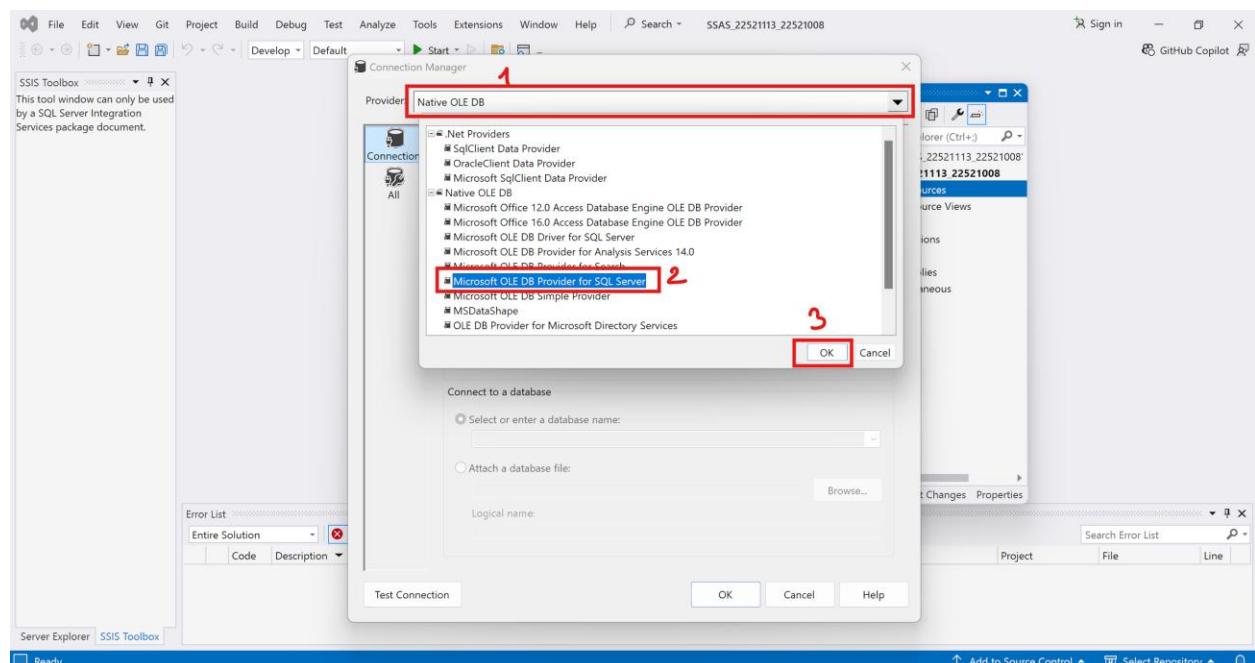
Hình 3. 8. Tạo kết nối với cơ sở dữ liệu được tạo từ quá trình SSIS.

Thông báo hiện lên yêu cầu chọn một nhà cung cấp khác từ danh sách các nhà cung cấp có sẵn cho **connection manager**. Chọn OK.

IS217 – Kho dữ liệu và OLAP



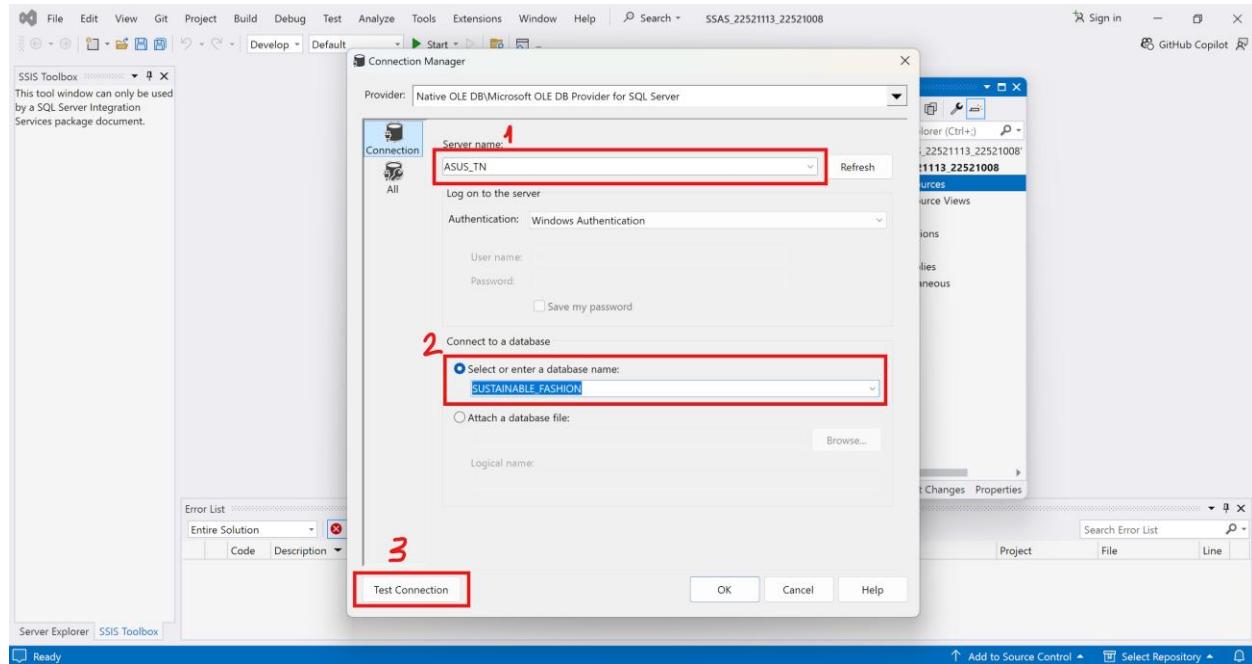
Bước 6: Chọn provider là Microsoft OLE DB Provider for SQL Server và bấm OK.



Hình 3. 9. Chon Provider

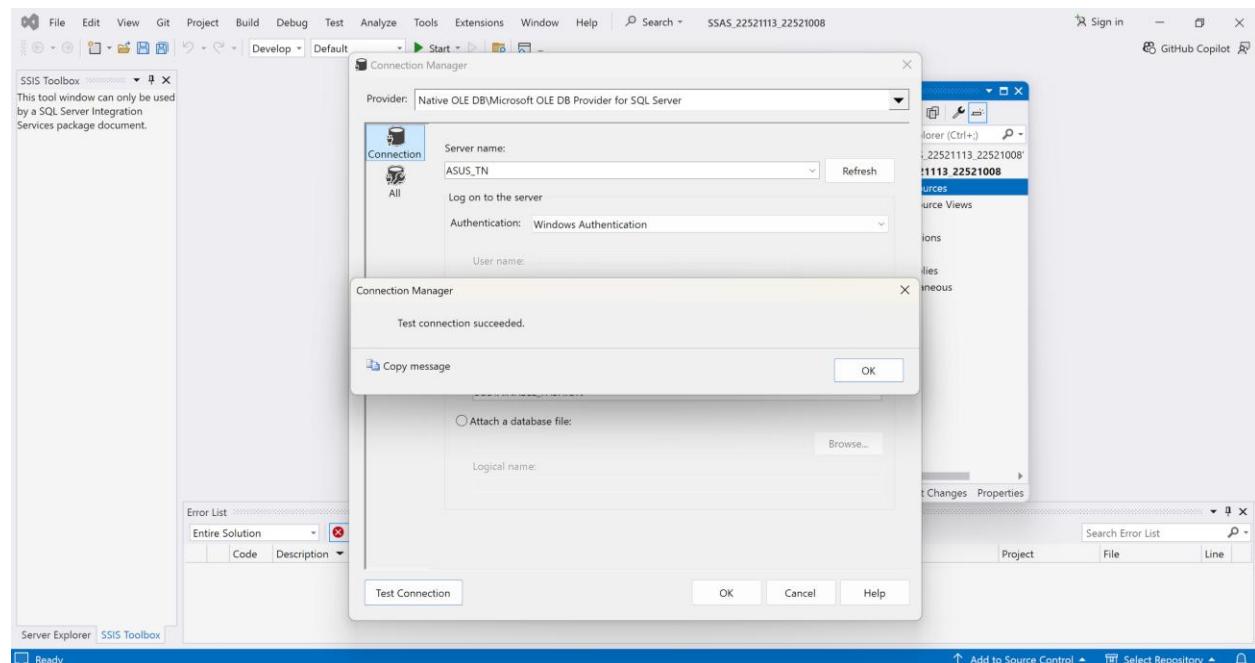
Bước 7: Nhập tên server và chọn database đã tạo (**SUSTAINABLE_FASHION**). Chọn Test Connection để kiểm tra kết nối.

IS217 – Kho dữ liệu và OLAP



Hình 3. 10. Chon server và Database.

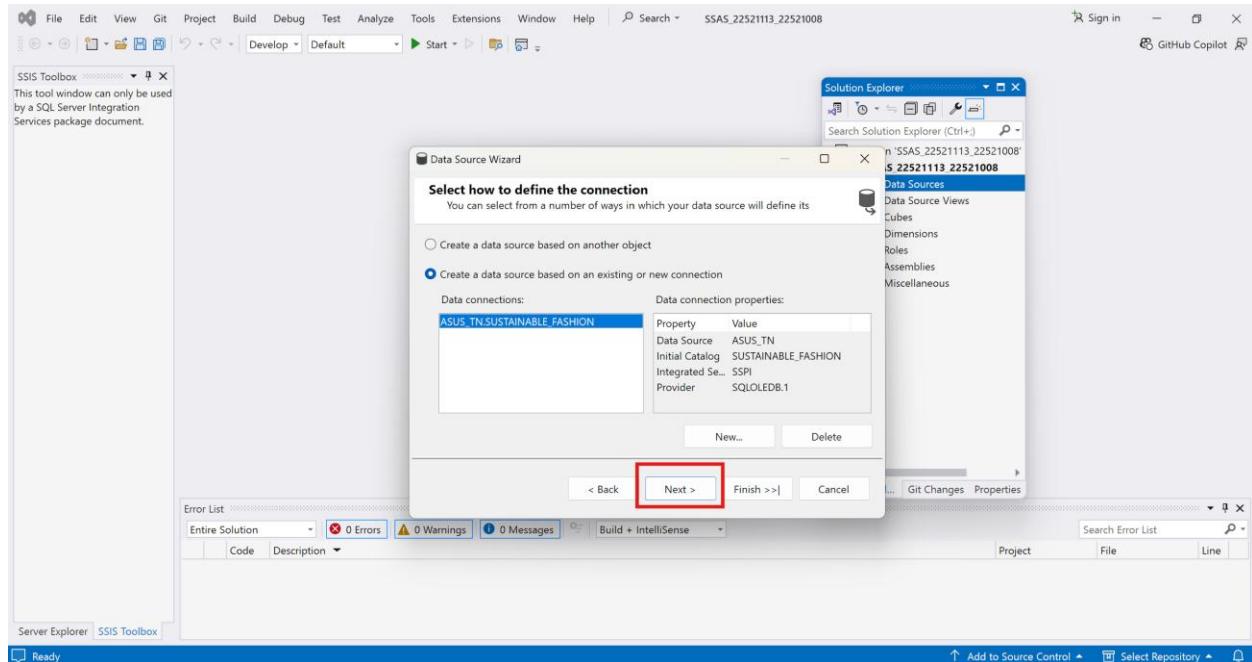
Kết nối thành công:



Hình 3. 11. Kết nối thành công

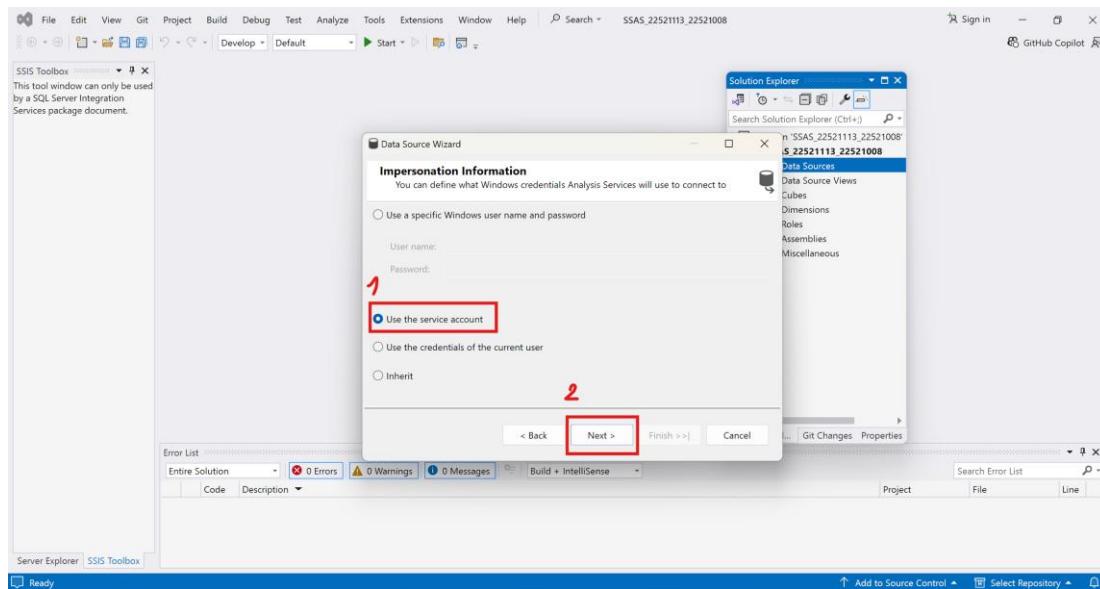
Bước 8: Chọn Data Connection vừa tạo, nhấn Next để tiếp tục.

IS217 – Kho dữ liệu và OLAP



Hình 3. 12. Chọn Data Connection đã tạo.

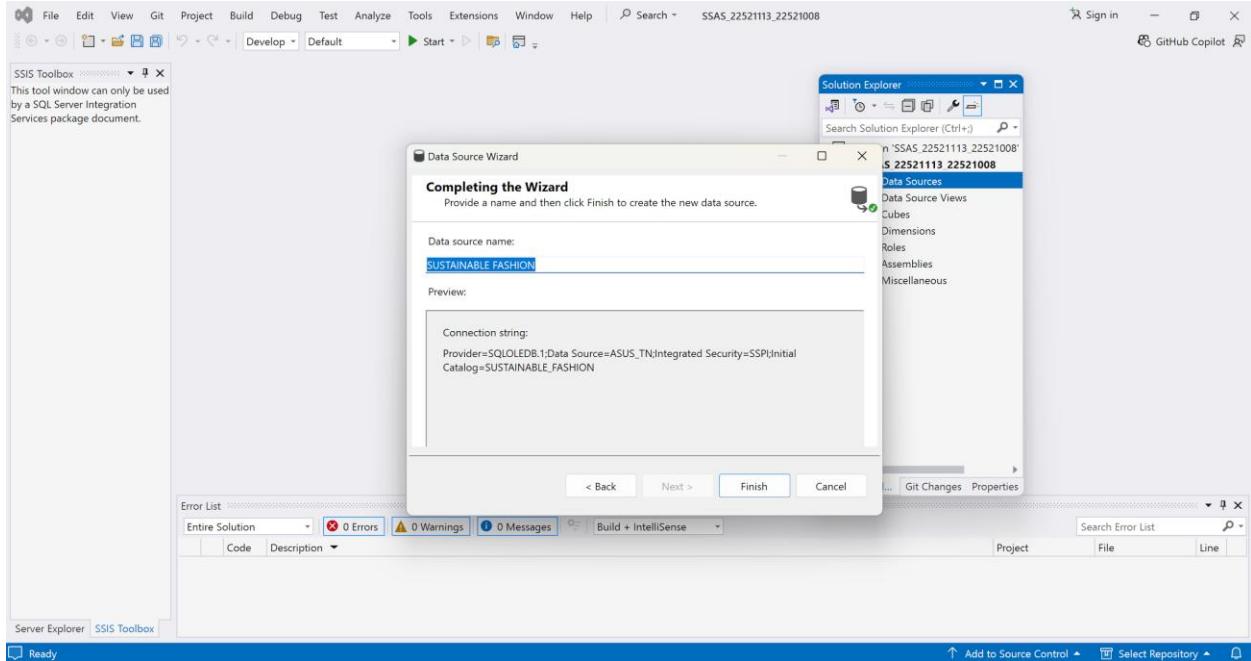
Bước 9: Ở bước Impersonation Information, chọn Use the service account. Sau đó chọn Next để tiếp tục.



Hình 3. 13. Chọn Use the service account.

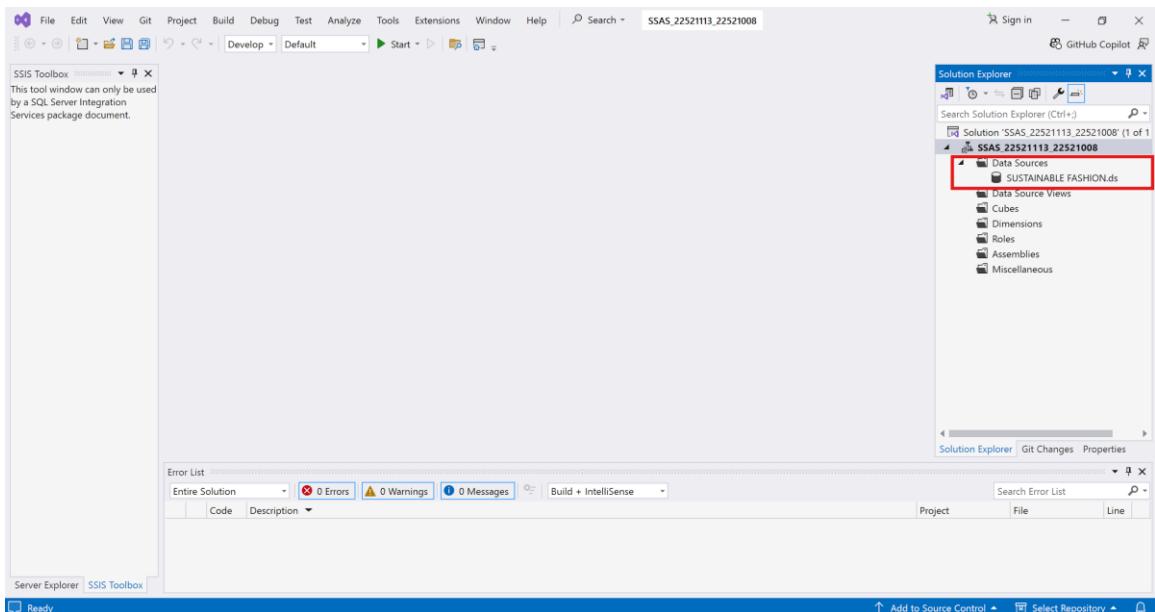
Bước 10: Nhấn Finish để hoàn tất quá trình xác định Data Source.

IS217 – Kho dữ liệu và OLAP



Hình 3. 14. Hoàn tất quá trình xác định Data Source.

Nếu thêm Data Source thành công, ta sẽ thấy tên Data Source vừa tạo (**SUSTAINABLE_FASHION**) hiển thị trong mục Data Source.

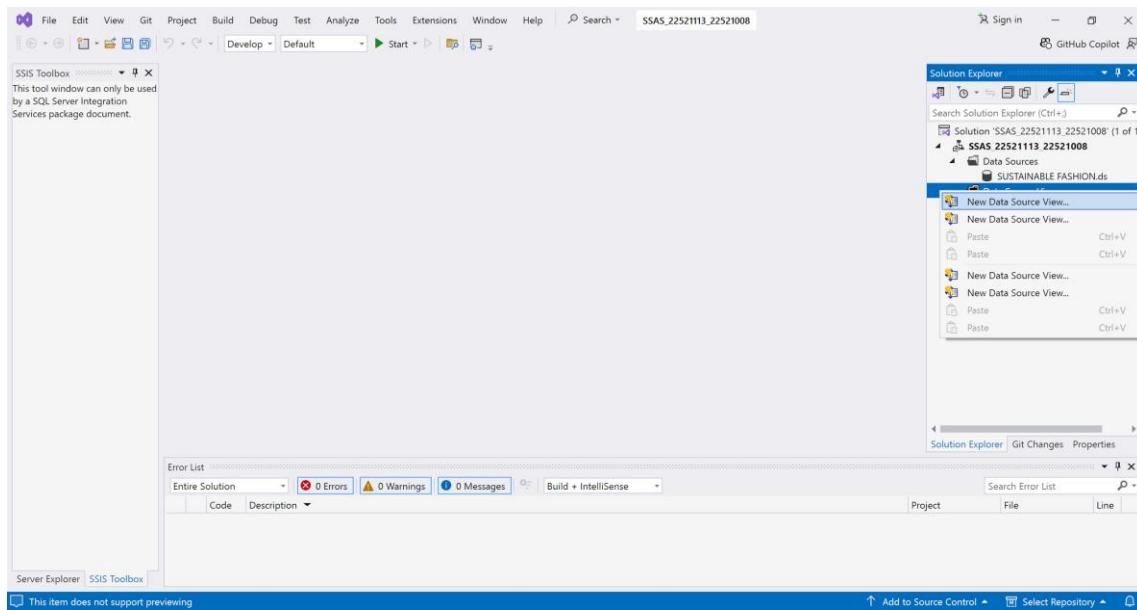


Hình 3. 15. Hoàn tất quá trình xác định Data Source.

3.1.2 Xác định khung nhìn dữ liệu nguồn (Data Source Views):

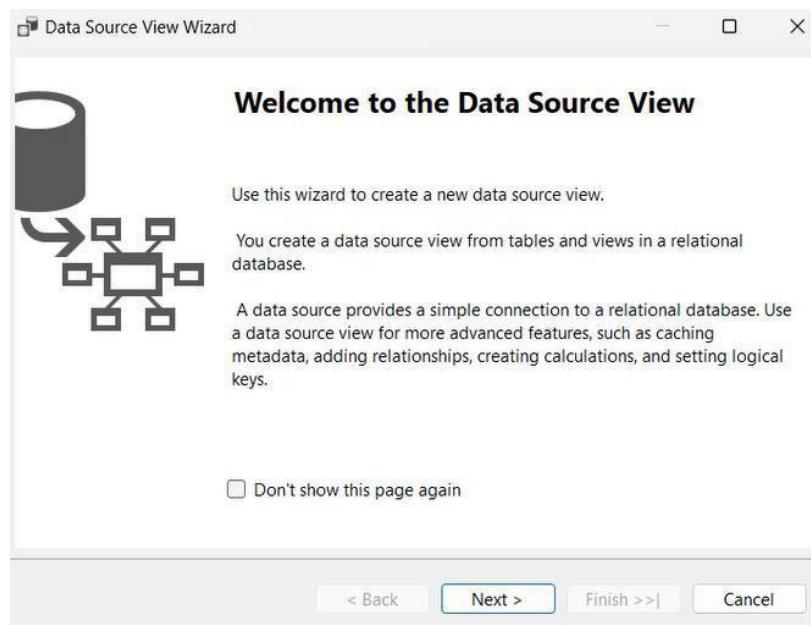
Bước 1: Trong Solution Explorer, click chuột phải vào Data Source Views và

chọn New Data Source View.



Hình 3. 16. Chọn New Data Source View từ Data Source Views.

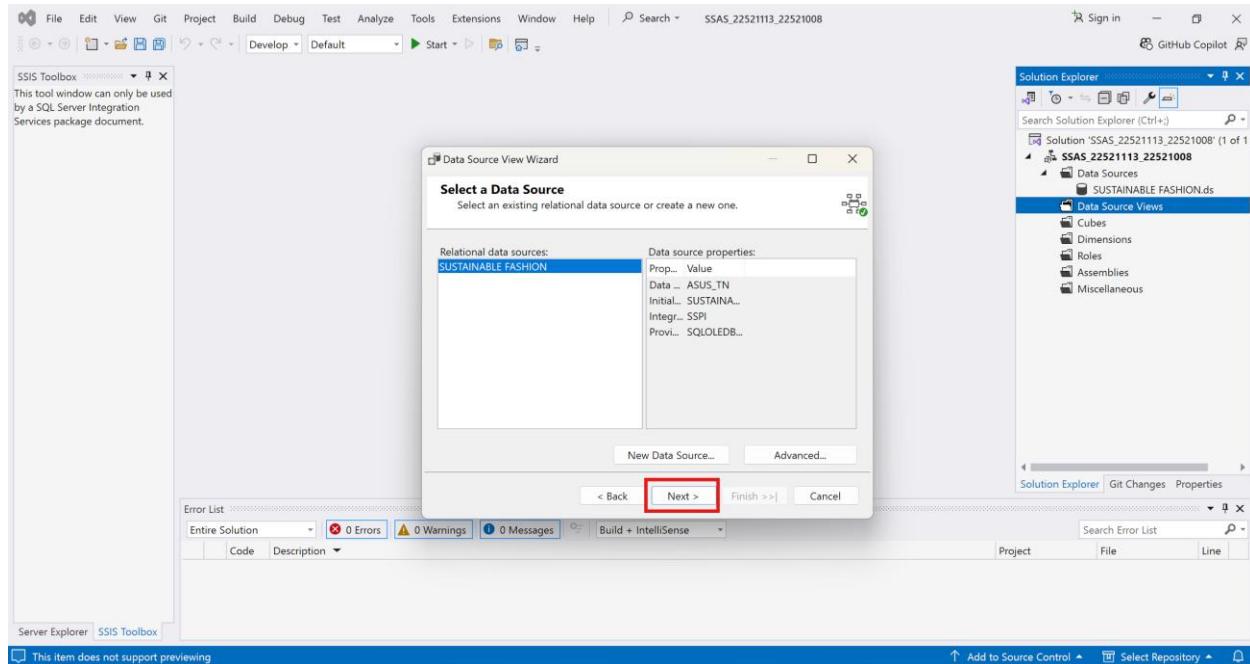
Bước 2: Cửa sổ **Data Source View Wizard** hiện lên. Nhấn **Next** để tiếp tục.



Hình 3. 17. Cửa sổ Data Source View Wizard hiện lên.

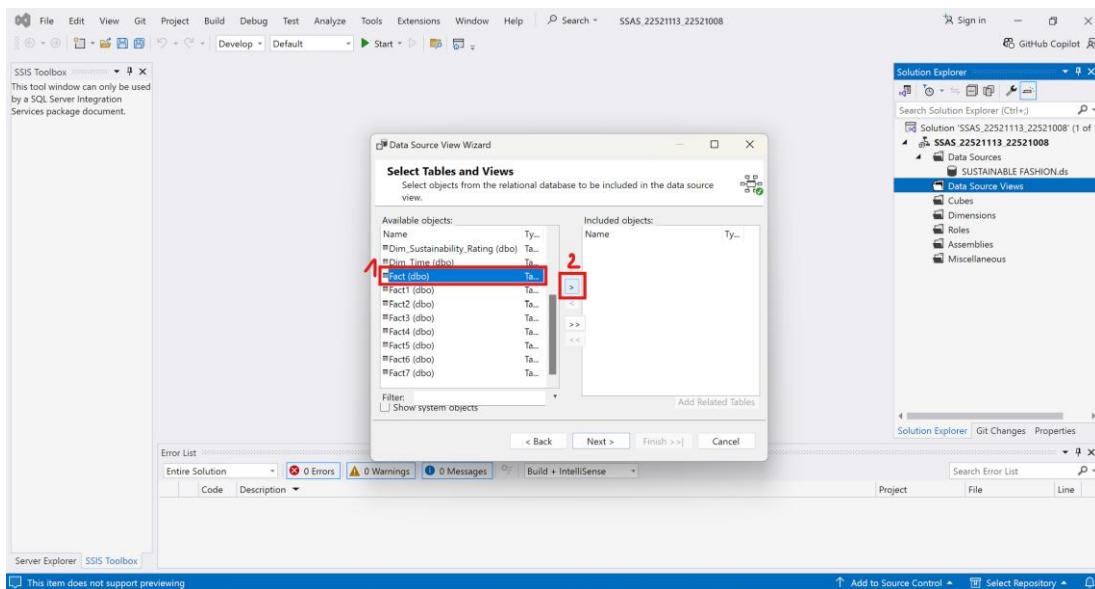
Bước 3: Chọn Data Source vừa tạo, nhấn **Next** để tiếp tục.

IS217 – Kho dữ liệu và OLAP



Hình 3. 18. Chọn Data Source đã tạo.

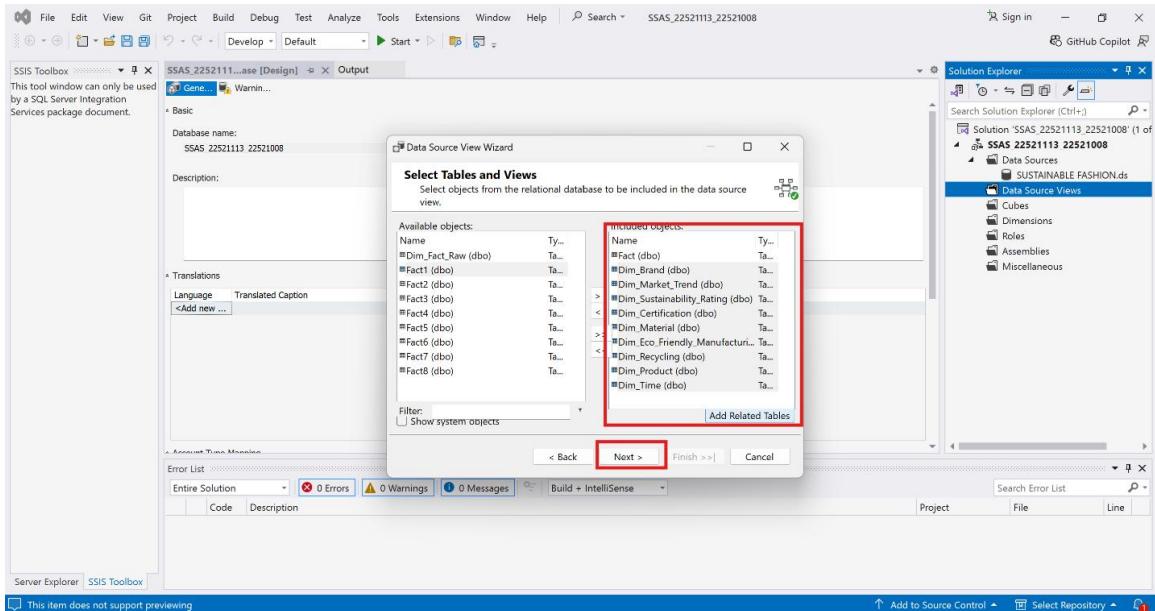
Bước 4: Chọn bảng Fact. Nhấn vào nút > để thêm vào Data Source View.



Hình 3. 19. Thêm bảng Fact vào Data Source View.

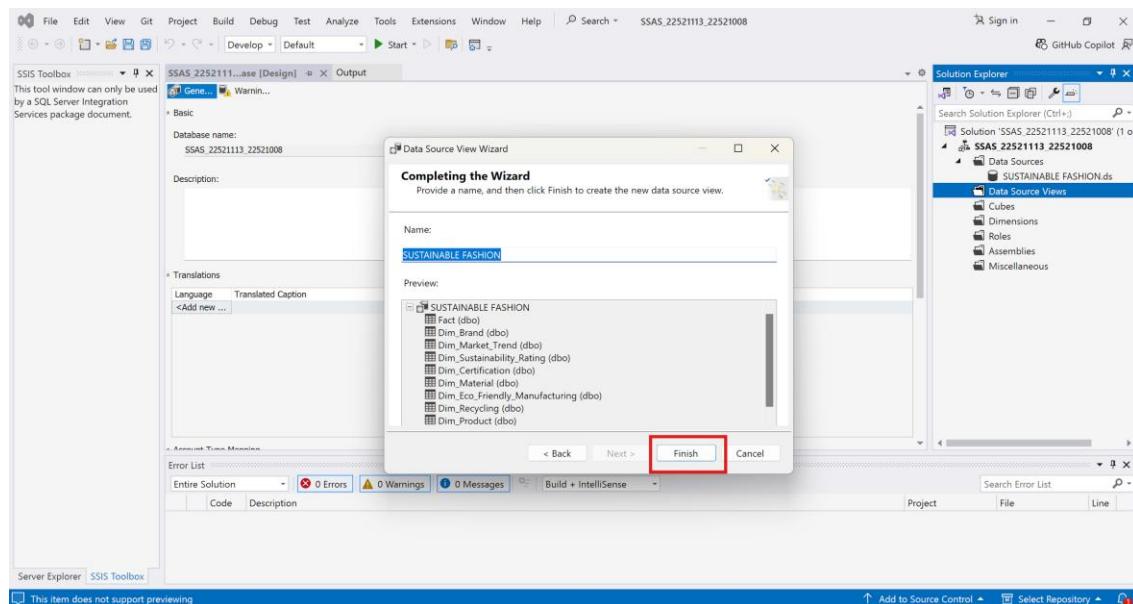
Bước 5: Chọn Add Related Tables, tiến hành thêm các bảng Dim vào Data Source View. Nhấn Next để tiếp tục.

IS217 – Kho dữ liệu và OLAP



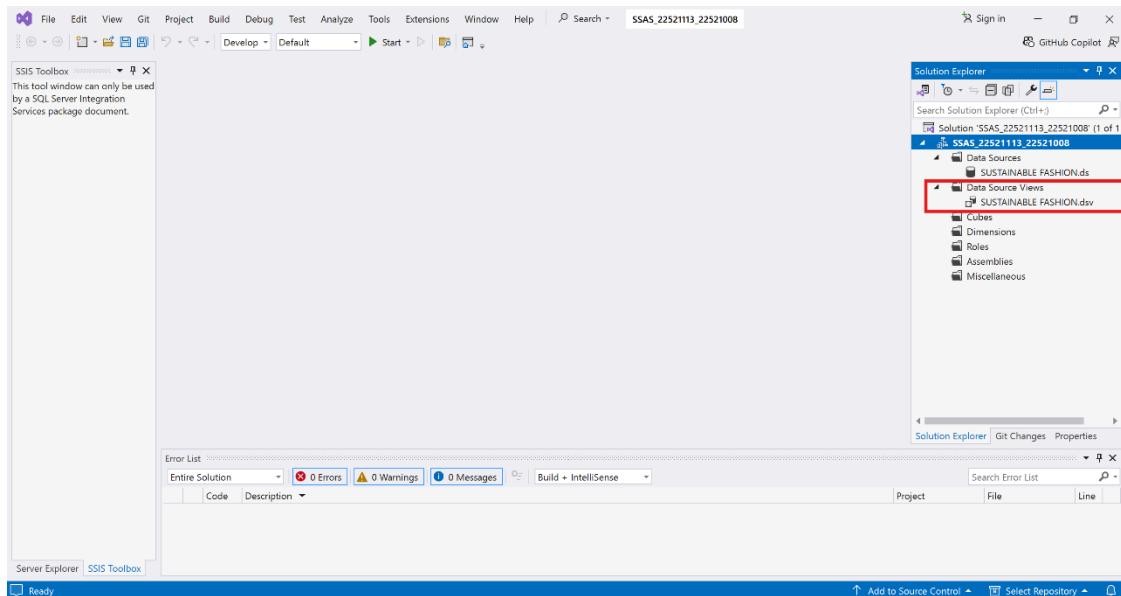
Hình 3. 20. Thêm bảng Dim vào Data Source View.

Bước 6: Nhấn Finish để hoàn tất quá trình xác định Data Source View.



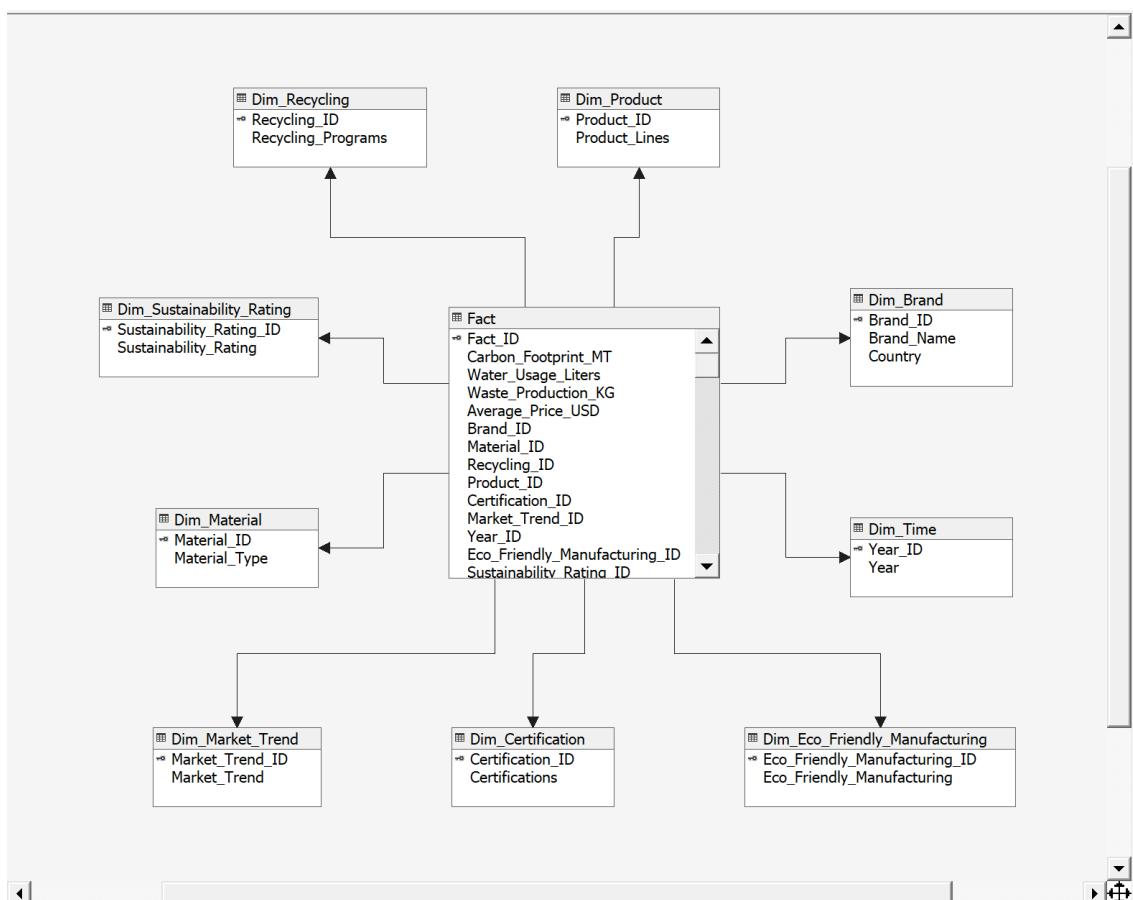
Hình 3. 21. Hoàn tất quá trình xác định Data Source View.

Nếu thêm Data Source View thành công, ta sẽ thấy SUSTAINABLE_FASHION trong mục Data Source View:



Hình 3. 22. Hoàn tất quá trình xác định Data Source View.

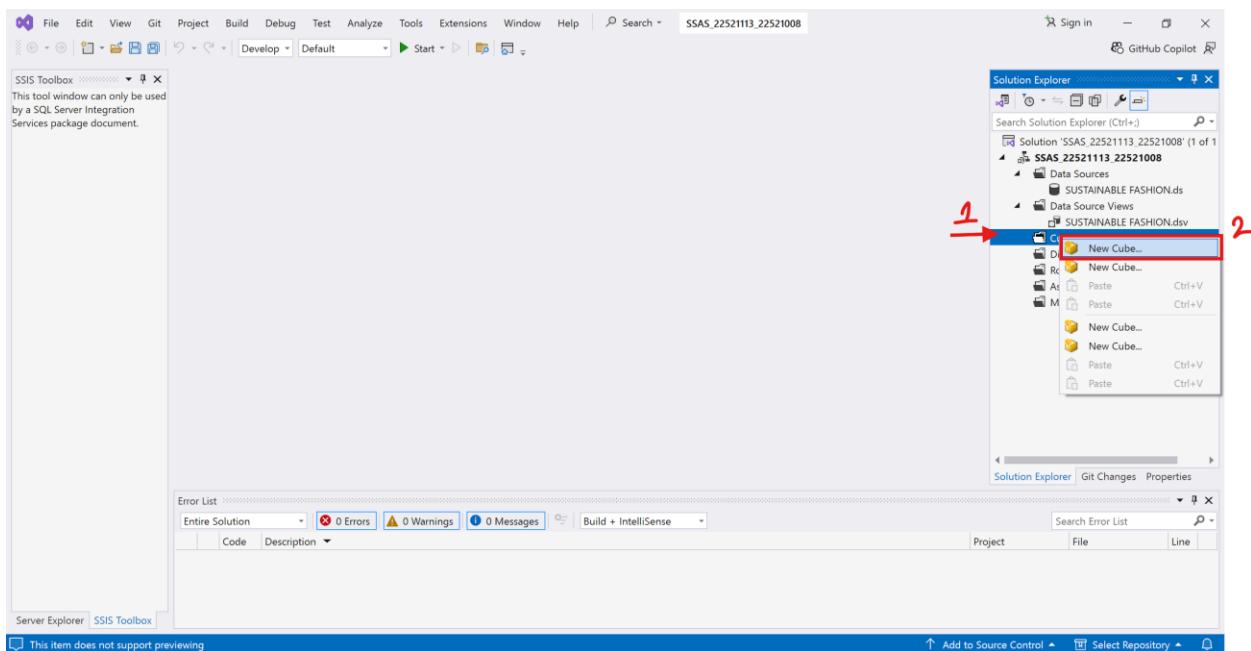
Bước 7: Kết quả của quá trình xác định Data Source View.



Hình 3. 23. Kết quả quá trình xác định Data Source View.

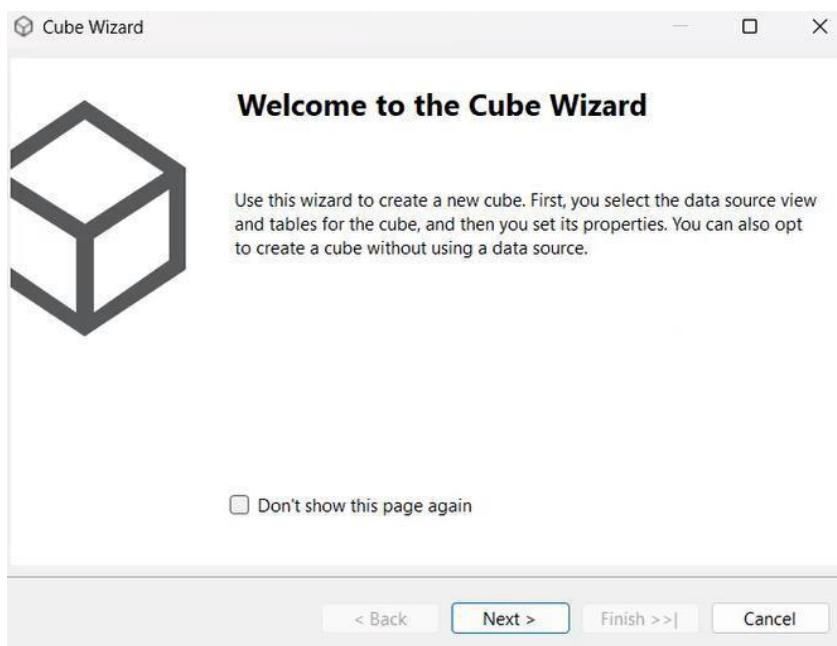
3.1.3 Xây dựng khối (Cube)

Bước 1: Trong Solution Explorer, click chuột phải vào **Cubes** và chọn **New Cube**.



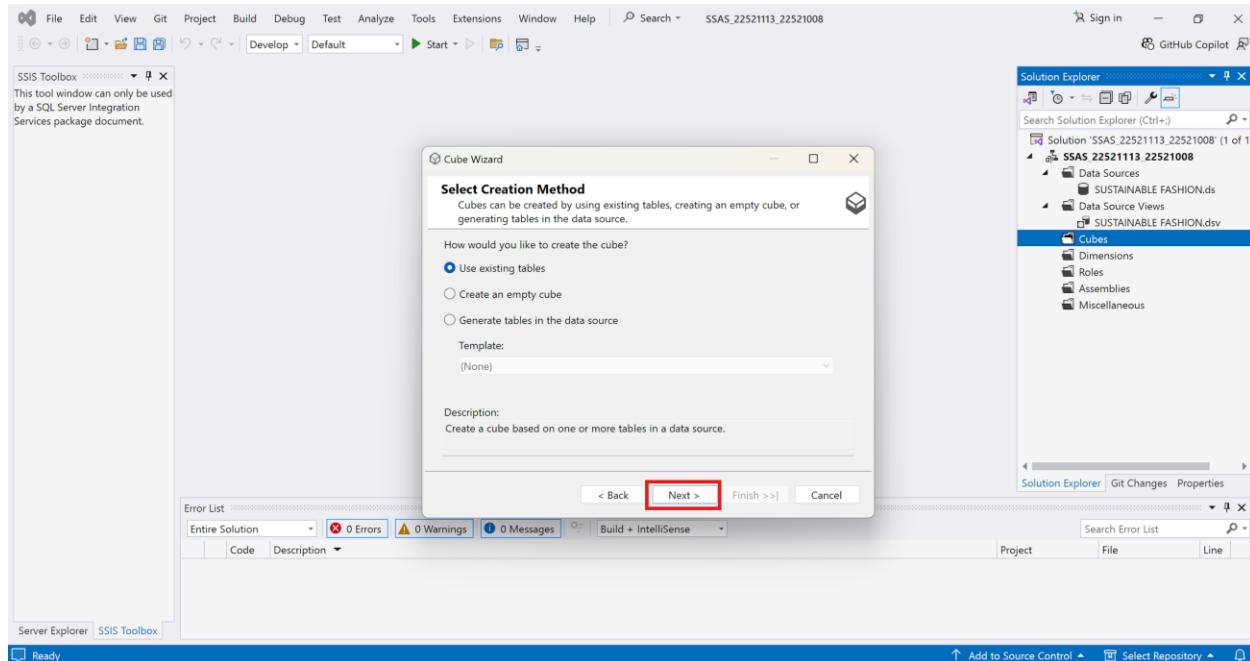
Hình 3. 24. Chọn New Cube từ Cubes.

Bước 2: Cửa sổ **Cube Wizard** hiện lên. Nhấn **Next** để tiếp tục.



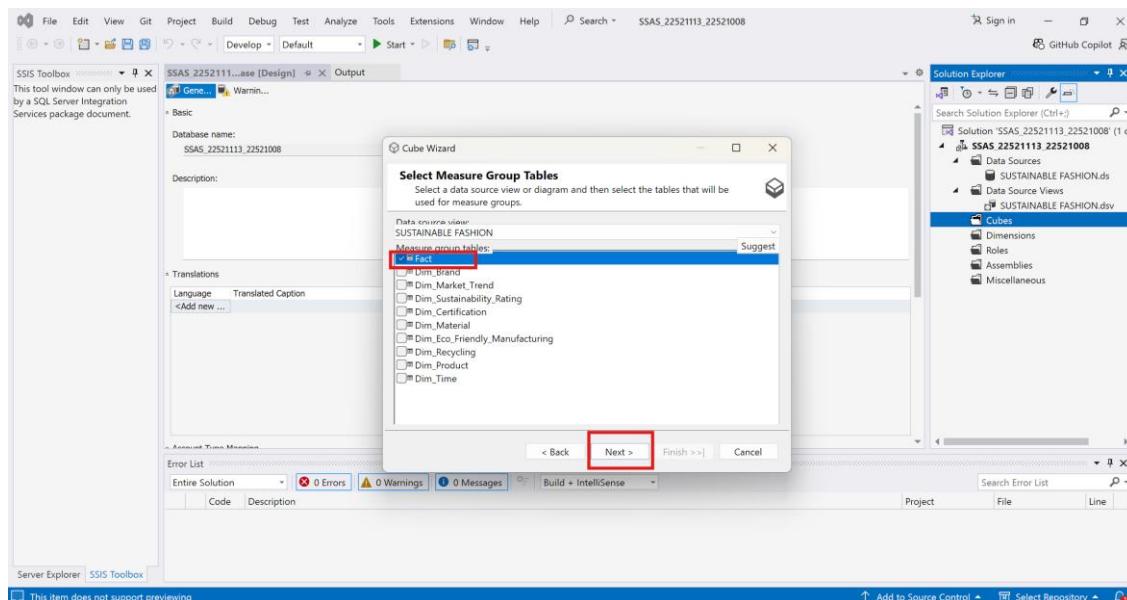
Hình 3. 25. Cửa sổ Cube Wizard hiện lên.

Bước 3: Chọn Use existing tables. Nhấn Next để tiếp tục.



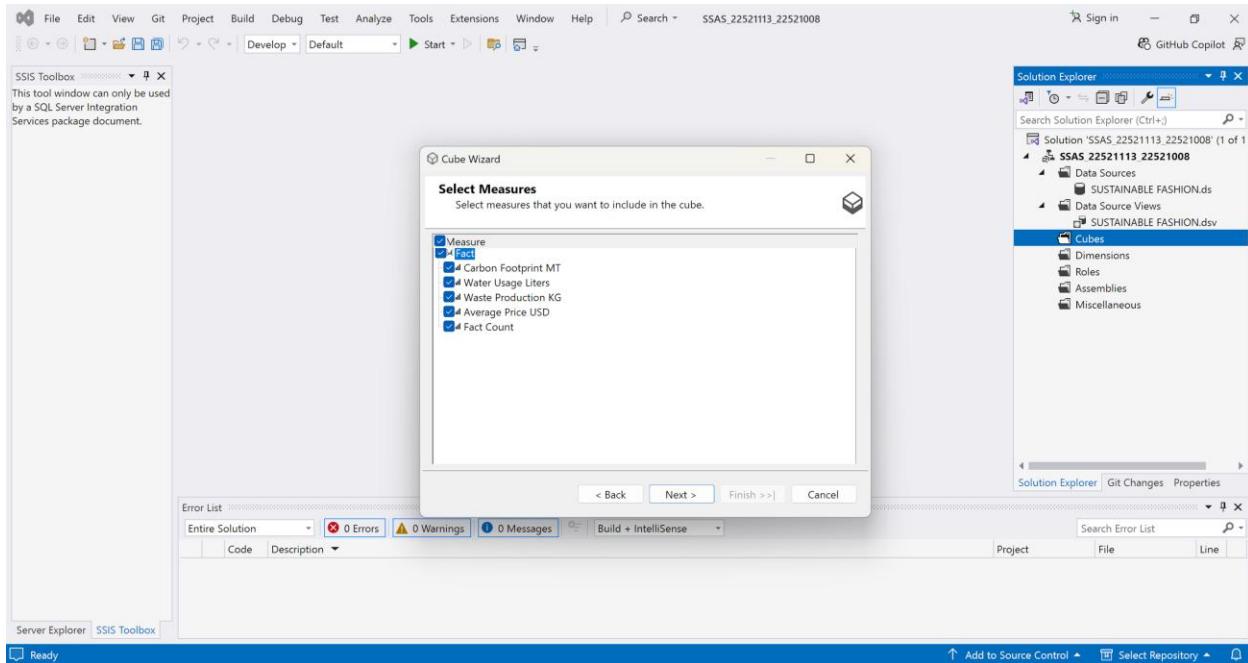
Hình 3. 26. Chọn Use existing tables.

Bước 4: Chọn bảng Fact để phân chia các measure group.



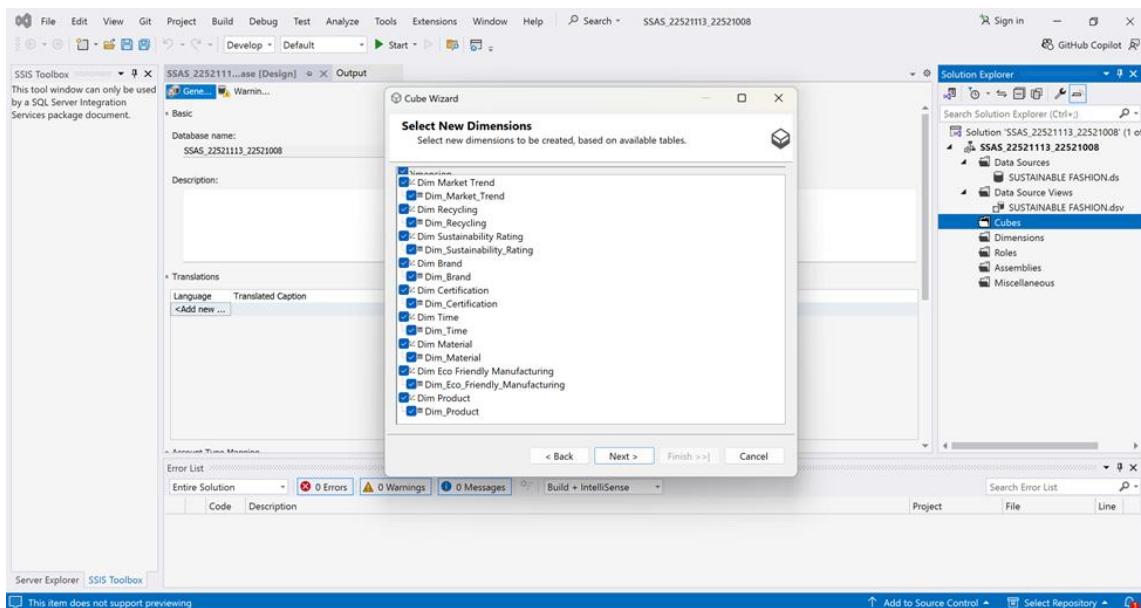
Hình 3. 27. Chọn bảng Fact để phân chia measures.

Bước 5: Chọn các độ đo để xuất. Nhấn Next để tiếp tục.



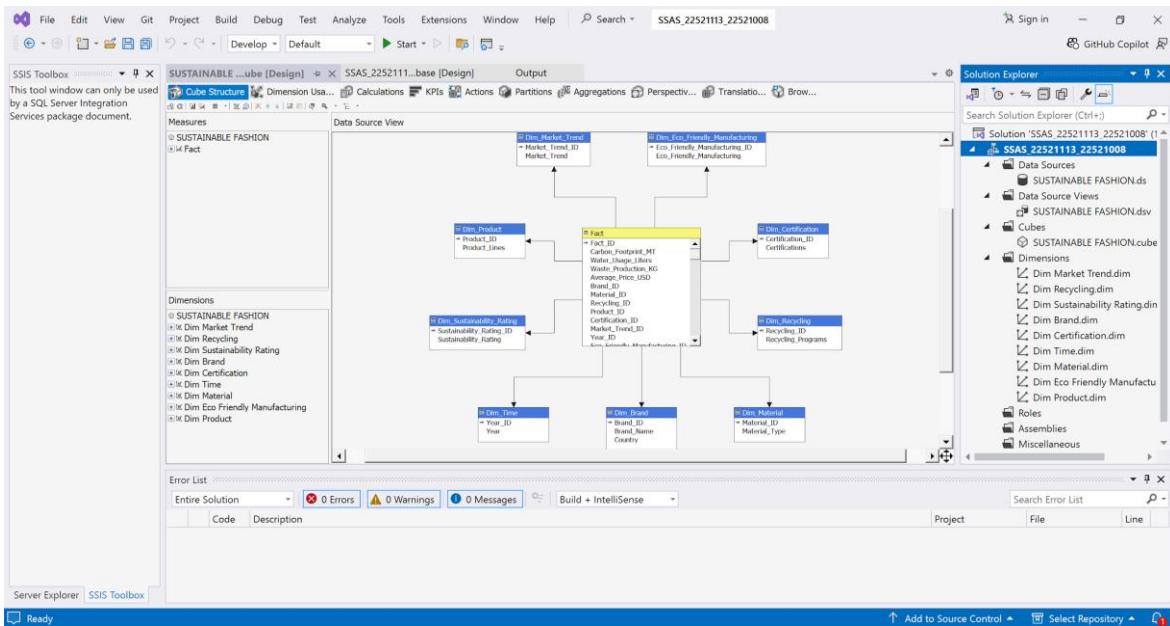
Hình 3. 28. Chọn độ đo để xuất.

Bước 6: Chọn các bảng Dim. Nhấn Next để tiếp tục. Và chọn Finish để kết thúc.



Hình 3. 29. Chọn các bảng Dim.

Bước 7: Kết quả của quá trình xây dựng khói (Cubes).

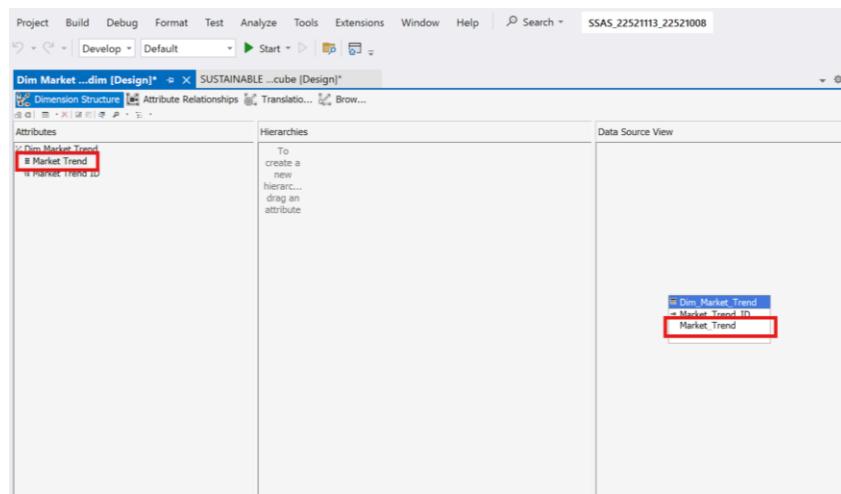


Hình 3. 30. Kết quả quá trình xây dựng cube.

3.1.4 Xác định các chiều (Dimensions)

Bước 1:

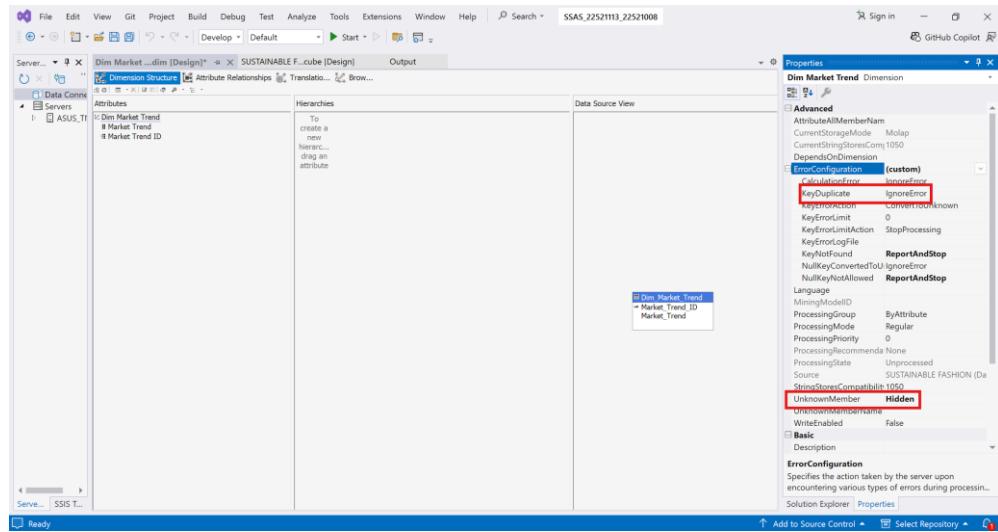
- Trong **Solution Explorer**, click vào **Dim Market_Trend.dim** từ **Dimensions**. Kéo thả thuộc tính *Market_Trend* từ **Data Source View** vào **Attributes**.



Hình 3. 31. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Market-Trend.

- Nháy chuột phải vào **Dim_Market_Trend** trong phần **Attributes**, chọn

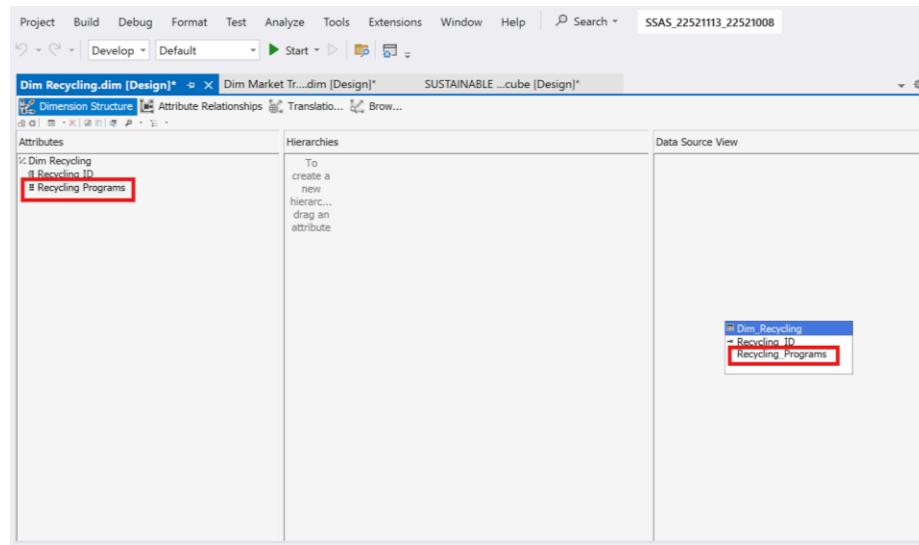
Properties. Trong cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration → KeyDuplicate là IgnoreError. Chỉnh sửa giá trị UnknownMember thành Hidden.



Hình 3.32 Cấu hình cho Dim Market Trend

Bước 2:

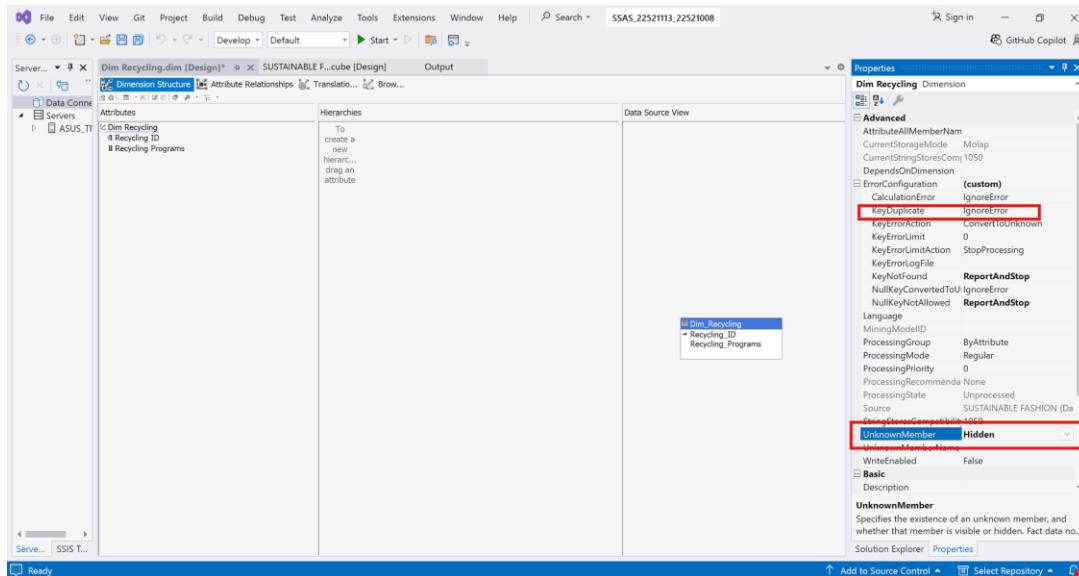
- Trong **Solution Explorer**, click vào **Dim Recycling.dim** từ **Dimensions**. Kéo thả thuộc tính *Recycling_Program* từ **Data Source View** vào **Attributes**.



Hình 3. 33. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Recycling.

- Nhấp chuột phải vào **Dim_Recycling** trong phần **Attributes**, chọn **Properties**.

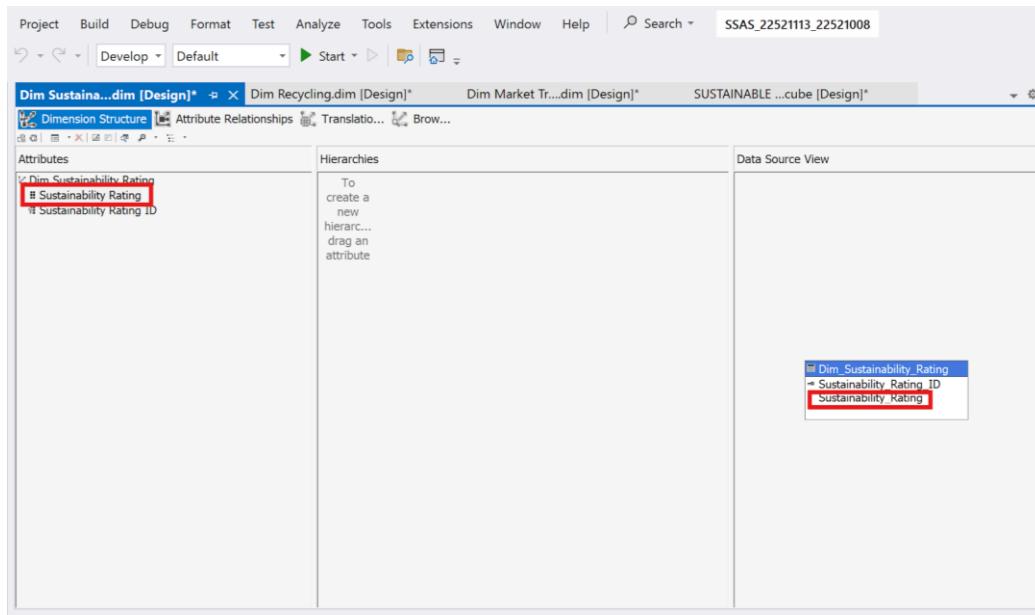
Trong cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration → KeyDuplicate là IgnoreError. Chỉnh sửa giá trị UnknownMember thành Hidden.



Hình 3.34 Cấu hình cho Dim Recycling

Bước 3:

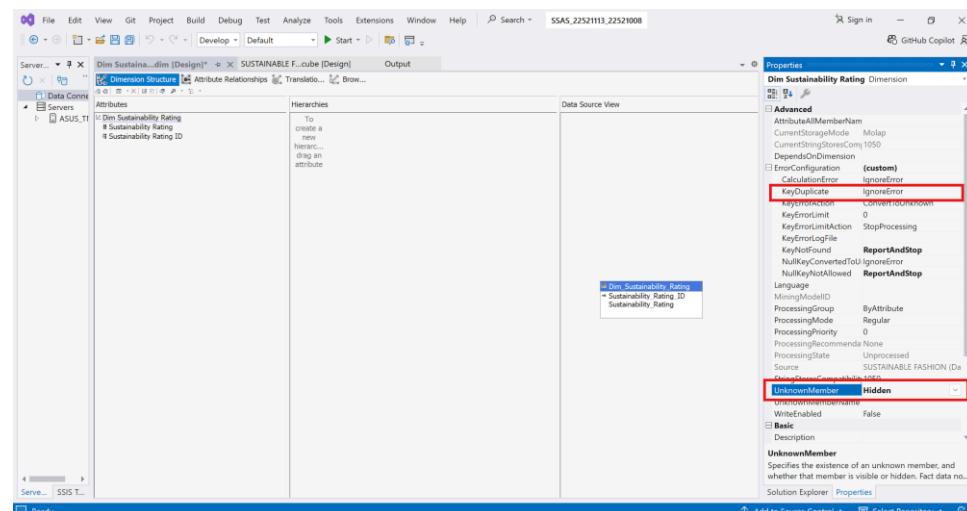
- Trong **Solution Explorer**, click vào **Dim Sustainability_Rating.dim** từ **Dimensions**. Kéo thả thuộc tính *Sustainability_Rating* từ **Data Source View** vào **Attributes**.



Hình 3. 35. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Sustainability Rating.

Bước 4:

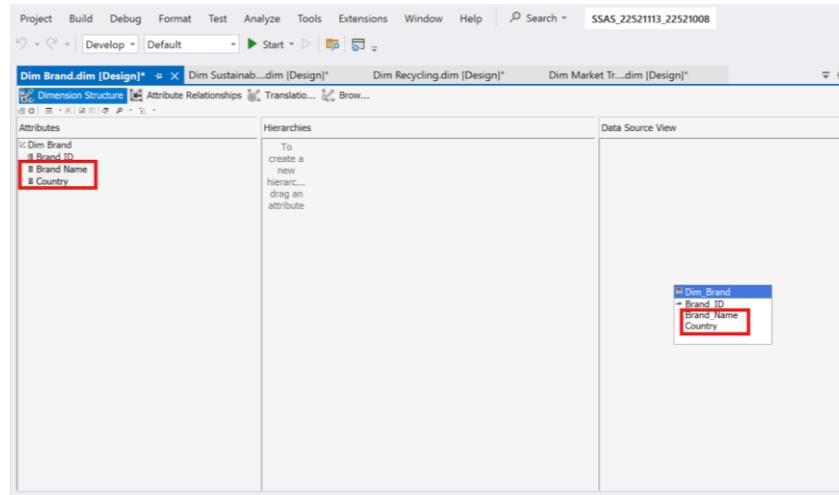
- Nháy chuột phải vào **Dim_Sustainability_Rating** trong phần **Attributes**, chọn **Properties**. Trong cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration → KeyDuplicate là IgnoreError. Chỉnh sửa giá trị UnknownMember thành Hidden.



Hình 3. 36. Cấu hình cho Dim Sustainability Rating.

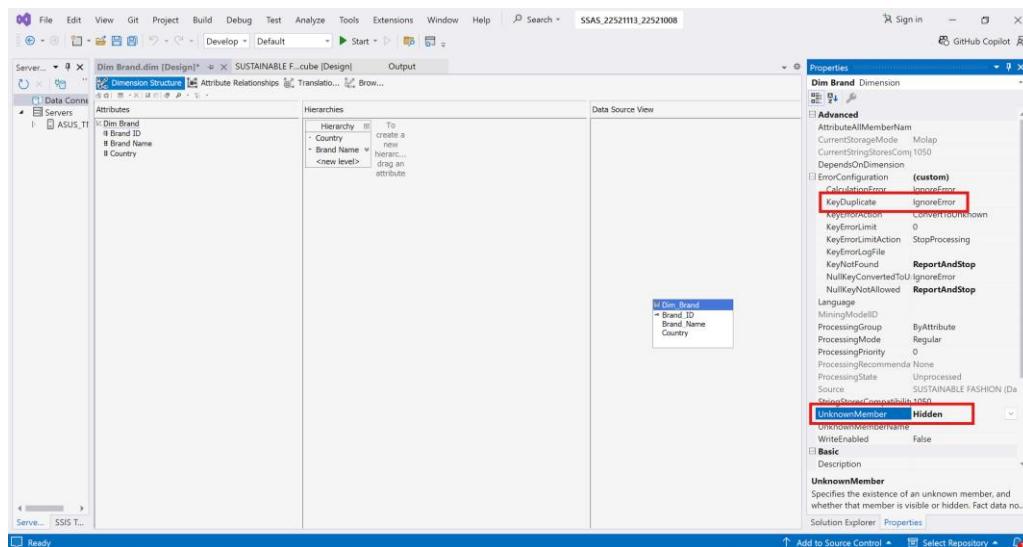
Bước 4:

- Trong **Solution Explorer**, click vào **Dim Brand.dim** từ **Dimensions**. Kéo thả các thuộc tính *Brand_Name*, *Country* từ **Data Source View** vào **Attributes**.



Hình 3. 37. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Brand

- Nháy chuột phải vào **Dim_Brand** trong phần **Attributes**, chọn **Properties**. Trong cửa sổ **Properties**, chỉnh sửa giá trị **ErrorConfiguration** → **KeyDuplicate** là **IgnoreError**. Chỉnh sửa giá trị **UnknownMember** thành **Hidden**.

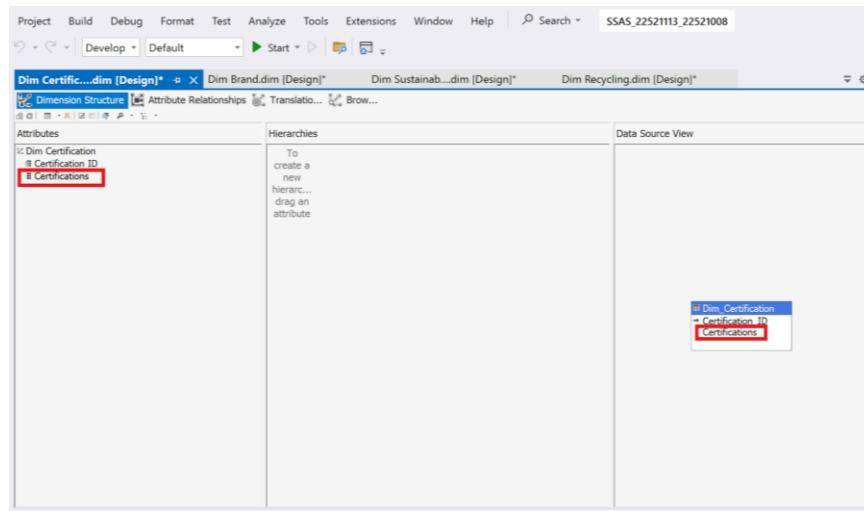


Hình 3. 38. Cấu hình cho Dim Brand.

Bước 5:

- Trong Solution Explorer, click vào Dim Certification.dim từ Dimensions.

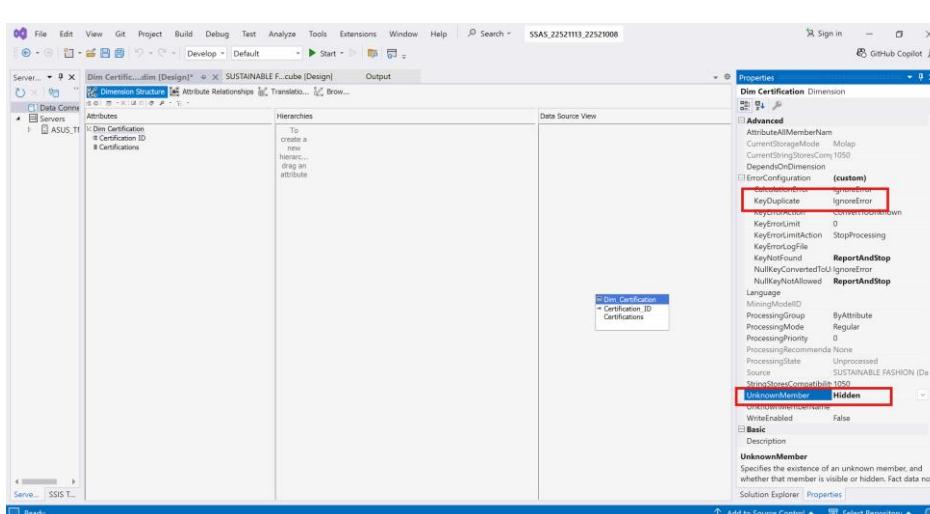
Kéo thả thuộc tính Certification từ Data Source View vào Attributes.



Hình 3. 39. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Certification

- Nhấp chuột phải vào **Dim_Certification** trong phần **Attributes**, chọn **Properties**.

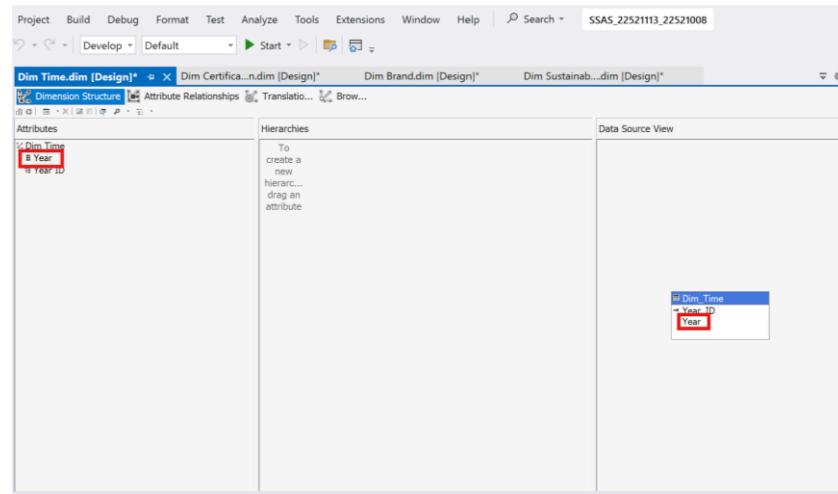
Trong cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration → KeyDuplicate là IgnoreError. Chỉnh sửa giá trị UnknownMember thành Hidden.



Hình 3. 40. Cấu hình cho Dim Certification

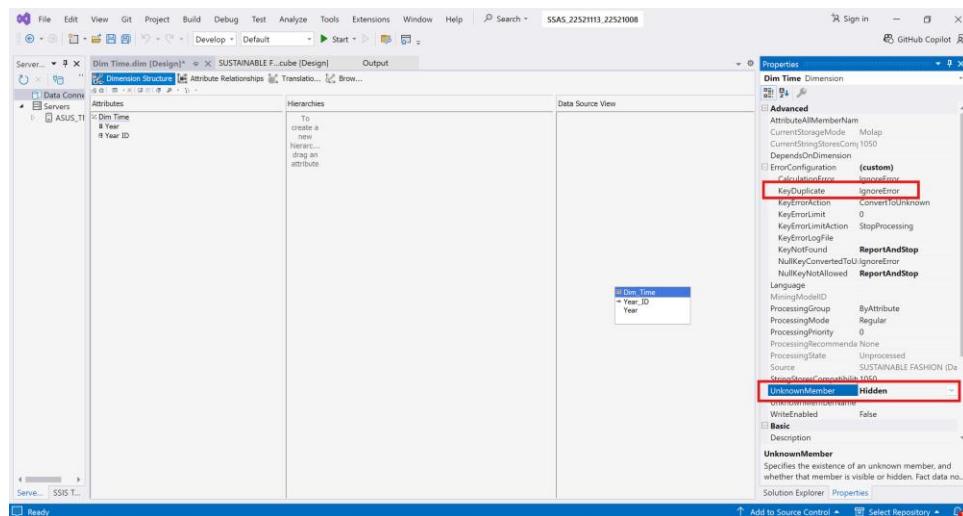
Bước 6:

- Trong **Solution Explorer**, click vào **Dim Time.dim** từ **Dimensions**. Kéo thả thuộc tính *Year* từ **Data Source View** vào **Attributes**.



Hình 3. 41. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Time.

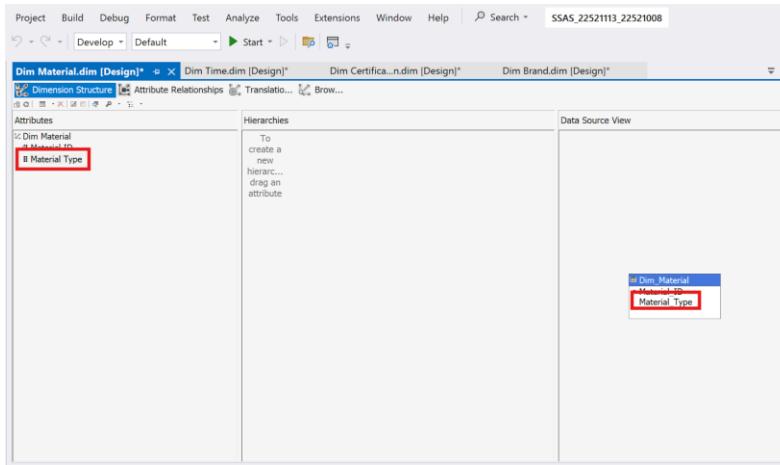
- Nháy chuột phải vào **Dim_Time** trong phần **Attributes**, chọn **Properties**. Trong cửa sổ **Properties**, chỉnh sửa giá trị **ErrorConfiguration** → **KeyDuplicate** là **IgnoreError**.
Chỉnh sửa giá trị **UnknownMember** thành **Hidden**.



Hình 3. 42. Cấu hình cho Dim Time.

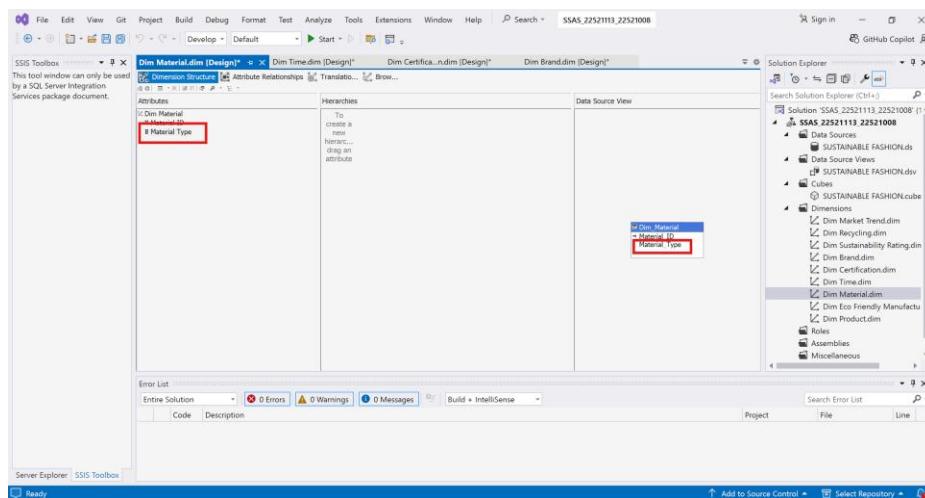
Bước 7:

- Trong **Solution Explorer**, click vào **Dim Material.dim** từ **Dimensions**. Kéo thả thuộc tính *Material_Type* từ **Data Source View** vào **Attributes**.



Hình 3. 43. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Material.

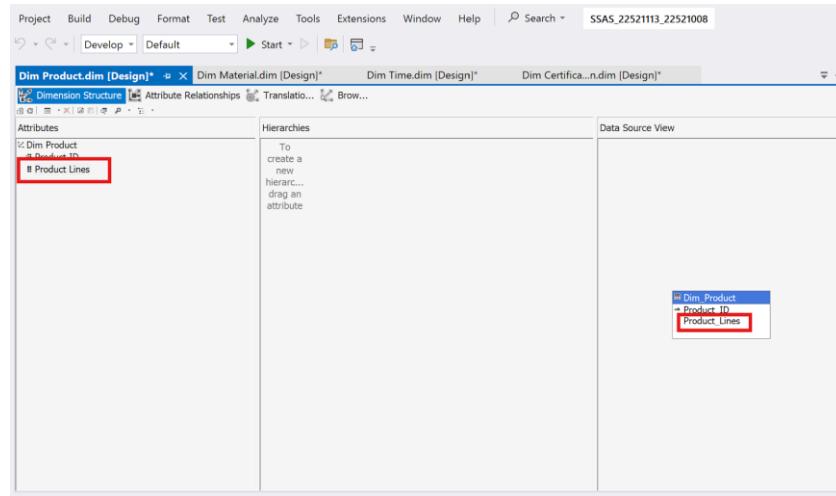
- Nháy chuột phải vào **Dim_Material** trong phần **Attributes**, chọn **Properties**. Trong cửa sổ **Properties**, chỉnh sửa giá trị **ErrorConfiguration** → **KeyDuplicate** là **IgnoreError**. Chỉnh sửa giá trị **UnknownMember** thành **Hidden**.



Hình 3. 44. Cấu hình cho Dim Material.

Bước 8:

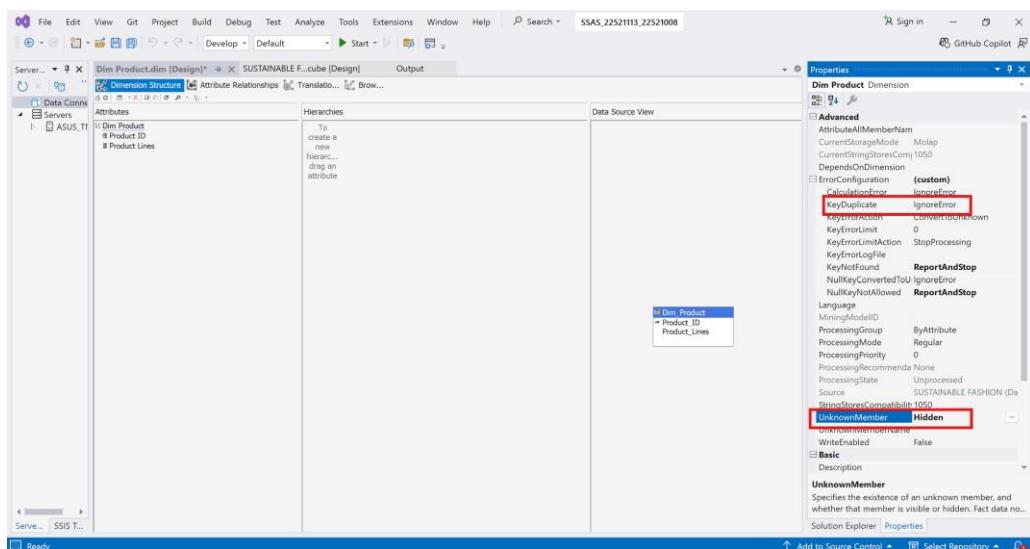
- Trong **Solution Explorer**, click vào **Dim Product.dim** từ **Dimensions**. Kéo thả thuộc tính *Product_Lines* từ **Data Source View** vào **Attributes**.



Hình 3. 45. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Product.

- Nháy chuột phải vào **Dim_Product** trong phần **Attributes**, chọn **Properties**.

Trong cửa sổ **Properties**, chỉnh sửa giá trị **ErrorConfiguration** → **KeyDuplicate** là **IgnoreError**. Chỉnh sửa giá trị **UnknownMember** thành **Hidden**.

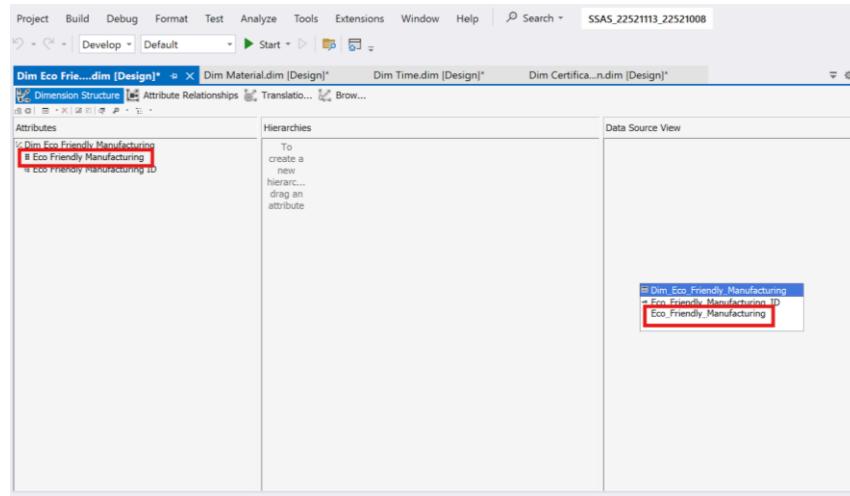


Hình 3. 46. Cấu hình cho Dim Product.

Bước 9:

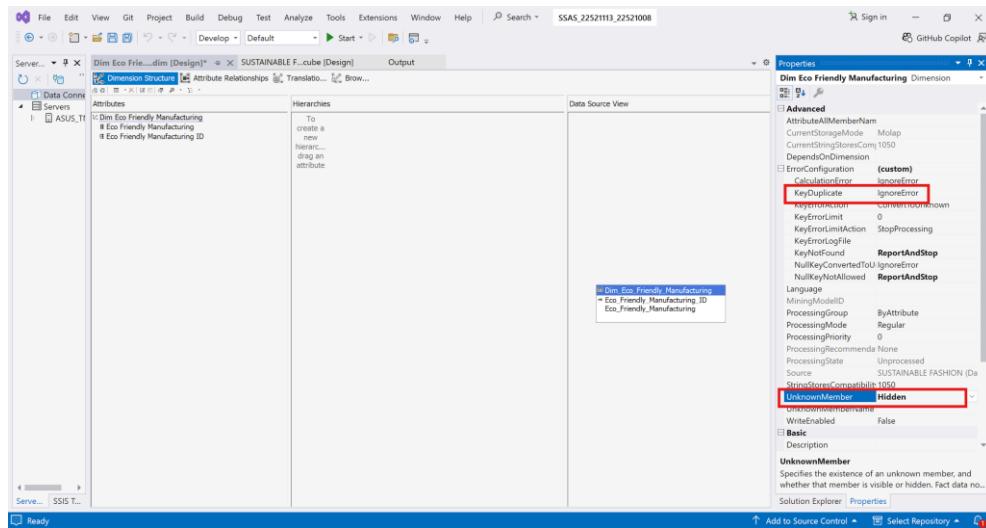
- Trong **Solution Explorer**, click vào **Dim Eco_Friendly_Manufacturing.dim** từ

Dimensions. Kéo thả thuộc tính *Eco_Friendly_Manufacturing* từ **Data Source View** vào **Attributes**.



Hình 3. 47. Kéo thả thuộc tính từ Data Source View vào Attributes cho Dim Eco_Friendly_Manufacturing

- Nhấp chuột phải vào **Dim_Eco-Friendly_Manufacturing** trong phần **Attributes**, chọn **Properties**. Trong cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration → KeyDuplicate là IgnoreError. Chỉnh sửa giá trị UnknownMember thành Hidden.

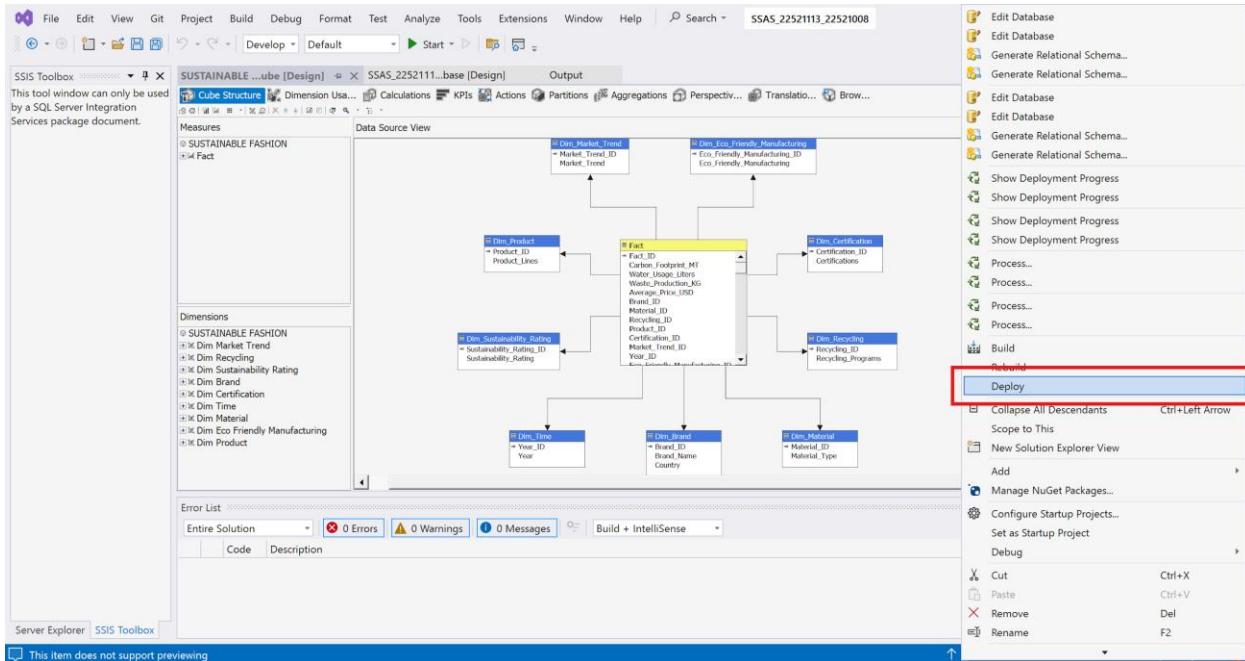


Hình 3. 48. Cấu hình cho Dim Eco_Friendly_Manufacturing

3.1.5 Chạy dự án SSAS

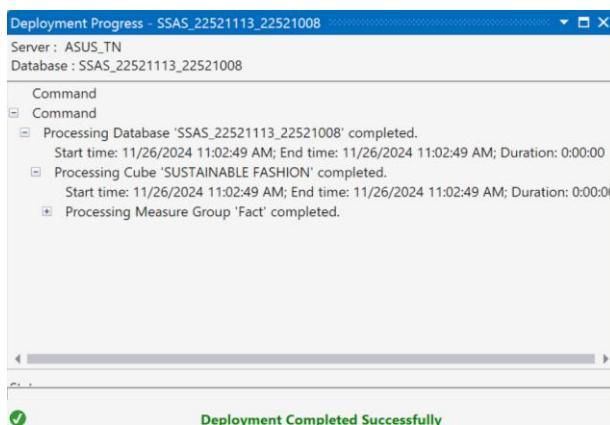
Bước 1:

- Thực hiện deploy project bằng cách click chuột phải vào project (SSAS_22521113_22521008) và chọn Deploy.



Hình 3. 49. Deploy dự án

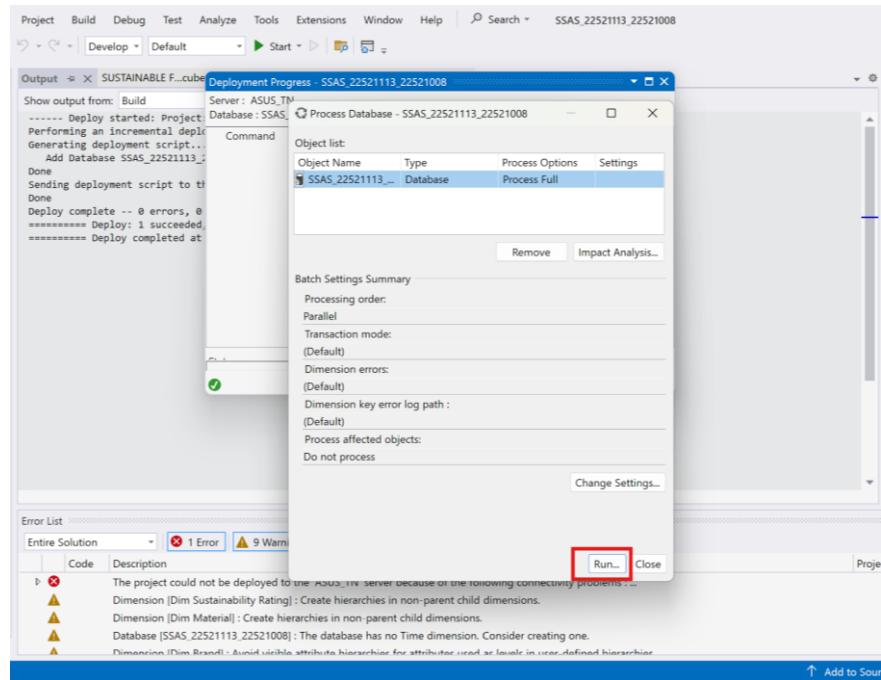
- Nếu thành công, hệ thống sẽ hiển thị cửa sổ với Status là Deployment Completed Successfully như hình dưới.



Hình 3. 50. Deploy Project thành công.

- Bước 2:** Màn hình Process Database hiển thị. Ta bấm nút Run để tiến hành Process dữ

liệu.



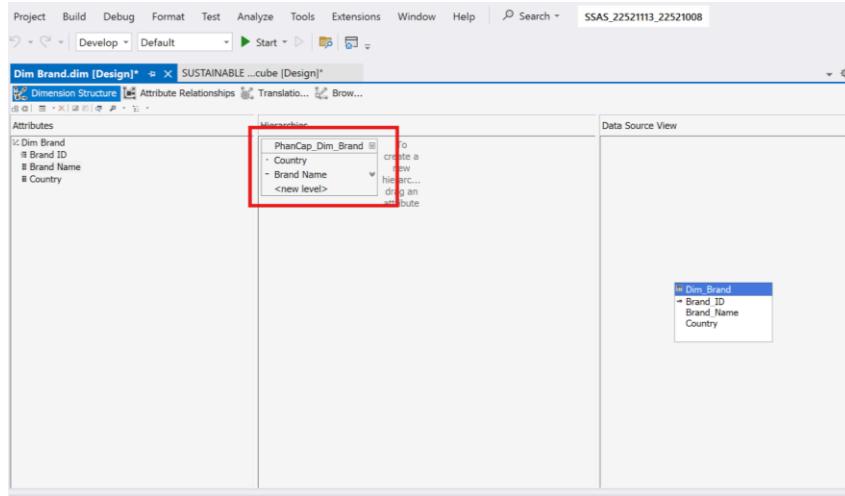
Hình 3. 51 Process dữ liệu

3.1.6 Tạo Hierarchies và định nghĩa Attribute Relationship

Nhóm sẽ tiến hành phân cấp cho bảng chiều: Dim_Brand.

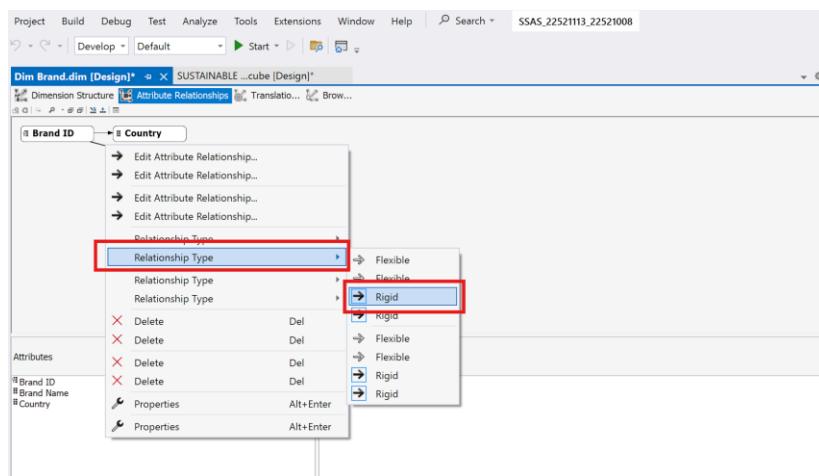
Bước 1: Ở folder Dimension, bấm vào bảng Dim_Brand để tiến hành phân cấp

Sau đó kéo thả những thuộc tính cần phân cấp từ **Attributes** qua của sổ **Hierarchies**, và theo thứ tự từ trên xuống là phân cấp từ cao nhất tới thấp nhất. Tại đây, ta tiến hành phân cấp theo thứ tự **Country => Brand_Name**. Và tiến hành đổi tên thành **PhanCap_Dim_Brand**.



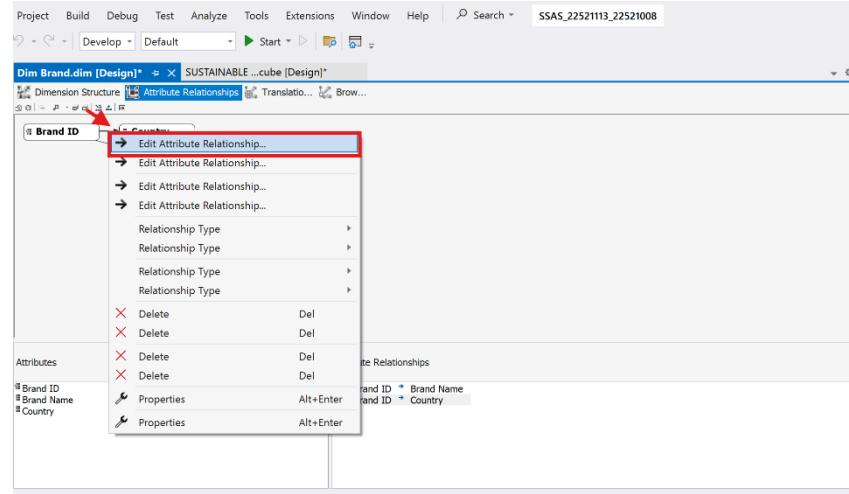
Hình 3. 52. Phân cấp cho Dim Brand.

Bước 2: Sau đó, qua tab **Attribute Relationships** để liên kết các mối quan hệ phân cấp. Điều chỉnh **Relationship Type** thành **Rigid**.



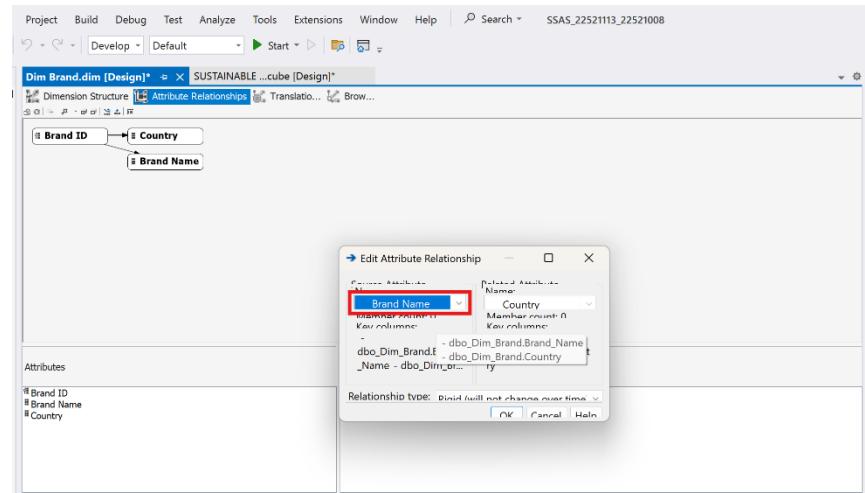
Hình 3. 53. Đổi loại quan hệ trong Dim Brand

Bước 3: Click phải chuột vào mũi tên nối từ **Brand_ID** đến **Country**. Chọn **Edit Attribute Relationship...**



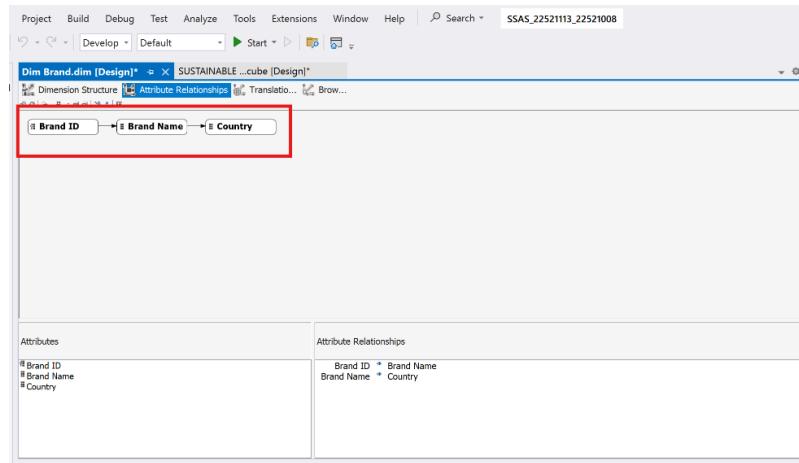
Hình 3. 54. Chọn Edit Attribute Relationship.

Bước 4: Trong khung Edit Attribute Relationship, đổi Brand_ID thành Brand_Name.



Hình 3. 55. Đổi Source Attribute

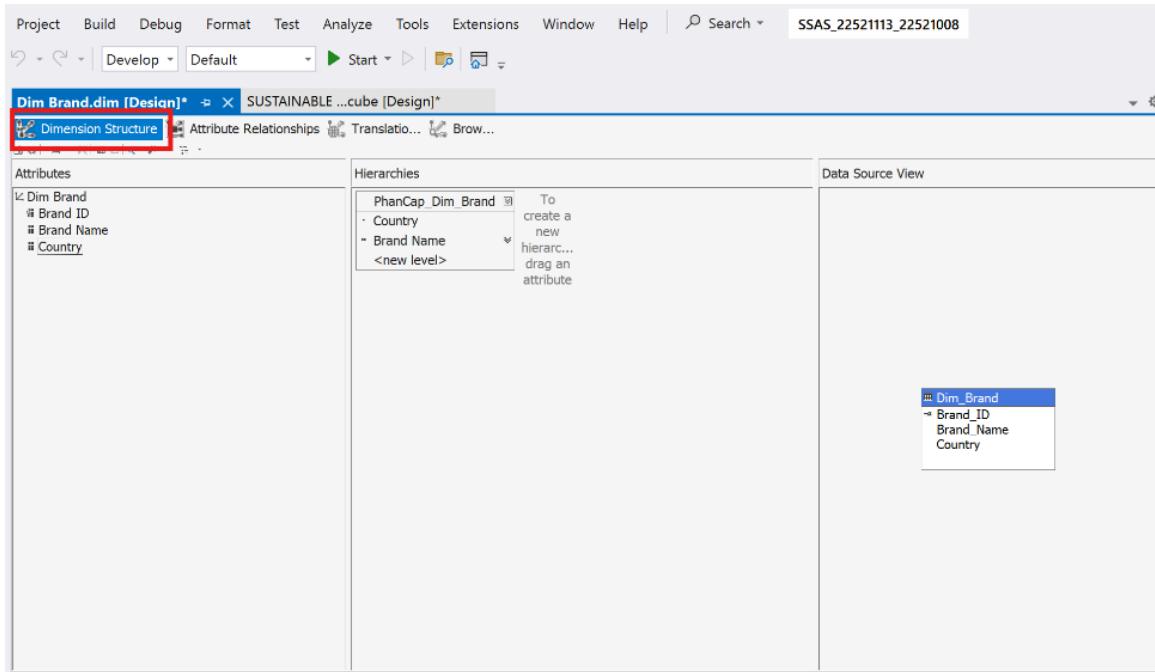
Kết quả thu được:



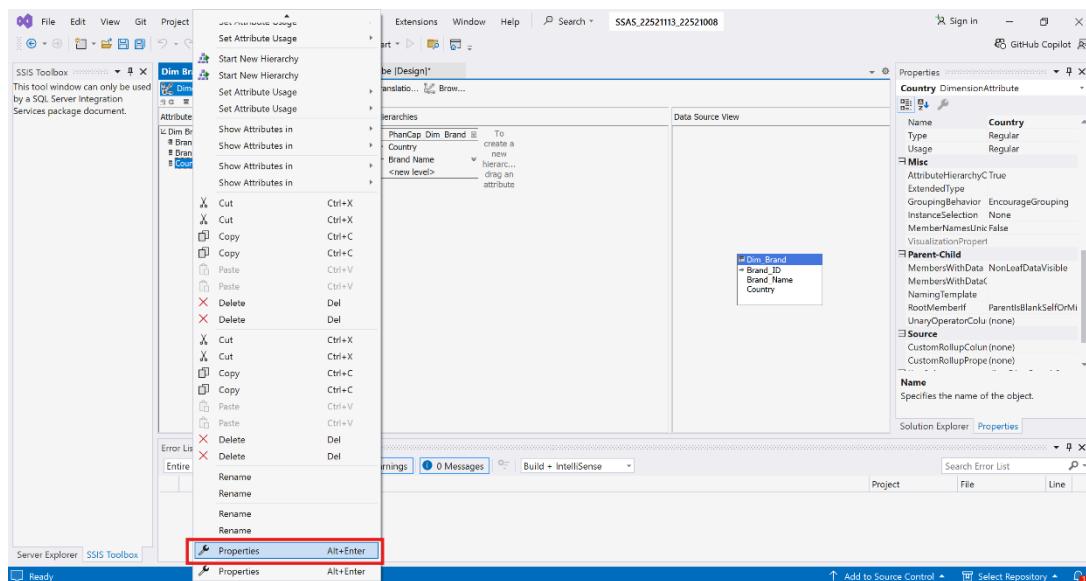
Hình 3. 56. Kết quả sau khi đổi source attribute

Bước 5: Chuyển qua tab Dimensions Structure. Ta tiến hành điều chỉnh lại từng cột thuộc tính. Tại cột Attributes, click phải chuột vào thuộc tính Country => Properties.

IS217 – Kho dữ liệu và OLAP



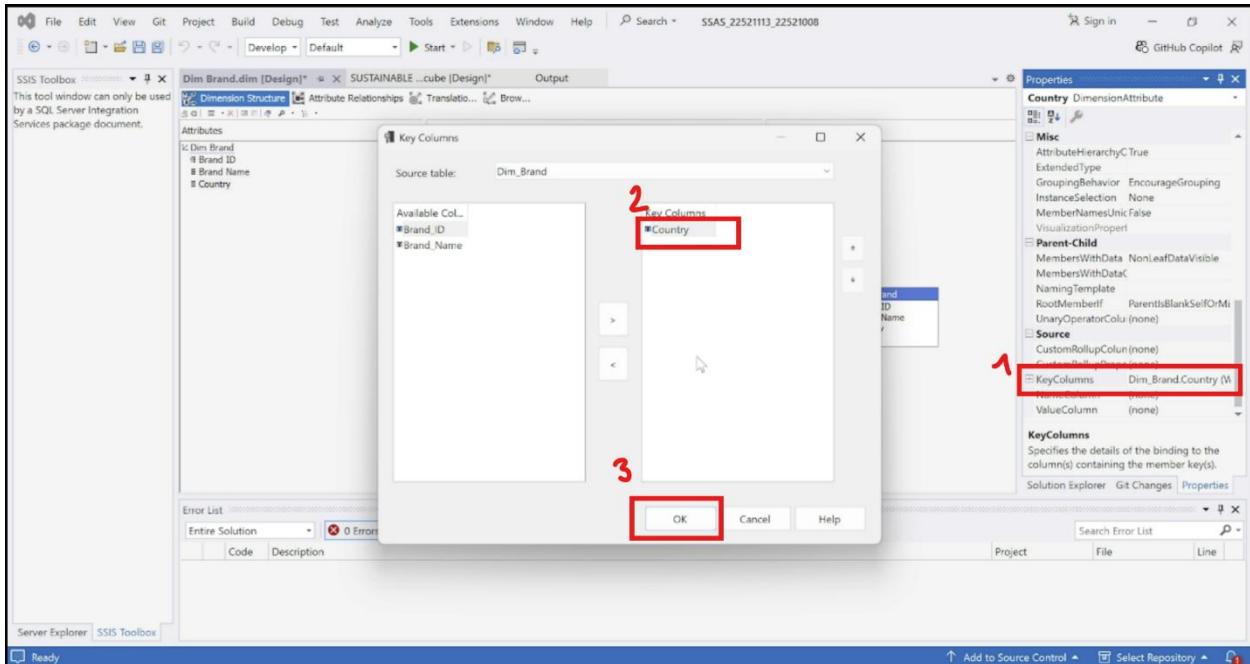
Hình 3.57 Kết quả sau khi điều chỉnh Attribute Relationship.



Hình 3. 58. Chọn properties để cấu hình Dim Brand

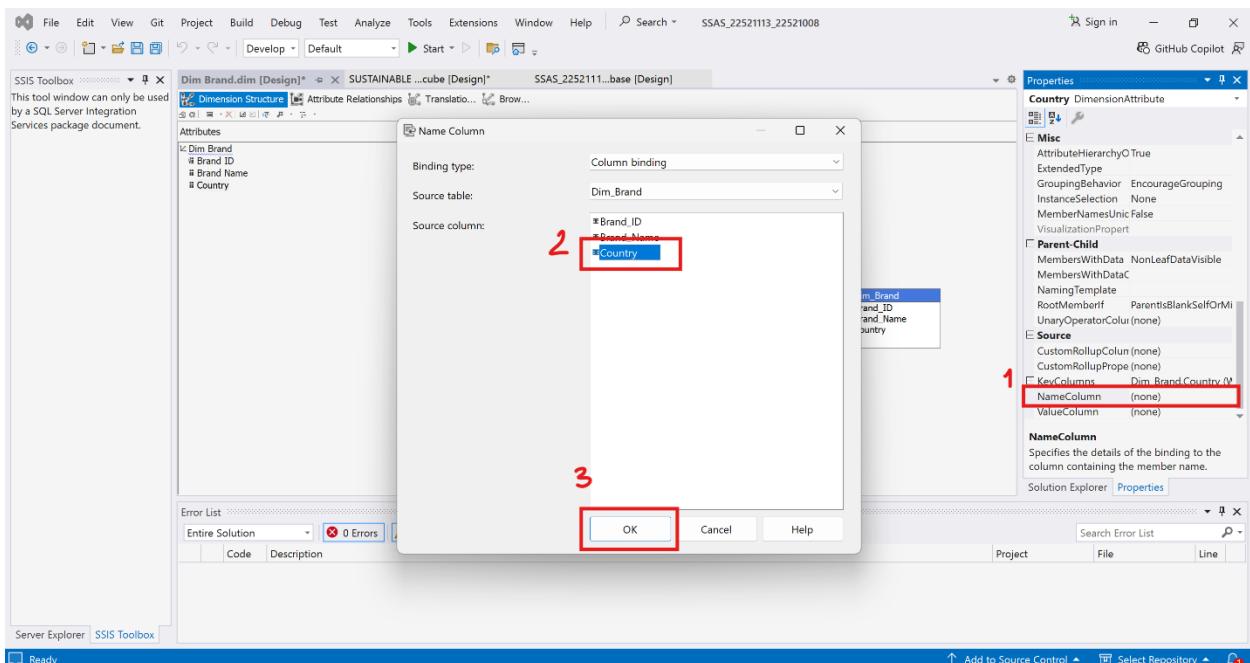
Bước 6: Ở **Properties**, ta chọn Key Column, và xuất hiện hộp thoại Key Column. Ta tiến hành điều chỉnh như sau:

IS217 – Kho dữ liệu và OLAP



Hình 3. 59. Cấu hình cột Country trong Dim Brand

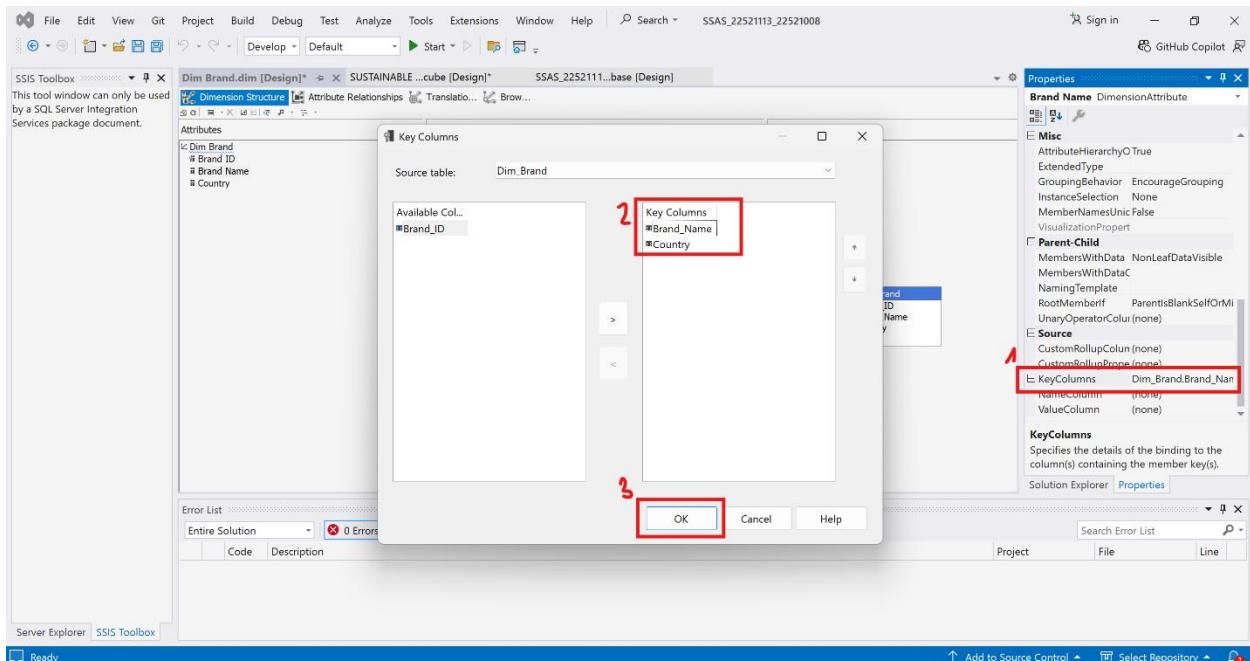
Đổi NameColumn có source column là Country:



Hình 3. 60. Cấu hình NameColumn cho thuộc tính Country.

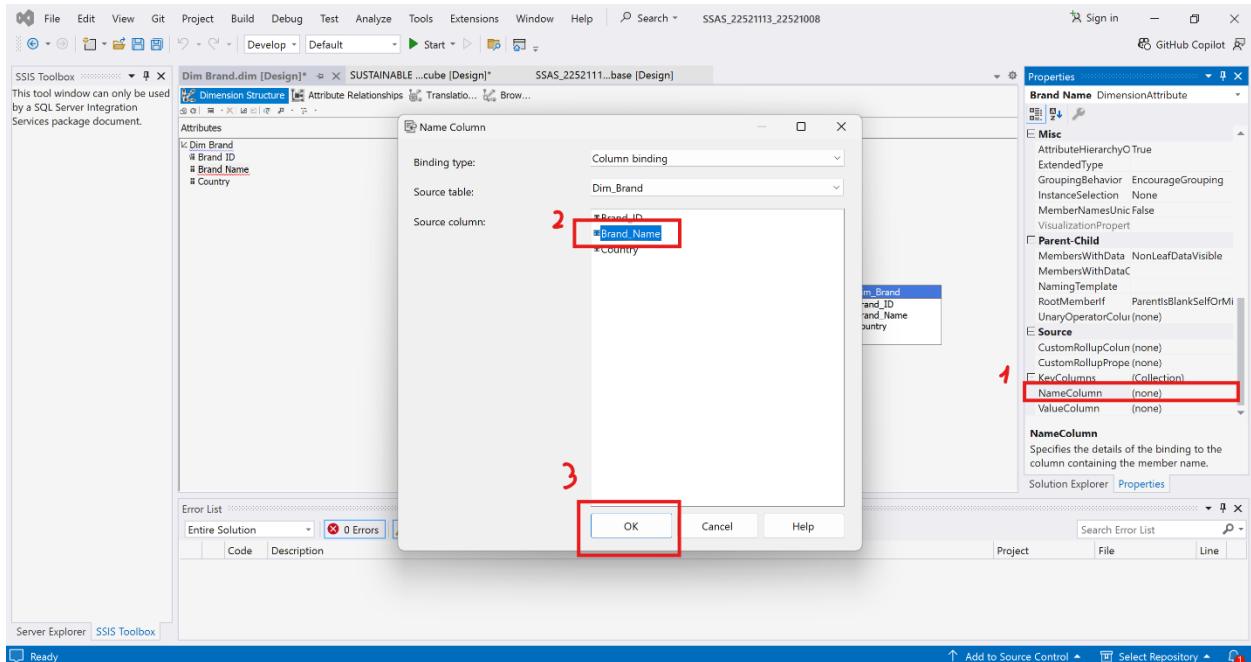
Bước 7: Lặp lại các bước tương tự như trên với thuộc tính Brand_Name:

Tuy nhiên key column sẽ là Brand_Name và Country vì thuộc tính **Brand_Name** là thuộc tính cấp nhỏ nhất sẽ lấy khóa dòng gồm chính nó và những thuộc tính cấp cao hơn.



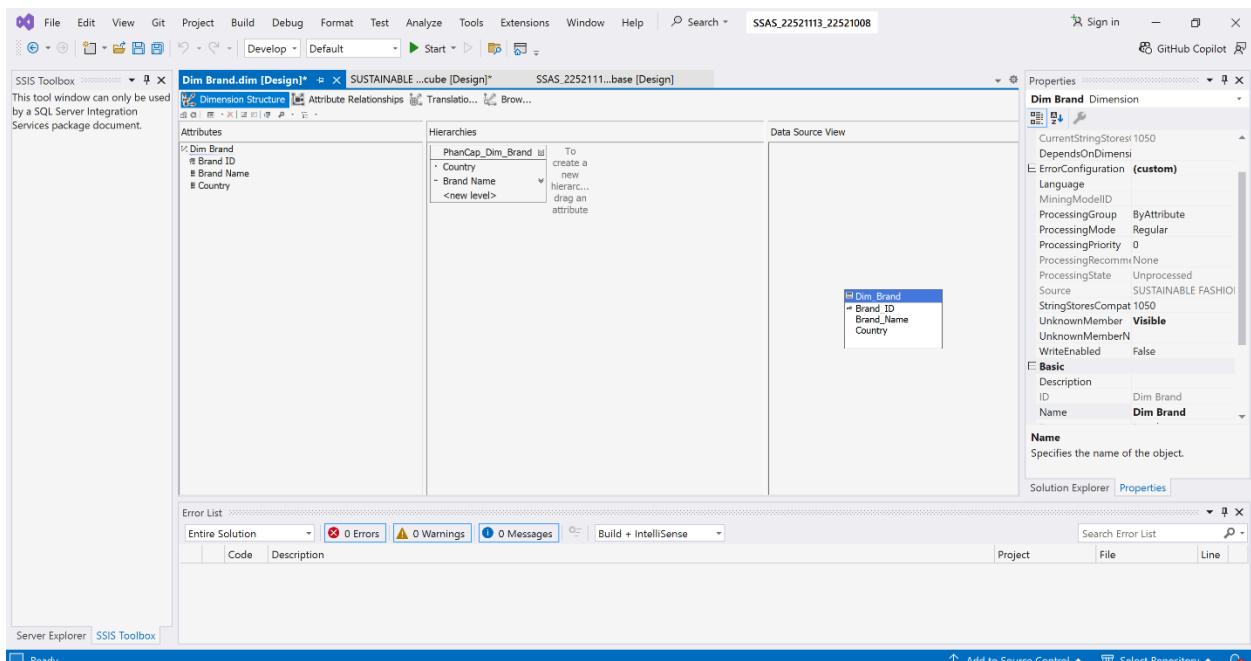
Hình 3. 61. Cấu hình cho KeyColumn cho thuộc tính Brand Name

IS217 – Kho dữ liệu và OLAP



Hình 3. 62. Cấu hình NameColumn cho thuộc tính Brand Name

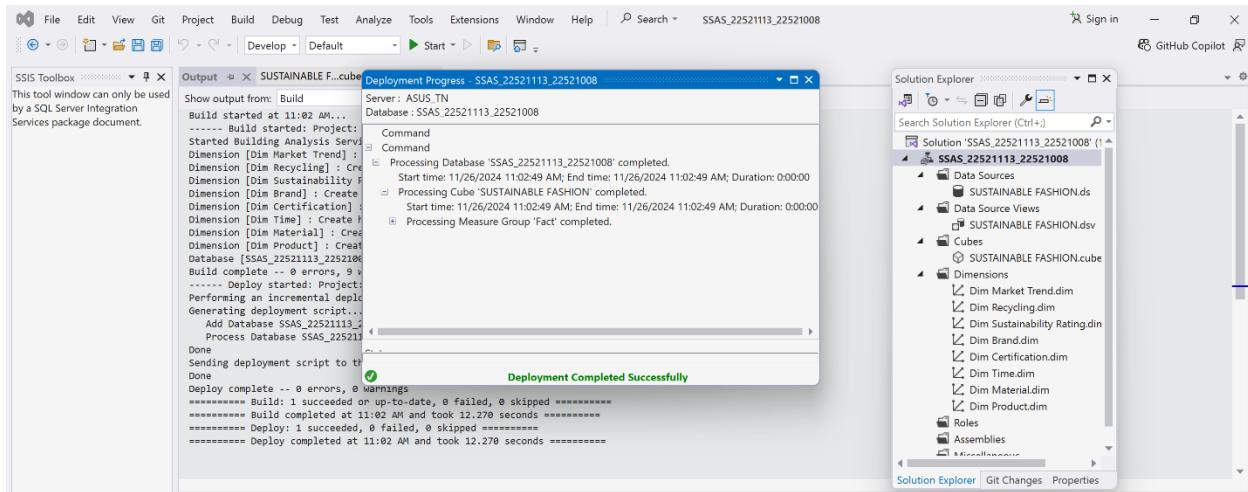
Cuối cùng ta thu được kết quả phân cấp như sau:



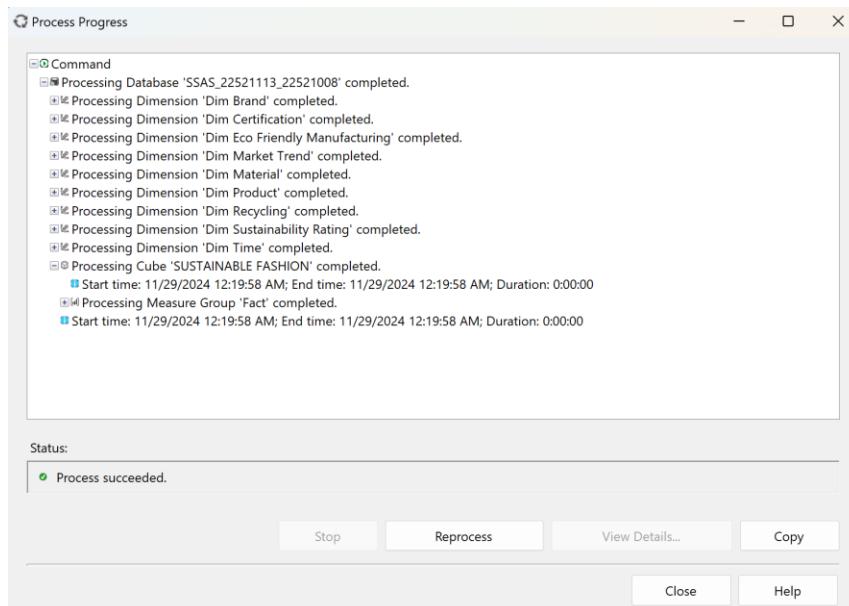
Hình 3. 63. Phân cấp cho Dim Brand thành công

3.1.7: Process lại Project

Nhấp chuột phải vào **SSAS_22521113_22521008** và chọn **Process...**

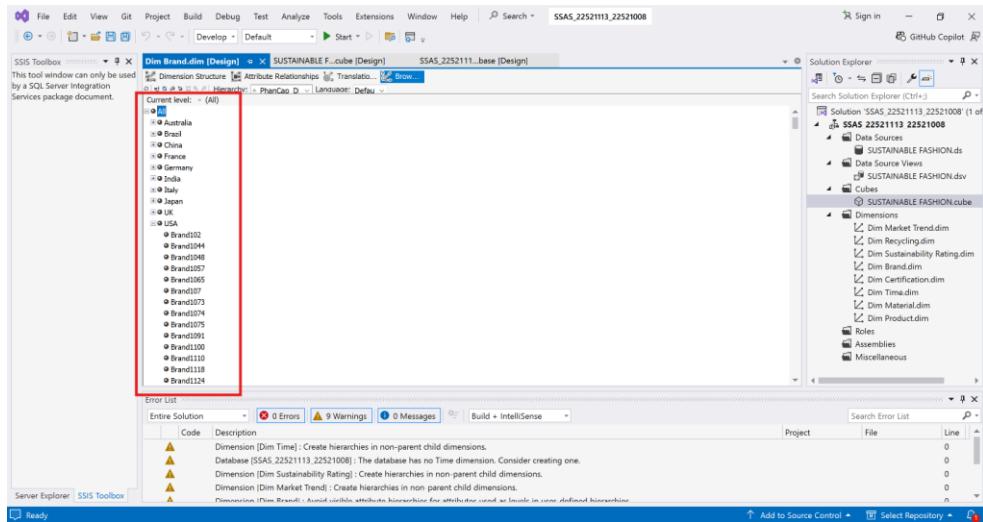


Hình 3. 64. Deploy Project thành công



Hình 3. 65. Deploy Project thành công

Click chuột vào **Dim_Brand** và chuyển sang tab **Browser** để xem kết quả phân cấp.



Hình 3. 66. Deploy Project thành công

3.2 Quá trình phân tích dữ liệu bằng công cụ SSAS trên các khối CUBE.

3.2.1 Câu truy vấn 1

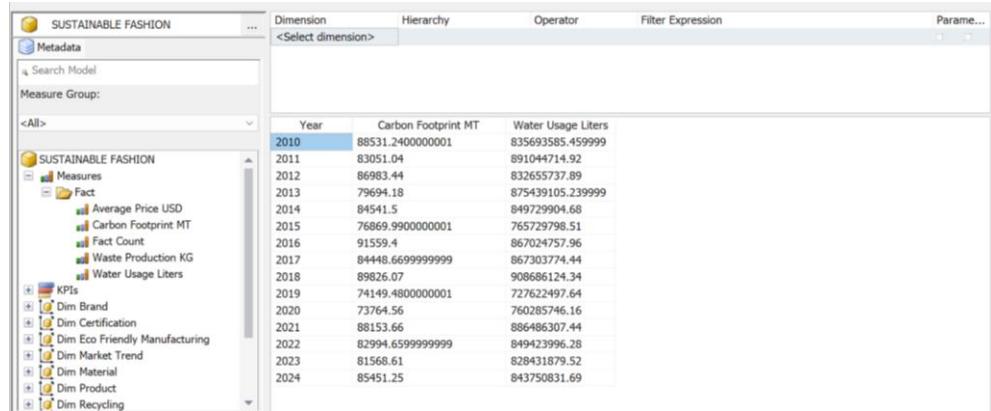
Nội dung câu truy vấn: Thông kê số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của ngành thời trang theo từng năm.

Nhận xét:

- Lượng carbon thải ra: dựa vào số liệu ta nhận thấy lượng carbon của ngành thời trang thải ra không khí ở mức cao, có sự tăng giảm xen kẽ, tức là nếu năm trước tăng thì năm sau sẽ giảm. Khi giảm thì giảm khá nhiều và khi tăng cũng tăng khá nhiều. Tình hình như thế cho thấy các quốc gia trên thế giới có sự chú trọng trong việc giảm thiểu lượng carbon thải ra tuy nhiên vẫn chưa được hiệu quả, và kiểm soát thiểu chật chẽ.
- Lượng nước sử dụng: lượng nước sử dụng luôn ở mức cao và có nhiều sự tăng giảm qua các năm. Điều cho thấy, tương tự như lượng khí thải carbon, cũng có sự chú trọng trong việc giảm mức sử dụng nhưng không đáng kể, cần nhiều sự điều chỉnh và quản lý nghiêm ngặt hơn.

Cách thực hiện:

- Kéo các thuộc tính Year từ bảng Dim Time, Carbon Footprint MT và Water Usage Liters từ bảng Fact cửa sổ thực thi truy vấn
- Kết quả thực thi :



The screenshot shows a data cube interface with a navigation pane on the left and a results grid on the right.

Navigation Pane (Left):

- SUSTAINABLE FASHION
- Metadata
- Measure Group: <All>
- SUSTAINABLE FASHION
 - Measures
 - Fact
 - Average Price USD
 - Carbon Footprint MT
 - Fact Count
 - Waste Production KG
 - Water Usage Liters
- KPIs
- Dim Brand
- Dim Certification
- Dim Eco Friendly Manufacturing
- Dim Market Trend
- Dim Material
- Dim Product
- Dim Recycling

Hình 3. 67. Kết quả câu truy vấn 1 dùng khối CUBE

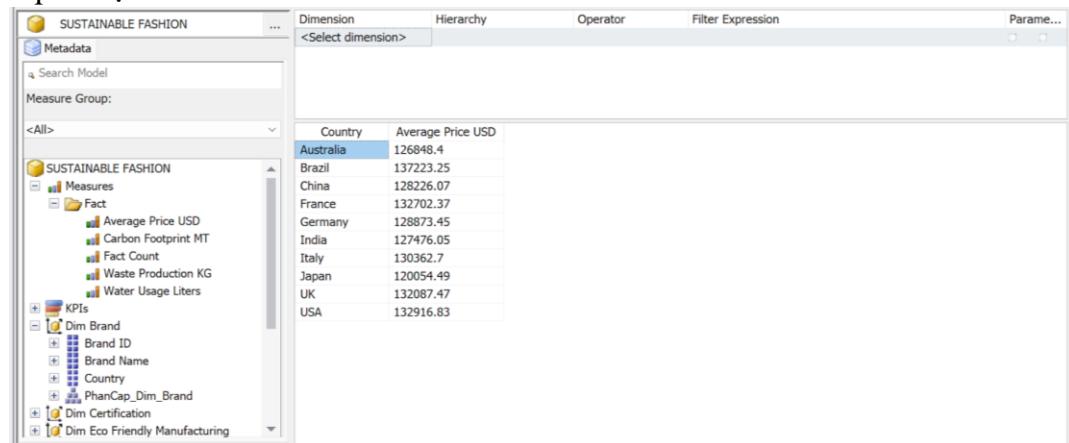
3.2.2 Câu truy vấn 2

Nội dung câu truy vấn: Thống kê tổng doanh thu của các thương hiệu ở mỗi nước.

Nhận xét: Kết quả thống kê cho thấy nước Brazil là nước có doanh thu cao nhất, có thể nói rằng người Brazil chi khá nhiều cho ngành thời trang.

Cách thực hiện:

- Kéo thả thuộc tính Country từ bảng Dim Brand và Average Price USD từ bảng Fact vào cửa sổ thực thi truy vấn.
- Kết quả thực thi:



The screenshot shows a data cube interface with a navigation pane on the left and a results grid on the right.

Navigation Pane (Left):

- SUSTAINABLE FASHION
- Metadata
- Measure Group: <All>
- SUSTAINABLE FASHION
 - Measures
 - Fact
 - Average Price USD
 - Carbon Footprint MT
 - Fact Count
 - Waste Production KG
 - Water Usage Liters
- KPIs
- Dim Brand
 - Brand ID
 - Brand Name
 - Country
 - PhanCap_Dim_Brand
- Dim Certification
- Dim Eco Friendly Manufacturing

Hình 3. 68. Kết quả câu truy vấn 2 dùng CUBE

3.2.3 Câu truy vấn 3

Nội dung câu truy vấn: Thống kê tổng số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của từng loại vật liệu đã thải ra từ năm 2010 đến 2024.

Nhận xét: với lượng khí thải carbon thải ra thì loại vật liệu “Recycled Polyester” đạt cao nhất. Với lượng nước sử dụng thì “Hemp” đạt mức cao nhất. Với lượng rác thải thì “Recycled Polyester” lại thêm một lần đạt cao nhất. Ngoài ra, khi nhìn vào bảng thống kê, ta có thể thấy loại vật liệu “Vegan Leather” gần như đạt mức thấp nhất ở cả 3 loại chỉ số. Từ đó, có thể cho rằng “Vegan Leather” là loại vật liệu vẫn gây hại cho môi trường nhưng ít hơn so với những vật liệu còn lại.

Cách thực hiện:

- Kéo thả thuộc tính Material Type từ bảng Dim Material; các thuộc tính Carbon Footprint MT, Water Usage Liters và Waste Production KG từ bảng Fact vào cửa sổ thực thi truy vấn.
- Kết quả thực thi:

The screenshot shows a Microsoft Analysis Services (SSAS) cube browser interface. On the left, there is a navigation pane with a tree view of the cube structure. The root node is 'SUSTAINABLE FASHION'. Under it, there are nodes for 'Measures' (Fact, Average Price USD, Carbon Footprint MT, Fact Count, Waste Production KG, Water Usage Liters), 'KPIs', and several dimension nodes (Dim Brand, Dim Certification, Dim Eco Friendly Manufacturing). The main area displays a query results grid with the following data:

Material Type	Carbon Footprint MT	Water Usage Liters	Waste Production KG
Bamboo Fabric	205415.61	2095858839.28	41856232.15
Hemp	213271.55	2127045944.36	4244672.85
Organic Cotton	204043.78	2043406142.82	41104045.75
Recycled Polyester	224853.65	2181995126.45	42545018.98
Tencel	203611.37	2118911371.74	40750944.8099999
Vegan Leather	200391.79	2022091337.52	41835605.7

Hình 3. 69. Kết quả thực thi câu truy vấn 3 dùng CUBE

3.2.4 Câu truy vấn 4

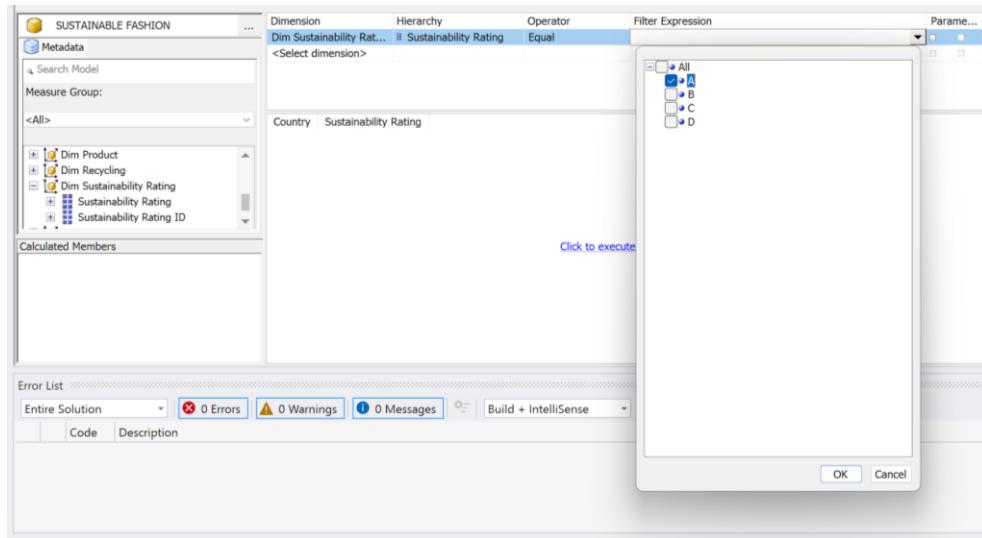
Nội dung câu truy vấn: Thống kê số lượng thương hiệu có đánh giá tiêu chuẩn (Sustainability_Rating) là A của mỗi nước trong năm 2014 và 2024.

Nhận xét: Dựa vào số lượng thương hiệu trong năm 2014 và 2024 ở mỗi nước, ta nhận thấy Australia, France, India, Japan, USA là các nước có số thương hiệu tăng sau 10 năm,

đặc biệt là Japan có số lượng tăng vượt bậc. Qua đó, có thể nói 5 nước này rất chú trọng trong việc phát triển các thương hiệu đạt tỉ lệ bền vững cao, góp phần làm giảm các tác động xấu của ngành thời trang đến môi trường. Tuy nhiên, vẫn còn khá nhiều nước chưa thật sự đẩy mạnh phong trào thời trang bền vững như China, số lượng giảm một cách báo động. Cần có những thúc đẩy để tiếp tục phát triển xu hướng thời trang bền vững cả ở những nước đã và đang trong quá trình thực hiện.

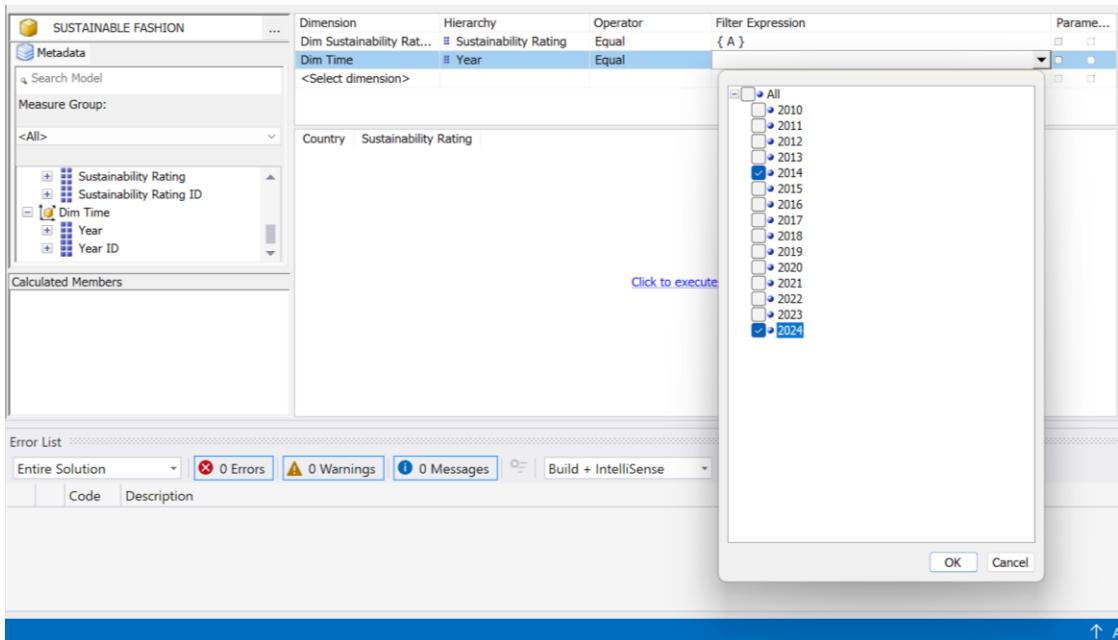
Cách thực hiện:

1. Kéo thuộc tính Country từ bảng Dim Brand và thuộc tính Sustainability Rating từ bảng Dim Sustainability Rating và cửa sổ thực thi truy vấn.
2. Kéo thuộc tính Sustainability Rating từ bảng Dim Sustainability Rating vào cửa sổ lọc dữ liệu và tại Filter Expression chọn mức tỉ lệ là A.



Hình 3. 70. Thao tác trên thuộc tính Sustainability Rating

3. Kéo thuộc tính Year từ bảng Dim Time vào cửa sổ lọc dữ liệu và tại Filter Expression chọn năm 2014 và 2024.



Hình 3. 71. Thao tác trên thuộc tính Year

- Kéo thuộc tính Fact Count từ bảng Fact vào cửa sổ thực thi truy vấn và tiến hành thực thi ta có kết quả:

Country	Sustainability Rating	Year	Fact Count
Australia	A	2014	7
Australia	A	2024	14
Brazil	A	2014	9
Brazil	A	2024	7
China	A	2014	11
China	A	2024	4
France	A	2014	5
France	A	2024	9
Germ...	A	2014	7
Germ...	A	2024	5
India	A	2014	5
India	A	2024	8
Italy	A	2014	10
Italy	A	2024	9
Japan	A	2014	4
Japan	A	2024	13
UK	A	2014	7
UK	A	2024	3
USA	A	2014	6
USA	A	2024	11

Hình 3. 72. Kết quả thực thi câu truy vấn 4 dùng CUBE

3.2.5 Câu truy vấn 5

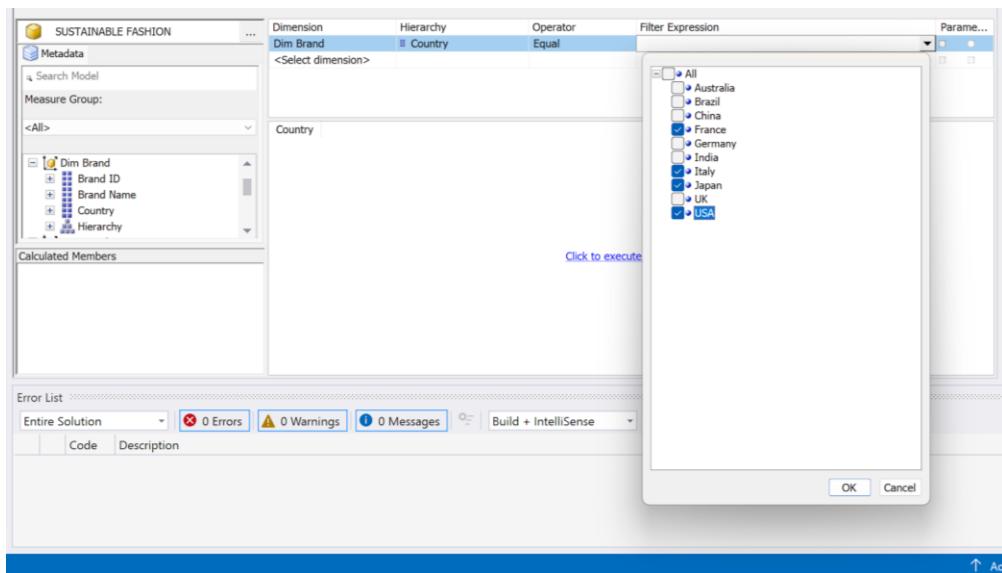
Nội dung câu truy vấn: Liệt kê số thương hiệu thuộc nước France, Italy, Japan và USA không có giấy chứng nhận và tổng lượng khí thải carbon mà các thương hiệu ấy đã thải ra.

Nhận xét: Dựa vào kết quả thực thi, ta có thể thấy trong bốn nước có nền thời trang phát

triển mạnh mẽ thì USA là nước có số lượng thương hiệu không có giấy chứng nhận nhiều nhất (107) và dẫn đến lượng khí thải carbon thải ra cũng ở mức cao. Trong khi đó, Italy có số lượng ít nhất nhưng vẫn ở một mức cao (83). Những con số này nói lên rằng, dù đây là những quốc gia kinh đô thời trang nhưng số lượng các thương hiệu đạt các chứng chỉ bền vững vẫn còn khá nhiều, dẫn đến góp một phần không ít vào sự đào thải khí carbon ra ngoài môi trường. Cần xem xét lại các chính xác đang quản lý để kịp thời điều chỉnh, giảm thiểu sự ảnh hưởng môi trường và thúc đẩy thời trang bền vững.

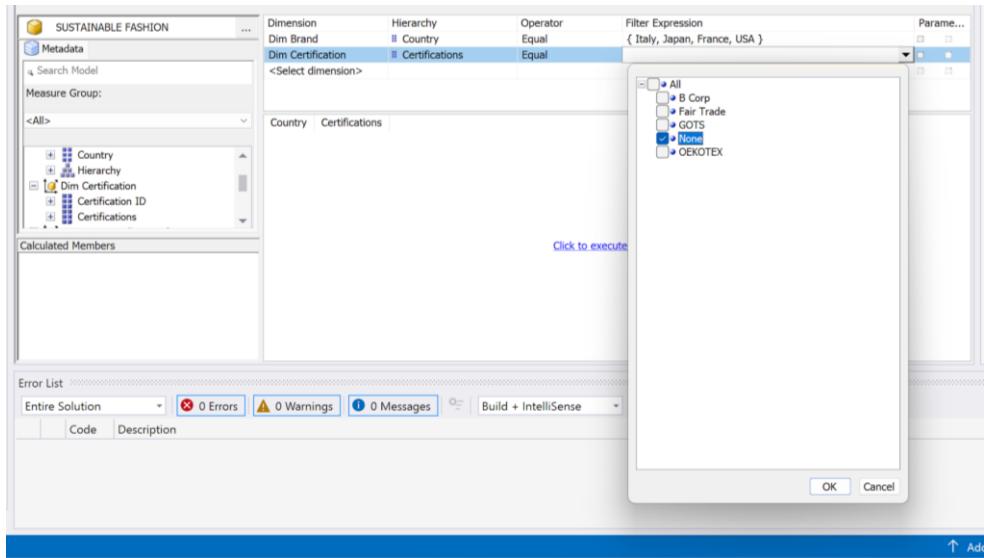
Cách thực hiện:

- Kéo thuộc tính Country từ bảng Dim Brand vào cửa sổ thực thi và vào cửa sổ lọc dữ liệu. Tại Filter Expression chọn các nước France, Japan, Italy và USA.



Hình 3. 73. Thao tác trên thuộc tính Country

- Kéo thuộc tính Certification từ bảng Dim Certification vào cửa sổ thực thi truy vấn và cửa sổ lọc dữ liệu. Tại Filter Expression chọn None (không có chứng chỉ).



Hình 3. 74. Thao tác trên thuộc tính Certifications

3. Kéo thuộc tính Fact Count và Carbon Footprint MT từ bảng Fact vào cửa sổ thực thi truy vấn.
4. Kết quả thực thi:

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Brand	Country	Equal	{ France, Italy, Japan, USA }	
Dim Certification	Certifications	Equal	{ None }	
<Select dimension>				
Country	Certifications	Fact Count	Carbon Footprint MT	
France	None	103	26306.45	
Italy	None	87	22182.96	
Japan	None	89	22786.68	
USA	None	107	26523.97	

Hình 3. 75. Kết quả thực thi câu truy vấn 5 dùng CUBE

3.2.6 Câu truy vấn 6

Nội dung câu truy vấn: Liệt kê 10 thương hiệu có giá trung bình cao nhất và tỉ lệ bền vững của chúng.

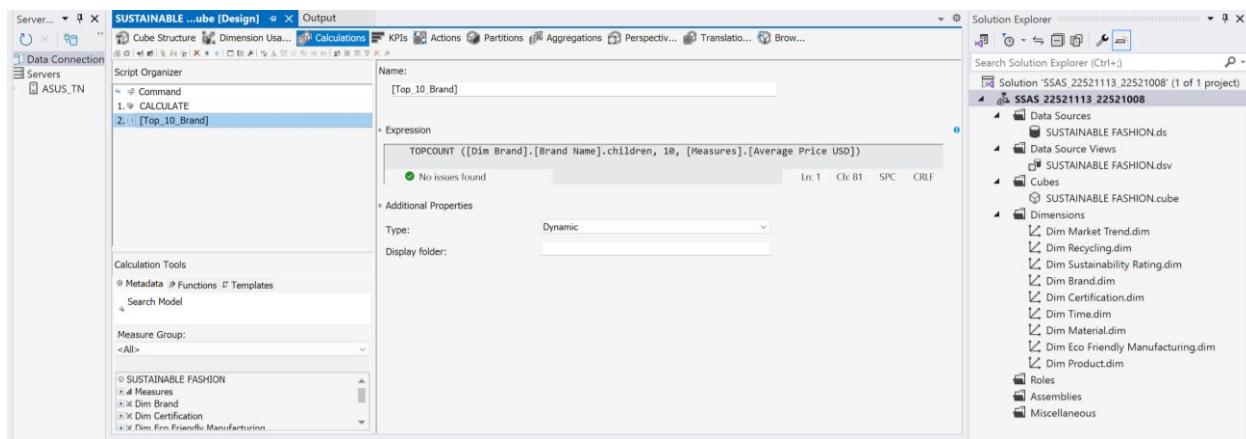
Nhận xét: Những thương hiệu có giá trung bình cao tức là đây là những thương hiệu xa xỉ, được đầu tư để sản xuất và dành cho những khách hàng thường lưu. Có một tín hiệu tốt là mười thương hiệu có giá trung bình cao nhất đều đạt tỉ lệ bền vững khá cao, có tới bốn

thương hiệu là đạt mức bền vững A, chỉ có một thương hiệu đạt tỉ lệ bền vững thấp nhất là D. Điều đó cho thấy những thương hiệu cao cấp và các khách hàng tiềm năng đều có gánh hướng tới sự bền vững cao trong việc sản xuất và lựa chọn những sản phẩm thời trang không gây hại cho môi trường. Góp phần thúc đẩy xu hướng ngày càng lan rộng và phát triển.

Cách thực hiện:

1. Tạo một nameset và đặt tên nameset đó là Top_10_Brand.

Câu truy vấn là: TOPCOUNT ([Dim Brand].[Brand Name].children, 10, [Measures].[Average Price USD]). Sau đó lưu lại.



Hình 3. 76. Tạo nameset Top_10_Brand

2. Sau khi tạo thành công nameset, ta sẽ thấy hiển thị trong bảng Dim Brand. Tiếp đó, kéo thuộc tính Brand Name từ bảng Dim Brand, thuộc tính Sustainability Rating từ bảng Dim Sustainability Rating, Average Price USD từ bảng Fact vào cửa sổ thực thi truy vấn.
3. Kéo thuộc tính nameset Top_10_Brand từ bảng Dim Brand vào cửa sổ lọc dữ liệu.

Kết quả thực thi:

The screenshot shows the SSAS Management Studio interface. On the left, the 'SUSTAINABLE ...ube [Design]' tab is active, displaying a query editor with MDX code. The main pane shows the results of a query, which is a table with columns: Brand Name, Sustainability Rating, and Average Price USD. The table contains 10 rows of data. The right pane, 'Solution Explorer', shows the project structure for 'SSAS 22521113_22521008' with various cubes, dimensions, and data sources listed.

Brand Name	Sustainability Rating	Average Price USD
Brand1425	B	499.73
Brand1579	A	499.55
Brand1753	A	499.87
Brand3904	D	499.54
Brand3942	C	499.94
Brand4402	B	499.43
Brand4419	A	499.84
Brand469	A	499.41
Brand4815	B	499.31
Brand929	A	499.85

Hình 3. 77. Kết quả thực thi câu truy vấn 6 bằng CUBE

3.2.7 Câu truy vấn 7

Nội dung câu truy vấn: Thống kê số lượng mỗi loại vật liệu theo từng xu hướng thị trường tăng (Growing) trong trường hợp sản xuất thân thiện và không thân thiện với môi trường.

Nhận xét: Từ kết quả truy vấn, ta thấy với mỗi loại vật liệu có xu hướng tăng thì số thương hiệu sử dụng để sản xuất theo hướng thân thiện với môi trường đều cao hơn hướng không thân thiện. Tuy nhiên, chênh lệch không quá nhiều, và số lượng sản xuất theo hướng không thân thiện vẫn còn khá cao. Đặc biệt chú ý đến các sản phẩm làm từ Tencel vì nhìn vào ta có thể thấy số lượng thương hiệu sử dụng vật liệu này để sản xuất theo hướng thân thiện lại thấp hơn hướng không thân thiện, cần điều chỉnh lại việc quản lý sản xuất đối với vật liệu này.

Cách thực hiện:

1. Kéo thả thuộc tính Material Type từ bảng Dim Material vào cửa sổ thực thi truy vấn.
2. Kéo thả thuộc tính Market Trend từ bảng Dim Market Trend vào cửa sổ thực thi cửa sổ lọc dữ liệu. Trong phần Filter Expression chọn Growing.
3. Kéo thuộc tính Eco Friendly Manufacturing từ bảng Dim Eco Friendly Manufacturing vào cửa sổ thực thi truy vấn.
4. Kéo thuộc tính Fact Count từ bảng Fact vào cửa sổ thực thi truy vấn.
5. Kết quả thực thi:

Material Type	Eco Friendly Manufacturing	Fact Count
Bamboo Fabric	No	115
Bamboo Fabric	Yes	137
Hemp	No	129
Hemp	Yes	135
Organic Cotton	No	123
Organic Cotton	Yes	150
Recycled Polyester	No	140
Recycled Polyester	Yes	126
Tencel	No	156
Tencel	Yes	128
Vegan Leather	No	121
Vegan Leather	Yes	152

Hình 3. 78. Thao tác trên thuộc tính Market Trend và kết quả thực thi câu truy vấn 7 dùng CUBE

3.2.8 Câu truy vấn 8

Nội dung câu truy vấn: Với từng loại chứng chỉ, khi đạt các mức độ bền vững khác nhau hãy thống kê số lượng thương hiệu theo từng xu hướng thị trường.

Nhận xét: Dựa vào kết quả, ta thấy đối với chứng chỉ GOTS và OEKOTEX đều có xu hướng là đứng im, nhưng số lượng thương hiệu đạt tỉ lệ bền vững A lại thấp nhất là đối với chứng chỉ OEKOTEX. Không có chứng chỉ và có chứng chỉ B Corp thì có xu hướng giảm, nhưng lại có tỉ lệ bền vững giữa các mức độ gần như bằng nhau. Đối với chứng chỉ Fair Trade thì xu hướng là tăng và tỉ lệ bền vững của chứng chỉ này phần lớn là ở mức cao. Có thể rút ra rằng nên cố gắng đẩy mạnh việc đạt chứng chỉ Fair Trade ở các thương hiệu và đánh giá lại chất lượng của các chứng chỉ còn lại.

Cách thực hiện:

1. Kéo thuộc tính Certifications từ bảng Dim Certification vào cửa sổ thực thi truy vấn.
2. Kéo thuộc tính Sustainability Rating của bảng Dim Sustainability Rating vào cửa sổ thực thi truy vấn

3. Kéo thuộc tính Market Trend từ bảng Dim Market Trend vào cửa sổ thực thi truy vấn.
4. Kéo Fact Count từ bảng Fact vào cửa sổ thực thi truy vấn.
5. Kết quả thực thi:

Certifications	Sustainability Rating	Market Trend	Fact Count
B Corp	A	Declining	89
B Corp	A	Growing	80
B Corp	A	Stable	74
B Corp	B	Declining	89
B Corp	B	Growing	85
B Corp	B	Stable	81
B Corp	C	Declining	80
B Corp	C	Growing	79
B Corp	C	Stable	97
B Corp	D	Declining	100
B Corp	D	Growing	68
B Corp	D	Stable	65
Fair Trade	A	Declining	77
Fair Trade	A	Growing	80
Env Trends	A	Declining	77

Hình 3. 79. Kết quả thực thi câu truy vấn 8 bằng CUBE

3.2.9 Câu truy vấn 9

Nội dung câu truy vấn: Với mỗi vật liệu hãy cho biết số lượng thương hiệu sử dụng cả chương trình tái chế và sản xuất thân thiện với môi trường theo từng xu hướng.

Nhận xét: Dựa vào kết quả ta thấy xu hướng không mấy khả quan, tất cả các loại vật liệu được sản xuất và tái chế thân thiện với môi trường đều có xu hướng giảm.

Cách thực hiện:

1. Kéo thuộc tính Material Type từ bảng Dim Material vào cửa sổ thực thi truy vấn.
2. Kéo thuộc tính Eco Friendly Manufacturing của bảng Dim Eco Friendly Manufacturing cửa sổ lọc dữ liệu. Tại Filter Expression chọn Yes (sản xuất thân thiện với môi trường)
3. Kéo thuộc tính Recycling Programs từ bảng Dim Recycling vào cửa sổ lọc dữ liệu. Tại Filter Expression chọn Yes (có tái chế).

4. Kéo thuộc tính Market Trend từ bảng Dim Market Trend vào cửa sổ thực thi truy vấn.
5. Kéo thuộc tính Fact Count từ bảng Fact vào cửa sổ thực thi truy vấn.
6. Kết quả truy vấn:

Table Data:

Material Type	Market Trend	Fact Count
Bamboo Fabric	Declining	77
Bamboo Fabric	Growing	70
Bamboo Fabric	Stable	71
Hemp	Declining	81
Hemp	Growing	62
Hemp	Stable	59
Organic Cotton	Declining	74
Organic Cotton	Growing	71
Organic Cotton	Stable	73
Recycled Polyester	Declining	76
Recycled Polyester	Growing	66
Recycled Polyester	Stable	77
Tencel	Declining	74
Tencel	Growing	68

Hình 3. 80. Kết quả thực thi câu truy vấn 9 bằng CUBE

3.2.10 Câu truy vấn 10

Nội dung câu truy vấn: Thông kê số lượng thương hiệu đạt các mức bền vững ở mỗi nước.

Nhận xét: Ta có thể thấy ở hai nước UK và Australia có tỉ lệ đạt mức bền vững A cao hơn các mức bền vững khác. Ngược lại China và USA lại có tỉ lệ đạt mức bền vững D khá cao.

Cách thực hiện:

1. Kéo thuộc tính Country từ bảng Dim Brand vào cửa sổ thực thi truy vấn.
2. Kéo thuộc tính Sustainability Rating từ bảng Dim Sustainability Rating vào cửa sổ thực thi truy vấn.
3. Kéo Fact Count từ bảng Fact vào cửa sổ thực thi truy vấn.
4. Kết quả thực thi:

The screenshot shows the Microsoft Analysis Services (SSAS) cube design interface. On the left, there's a navigation pane with 'SUSTAINABLE FASHION' selected. Under 'Measure Group', there are several items: 'Average Price USD', 'Carbon Footprint MT', 'Fact Count', 'Waste Production KG', and 'Water Usage Liters'. Below this is a 'KPIs' section with 'Dim Brand' and 'Top 10_ Brand'. Under 'Calculated Members', there are 'Brand ID' and 'Brand Name'. The main area displays a query results grid with columns: Country, Sustainability Rating, and Fact Count. The data includes rows for Australia (A, B, C, D), Brazil (A, B, C, D), China (A, B, C, D), and France (A, B, C). The Fact Count values range from 109 to 138.

Country	Sustainability Rating	Fact Count
Australia	A	124
Australia	B	120
Australia	C	133
Australia	D	122
Brazil	A	109
Brazil	B	135
Brazil	C	131
Brazil	D	125
China	A	114
China	B	119
China	C	131
China	D	138
France	A	132
France	B	116
France	C	121

Hình 3.81. Kết quả thực thi câu truy vấn 10 dùng CUBE

3.3 Quá trình phân tích dữ liệu bằng công cụ SSAS trên Excel

3.3.1 Câu truy vấn 1

Nội dung câu truy vấn: Thống kê số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của ngành thời trang theo từng năm.

Cách thực hiện:

1. Kéo thuộc tính Year vào Rows.
2. Kéo thuộc tính Carbon Footprint MT, Water Usage Liters, Waste Production KG từ bảng Fact vào Values.

The screenshot shows the 'PivotTable Fields' ribbon in Excel. The 'Filters' section is empty. The 'Columns' section has a dropdown set to 'Σ Values'. The 'Rows' section has a dropdown set to 'Year'. The 'Values' section contains three dropdowns: 'Carbon Footprint MT', 'Water Usage Liters', and 'Waste Production KG'.

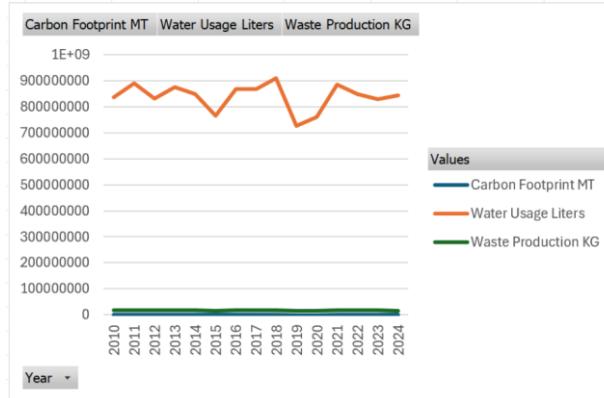
Hình 3. 82. Thao tác câu truy vấn 1 bằng Excel

Kết quả:

Row Labels	Carbon Footprint MT	Water Usage Liters	Waste Production KG
2010	88531.24	835693585.5	16835398.18
2011	83051.04	891044714.9	17376872.71
2012	86983.44	832655737.9	16930701.3
2013	79694.18	875439105.2	17131906.52
2014	84541.5	849729904.7	16470282.89
2015	76869.99	765729798.5	15365043.93
2016	91559.4	867024758	18411242.72
2017	84448.67	867303774.4	17342478.44
2018	89826.07	908686124.3	17844880.85
2019	74149.48	727622497.6	14809252.58
2020	73764.56	760285746.2	14498320.14
2021	88153.66	886486307.4	18038963.25
2022	82994.66	849423996.3	16392074.49
2023	81568.61	828431879.5	16968209.95
2024	85451.25	843750831.7	16120892.29
Grand Total	1251587.75	12589308762	250536520.2

Hình 3. 83. Kết quả thực thi câu truy vấn 1 bằng Excel

3. Tạo pivot chart loại đường để biểu diễn sự thay đổi của các giá trị qua các năm.



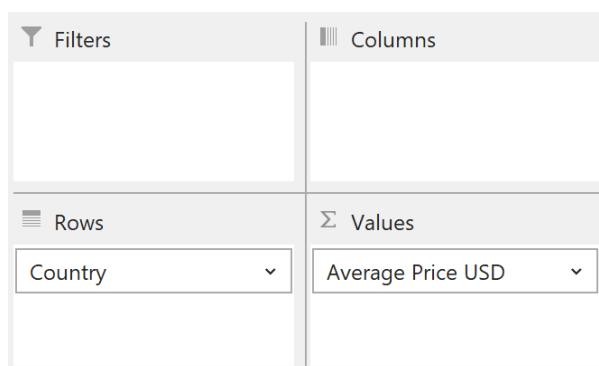
Hình 3. 84. Biểu đồ kết quả câu truy vấn 1 bằng Excel

3.3.2 Câu truy vấn 2

Nội dung câu truy vấn: Thống kê tổng doanh thu của các thương hiệu ở các nước.

Cách thực hiện:

- Kéo thuộc tính Country vào Rows.
- Kéo thuộc tính Average Price USD vào Values.



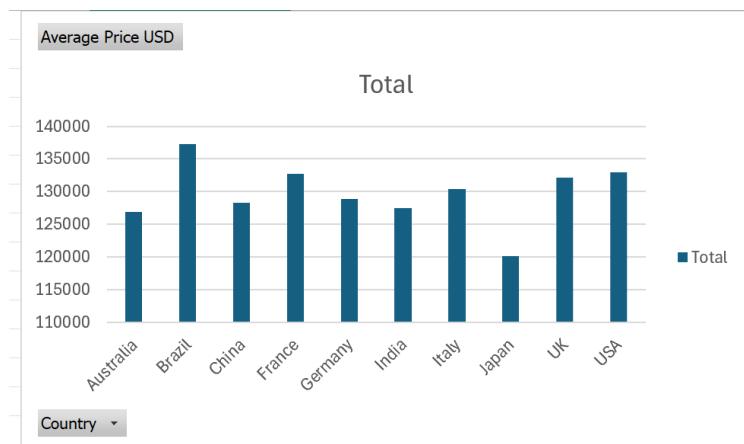
Hình 3. 85. Thao tác câu truy vấn 2 bằng Excel

Kết quả:

Row Labels	Average Price USD
Australia	126848.4
Brazil	137223.25
China	128226.07
France	132702.37
Germany	128873.45
India	127476.05
Italy	130362.7
Japan	120054.49
UK	132087.47
USA	132916.83
Grand Total	1296771.08

Hình 3. 86. Kết quả thực thi câu truy vấn 2 bằng Excel

3. Tạo Pivot Chart dạng cột để nhận thấy sự chênh lệch.



Hình 3. 87. Biểu đồ kết quả câu truy vấn 2 bằng Excel

3.3.3 Câu truy vấn 3

Nội dung câu truy vấn: Thông kê tổng số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của từng loại vật liệu đã thải ra từ năm 2010 đến 2024.s

Cách thực hiện:

1. Kéo thuộc tính vào Material Type vào Rows.

2. Kéo thuộc tính Carbon Footprint MT, Waste Production KG, Water Usage Liters vào Values.

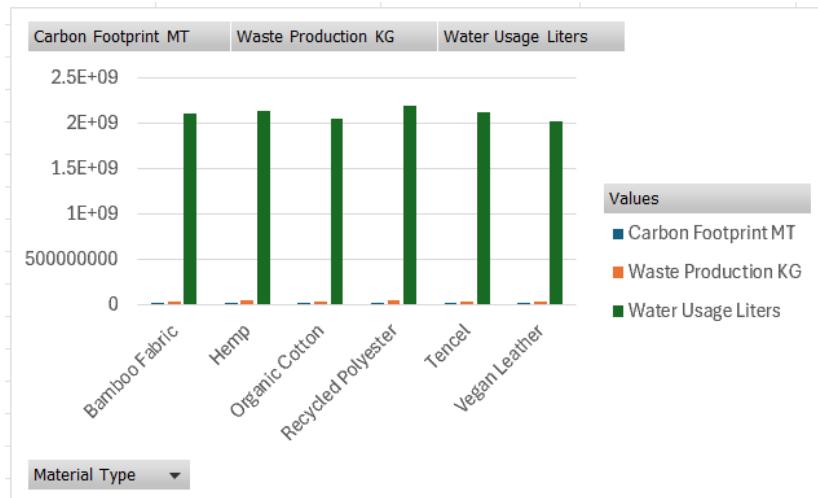
Hình 3. 88. Thao tác câu truy vấn 3 bằng Excel

Kết quả hiển thị:

Row Labels	Carbon Footprint MT	Waste Production KG	Water Usage Liters
Bamboo Fabric	205415.61	41856232.15	2095858839
Hemp	213271.55	42444672.85	2127045944
Organic Cotton	204043.78	41104045.75	2043406143
Recycled Polyester	224853.65	42545018.98	2181995126
Tencel	203611.37	40750944.81	2118911372
Vegan Leather	200391.79	41835605.7	2022091338
Grand Total	1251587.75	250536520.2	12589308762

Hình 3. 89. Kết quả thực thi câu truy vấn 3 bằng Excel

3. Tạo Pivot Chart dạng cột



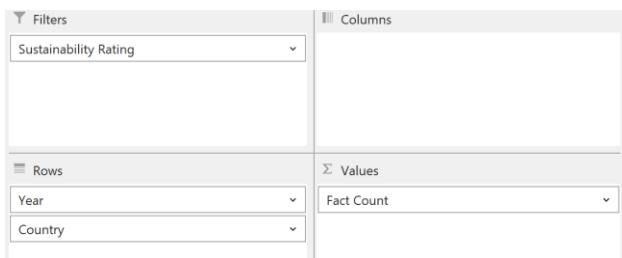
Hình 3. 90. Biểu đồ kết quả câu truy vấn 3 bằng Excel

3.3.4 Câu truy vấn 4

Nội dung câu truy vấn: Thống kê số lượng thương hiệu thời trang có đánh giá tiêu chuẩn (Sustainability_Rating) là A của mỗi nước trong năm 2014 và 2024.

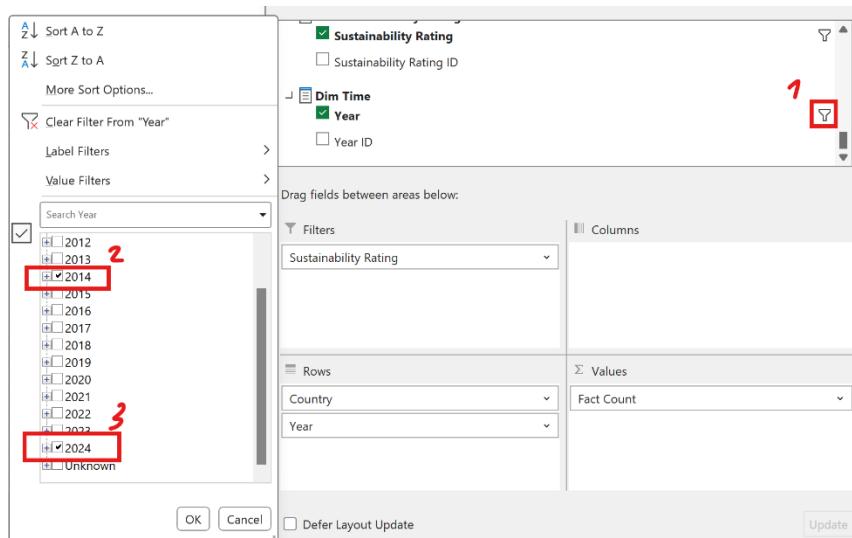
Cách thực hiện:

1. Kéo thuộc tính Sustainability Rating vào Filters.
2. Kéo thuộc tính Year và Country vào Rows.
3. Kéo Fact Count vào Values.



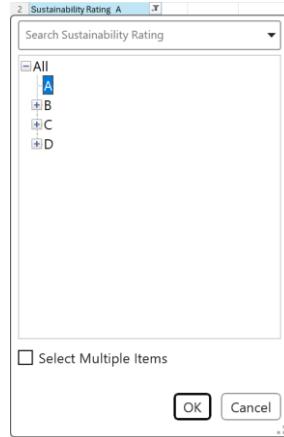
Hình 3. 91. Thao tác câu truy vấn 4 bằng Excel

4. Ở thuộc tính Year, chọn vào biểu tượng hình cái phễu, cửa sổ hiện ra tích chọn năm 2014 và 2024.



Hình 3.92. Lọc thuộc tính Year

5. Trong phần lọc tỉ lệ bền vững (Sustainability Rating), bỏ chọn Select Multiple Items (để chỉ được chọn một giá trị). Sau đó chọn giá trị A.



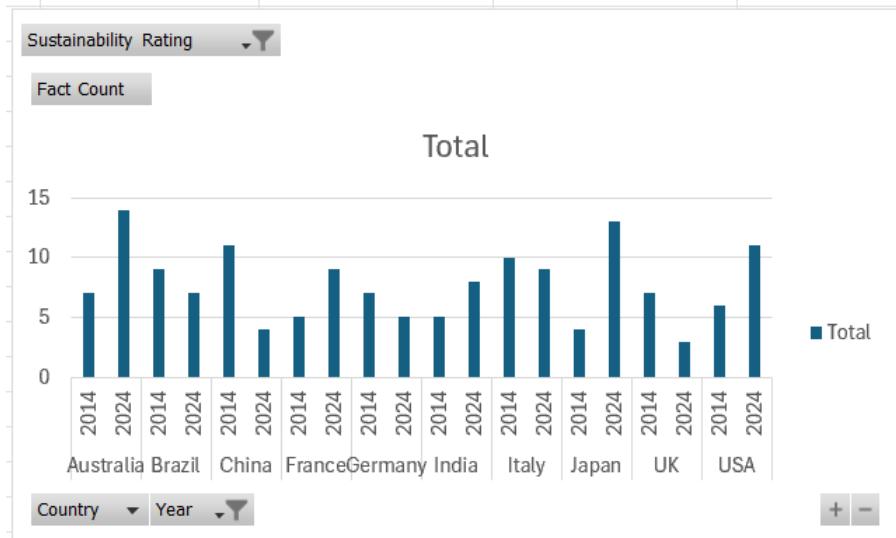
Hình 3. 93. Lọc thuộc tính Sustainability Rating

Kết quả hiển thị:

Sustainability Rating A	
Row Labels	Fact Count
Australia	
2014	7
2024	14
Brazil	
2014	9
2024	7
China	
2014	11
2024	4
France	
2014	5
2024	9
Germany	
2014	7
2024	5
India	
2014	5
2024	8
Italy	
2014	10
2024	9
Japan	
2014	4
2024	13
UK	
2014	7
2024	3
USA	
2014	6
2024	11
Grand Total	
154	

Hình 3. 94. Kết quả thực thi câu truy vấn 4 bằng Excel

6. Tạo Pivot Chart dạng cột để biểu diễn rõ hơn sự thay đổi về số lượng thương hiệu đạt tỉ lệ bền vững cao sau 10 năm của mỗi quốc gia.



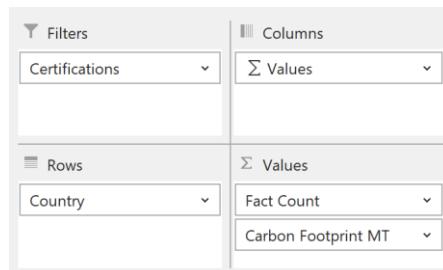
Hình 3. 95. Biểu đồ kết quả câu truy vấn 4 bằng Excel

3.3.5 Câu truy vấn 5

Nội dung câu truy vấn: Liệt kê số thương hiệu thuộc nước France, Italy, Japan và USA không có giấy chứng nhận và tổng lượng khí thải carbon mà các nhãn hàng ấy đã thải ra.

Cách thực hiện:

1. Kéo thả thuộc tính Certification vào Filters.
2. Kéo thả thuộc tính Country vào Columns.
3. Kéo thả thuộc tính Fact Count và Carbon Footprint MT vào Values.



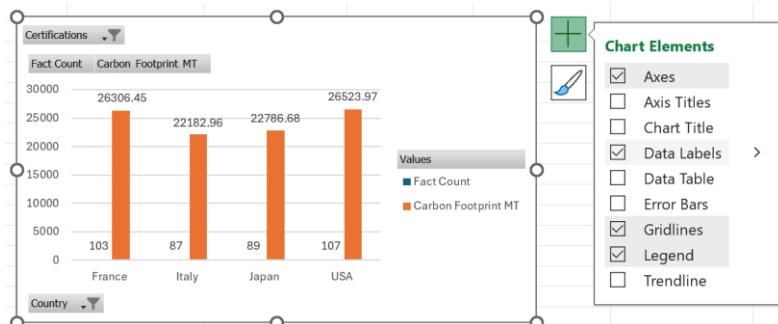
Hình 3. 96. Thao tác câu truy vấn 5 bằng Excel

4. Trong phần lọc Certifications chọn giá trị None (nghĩa là không có chứng chỉ). Và kết quả hiển thị như sau.

Certifications	None	
Row Labels		Fact Count Carbon Footprint MT
France	103	26306.45
Italy	87	22182.96
Japan	89	22786.68
USA	107	26523.97
Grand Total	386	97800.06

Hình 3. 97. Kết quả câu truy vấn 5 bằng Excel

5. Tạo Pivot Chart dạng cột để thấy sự chênh lệch lượng carbon thải ra của các thương hiệu không có những chứng chỉ ở các 4 nước hàng đầu về thời trang. Click vào biểu tượng và tick chọn Data Label để hiển thị giá trị trong biểu đồ.



Hình 3. 98. Biểu đồ kết quả câu truy vấn 5 bằng Excel

3.3.6 Câu truy vấn 6

Nội dung câu truy vấn: Liệt kê 10 thương hiệu có giá trung bình cao nhất và tỉ lệ bền vững của chúng.

Cách thực hiện:

- Kéo thuộc tính Brand Name vào Rows.
- Kéo thuộc tính Sustainability Rating vào Columns.
- Kéo Average Price USD vào Values.

Hình 3. 99. Thao tác câu truy vấn 6 bằng Excel

4. Click vào biểu tượng cái phễu ở thuộc tính Brand Name.

Hình 3. 100. Thao tác trên thuộc tính Brand Name

5. Ở cửa sổ hiện ra chọn Value Filters, chọn Top 10...

Hình 3. 111. Lấy ra top 10 thương hiệu

6. Cửa sổ mới hiện ra, chọn các giá trị như hình để lấy ra 10 thương hiệu có giá trung bình cao nhất.

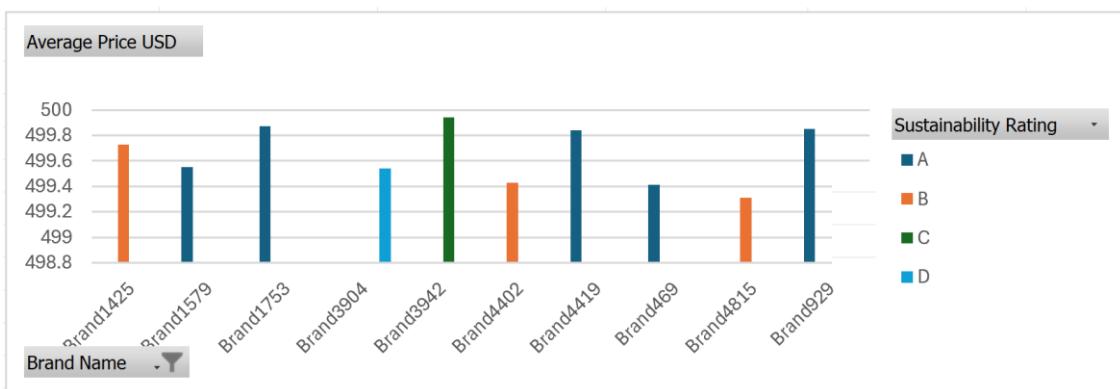
Hình 3. 112. Điều chỉnh giá trị để lấy ra 10 thương hiệu có giá trị trung bình cao nhất

Kết quả hiển thị như hình:

Row Labels	Average Price USD				Grand Total
	A	B	C	D	
Brand1425		499.73			499.73
Brand1579		499.55			499.55
Brand1753		499.87			499.87
Brand3904				499.54	499.54
Brand3942			499.94		499.94
Brand4402		499.43			499.43
Brand4419		499.84			499.84
Brand469		499.41			499.41
Brand4815			499.31		499.31
Brand929		499.85			499.85
Grand Total	2498.52	1498.47	499.94	499.54	4996.47

Hình 3. 113. Kết quả thực thi câu truy vấn 6 bằng Excel

7. Tạo Pivot Chart dạng cột để thể hiện số lượng thương hiệu mỗi loại tỉ lệ bền vững.



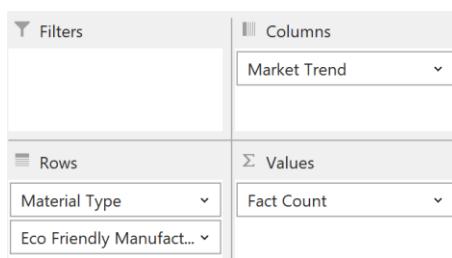
Hình 3. 114. Biểu đồ thực thi câu truy vấn 6 bằng Excel

3.3.7 Câu truy vấn 7

Nội dung câu truy vấn: Thống kê số lượng mỗi loại vật liệu theo từng xu hướng thị trường tăng (Growing) trong trường hợp sản xuất thân thiện với môi trường và không.

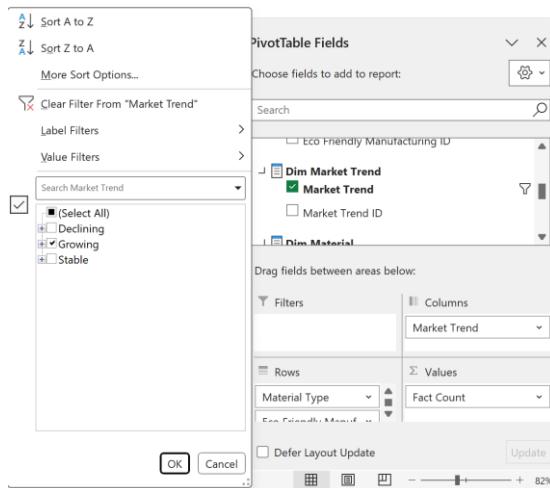
Cách thực hiện:

- Kéo thuộc tính Material Type và Eco Friendly Manufacturing vào Rows.
- Kéo thuộc tính Market Trend vào Columns.
- Kéo Fact Count vào Values.



Hình 3. 115. Thao tác câu truy vấn 7 bằng Excel

4. Tiến hành lọc thuộc tính Market Trend để lấy giá trị Growing.



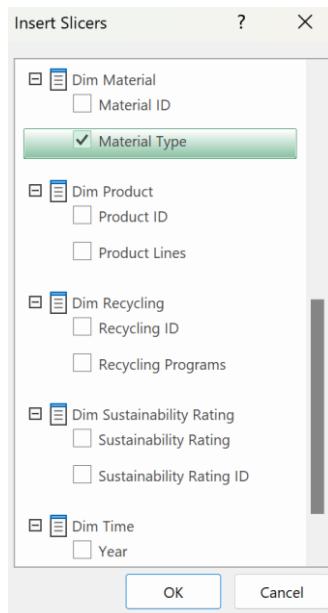
Hình 3. 116. Lọc thuộc tính Market Trend

Kết quả hiển thị như hình:

Fact Count	Column Labels	Grand Total
Row Labels	Growing	
Bamboo Fabric		
No	115	115
Yes	137	137
Hemp		
No	129	129
Yes	135	135
Organic Cotton		
No	123	123
Yes	150	150
Recycled Polyester		
No	140	140
Yes	126	126
Tencel		
No	156	156
Yes	128	128
Vegan Leather		
No	121	121
Yes	152	152
Grand Total	1612	1612

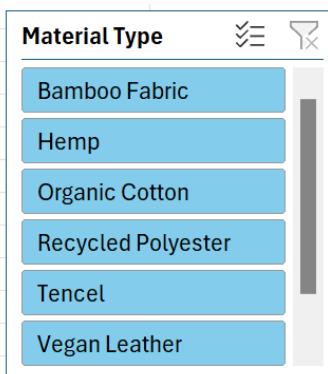
Hình 3. 117. Kết quả thực thi câu truy vấn 7 bằng Excel

5. Tạo slicer bằng cách chọn mục Items, sau đó chọn Slicer. Ở cửa sổ hiện lên, tick chọn Material Type và nhấn OK.



Hình 3. 118. Thao tác tạo Slicer cho thuộc tính Material Type

Màn hình hiển thị Slicer như hình:



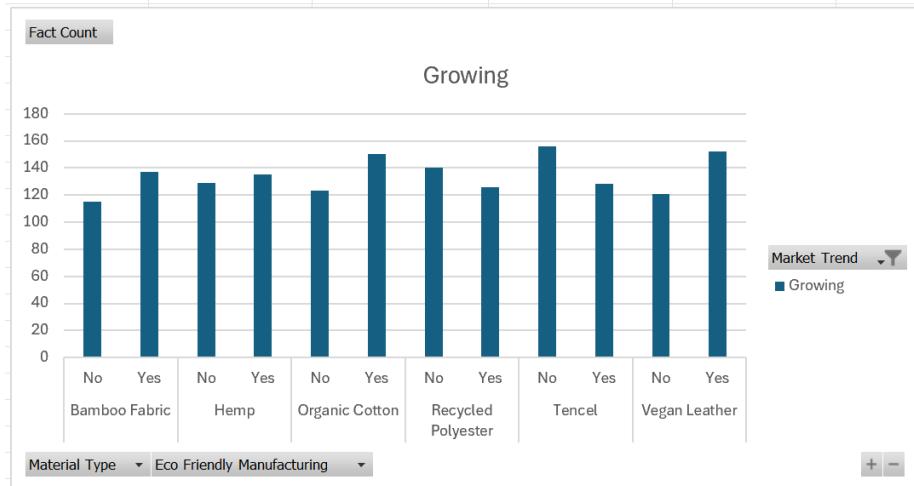
Hình 3. 119. Slider cho thuộc tính Material Type

Tương tự, tạo Slicer cho thuộc tính Eco Friendly Manufacturing.



Hình 3. 120. Slider cho thuộc tính Eco Friendly Manufacturing

6. Bước 5: Tạo Pivot Chart dạng cột để thể hiện sự chênh lệch và xu hướng của các loại vật liệu.



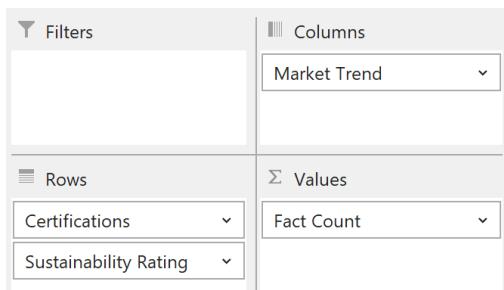
Hình 3. 121. Biểu đồ kết quả thực thi câu truy vấn 7 bằng Excel

3.3.8 Câu truy vấn 8

Nội dung câu truy vấn: Với từng loại chứng chỉ, khi đạt các mức độ bền vững khác nhau hãy thống kê số lượng thương hiệu theo từng xu hướng thị trường.

Cách thực hiện:

1. Kéo thuộc tính Certifications và Sustainability Rating vào Rows.
2. Kéo thuộc tính Market Trend vào Columns.
3. Kéo Fact Count vào Values.



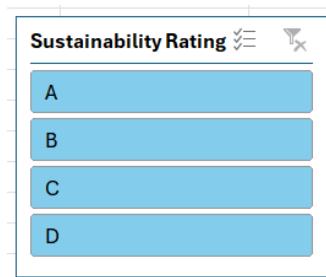
Hình 3. 122. Thao tác câu truy vấn 8 bằng Excel

Kết quả hiển thị như hình:

Fact Count	Column Labels			
Row Labels	Declining	Growing	Stable	Grand Total
■ B Corp				
A	89	80	74	243
B	89	85	81	255
C	80	79	97	256
D	100	68	65	233
■ Fair Trade				
A	77	80	67	224
B	86	74	73	233
C	61	78	91	230
D	73	102	69	244
■ GOTS				
A	85	74	100	259
B	90	79	94	263
C	91	81	98	270
D	91	78	89	258
■ None				
A	88	69	92	249
B	95	87	74	256
C	84	78	86	248
D	89	79	83	251
■ OEKOTEX				
A	65	94	97	256
B	86	94	91	271
C	82	72	84	238
D	78	81	104	263
Grand Total	1679	1612	1709	5000

Hình 3. 123. Kết quả thực thi câu truy vấn 8 bằng Excel

4. Thực hiện tạo Slicer cho thuộc tính Sustainability Rating và Certifications.

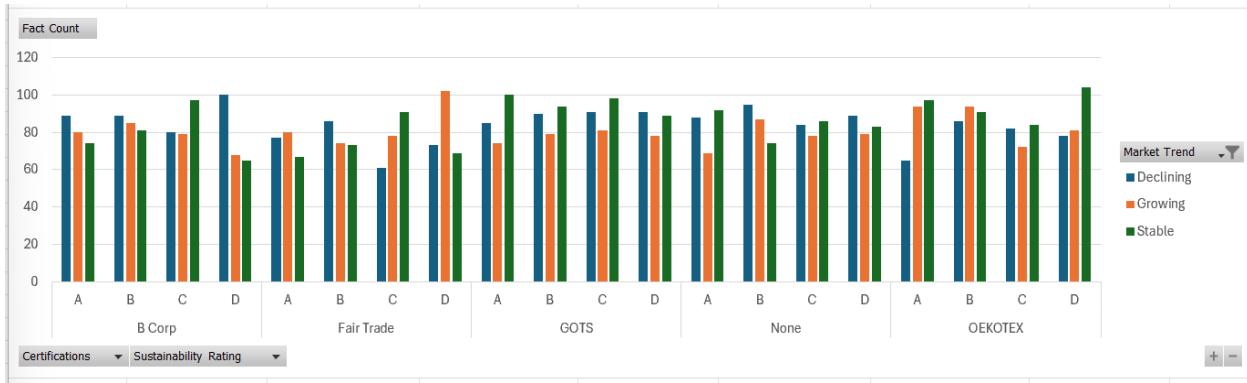


Hình 3. 124. Slicer Sustainability Rating



Hình 3. 125. Slicer Certifications

5. Tạo Pivot Chart để xem xét sự chênh lệch về xu hướng giữa các loại chứng chỉ.



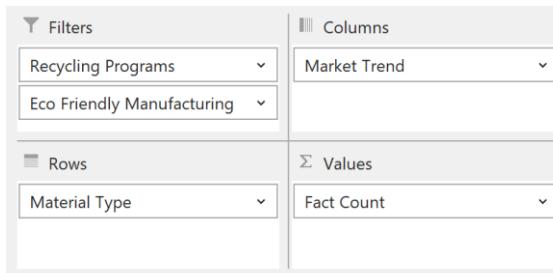
Hình 3. 126. Biểu đồ kết quả thực thi câu truy vấn 8 bằng Excel

3.3.9 Câu truy vấn 9

Nội dung câu truy vấn: Với mỗi vật liệu hãy cho biết số lượng thương hiệu sử dụng cả chương trình tái chế và sản xuất thân thiện với môi trường theo từng xu hướng.

Cách thực hiện:

- Kéo thuộc tính Recycling và Eco Friendly Manufacturing vào Filter.
- Kéo thuộc tính Market Trend vào Columns.
- Kéo Fact Count vào Values.



Hình 3. 127. Thao tác câu truy vấn 9 bằng Excel

- Chọn giá trị của thuộc tính Recycling và Eco Friendly Manufacturing là Yes.

Recycling Programs	Yes
Eco Friendly Manufacturing	Yes

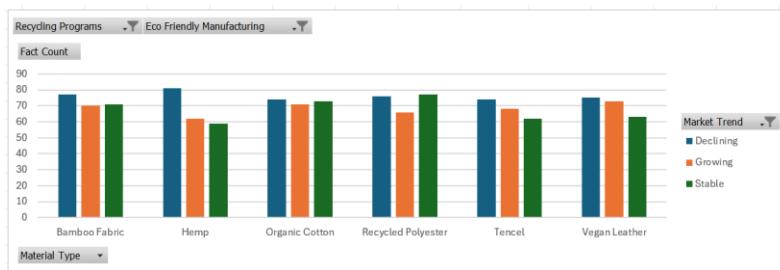
Hình 3. 128. Lọc thuộc tính Recycling Programs và Eco Friendly Manufacturing

Kết quả hiển thị như hình:

Fact Count	Column Labels				Grand Total
	Declining	Growing	Stable	Grand Total	
Bamboo Fabric		77	70	71	218
Hemp		81	62	59	202
Organic Cotton		74	71	73	218
Recycled Polyester		76	66	77	219
Tencel		74	68	62	204
Vegan Leather		75	73	63	211
Grand Total	457	410	405	1272	

Hình 3. 129. Kết quả thực thi câu truy vấn 9 bằng Excel

5. Tạo Pivot Chart dạng cột để thể hiện xu hướng của mỗi loại vật liệu.



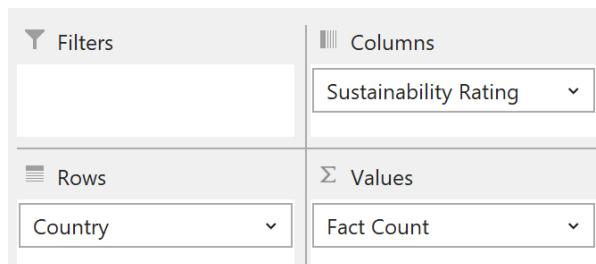
Hình 3. 130. Biểu đồ kết quả thực thi câu truy vấn 9 bằng Excel

3.3.10 Câu truy vấn 10

Nội dung câu truy vấn: Thống kê số lượng thương hiệu đạt các mức bền vững ở mỗi nước.

Cách thực hiện:

- Kéo thuộc tính Country vào Rows.
- Kéo thuộc tính Sustainability Rating vào Columns.
- Kéo Fact Count vào Values.



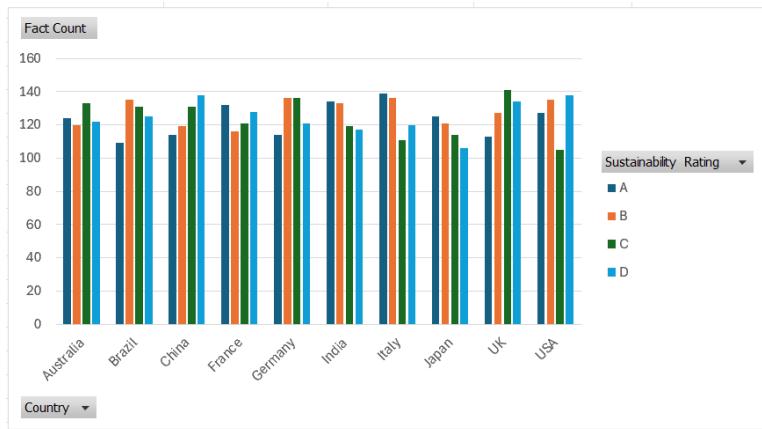
Hình 3. 131. Thao tác câu truy vấn 10 bằng Excel

Kết quả hiển thị như trên hình:

Row Labels	A	B	C	D	Grand Total	
Australia		124	120	133	122	499
Brazil		109	135	131	125	500
China		114	119	131	138	502
France		132	116	121	128	497
Germany		114	136	136	121	507
India		134	133	119	117	503
Italy		139	136	111	120	506
Japan		125	121	114	106	466
UK		113	127	141	134	515
USA		127	135	105	138	505
Grand Total	1231	1278	1242	1249	5000	

Hình 3. 132. Kết quả thực thi câu truy vấn 10 bằng Excel

4. Tạo Pivot Chart.



Hình 3. 133. Biểu đồ kết quả câu truy vấn 10 bằng Excel

*Nhận xét: Kết quả các câu truy vấn bằng Excel cho ra kết quả giống với các câu truy vấn bằng CUBE.

3.4 Quá trình phân tích dữ liệu bằng ngôn ngữ MDX

3.4.1 Câu truy vấn 1:

Nội dung câu truy vấn: Thông kê số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của ngành thời trang theo từng năm.

Câu lệnh truy vấn bằng MDX:

SELECT

```
{[Measures].[Carbon Footprint MT], [Measures].[Water UsageLiters],
[Measures].[Waste Production KG]} on columns,
```

non empty({[Dim Time].[Year].children})on rows

FROM [SUSTAINABLE FASHION];

Kết quả câu truy vấn:

```

/*Câu 1. Thống kê số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của ngành thời trang theo từng năm.*/
SELECT {[Measures].[Carbon Footprint MT], [Measures].[Water Usage Liters],[Measures].[Waste Production KG]} on columns,
non empty({[Dim Time].[Year].children} ) on rows
FROM [SUSTAINABLE FASHION];

```

Year	Carbon Footprint MT	Water Usage Liters	Waste Production KG
2010	88531.24	83563365.459999	16833818
2011	83018340000001	891044714.92	17336712.71
2012	82301803.44	829301737.89	16833813
2013	79884180000001	876429171.12	17133504.55
2014	84541.5	84972904.68	16473262.89
2015	78601.99	76572798.51	15380404.93
2016	91504.2	86702179.86	18411242.72
2017	88926.47	87730371.64	17241144.64
2018	89826.97	905086124.34	1784480.85
2019	74164800000001	72782497.64	14890252.58
2020	73716.86	790285744.19	14490323.14
2021	81506.44	8580351.4	16208932.35
2022	8298459999999999	849423995.279999	1532014.49
2023	81506.61	828431676.52	16582030.95
2024	85401.26	84576031.69	16120892.29

Hình 3. 134. Kết quả truy vấn câu 1 bằng MDX

3.4.2 Câu truy vấn 2:

Nội dung câu truy vấn: Thống kê tổng doanh thu của các thương hiệu thời trang ở các nước.

Câu lệnh truy vấn bằng MDX:

SELECT

{[Measures].[Average Price USD]} on columns,

non empty ([Dim Brand].[Country].children) on rows

FROM [SUSTAINABLE FASHION];

Kết quả câu truy vấn:

The screenshot shows the MDXQuery2.mdx query editor window. The left pane displays the cube structure under 'Cube: SUSTAINABLE FASHION'. The right pane shows the MDX query and its results.

```

/*Câu 2: Thống kê tổng doanh thu của các nhãn thời trang ở các nước*/
SELECT {[Measures].[Average Price USD]} ON COLUMNS,
NON EMPTY ([Dim Brand].[Country].children) ON ROWS
FROM [SUSTAINABLE FASHION];
  
```

The results table shows Average Price USD for various countries:

Country	Average Price USD
Australia	129448.4
Brazil	137223.25
China	128226.07
France	132702.37
Germany	133974.58
India	127476.08
Italy	130362.7
Japan	120554.49
UK	132087.47
USA	132916.83

At the bottom, it says 'Query executed successfully.'

Hình 3. 135. Kết quả truy vấn câu 2 bằng MDX

3.4.3 Câu truy vấn 3:

Nội dung câu truy vấn: Thống kê tổng số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của từng loại vật liệu đã thải ra từ năm 2010 đến 2024.

Câu lệnh truy vấn bằng MDX:

SELECT

```

{[Measures].[Carbon Footprint MT],
[Measures].[Water Usage Liters],
[Measures].[Waste Production KG]} ON COLUMNS,
NON EMPTY ([Dim Material].[Material Type].Children) *
([Dim Time].[Year].children) ON ROWS
  
```

FROM [SUSTAINABLE FASHION];

Kết quả câu truy vấn:

The screenshot shows the MDXQuery2.mdx interface. On the left, the cube 'SUSTAINABLE FASHION' is selected. The 'Measure Group' dropdown shows '<All>'. The 'Measures' section includes 'Fact' (with 'Average Price USD', 'Carbon Footprint MT', 'Fact Count', 'Waste Production KG', 'Water Usage Liters') and 'KPIs' (with 'Brand_Waste_Under_10000', 'Top_10_Brand', 'Brand ID', 'Brand Name', 'Country', 'PharCap_Dim_Brand', 'Dim_Certification', 'Certification_ID', 'Certifications', 'Eco_Friendly_Manufacturing', 'Eco_Friendly_Manufacturing_ID', 'Dim_Market_Trend', 'Market_Trend_ID', 'Dim_Material', 'Material_ID', 'Material_Type', 'Dim_Product', 'Product_ID', 'Product_Lines'). The main area displays the MDX query:

```

/*Câu 3: Thống kê tổng số lượng khí nhà kính (Carbon_Footprint_MT),  
lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG)  
của từng loại vật liệu đã thải ra từ năm 2010 đến 2024.*/
SELECT {[Measures].[Carbon Footprint MT],  
[Measures].[Water Usage Liters],  
[Measures].[Waste Production KG]} ON COLUMNS,  
NON EMPTY ([Dim Material].[Material Type].Children) *  
([Dim Time].[Year].Children) ON ROWS  
FROM [SUSTAINABLE FASHION];

```

The results grid shows data for Bamboo Fabric from 2010 to 2018, including Carbon Footprint MT, Water Usage Liters, and Waste Production KG. A message bar at the bottom indicates 'Query executed successfully.'

Hình 3. 136. Kết quả truy vấn câu 3 bằng MDX

3.4.4 Câu truy vấn 4:

Nội dung câu truy vấn: Thống kê số lượng thương hiệu thời trang có đánh giá tiêu chuẩn (Sustainability_Rating) là A của mỗi nước trong năm 2014 và 2024.

Câu lệnh truy vấn bằng MDX:

SELECT

{[Measures].[Fact Count]} ON COLUMNS,

NON EMPTY([Dim Brand].[Country].children) * {[Dim Time].[Year].[2014],[Dim Time].[Year].[2024]} ON ROWS

FROM

[SUSTAINABLE FASHION]

WHERE

[Dim Sustainability Rating].[Sustainability Rating].[A]);

Kết quả câu truy vấn

```

Câu Lệnh MDX.mdx ...(ASUS_TN\22521)* # X
Cube: SUSTAINABLE FASHION
Measure Group: <All>
<All>
    SUSTAINABLE FASHION
    al Measures
    KPIs
    Dim Brand
        Brand_Waste_Under_10000
        Top_10_Brand
        Brand ID
        Brand Name
        Country
        PhanCap_Dim_Brand
        Dim Certification
        Dim Eco Friendly Manufacturing
        Dim Market Trend
        Dim Material
        Dim Product
        Dim Recycling
        Dim Sustainability Rating
            Sustainability Rating
            Sustainability Rating ID
        Dim Time
        Year
        Year ID
    Dim Time
    Year
    Year ID

    /Câu 4: Thông kê số lượng nhãn hàng thời trang có đánh giá tiêu chuẩn (Sustainability_Rating) là A của mỗi nước trong năm 2014 và 2024. /
    SELECT {[Measures].[Fact Count]} ON COLUMNS,
    NON EMPTY {[Dim Brand].[Country].children} * {[Dim Time].[Year].[2014],[Dim Time].[Year].[2024]} ON ROWS
    FROM [SUSTAINABLE FASHION]
    WHERE ([Dim Sustainability Rating].[Sustainability Rating].[A]);

```

Messages	Results	
	Fact Count	
Australia	2014	7
Australia	2024	14
Brazil	2014	9
Brazil	2024	7
China	2014	11
China	2024	4
France	2014	5
France	2024	9
Germany	2014	7
Germany	2024	5
India	2014	5
India	2024	8
Italy	2014	10
Italy	2024	9
Japan	2014	4
Japan	2024	13
UK	2014	7

Hình 3. 137. Kết quả truy vấn câu 4 bằng MDX

3.4.5 Câu truy vấn 5:

Nội dung câu truy vấn: Liệt kê số thương hiệu thuộc nước France, Italy, Japan và USA không có giấy chứng nhận và tổng lượng khí thải carbon mà các nhãn hàng áy đã thải ra.

Câu lệnh truy vấn bằng MDX:

SELECT

```

{[Measures].[Carbon Footprint MT],[Measures].[Fact Count]} ON COLUMNS,
{[Dim Brand].[Country].[France],
[Dim Brand].[Country].[Italy],
[Dim Brand].[Country].[Japan],
[Dim Brand].[Country].[USA]} ON ROWS

```

FROM

[SUSTAINABLE FASHION]

WHERE

([Dim Certification].[Certifications].[None]);

Kết quả câu truy vấn:

The screenshot shows the SSAS MDX Query Editor interface. On the left, there is a navigation pane with a tree view of the cube structure. The selected node is 'SUSTAINABLE FASHION'. The main area contains an MDX query and its execution results.

```

/*Câu 5: Liệt kê số nhãn hàng thuộc nước France, Italy, Japan và USA không có giấy chứng nhận
và lượng khí thải carbon mà nhãn hàng ấy đã thải ra.*/
SELECT {[Measures].[Carbon Footprint MT],[Measures].[Fact Count]} ON COLUMNS,
       {[Dim Brand].[Country].[France],
        [Dim Brand].[Country].[Italy],
        [Dim Brand].[Country].[Japan],
        [Dim Brand].[Country].[USA]} ON ROWS
FROM   [SUSTAINABLE FASHION]
WHERE  ([Dim Certification].[Certifications].[None]);
  
```

The results table shows the following data:

	Carbon Footprint MT	Fact Count
France	26306.45	103
Italy	22182.96	87
Japan	22786.68	89
USA	26523.97	107

At the bottom left, a green checkmark indicates "Query executed successfully". At the bottom right, the status bar shows "ASUS_TN ASUS_TN\22521 SSAS_22521113_22521008 00:00:01".

Hình 3. 138. Kết quả truy vấn câu 5 bằng MDX

3.4.6 Câu truy vấn 6:

Nội dung câu truy vấn: Liệt kê 10 thương hiệu có giá trung bình cao nhất và tỉ lệ bền vững của chúng.

Câu lệnh truy vấn bằng MDX:

```

SELECT
    NON EMPTY
        {[Measures].[Average Price USD]} ON COLUMNS,
    NON EMPTY(TOPCOUNT( [Dim Brand].[Brand Name].children, 10,
        [Measures].[Average Price USD] )*
            [Dim Sustainability Rating].[Sustainability Rating].children) ON ROWS
FROM
    [SUSTAINABLE FASHION];
  
```

Kết quả câu truy vấn:

		Average Price USD
Brand3942	C	499.94
Brand1753	A	499.87
Brand929	A	499.85
Brand4419	A	499.84
Brand1425	B	499.73
Brand1579	A	499.55
Brand3904	D	499.54
Brand4402	B	499.43
Brand469	A	499.41
Brand4815	B	499.31

Hình 3. 139. Kết quả truy vấn câu 6 bằng MDX

3.4.7 Câu truy vấn 7:

Nội dung câu truy vấn: Thông kê số lượng mỗi loại vật liệu theo từng xu hướng thị trường tăng (Growing) trong trường hợp sản xuất thân thiện với môi trường và không.

Câu lệnh truy vấn bằng MDX:

```

SELECT
{[Measures].[Fact Count]*

[Dim Market Trend].[Market Trend].children } ON COLUMNS,
NON EMPTY(([Dim Material].[Material Type].children) *

{[Dim Eco Friendly Manufacturing].[Eco Friendly Manufacturing].children}) ON
ROWS
FROM
[SUSTAINABLE FASHION]
WHERE [Dim Market Trend].[Market Trend].[Growing];

```

Kết quả câu truy vấn:

	Messades	Results
		Fact Count
Bamboo Fabric	No	115
Bamboo Fabric	Yes	137
Hemp	No	129
Hemp	Yes	135
Organic Cotton	No	123
Organic Cotton	Yes	150
Recycled Polyester	No	140
Recycled Polyester	Yes	126
Tencel	No	156
Tencel	Yes	128
Vegan Leather	No	121
Vegan Leather	Yes	152

Hình 3. 140. Kết quả truy vấn câu 7 bằng MDX

3.4.8 Câu truy vấn 8:

Nội dung câu truy vấn: Với từng loại chứng chỉ, khi đạt các mức độ bền vững khác nhau hãy thống kê số lượng thương hiệu theo từng xu hướng thị trường.

Câu lệnh truy vấn bằng MDX:

SELECT

```
{[Measures].[Fact Count]
*{[Dim Sustainability Rating].[Sustainability Rating].children}
*[Dim Market Trend].[Market Trend].children} ON COLUMNS,
NON EMPTY {[Dim Certification].[Certifications].children} ON ROWS
```

FROM

[SUSTAINABLE FASHION];

Kết quả câu truy vấn:

	Fact Count											
	A	A	A	B	B	B	C	C	C	D	D	D
	Declining	Growing	Stable									
B Corp	89	80	74	89	85	81	80	79	97	100	68	65
Fair Trade	77	80	67	86	74	73	61	78	91	73	102	69
GOTS	85	74	100	90	79	94	91	81	98	91	78	89
None	88	69	92	95	87	74	84	78	86	89	79	83
OEKOTEX	65	94	97	86	94	91	82	72	84	78	81	104

Hình 3. 141. Kết quả truy vấn câu 8 bằng MDX

3.4.9 Câu truy vấn 9:

Nội dung câu truy vấn: Với mỗi vật liệu hãy cho biết số lượng thương hiệu sử dụng cả chương trình tái chế và sản xuất thân thiện với môi trường theo từng xu hướng.

Câu lệnh truy vấn bằng MDX:

SELECT

```
{[Measures].[Fact Count]*[Dim Market Trend].[Market Trend].children} ON
COLUMNS,
{[Dim Material].[Material Type].children} ON ROWS
FROM
[SUSTAINABLE FASHION]
WHERE
([Dim Recycling].[Recycling Programs].&[Yes],
[Dim Eco Friendly Manufacturing].[Eco Friendly Manufacturing].[Yes]);
```

Kết quả câu truy vấn:

	Fact Count	Fact Count	Fact Count
	Declining	Growing	Stable
Bamboo Fabric	77	70	71
Hemp	81	62	59
Organic Cotton	74	71	73
Recycled Polyester	76	66	77
Tencel	74	68	62
Vegan Leather	75	73	63

Hình 3. 142. Kết quả truy vấn câu 9 bằng MDX

3.4.10 Câu truy vấn 10:

Nội dung câu truy vấn: Thông kê số lượng thương hiệu đạt các mức bền vững ở mỗi nước.

Câu lệnh truy vấn bằng MDX:

SELECT

```
{[Measures].[Fact Count]*[Dim Sustainability Rating].[Sustainability
Rating].children} ON COLUMNS,
{[Dim Brand].[Country].children } ON ROWS
FROM
[SUSTAINABLE FASHION];
```

Kết quả câu truy vấn:

	Fact Count	Fact Count	Fact Count	Fact Count
	A	B	C	D
Australia	124	120	133	122
Brazil	109	135	131	125
China	114	119	131	138
France	132	116	121	128
Germany	114	136	136	121
India	134	133	119	117
Italy	139	136	111	120
Japan	125	121	114	106
UK	113	127	141	134
USA	127	135	105	138

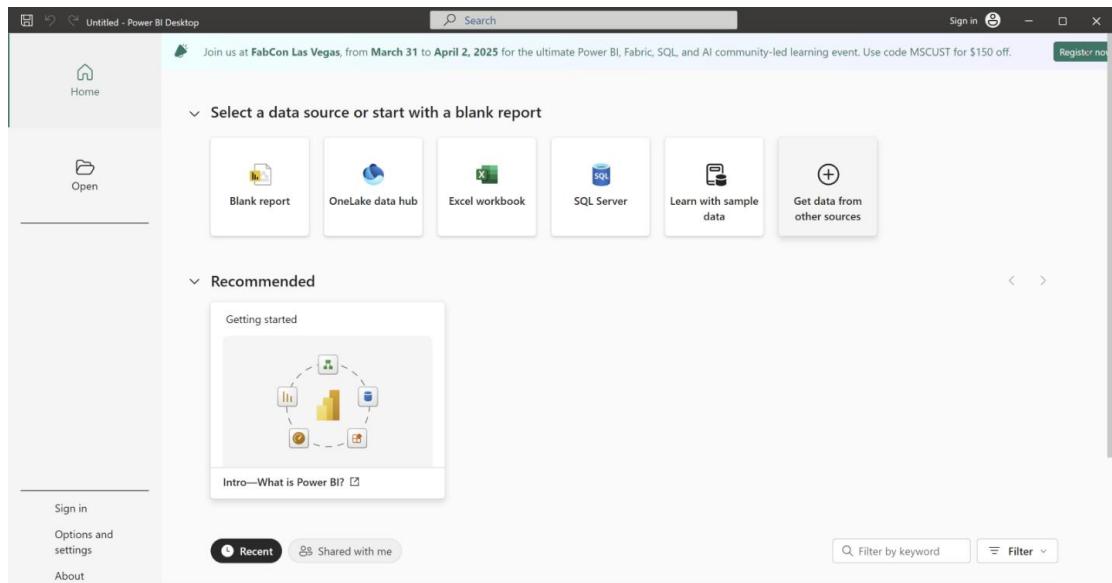
Hình 3. 143. Kết quả truy vấn câu 10 bằng MDX

*Nhận xét: Kết quả các câu truy vấn bằng MDX cho ra kết quả giống với các câu truy vấn bằng CUBE và Excel.

3.5 Quá trình lập báo biểu bằng công cụ Power BI

3.5.1 Tạo mới project và kết nối với cơ sở dữ liệu trên BI.

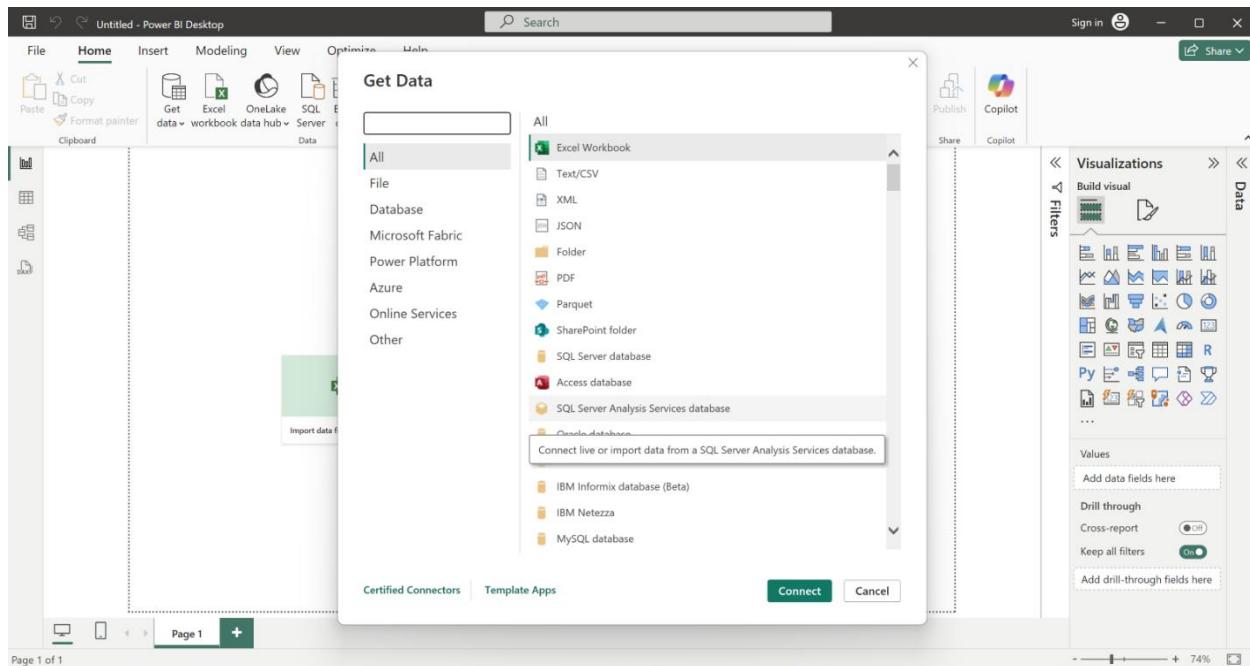
- Khởi chạy phần mềm Power BI desktop và chọn Get Data.



Hình 3. 144. Khởi chạy phần mềm Power BI desktop và chọn Get Data.

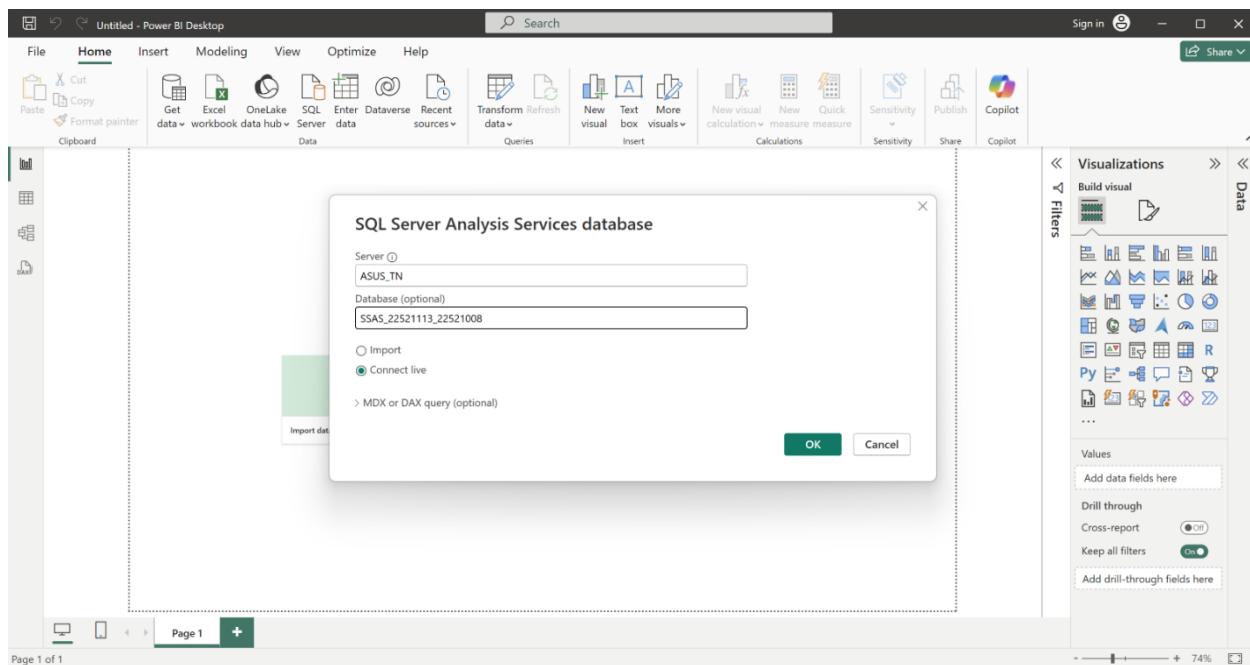
- Tiếp theo lựa chọn Get data => SQL Server Analysis Services database. Sau đó bấm Connect.

IS217 – Kho dữ liệu và OLAP



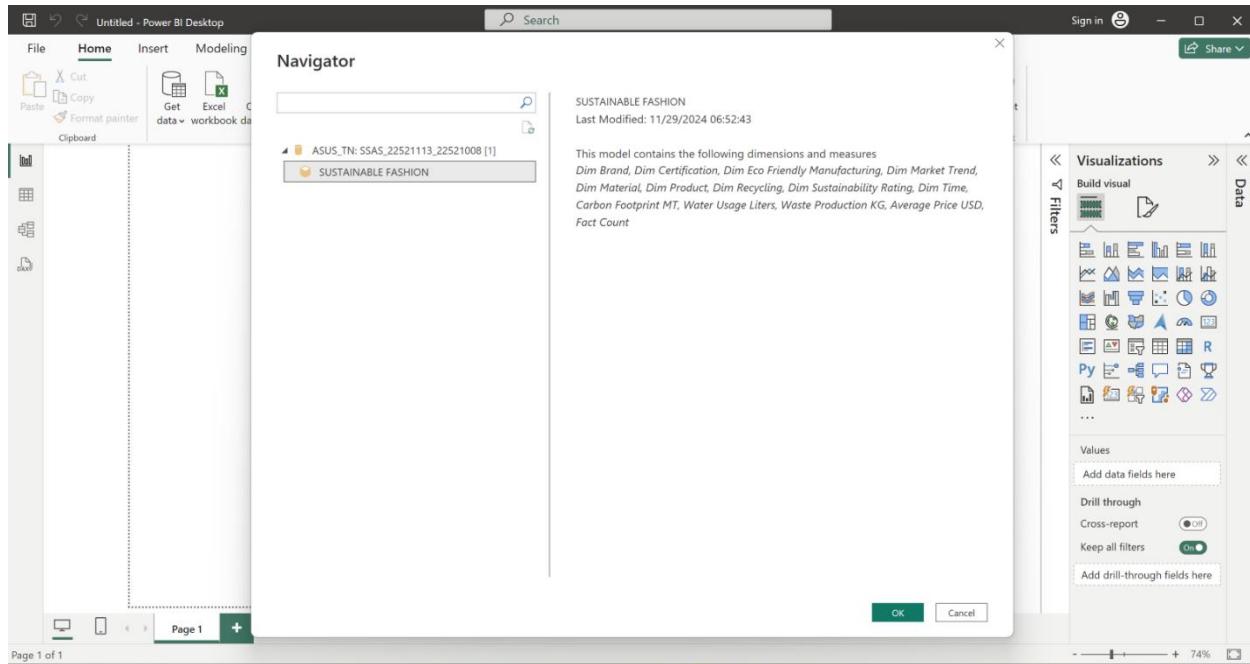
Hình 3. 145. Lựa chọn Get data và SQL Server Analysis Services database

Tiếp theo điền tên Server và Database. Sau đó bấm OK để tạo kết nối.



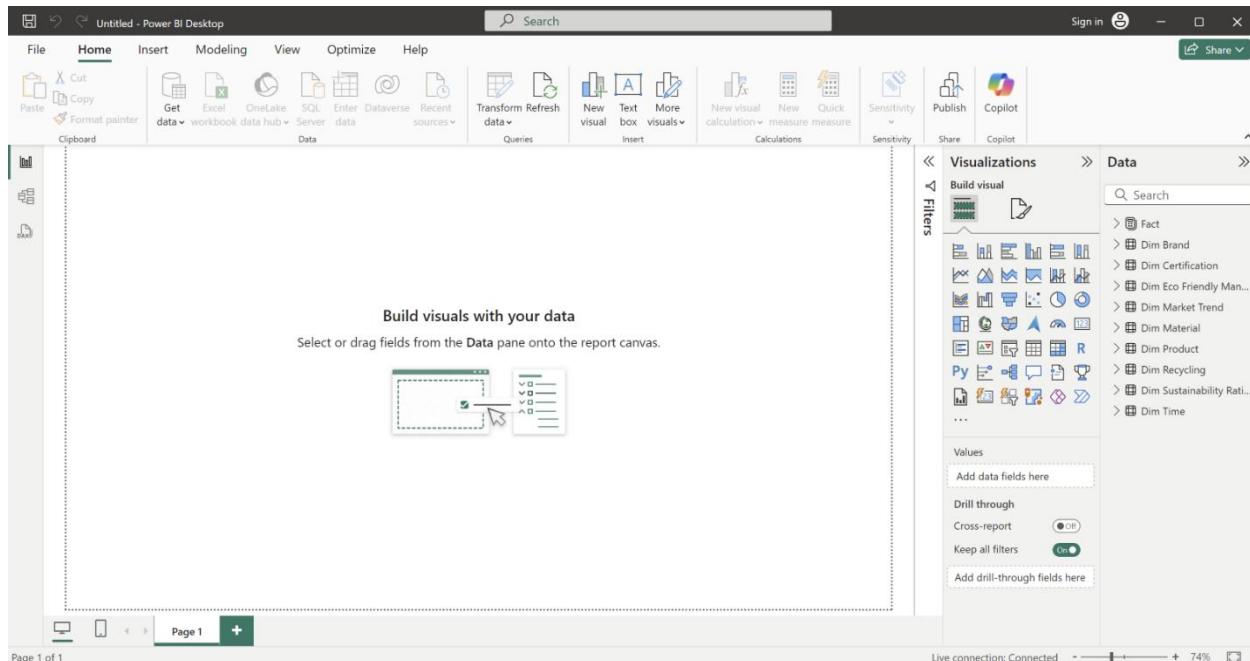
Hình 3. 146. Điền tên Server và Database

Tiếp theo lựa chọn database sau đó bấm OK.



Hình 3. 147. Lựa chọn database

- Màn hình BI sau khi tạo kết nối thành công.

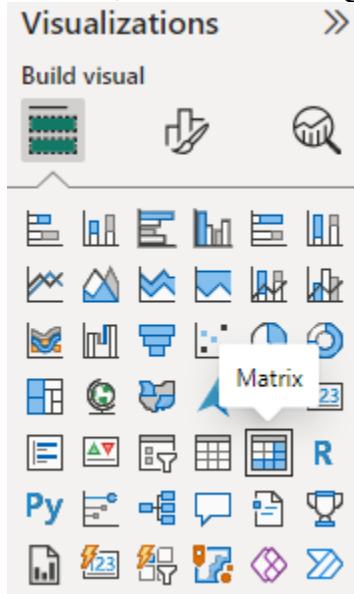


Hình 3. 148. Màn hình BI sau khi tạo kết nối thành công.

3.5.2 Report 1

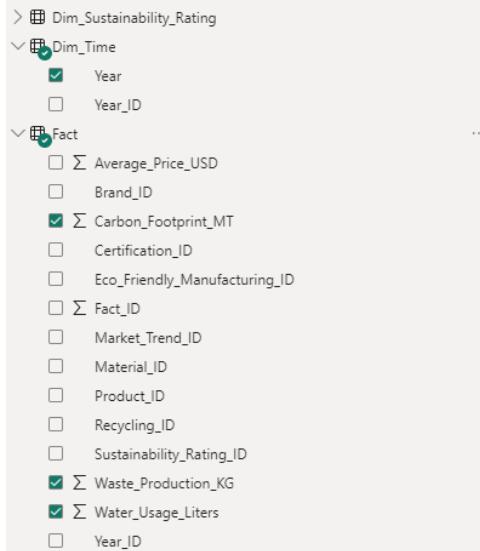
Nội dung: Thống kê số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của ngành thời trang theo từng năm.

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



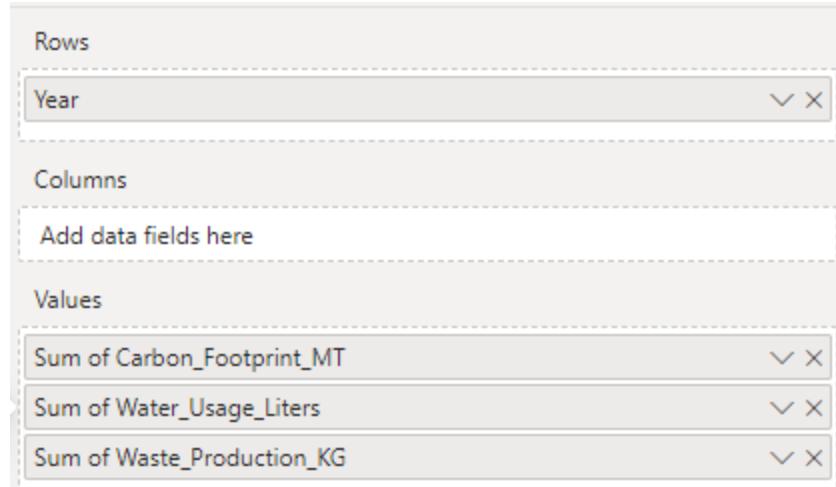
Hình 3. 149. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn.



Hình 3. 150. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị Rows, Values trong Visualizations



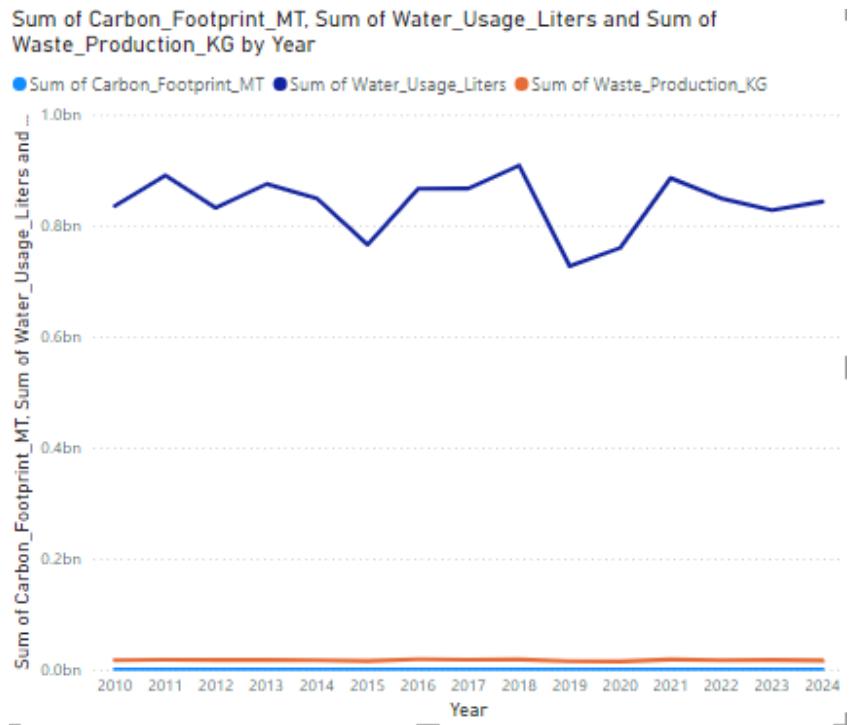
Hình 3. 151. Chính sửa hiển thị Rows, Values trong hộp thoại Visualizations

Kết quả dạng Matrix:

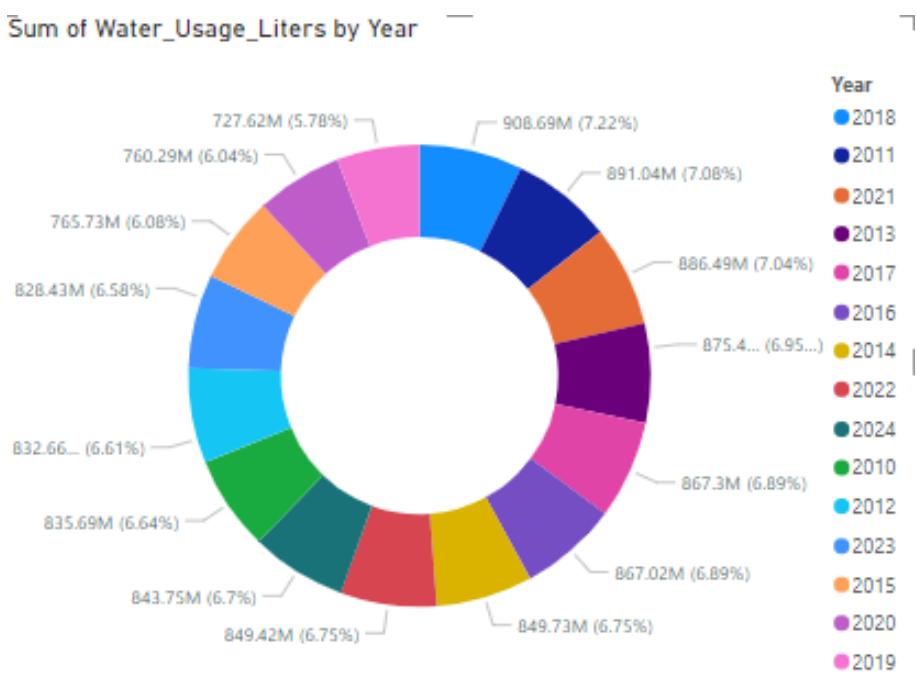
Year	Sum of Carbon_Footprint_MT	Sum of Water_Usage_Liters	Sum of Waste_Production_KG
2018	89,826.07	908,686,124.34	17,844,880.85
2011	83,051.04	891,044,714.92	17,376,872.71
2021	88,153.66	886,486,307.44	18,038,963.25
2013	79,694.18	875,439,105.24	17,131,906.52
2017	84,448.67	867,303,774.44	17,342,478.44
2016	91,559.40	867,024,757.96	18,411,242.72
2014	84,541.50	849,729,904.68	16,470,282.89
2022	82,994.66	849,423,996.28	16,392,074.49
2024	85,451.25	843,750,831.69	16,120,892.29
2010	88,531.24	835,693,585.46	16,835,398.18
2012	86,983.44	832,655,737.89	16,930,701.30
2023	81,568.61	828,431,879.52	16,968,209.95
2015	76,869.99	765,729,798.51	15,365,043.93
2020	73,764.56	760,285,746.16	14,498,320.14
2019	74,149.48	727,622,497.64	14,809,252.58
Total	1,251,587.75	12,589,308,762.17	250,536,520.24

Hình 3. 152. Biểu đồ Matrix của Report1

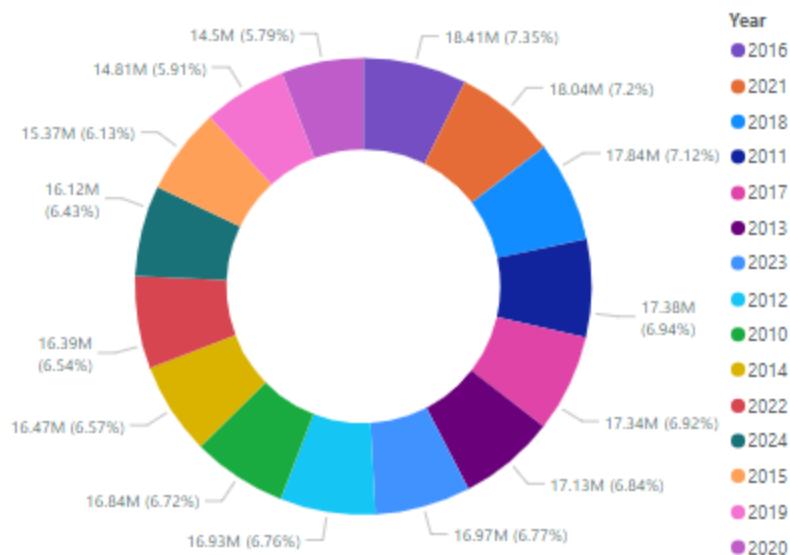
Kết quả dạng Line Chart:



Kết quả dạng Donut Chart:

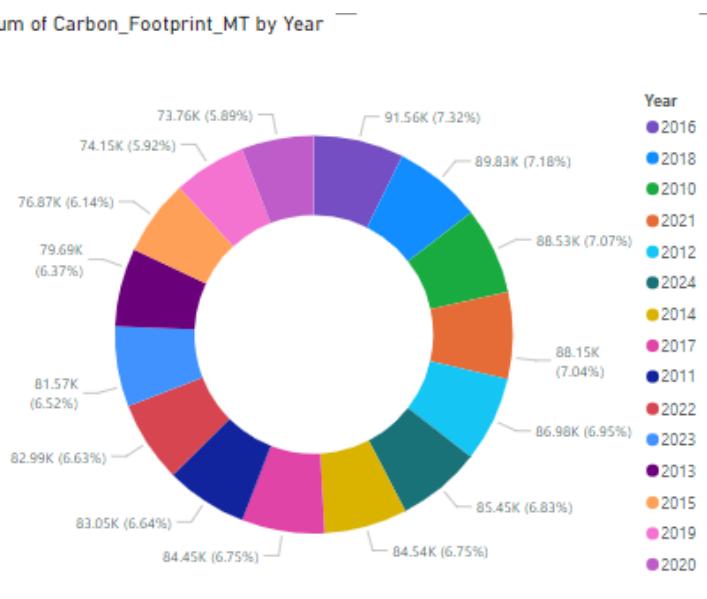


Sum of Waste_Production_KG by Year



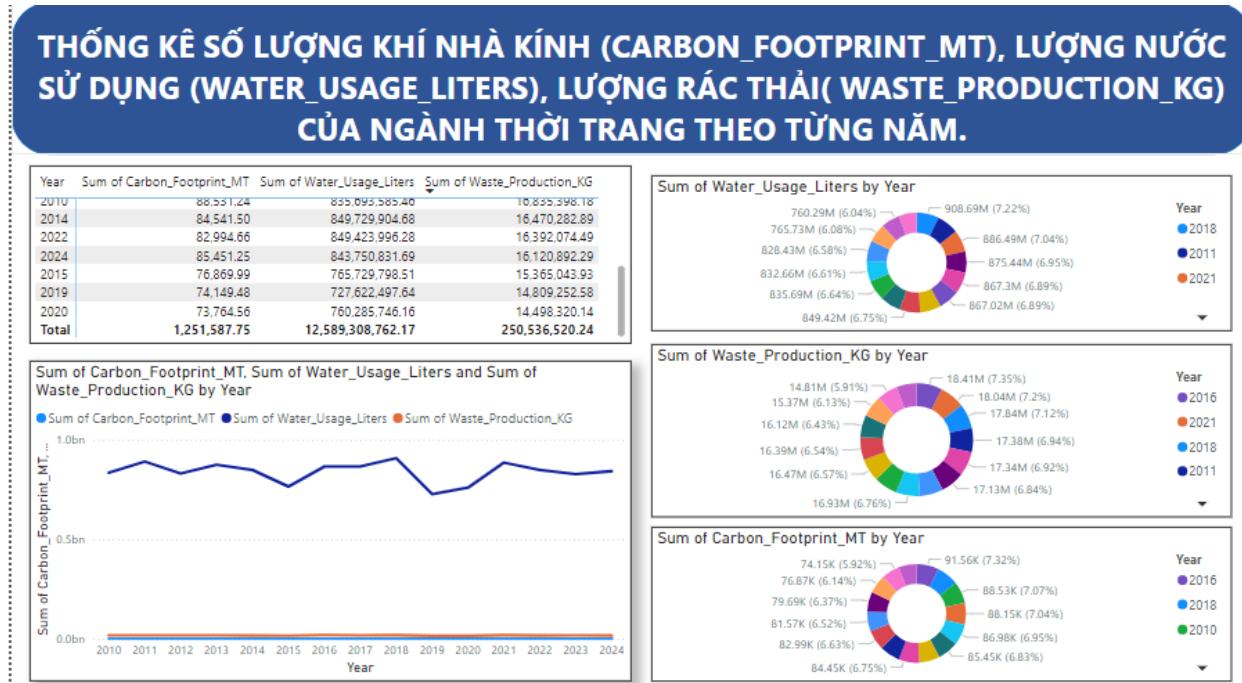
Hình 3. 155. Biểu đồ Donut Chart của Report 1 thứ 2

Sum of Carbon_Footprint_MT by Year



Hình 3. 156. Biểu đồ Donut Chart của Report 1 thứ 3

Sau đó ta thu được kết quả:

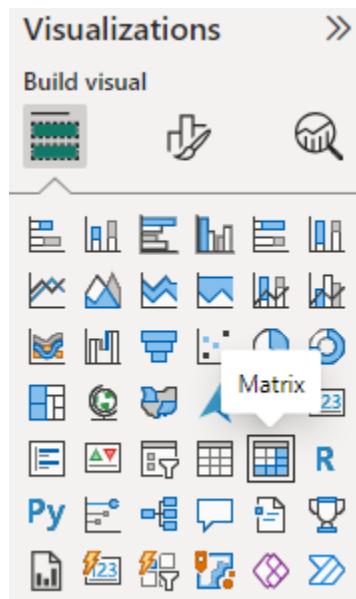


Hình 3. 157. Kết quả Report 1

3.5.3 Report 2

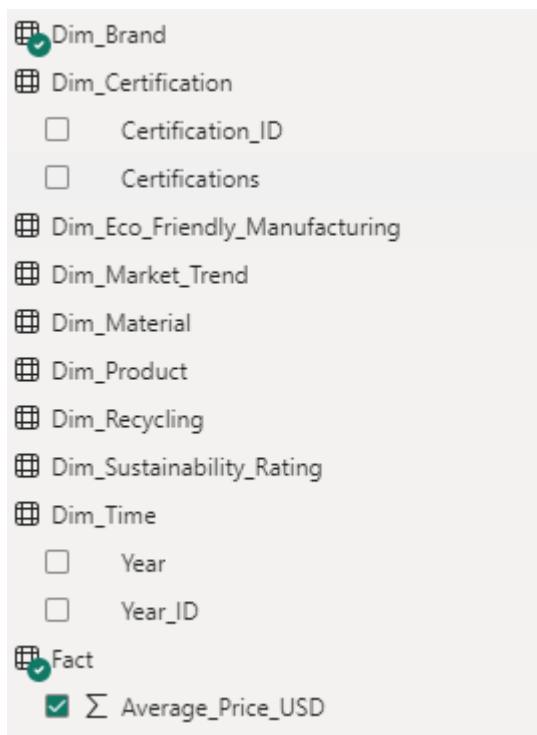
Nội dung: Thống kê tổng doanh thu của các thương hiệu ở các nước.

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 158. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn.



Hình 3. 159. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị Rows, Values trong Visualizations



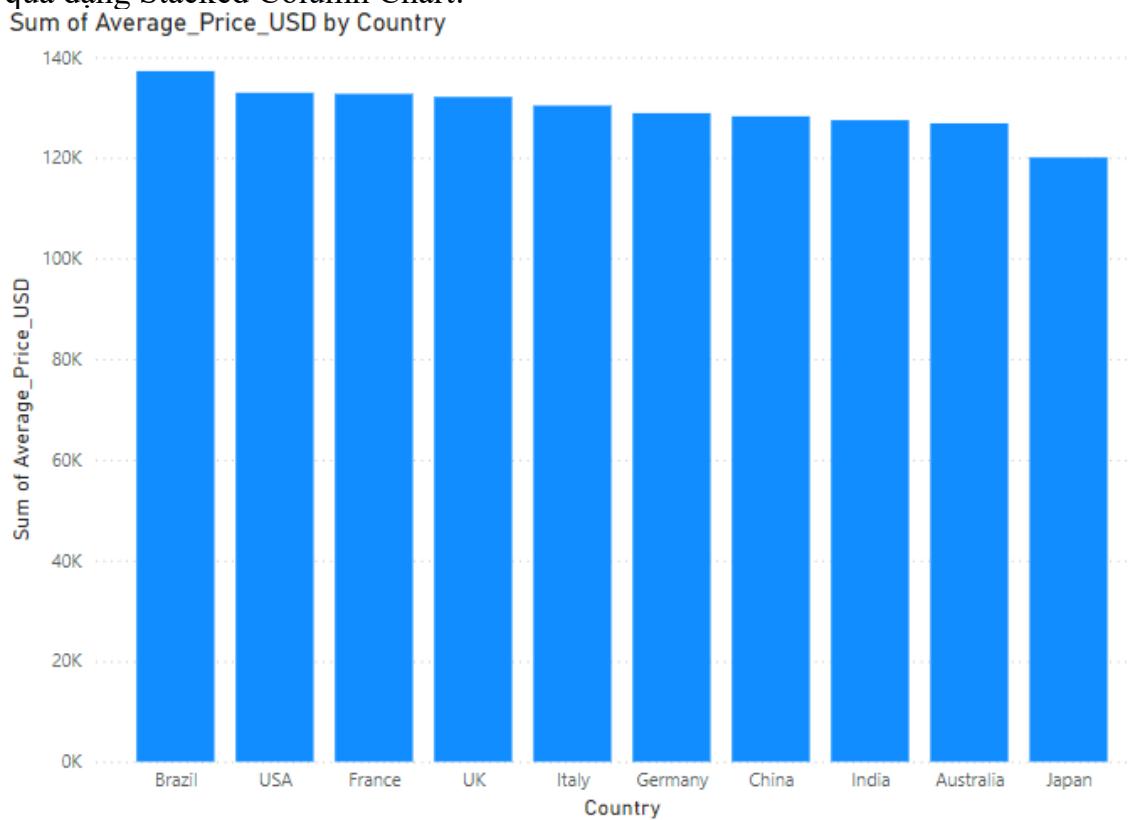
Hình 3. 160. Chính sửa hiển thị Rows, Values trong hộp thoại Visualizations

Kết quả dạng Matrix:

Country	Sum of Average_Price_USD
Australia	126,848.40
Brazil	137,223.25
China	128,226.07
France	132,702.37
Germany	128,873.45
India	127,476.05
Italy	130,362.70
Japan	120,054.49
UK	132,087.47
USA	132,916.83
Total	1,296,771.08

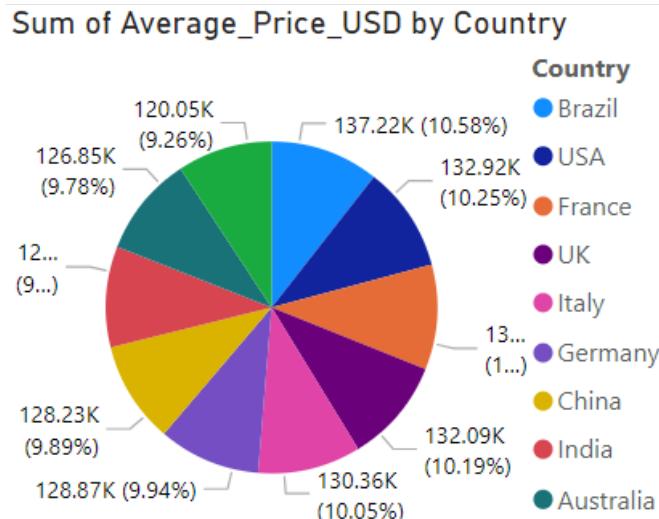
Hình 3. 161. Biểu đồ Matrix của Report 2

Kết quả dạng Stacked Column Chart:



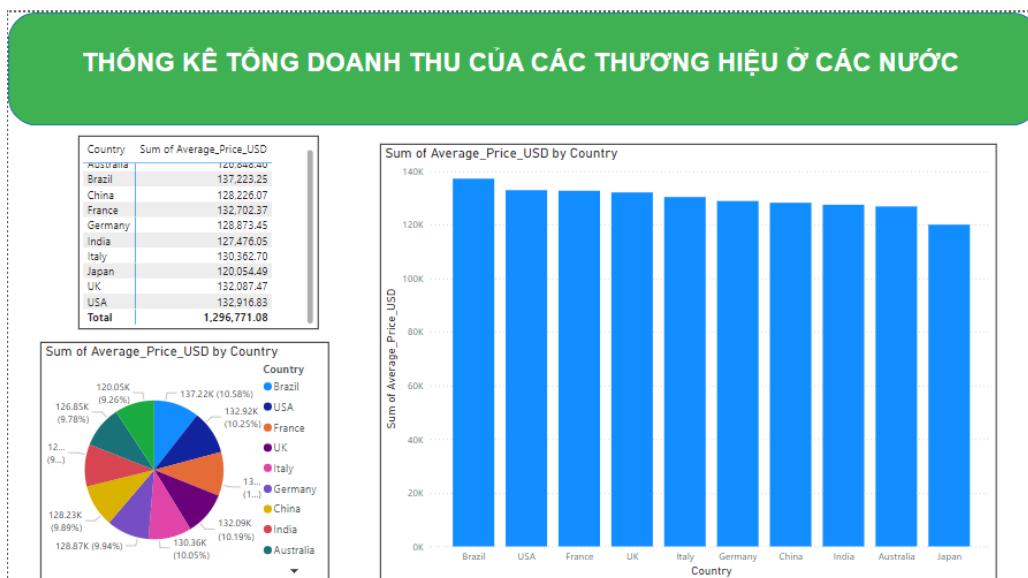
Hình 3. 162. Biểu đồ Stacked Column Chart của Report 2

Kết quả dạng Pie Chart:



Hình 3. 163. Biểu đồ Pie Chart của Report 2

Sau đó ta thu được kết quả:

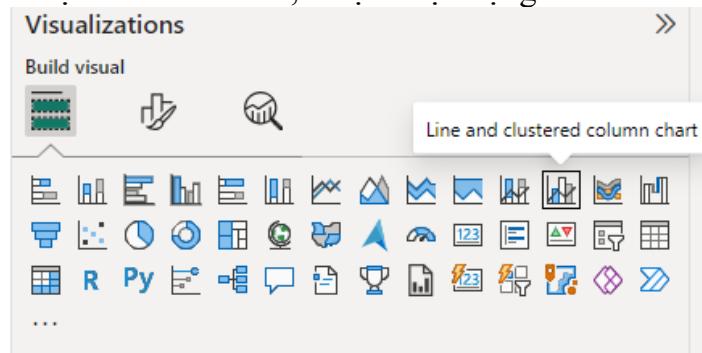


Hình 3. 164. Kết quả Report 2

3.5.4 Report 3

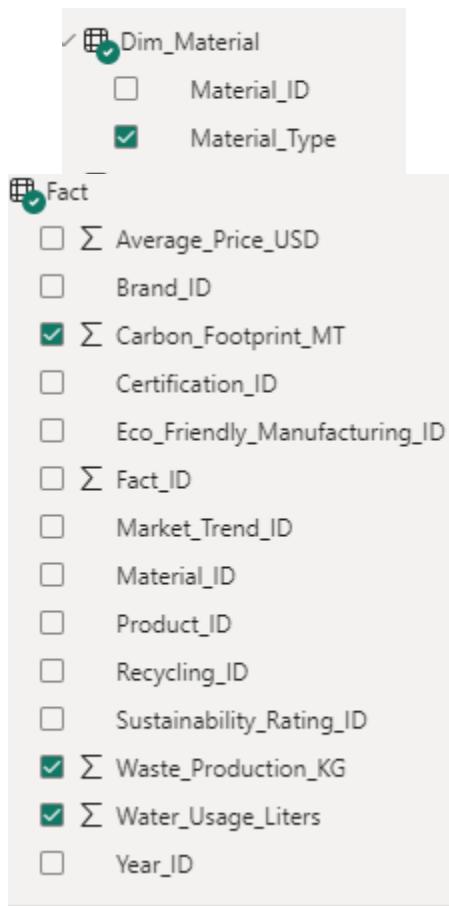
Thống kê tổng số lượng lượng khí nhà kính (Carbon_Footprint_MT), lượng nước sử dụng (Water_Usage_Liters), lượng rác thải (Waste_Production_KG) của từng loại vật liệu đã thải ra từ năm 2010 đến 2024

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 165. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn.



Hình 3. 166. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị X-ais, Y-ais trong Visualizations

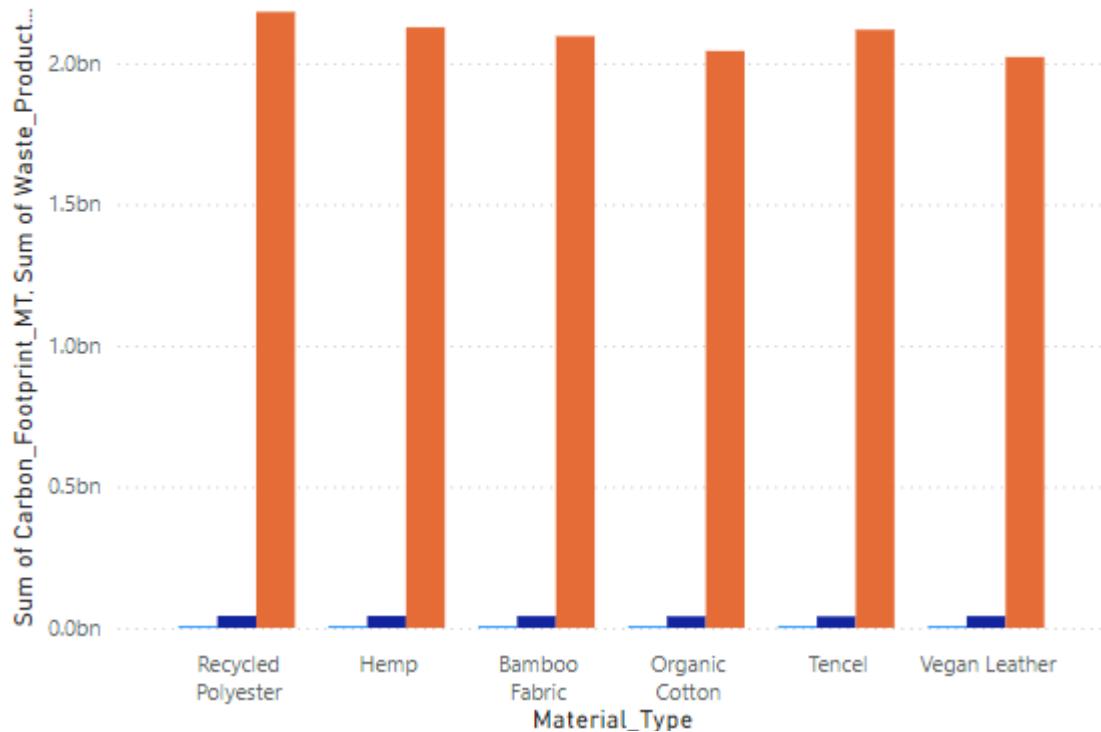


Hình 3. 167. Chính sửa hiển thị X-ais, Y- ais trong hộp thoại Visualizations

Kết quả dạng Line and Stacked Column Chart:

Sum of Carbon_Footprint_MT, Sum of Waste_Production_KG and Sum of Water_Usage_Liters by Material_Type

● Sum of Carbon_Footprint_MT ● Sum of Waste_Production_KG ● Sum of Water_Usage_Liters



Hình 3. 168. Biểu đồ Line and Stacked Column Chart của Report 3

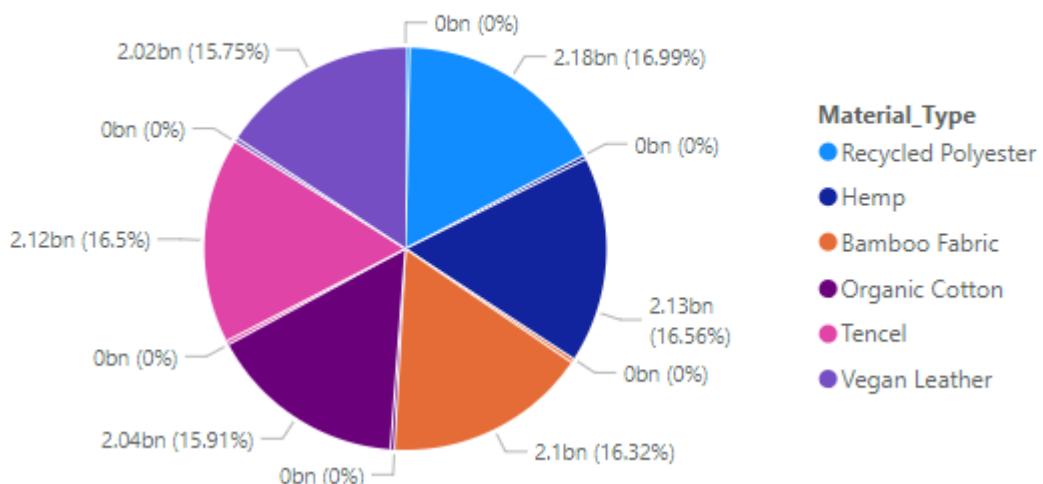
Kết quả dạng Matrix:

Material_Type	Sum of Carbon_Footprint_MT	Sum of Waste_Production_KG	Sum of Water_Usage_Liters
Recycled Polyester	224,853.65	42,545,018.98	2,181,995,126.45
Hemp	213,271.55	42,444,672.85	2,127,045,944.36
Bamboo Fabric	205,415.61	41,856,232.15	2,095,858,839.28
Organic Cotton	204,043.78	41,104,045.75	2,043,406,142.82
Tencel	203,611.37	40,750,944.81	2,118,911,371.74
Vegan Leather	200,391.79	41,835,605.70	2,022,091,337.52
Total	1,251,587.75	250,536,520.24	12,589,308,762.17

Hình 3. 169. Biểu đồ Matrix của Report 2

Kết quả dạng Pie Chart:

Sum of Carbon_Footprint_MT, Sum of Waste_Production_KG and Sum of Water_Usage_Liters by Material_Type



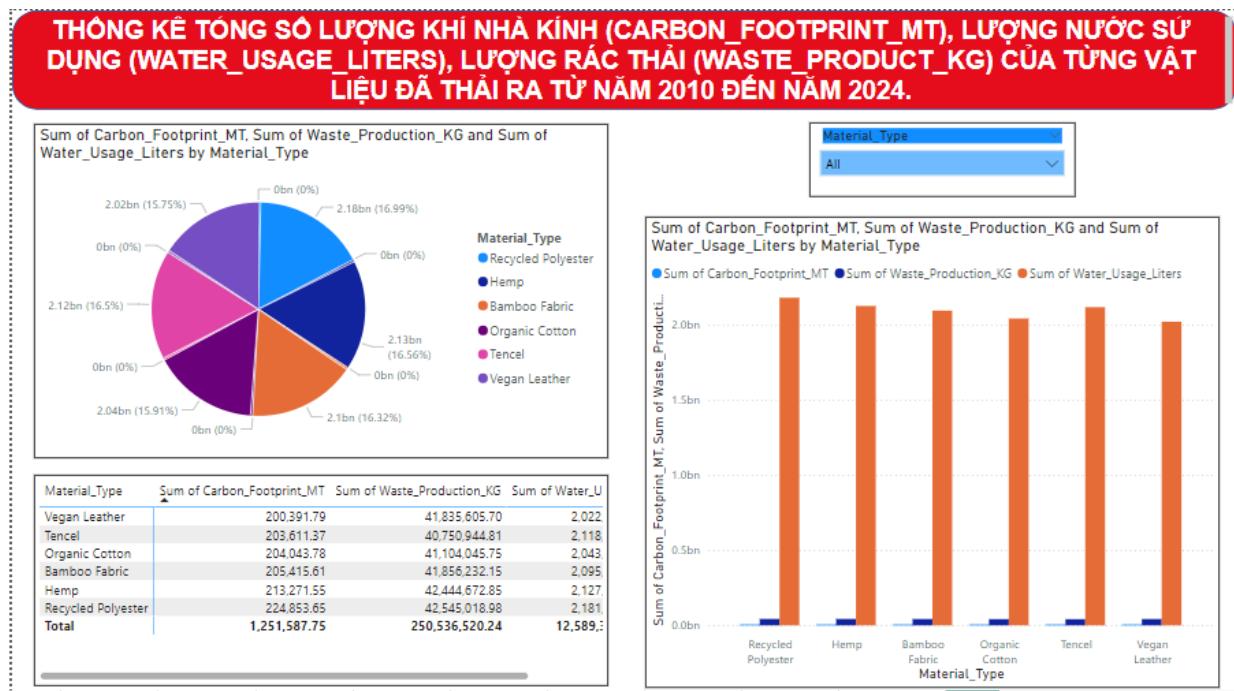
Hình 3. 170. Biểu đồ Pie Chart của Report 3

Tạo Slicer điều chỉnh thuộc tính Material_Type:



Hình 3. 171. Slicer điều chỉnh thuộc tính Material_Type của Report 3

Sau đó ta thu được kết quả:

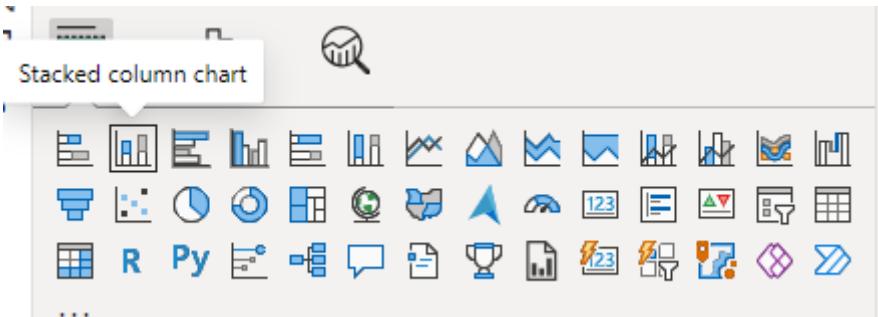


Hình 3. 172. Kết quả Report 3

3.5.5 Report 4

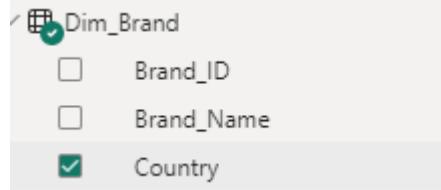
Thống kê số lượng nhãn hàng thời trang có đánh giá tiêu chuẩn (Sustainability_Rating) là A của mỗi nước trong năm 2014 và 2024

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 173. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn.



The screenshot shows the Data pane in Power BI. At the top, there is a section for the 'Dim_Time' dimension, which contains two items: 'Year' (selected) and 'Year_ID'. Below this, there is a section for the 'Fact' measure group, which contains several measures: 'Average_Price_USD', 'Brand_ID', 'Carbon_Footprint_MT', 'Certification_ID', 'Eco_Friendly_Manufacturing_ID', 'Fact_ID' (selected), and 'Market_Trend_ID'. The selected items ('Year' and 'Fact_ID') have green checkmarks next to them.

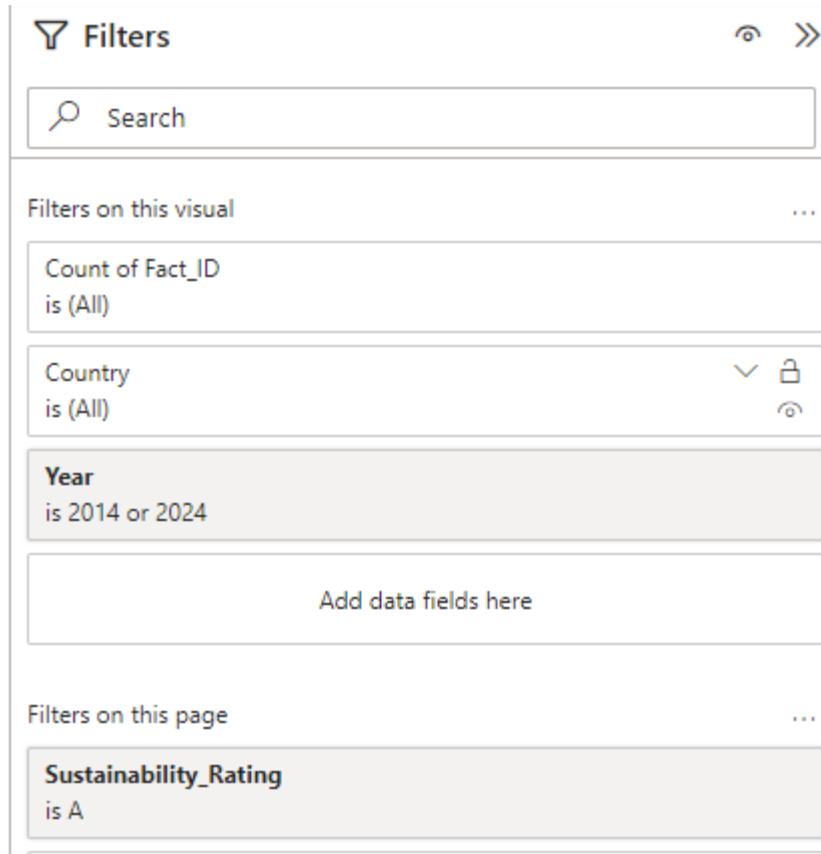
Hình 3. 174. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chỉnh sửa lại hiển thị X-axis, Y-axis trong Visualizations

The screenshot shows the Visualizations pane. It has two sections: 'X-axis' and 'Y-axis'. Under 'X-axis', there are two items: 'Year' and 'Country'. Under 'Y-axis', there is one item: 'Count of Fact_ID'. Each item has a dropdown arrow and an 'X' button to its right.

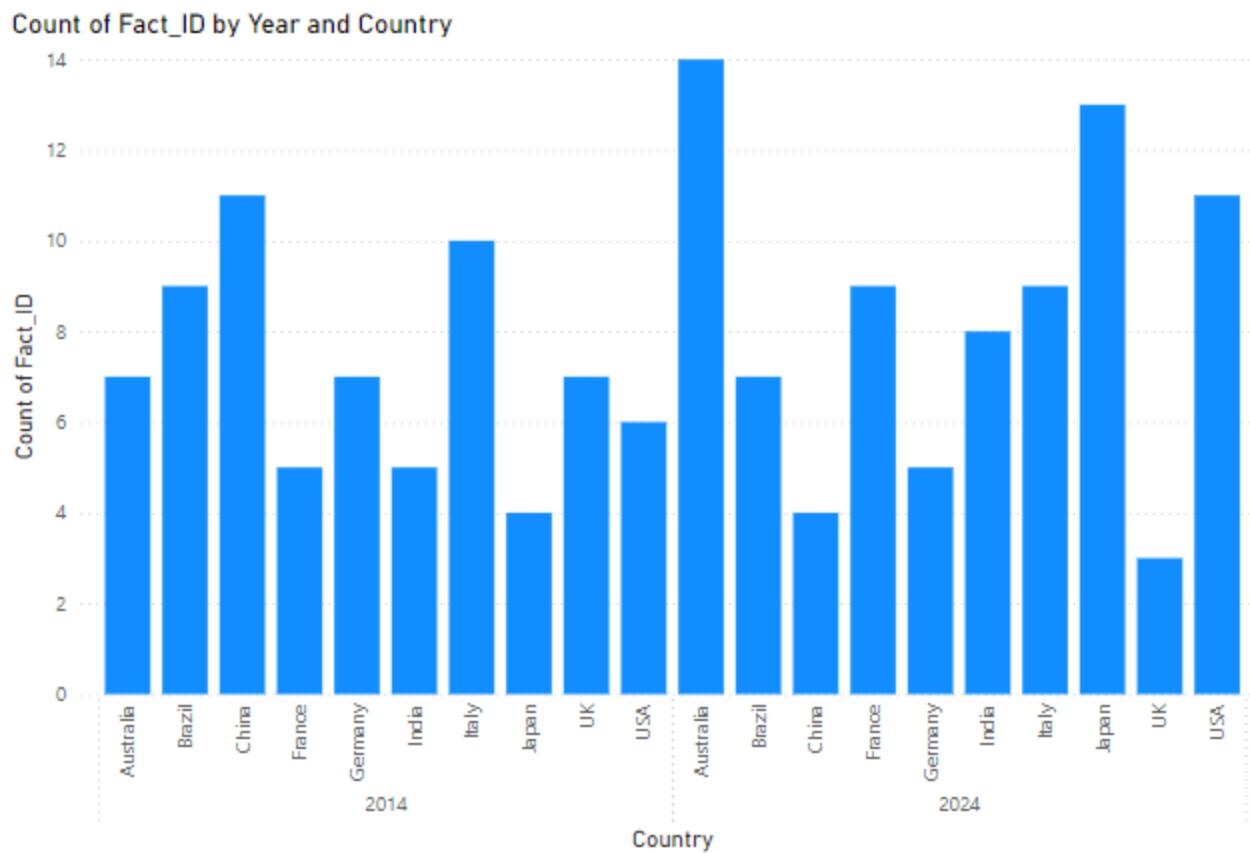
Hình 3. 175. Chỉnh sửa hiển thị X-axis, Y- axis trong hộp thoại Visualizations

Bước 4: Chọn điều kiện lọc trong Filters



Hình 3. 176. Điều kiện lọc trong Filters

Kết quả dạng Stacked Column Chart:



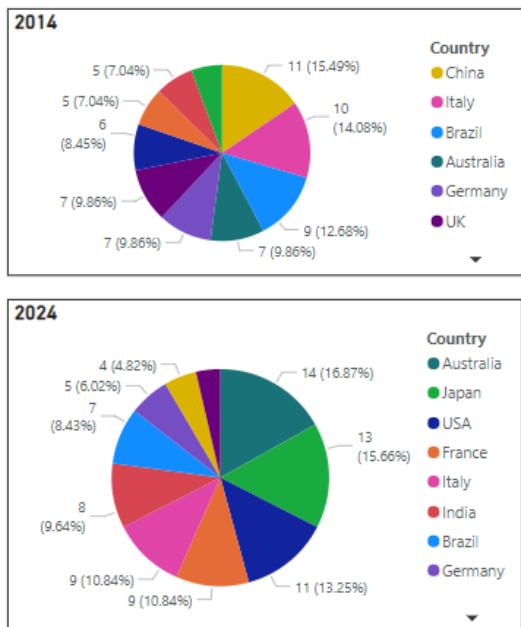
Hình 3. 177. Biểu đồ Stacked Column Chart của Report 4

Kết quả dạng Matrix:

Year	Count of Fact_ID
2014	71
Australia	7
Brazil	9
China	11
France	5
Germany	7
India	5
Italy	10
Japan	4
UK	7
USA	6
2024	83
Australia	14
Brazil	7
China	4
France	9
Germany	5
India	8
Italy	9
Japan	13
UK	3
USA	11
Total	154

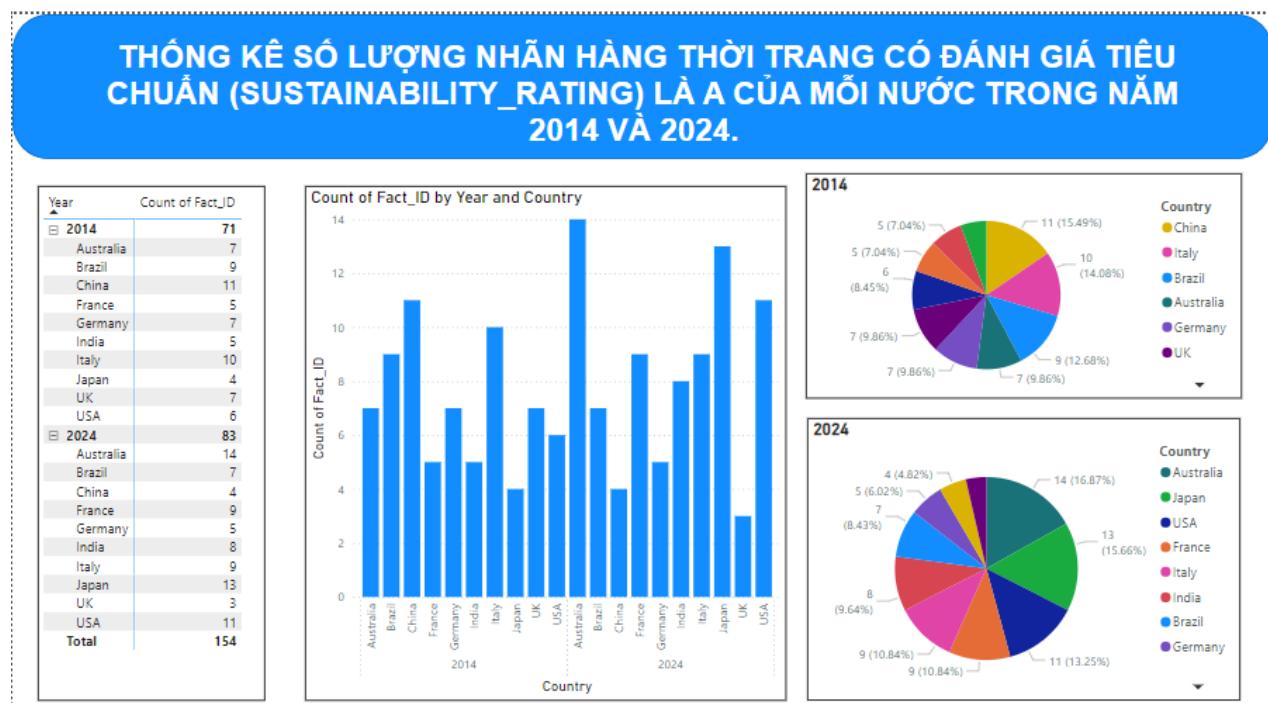
Hình 3. 178. Biểu đồ Matrix của Report 4

Kết quả dạng Pie Chart:



Hình 3. 179. Biểu đồ Pie Chart của Report 4

Sau đó ta thu được kết quả:

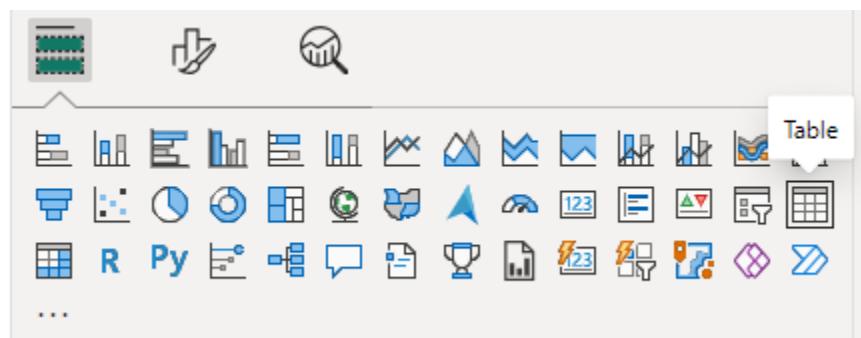


Hình 3. 180. Kết quả Report 4

3.5.6 Report 5

Liệt kê số nhãn hàng thuộc nước France, Italy, Japan và USA không có giấy chứng nhận và tổng lượng khí thải carbon mà các nhãn hàng ấy đã thải ra

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 181. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn.

The image shows the 'Data' pane in Power BI. It displays two tables: 'Dim_Brand' and 'Fact'. Under 'Dim_Brand', the 'Country' column is checked. Under 'Fact', the 'Carbon_Footprint_MT' and 'Fact_ID' columns are checked. Other columns like 'Brand_ID', 'Brand_Name', 'Year_ID', 'Average_Price_USD', 'Certification_ID', and 'Eco_Friendly_Manufacturing_ID' are not selected.

Hình 3. 182. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations



Hình 3. 183. Chỉnh sửa hiển thị Columns trong hộp thoại Visualizations

Bước 4: Chọn điều kiện lọc trong Filters

Filters

Search

Filters on this visual

- Count of Fact_ID is (All)
- Country is France, Italy, USA, or Japan
- Sum of Carbon_Footprint_MT is (All)

Add data fields here

Filters on this page

- Certifications is None

Hình 3. 184. Điều kiện lọc trong Filters

Kết quả dạng Table:

Country	Count of Fact_ID	Sum of Carbon_Footprint_MT
France	103	26,306.45
Italy	87	22,182.96
Japan	89	22,786.68
USA	107	26,523.97
Total	386	97,800.06

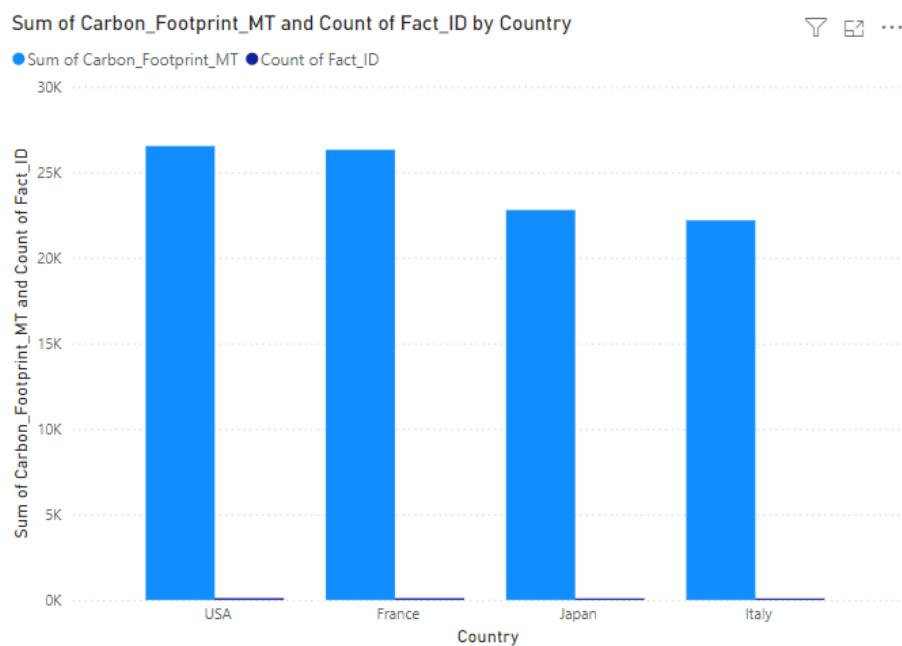
Hình 3. 185. Biểu đồ Table của Report 5

Kết quả dạng Matrix:

Country	Count of Fact_ID	Sum of Carbon_Footprint_MT
France	103	26,306.45
Italy	87	22,182.96
Japan	89	22,786.68
USA	107	26,523.97
Total	386	97,800.06

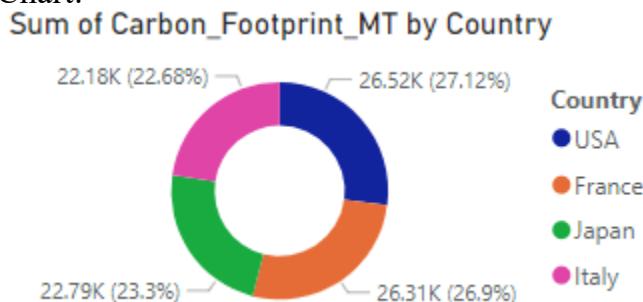
Hình 3. 186. Biểu đồ Matrix của Report 5

Kết quả dạng Clustered Column Chart:



Hình 3. 187. Biểu đồ Clustered Column Chart của Report 5

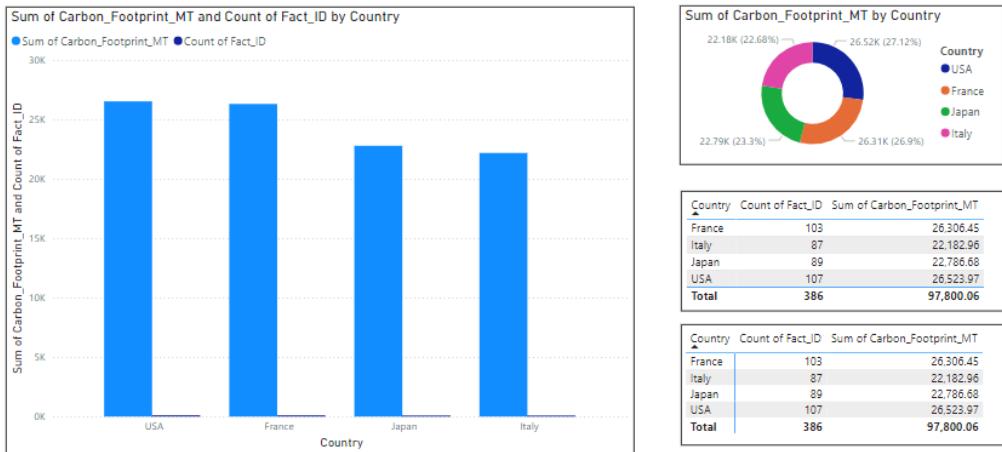
Kết quả dạng Donut Chart:



Hình 3. 188. Biểu đồ Chart của Report 5

Sau đó ta thu được kết quả:

**LIỆT KÊ SỐ NHÃN HÀNG THUỘC NƯỚC FRANCE, ITALY, JAPAN AND USA
KHÔNG CÓ GIẤY CHỨNG NHẬN VÀ TỔNG LƯỢNG KHI THÁI CARBON MÀ CÁC
NHÃN HÀNG ẤY ĐÃ THÁI RA.**

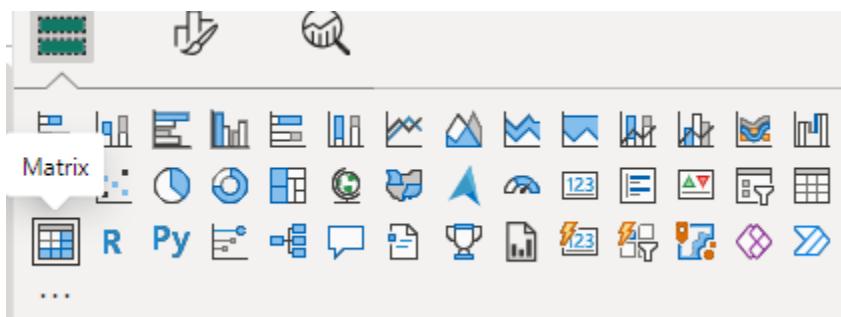


Hình 3. 189. Kết quả Report 5

3.5.7 Report 6

Liệt kê 10 cửa hàng có giá trung bình cao nhất và tỉ lệ bền vững của chúng.

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 190. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn

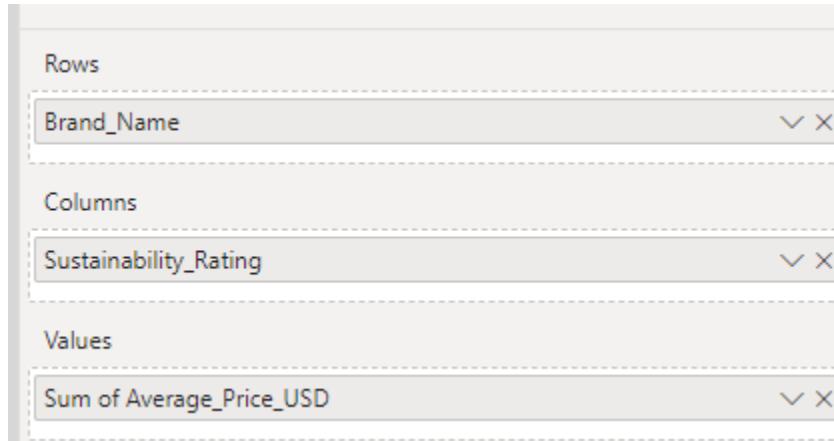
The screenshot shows the 'Data' pane in Power BI. It displays two sections:

- Dim_Brand:** A dimension table with columns 'Brand_ID' and 'Brand_Name'. 'Brand_Name' is checked as a selected column.
- Dim_Sustainability_Rating:** A dimension table with columns 'Sustainability_Rating'. This column is checked as a selected column.



Hình 3. 191. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations



Hình 3. 192. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations

Bước 4: Chọn điều kiện lọc trong Filters



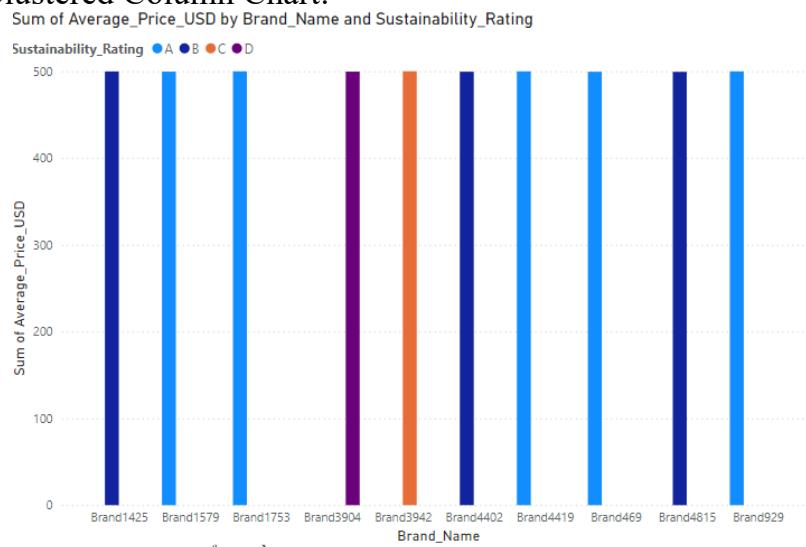
Hình 3. 193. Điều kiện lọc trong Filters

Kết quả dạng Matrix:

Brand_Name	A	B	C	D	Total
Brand1425		499.73			499.73
Brand1579	499.55				499.55
Brand1753	499.87				499.87
Brand3904			499.54		499.54
Brand3942		499.94			499.94
Brand4402		499.43			499.43
Brand4419	499.84				499.84
Brand469	499.41				499.41
Brand4815		499.31			499.31
Brand929	499.85				499.85
Total	2,498.52	1,498.47	499.94	499.54	4,996.47

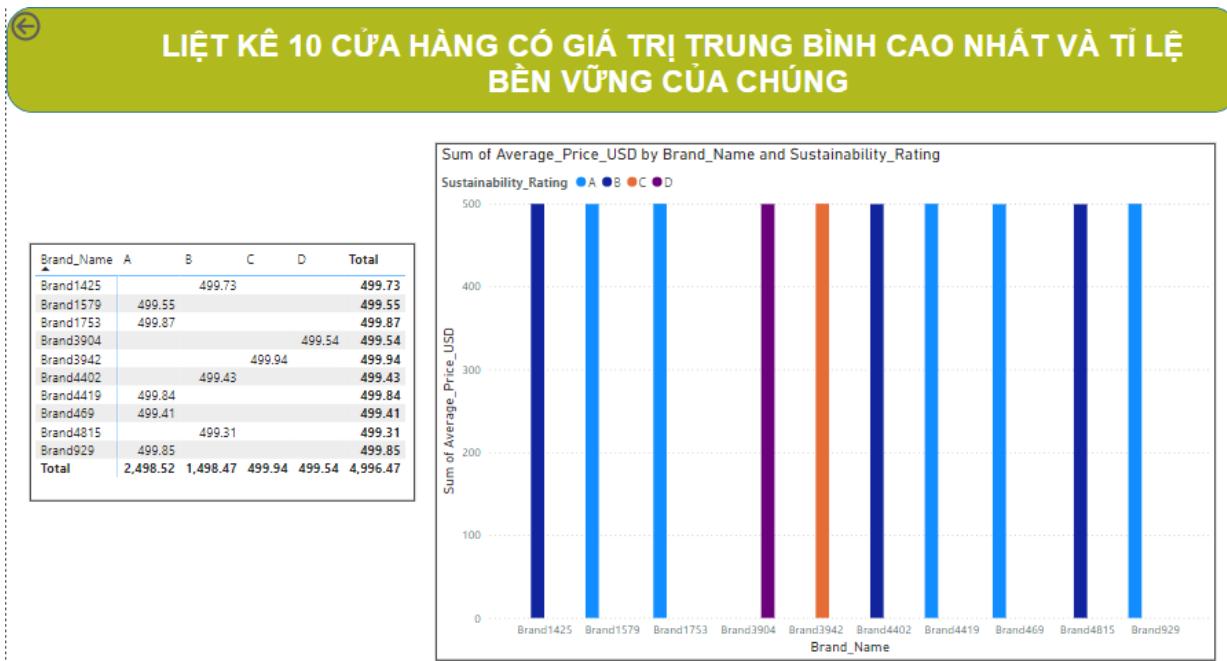
Hình 3. 194. Biểu đồ Matrix của Report 6

Kết quả dạng Clustered Column Chart:



Hình 3. 195. Biểu đồ Clustered Column Chart của Report 6

Sau đó ta thu được kết quả:

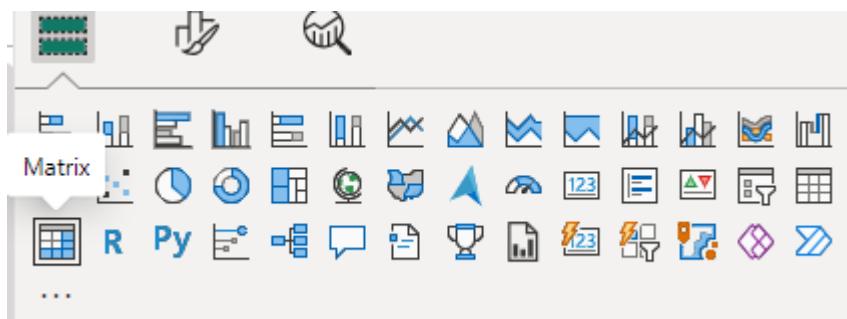


Hình 3. 196. Kết quả Report 6

3.5.8 Report 7

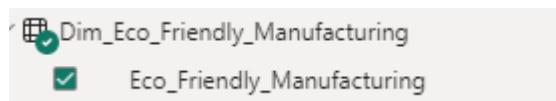
Thống kê số lượng mỗi loại vật liệu theo từng xu hướng thị trường tăng (Growing) trong trường hợp sản xuất thân thiện với môi trường và không.

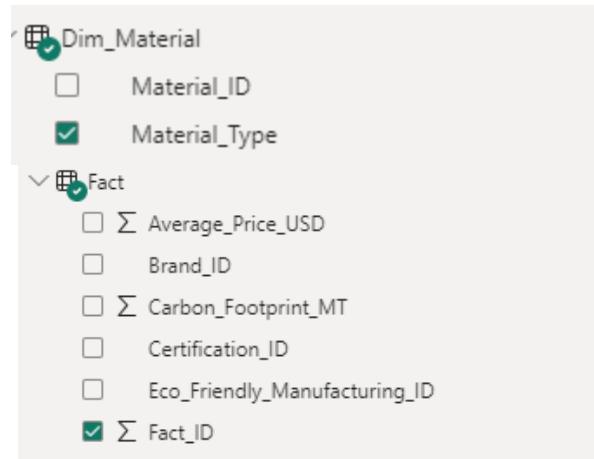
Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 197. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

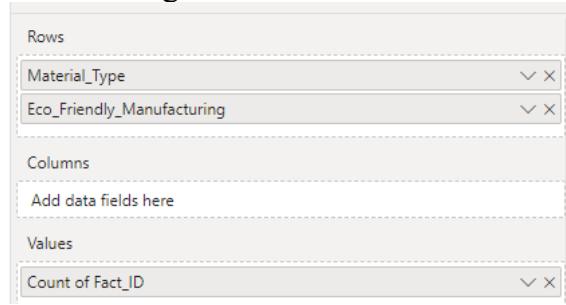
Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn





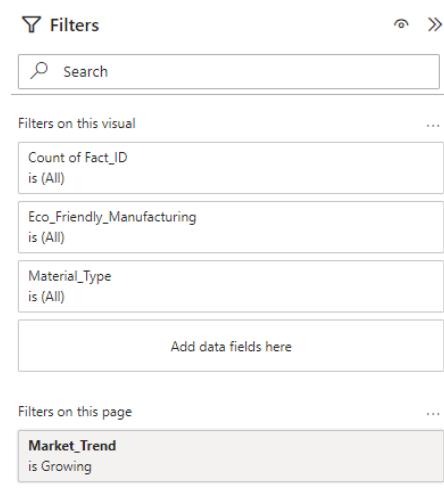
Hình 3. 198. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations



Hình 3. 199. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations

Bước 4: Chọn điều kiện lọc trong Filters



Hình 3. 200. Điều kiện lọc trong Filters

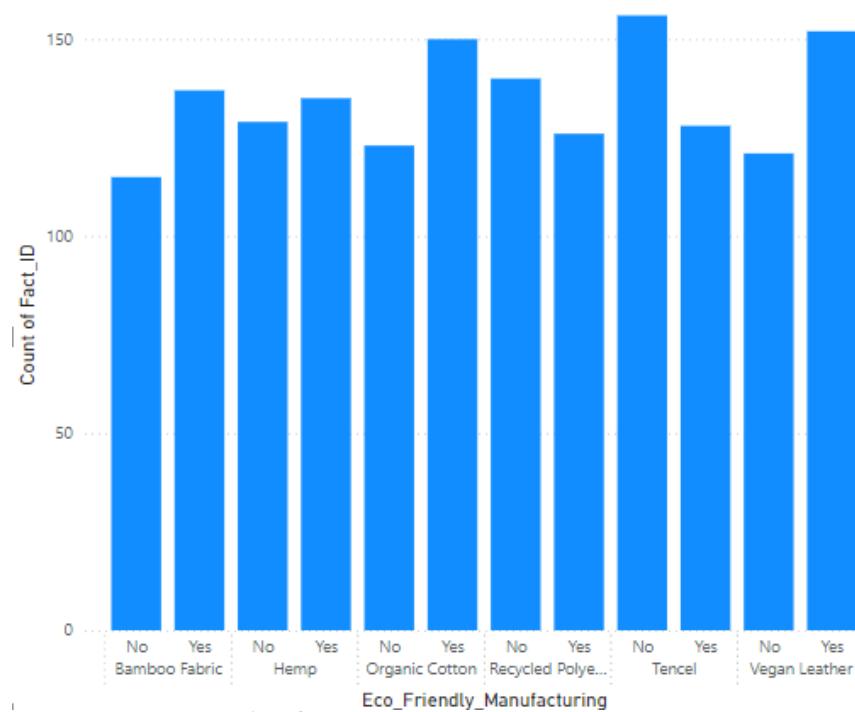
Kết quả dạng Matrix:

Material_Type	Count of Fact_ID
■ Bamboo Fabric	252
No	115
Yes	137
■ Hemp	264
No	129
Yes	135
■ Organic Cotton	273
No	123
Yes	150
■ Recycled Polyester	266
No	140
Yes	126
■ Tencel	284
No	156
Yes	128
■ Vegan Leather	273
No	121
Yes	152
Total	1612

Hình 3. 201. Biểu đồ Matrix của Report 7

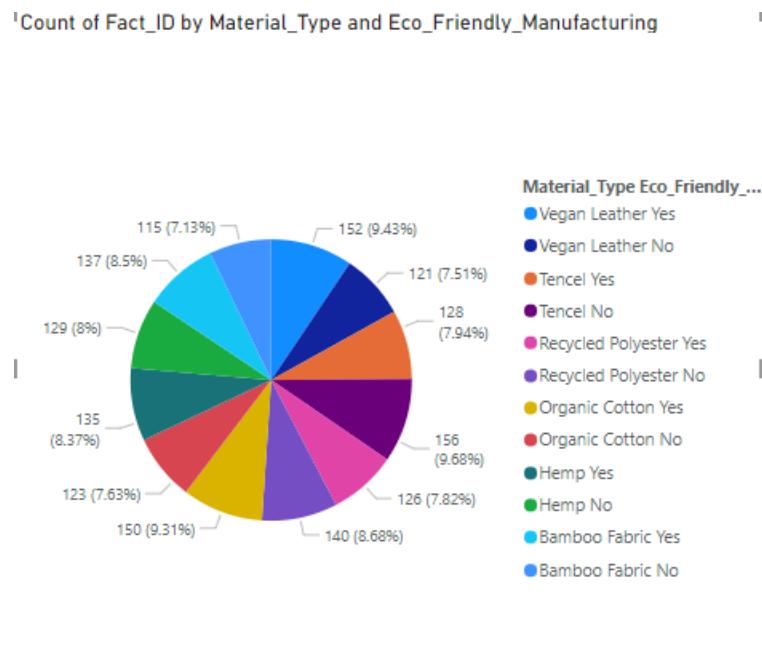
Kết quả dạng Stacked Column Chart:

Count of Fact_ID by Material_Type and Eco_Friendly_Manufacturing



Hình 3. 202. Biểu đồ Stacked Column Chart của Report 7

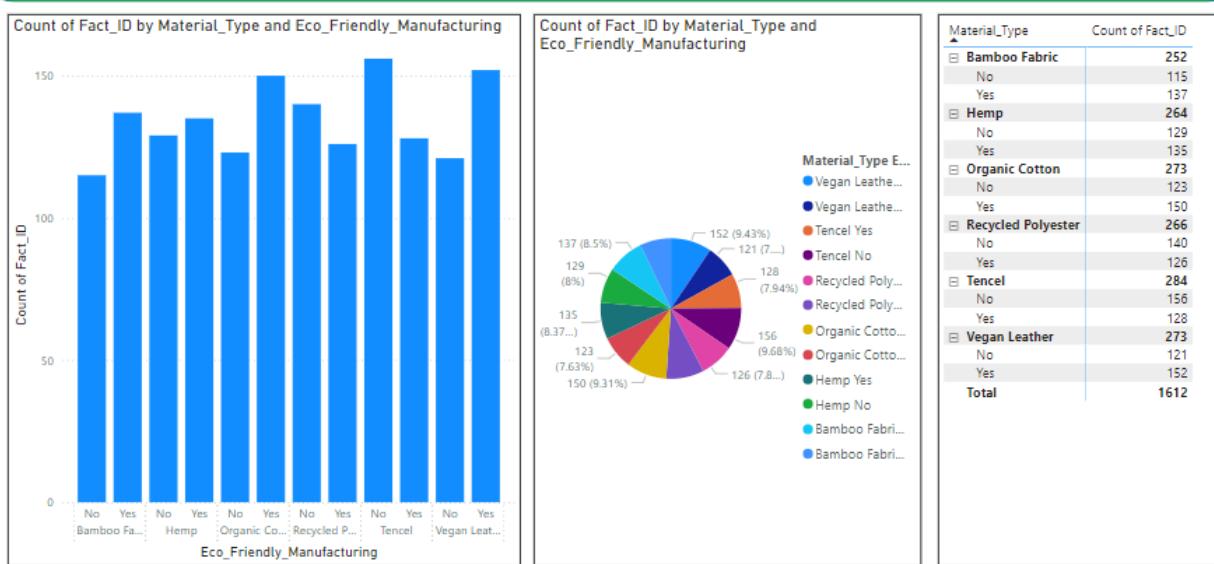
Kết quả dạng Pie Chart:



Hình 3. 203. Biểu đồ Pie Chart của Report 7

Sau đó ta thu được kết quả:

THỐNG KÊ SỐ LƯỢNG MỐI LOẠI VẬT LIỆU THEO TƯNG XU HƯỚNG THỊ TRƯỜNG (GROWING) TRONG TRƯỜNG HỢP SẢN XUẤT THÀN THIỆN VỚI MÔI TRƯỜNG VÀ KHÔNG.



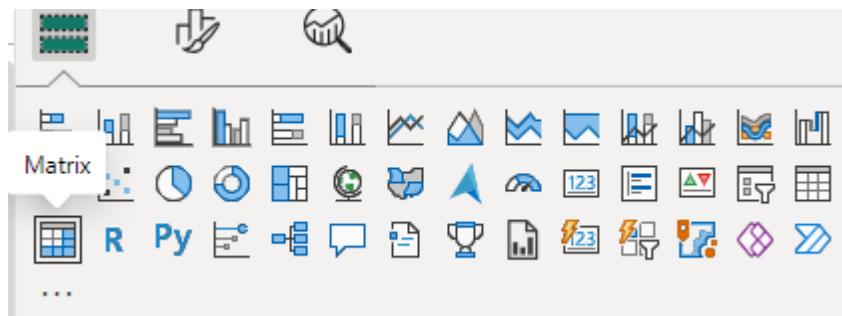
Hình 3. 204. Kết quả Report 7

3.5.9 Report 8

Với từng loại chứng chỉ, khi đạt các mức độ bền vững khác nhau hãy thống kê số lượng

cửa hàng theo từng xu hướng thị trường.

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 205. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn

A screenshot of the Microsoft Power BI Data pane. It displays a hierarchical list of data items. Under 'Dim_Certification', 'Certifications' is checked. Under 'Dim_Eco_Friendly_Manufacturing', both items are unchecked. Under 'Dim_Market_Trend', 'Market_Trend' is checked. Under 'Dim_Material', both items are unchecked. Under 'Dim_Product' and 'Dim_Recycling', both items are grayed out. Under 'Dim_Sustainability_Rating', 'Sustainability_Rating' is checked. Under 'Fact', several items are listed: 'Σ Average_Price_USD', 'Brand_ID', 'Σ Carbon_Footprint_MT', 'Certification_ID', 'Eco_Friendly_Manufacturing_ID', and 'Σ Fact_ID'. 'Σ Fact_ID' is checked.

Hình 3. 206. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations

The screenshot shows the 'Visualizations' configuration dialog with three main sections: Rows, Columns, and Values.

- Rows:** Contains two items: 'Certifications' and 'Sustainability_Rating'.
- Columns:** Contains one item: 'Market_Trend'.
- Values:** Contains one item: 'Count of Fact_ID'.

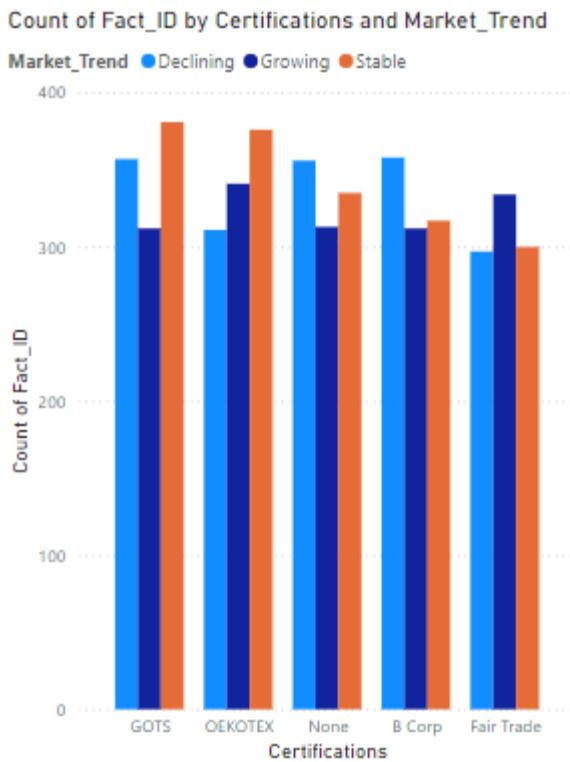
Hình 3. 207. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations

Kết quả dạng Matrix:

	Certifications	Declining	Growing	Stable	Total
▪ B Corp		358	312	317	987
A		89	80	74	243
B		89	85	81	255
C		80	79	97	256
D		100	68	65	233
▪ Fair Trade		297	334	300	931
A		77	80	67	224
B		86	74	73	233
C		61	78	91	230
D		73	102	69	244
▪ GOTS		357	312	381	1050
A		85	74	100	259
B		90	79	94	263
C		91	81	98	270
D		91	78	89	258
▪ None		356	313	335	1004
A		88	69	92	249
B		95	87	74	256
C		84	78	86	248
D		89	79	83	251
▪ OEKOTEX		311	341	376	1028
A		65	94	97	256
B		86	94	91	271
C		82	72	84	238
D		78	81	104	263
Total		1679	1612	1709	5000

Hình 3. 208. Biểu đồ Matrix của Report 8

Kết quả dạng Clustered Column Chart:



Hình 3. 209. Biểu đồ Clustered Column Chart của Report 8

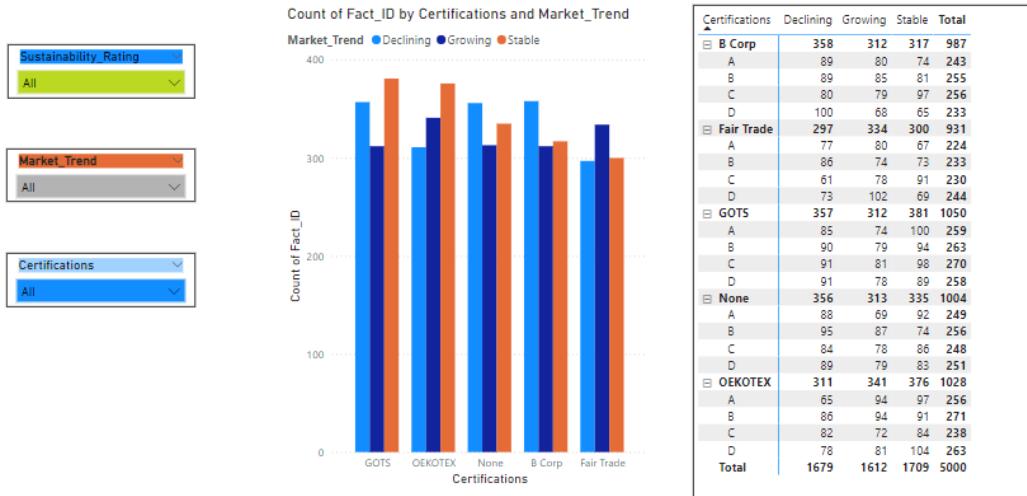
Tạo Slicer điều chỉnh thuộc tính Market_Trend, Sustainability_Rating và Certifications:



Hình 3. 210. Slicer điều chỉnh thuộc tính của Report 8

Sau đó ta thu được kết quả:

**VỚI TÙNG LOẠI CHỨNG CHỈ, KHI ĐẠT CÁC MỨC ĐỘ BỀN VỮNG KHÁC NHAU
HÃY THỐNG KÊ SỐ LƯỢNG CỦA HÀNG THEO XU HƯỚNG THỊ TRƯỜNG**

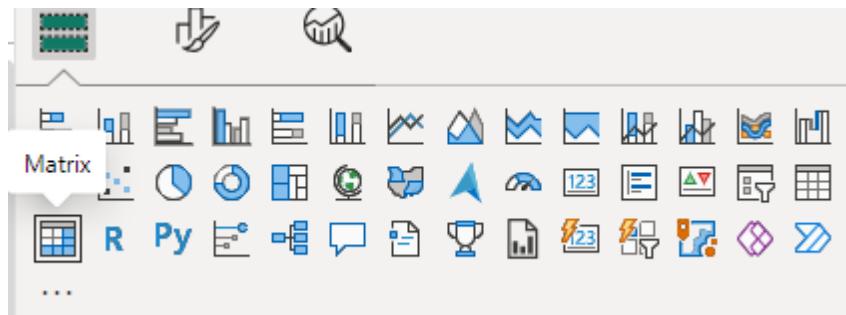


Hình 3. 211. Kết quả Report 8

3.5.10 Report 9

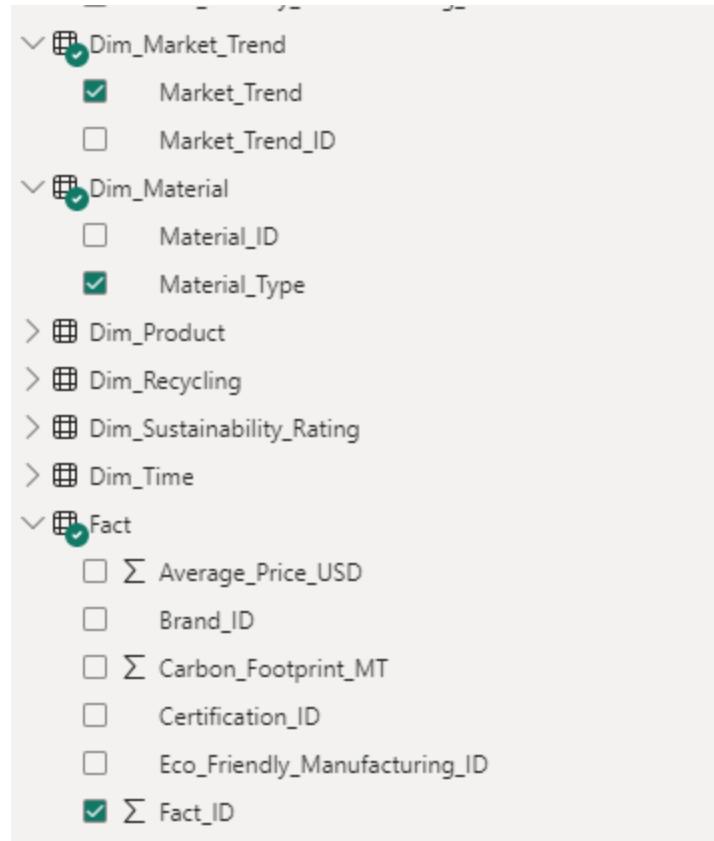
Với mỗi vật liệu hãy cho biết số lượng thương hiệu sử dụng cả chương trình tái chế và sản xuất thân thiện với môi trường theo từng xu hướng

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 212. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn



Hình 3. 213. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations

The screenshot shows the Visualizations interface with the following settings:

- Rows: Material_Type
- Columns: Market_Trend
- Values: Count of Fact_ID

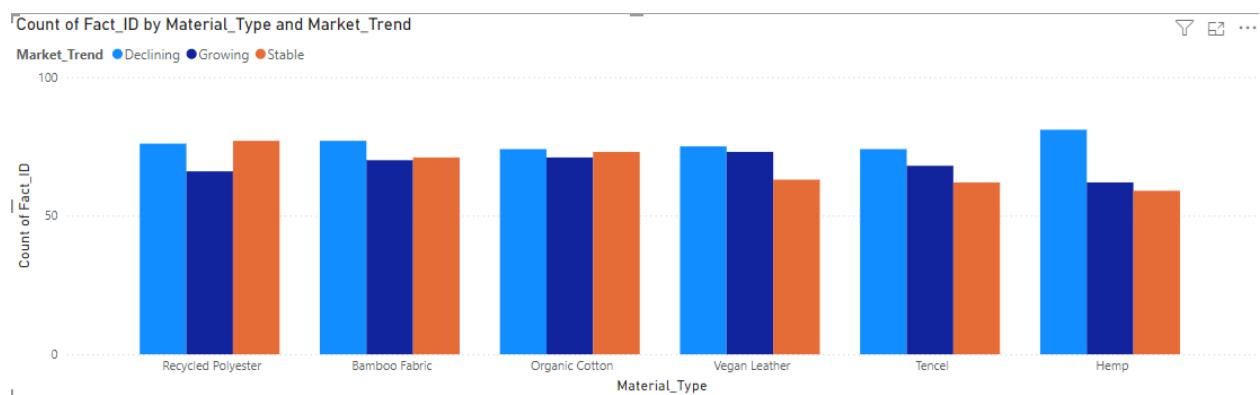
Hình 3. 214. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations

Kết quả dạng Matrix:

Material_Type	Declining	Growing	Stable	Total
Bamboo Fabric	77	70	71	218
Hemp	81	62	59	202
Organic Cotton	74	71	73	218
Recycled Polyester	76	66	77	219
Tencel	74	68	62	204
Vegan Leather	75	73	63	211
Total	457	410	405	1272

Hình 3. 215. Biểu đồ Matrix của Report 9

Kết quả dạng Clustered Column Chart:



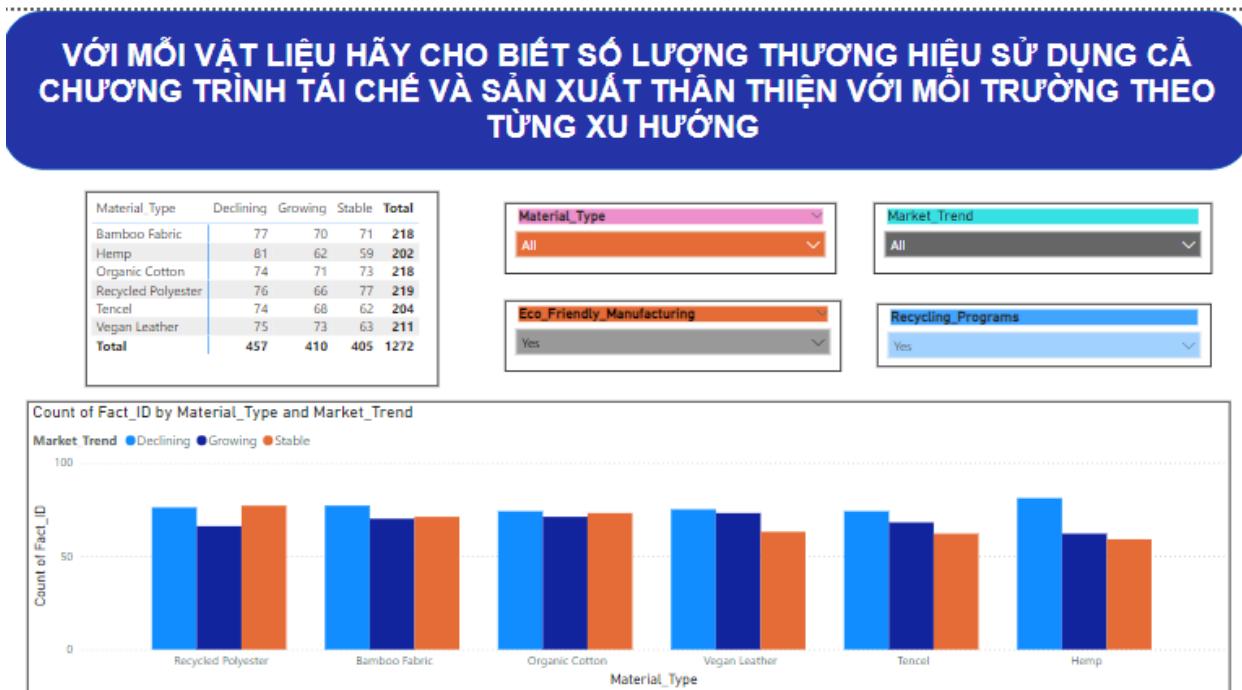
Hình 3. 216. Biểu đồ Clustered Column Chart của Report 9

Tạo Slicer điều chỉnh thuộc tính Market_Trend, Eco_Friendly_Manufacturing, Material_Type và Recycling_Programs:



Hình 3. 217. Slicer điều chỉnh thuộc tính của Report 9

Sau đó ta thu được kết quả:

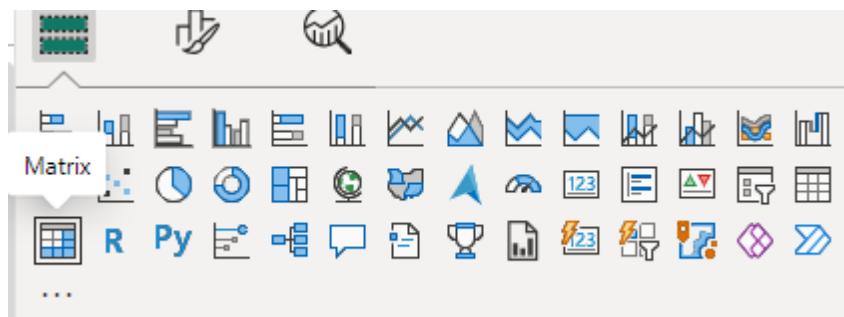


Hình 3. 218. Kết quả Report 9

3.5.11 Report 10

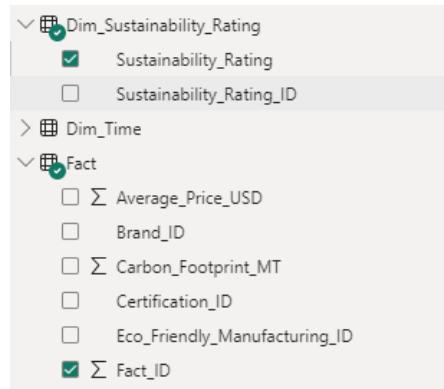
Thống kê số lượng thương hiệu đạt các mức bền vững ở mỗi nước

Bước 1: Trong hộp thoại Visualizations, ta lựa chọn dạng biểu đồ mà muốn hiển thị



Hình 3. 219. Lựa chọn dạng biểu đồ trong hộp thoại Visualizations

Bước 2: Chọn những thông tin cần thiết trong hộp thoại Data để phục vụ việc truy vấn



Hình 3. 220. Chọn những thông tin cần thiết trong hộp thoại Data

Bước 3: Chính sửa lại hiển thị trong Visualizations

The screenshot shows the Visualizations configuration window with the following settings:

- Country** (selected)
- Columns** (selected):
 - Sustainability_Rating
- Values** (selected):
 - Count of Fact_ID

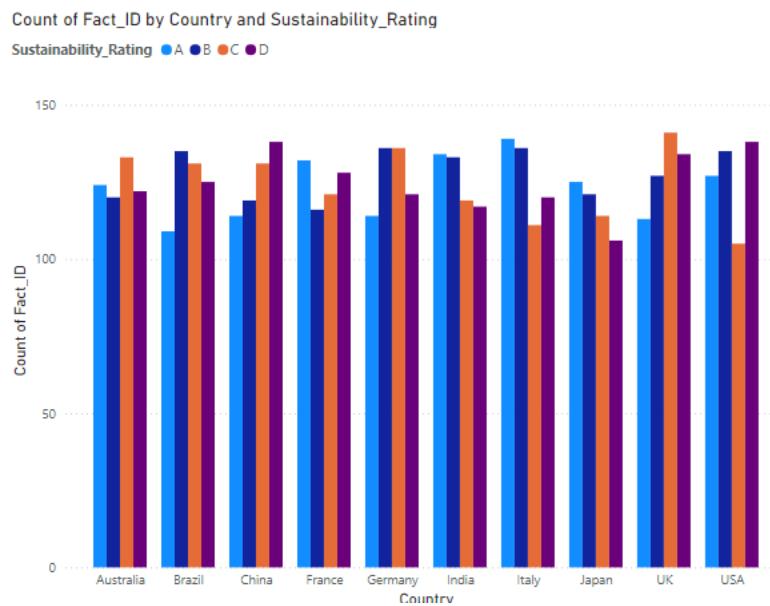
Hình 3. 221. Chính sửa hiển thị Rows, Columns và Values trong hộp thoại Visualizations

Kết quả dạng Matrix:

Country	A	B	C	D	Total
Australia	124	120	133	122	499
Brazil	109	135	131	125	500
China	114	119	131	138	502
France	132	116	121	128	497
Germany	114	136	136	121	507
India	134	133	119	117	503
Italy	139	136	111	120	506
Japan	125	121	114	106	466
UK	113	127	141	134	515
USA	127	135	105	138	505
Total	1231	1278	1242	1249	5000

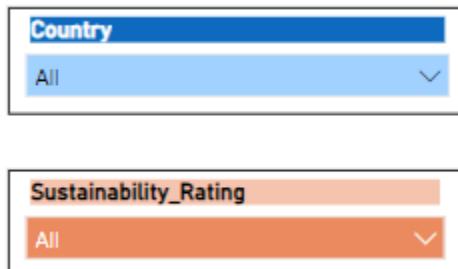
Hình 3. 222. Biểu đồ Matrix của Report 10

Kết quả dạng Clustered Column Chart:



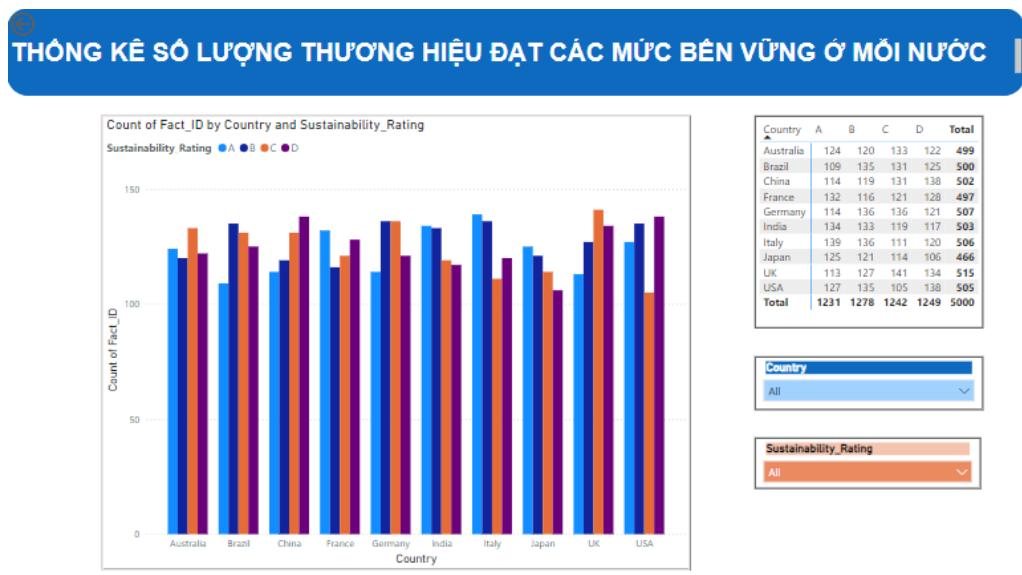
Hình 3. 223. Biểu đồ Clustered Column Chart của Report 10

Tạo Slicer điều chỉnh thuộc tính Country, Sustainability_Rating:



Hình 3. 224. Slicer điều chỉnh thuộc tính của Report 10

Sau đó ta thu được kết quả:



Hình 3. 225. Kết quả Report 10

*Nhân xét: Kết quả các câu truy vấn bằng Power BI cho ra kết quả giống với các câu truy vấn bằng CUBE, Excel, MDX.

4. QUÁ TRÌNH KHAI THÁC DỮ LIỆU (DATA MINING)

4.1 Tiết xử lý dữ liệu

Bước 1: Import các thư viện cần thiết và đọc file csv chứa dữ liệu đã được xử lý cơ bản ở phần SSIS.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.preprocessing import LabelEncoder

df=pd.read_csv('/content/drive/MyDrive/Nam3/OLAP/Do_An_OLAP/dataset_cleaned.csv')
```

Hình 4. 01. Import thư viện và đọc file csv chứa dữ liệu gốc

```
[ ] df.head()
```

	Brand_ID	Brand_Name	Country	Year	Sustainability_Rating	Material_Type	Eco_Friendly_Manufacturing	Carbon_Footprint_MT	Water_Usage_Liters	Waste_Production_KG	Region
0	BRAND0001	Brand1	Australia	2018	D	Tencel	No	1.75	451152.79	97844.11	Oceania
1	BRAND0002	Brand2	Japan	2015	D	Vegan Leather	Yes	124.39	1951566.31	37267.75	Asia
2	BRAND0003	Brand3	USA	2024	A	Vegan Leather	No	336.66	467454.52	38385.92	North America
3	BRAND0004	Brand4	Italy	2023	D	Bamboo Fabric	No	152.04	899576.90	32665.45	Europe
4	BRAND0005	Brand5	USA	2016	D	Bamboo Fabric	Yes	415.63	1809219.90	37295.47	North America

Hình 4. 02. Dữ liệu ban đầu

Bước 2: Kiểm tra số giá trị NULL trong tập dữ liệu.

Ta nhận thấy cột ‘Certification’ có 1004 giá trị NULL nên ta cần xử lý các giá này.

	df.isnull().sum()
Brand_ID	0
Brand_Name	0
Country	0
Year	0
Sustainability_Rating	0
Material_Type	0
Eco_Friendly_Manufacturing	0
Carbon_Footprint_MT	0
Water_Usage_Liters	0
Waste_Production_KG	0
Recycling_Programs	0
Product_Lines	0
Average_Price_USD	0
Market_Trend	0
Certifications	1004

Hình 4. 03. Số giá trị NULL trong tập dữ liệu ban đầu.

Bước 3: Tiến hành thay các giá trị NULL thành None trong cột Certification. Sau khi kiểm tra lại thì không còn giá trị NULL nào trong tập dữ liệu.

```
[ ] df['Certifications'].fillna('None', inplace=True)
print(df.isnull().sum())

→ Country          0
Sustainability_Rating 0
Material_Type        0
Eco_Friendly_Manufacturing 0
Carbon_Footprint_MT 0
Water_Usage_Liters   0
Waste_Production_KG 0
Recycling_Programs  0
Product_Lines        0
Average_Price_USD   0
Market_Trend         0
Certifications      0
dtype: int64
```

Hình 4. 04. Số giá trị NULL sau khi xử lý.

Country	Sustainability_Rating	Material_Type	Eco_Friendly_Manufacturing	Carbon_Footprint_MT	Water_Usage_Liters	Waste_Production_KG	Recycling_Programs	Product_Lines	Average_Price_USD	Market_Trend	Certifications
Australia	D	Tencel	No	1.75	4511152.79	97844.11	No	2	38.33	Growing	GOTS
Japan	D	Vegan Leather	Yes	124.39	1951566.31	37267.75	No	15	250.07	Growing	GOTS
USA	A	Vegan Leather	No	338.66	467454.52	38385.92	No	2	146.16	Growing	B Corp
Italy	D	Bamboo Fabric	No	152.04	899576.90	32665.45	No	13	165.52	Stable	OEKOTEX
USA	D	Bamboo Fabric	Yes	415.63	1809219.90	37265.47	Yes	19	211.63	Stable	Fair Trade
...
Brazil	A	Organic Cotton	Yes	423.73	251736.61	50532.28	No	1	490.95	Stable	None
Germany	C	Bamboo Fabric	Yes	68.24	2342347.78	64214.57	Yes	13	53.92	Stable	Fair Trade
Japan	D	Vegan Leather	Yes	321.12	317800.67	81467.21	Yes	12	431.32	Stable	None
India	A	Tencel	No	124.83	927539.68	34506.22	Yes	20	322.95	Declining	Fair Trade
Italy	B	Vegan Leather	No	192.11	3402552.70	81958.82	Yes	11	205.62	Growing	None

Hình 4. 05. Dữ liệu sau khi xử lý giá trị NULL.

Bước 4: Loại bỏ các thuộc tính không cần thiết cho quá trình phân loại.

Vì các thuộc tính liên quan đến mã thương hiệu, tên thương hiệu và năm không liên quan đến mục đích phân tích nên ta tiến hành loại bỏ.

```
df=df.drop(['Brand_ID', 'Brand_Name', 'Year'], axis=1)
```

Hình 4. 06. Xóa bỏ cột Brand_ID, Brand_Name và Year.

Bước 5: Mã hóa giá trị của các thuộc tính có kiểu là object theo dạng label encoder và thêm các cột chứa dữ liệu đã mã hóa vào tập dữ liệu.

IS217 – Kho dữ liệu và OLAP

```
from sklearn.preprocessing import LabelEncoder

# Tạo danh sách các cột cần mã hóa
columns_to_encode = ['Country', 'Sustainability_Rating', 'Material_Type','Eco_Friendly_Manufacturing','Recycling_Programs','Market_Trend','Certifications']

# Tạo đối tượng LabelEncoder
label_encoders = {}

# Áp dụng LabelEncoder cho từng cột
for column in columns_to_encode:
    le = LabelEncoder()
    df[column + '_LabelEncoded'] = le.fit_transform(df[column])
    label_encoders[column] = le
```

Hình 4. 07. Mã hóa giá trị của các thuộc tính.

Bước 6: Xóa các cột thuộc tính trước khi được mã hóa.

```
columns_to_drop = ['Country', 'Sustainability_Rating', 'Material_Type','Eco_Friendly_Manufacturing','Recycling_Programs','Market_Trend','Certifications']
df = df.drop(columns_to_drop, axis=1)
```

Hình 4. 08. Xóa các thuộc tính trước khi mã hóa.

Bước 7: Kiểm tra dữ liệu sau khi đã mã hóa.

	Carbon_Footprint_MT	Water_Usage_Liters	Waste_Production_KG	Product_Lines	Average_Price_USD	Country_LabelEncoded	Sustainability_Rating_LabelEncoded	Material_Type_LabelEncoded	Eco_Friendly_Manufacturing_LabelEncoded	Recycling_Programs_LabelEncoded	Market_Trend_LabelEncoded	Certifications_LabelEncoded
0	1.75	451102.79	97844.11	2	38.33	0	3	4	0	0	1	2
1	124.39	1951566.31	37267.75	15	250.07	7	3	5	1	0	1	2
2	336.66	467454.52	38385.92	2	146.16	9	0	5	0	0	1	0
3	152.04	899576.90	32665.45	13	165.02	6	3	0	0	0	2	4
4	415.63	1809219.90	37295.47	19	219.63	9	3	0	1	1	2	1
...
4995	423.73	251736.61	50332.28	1	490.95	1	0	2	1	0	2	3
4996	68.24	2342547.78	64214.57	13	53.92	4	2	0	1	1	2	1
4997	321.12	317800.67	81467.21	12	431.32	7	3	5	1	1	2	3
4998	124.83	927539.68	34506.22	20	322.95	5	0	4	0	1	0	1
4999	192.11	3402552.70	81958.82	11	209.62	6	1	5	0	1	1	3

Hình 4. 09. Dữ liệu sau khi được mã hóa.

Bước 8: Chuẩn hóa giá trị của các thuộc tính thuộc kiểu float hoặc in để làm giảm khoảng giá trị.

```
[ ] from sklearn.preprocessing import MinMaxScaler

# Chọn các cột số để chuẩn hóa
numeric_columns = ['Carbon_Footprint_MT', 'Water_Usage_Liters', 'Waste_Production_KG', 'Average_Price_USD']

# Tạo một đối tượng MinMaxScaler với phạm vi từ 0 đến 1
scaler = MinMaxScaler(feature_range=(0, 1))

# Áp dụng MinMaxScaler cho các cột số
df[numeric_columns] = scaler.fit_transform(df[numeric_columns])
```

Hình 4. 10. Chuẩn hóa giá trị của các thuộc tính.

Bước 9: Kiểm tra dữ liệu sau khi được chuẩn hóa.

	Carbon_Footprint_Mt	Water_Usage_Liters	Waste_Production_KG	Product_Lines	Average_Price_USD	Country_LabelEncoded	Sustainability_Rating_LabelEncoded	Material_Type_LabelEncoded	Eco_Friendly_Manufacturing_LabelEncoded
0	0.001423	0.901314	0.978733	2	0.038152	0	3	4	
1	0.247249	0.384173	0.366365	15	0.479351	7	3	5	
2	0.672733	0.084321	0.377669	2	0.262835	9	0	5	
3	0.302672	0.171628	0.319840	13	0.303176	6	3	0	
4	0.831025	0.355413	0.366645	19	0.399254	9	3	0	
...
4995	0.847261	0.040738	0.500457	1	0.981268	1	0	2	
4996	0.134699	0.463127	0.638771	13	0.070637	4	2	0	
4997	0.641584	0.054085	0.813179	12	0.857018	7	3	5	
4998	0.248131	0.177278	0.338449	20	0.631209	5	0	4	
4999	0.382990	0.677332	0.818148	11	0.386731	6	1	5	

Hình 4. 11. Dữ liệu sau khi được chuẩn hóa.

Bước 10: Lưu dữ liệu đã được xử lý.

```
[19] df.to_csv('/content/drive/MyDrive/Nam3/OLAP/Do_An_OLAP/data_preprocessing.csv', index=False)
```

Hình 4. 12. Lưu dữ liệu đã được xử lý.

4.2 Ứng dụng mô hình thuật toán

4.2.1 Chuẩn bị dữ liệu đưa vào mô hình

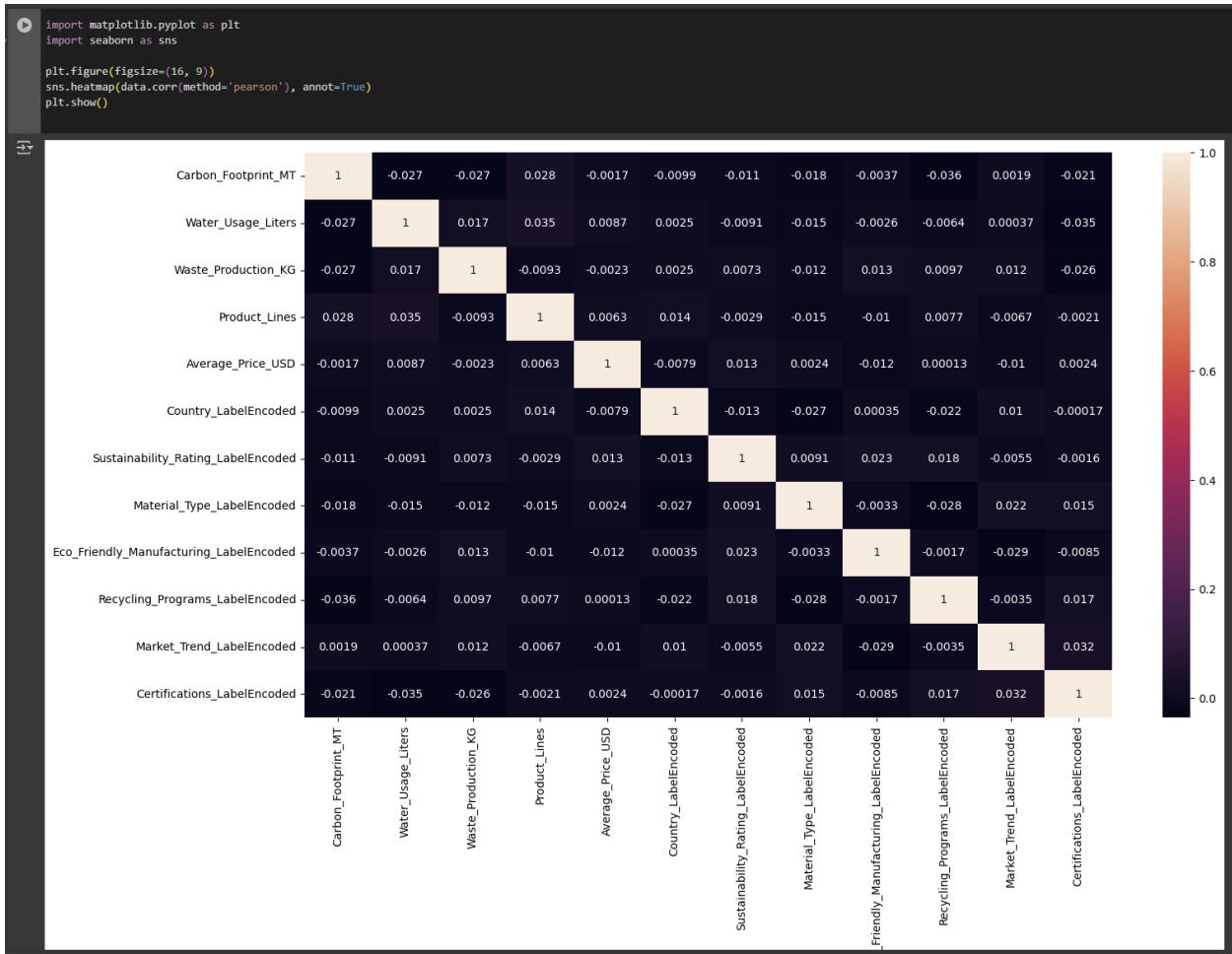
Bước 1: Đọc dữ liệu đã được tiền xử lý trước đó.

```
[20] data=pd.read_csv('/content/drive/MyDrive/Nam3/OLAP/Do_An_OLAP/data_preprocessing.csv')
```

Hình 4. 13. Đọc dữ liệu đã lưu trước đó.

Bước 2: Vẽ ma trận tương quan giữa các thuộc tính.

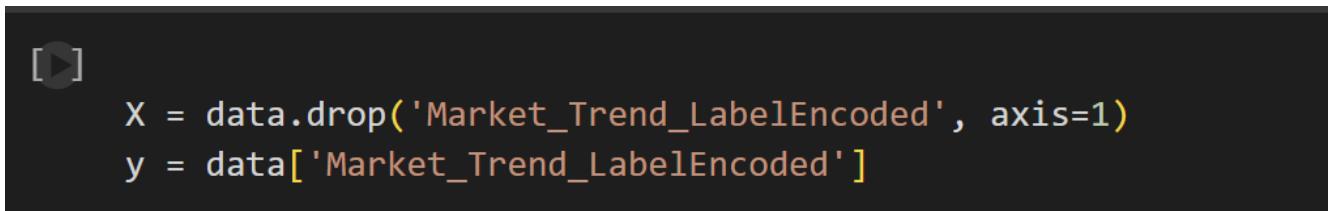
IS217 – Kho dữ liệu và OLAP



Hình 4. 14. Ma trận tương quan giữa các thuộc tính

Nhân xét: Dựa vào kết quả ma trận ta có thể thấy các thuộc tính có độ tương quan khá thấp với nhau, không có cặp thuộc tính khác nhau nào đạt độ tương quan 0.1.

Bước 3: Tách dữ liệu thành 2 phần: thuộc tính chứa nhãn cần dự đoán (y) các thuộc tính hỗ trợ dự đoán nhãn (X).

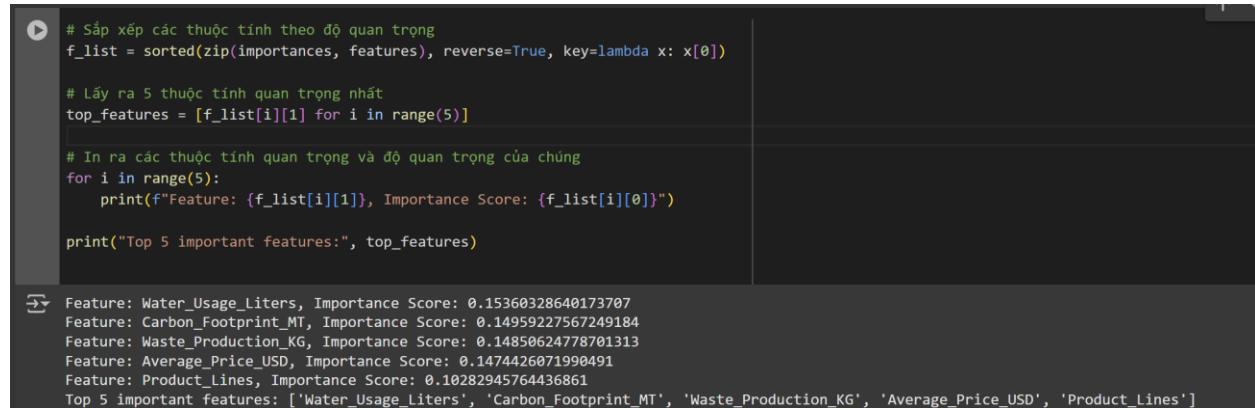


Hình 4. 15. Tách dữ liệu thành hai phần

Bước 4: Tính toán độ quan trọng của các thuộc tính và chọn ra 5 thuộc tính quan trọng

nhất.

Top 5 thuộc tính bao gồm: Water_Usage_Liters, Carbon_Footprint_MT, Waste_Production_KG, Average_Price_USD, Product_Lines.



```
# Sắp xếp các thuộc tính theo độ quan trọng
f_list = sorted(zip(importances, features), reverse=True, key=lambda x: x[0])

# Lấy ra 5 thuộc tính quan trọng nhất
top_features = [f_list[i][1] for i in range(5)]

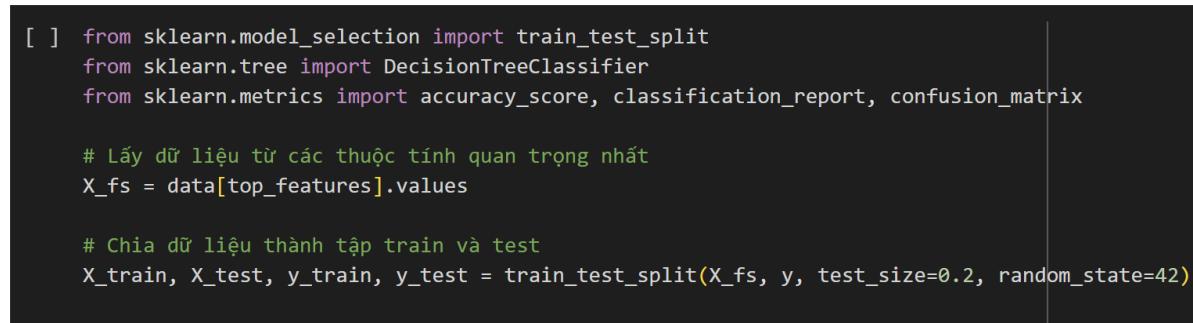
# In ra các thuộc tính quan trọng và độ quan trọng của chúng
for i in range(5):
    print(f"Feature: {f_list[i][1]}, Importance Score: {f_list[i][0]}")

print("Top 5 important features:", top_features)
```

```
Feature: Water_Usage_Liters, Importance Score: 0.15360328640173707
Feature: Carbon_Footprint_MT, Importance Score: 0.14959227567249184
Feature: Waste_Production_KG, Importance Score: 0.14850624778701313
Feature: Average_Price_USD, Importance Score: 0.1474426071990491
Feature: Product_Lines, Importance Score: 0.10282945764436861
Top 5 important features: ['Water_Usage_Liters', 'Carbon_Footprint_MT', 'Waste_Production_KG', 'Average_Price_USD', 'Product_Lines']
```

Hình 4. 16. Tính độ quan trọng và tìm ra 5 thuộc tính quan trọng nhất.

Bước 5: Lấy dữ liệu của top 5 thuộc tính quan trọng nhất vừa tìm được và chia tập dữ liệu thành train:test theo tỉ lệ 80:20



```
[ ] from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

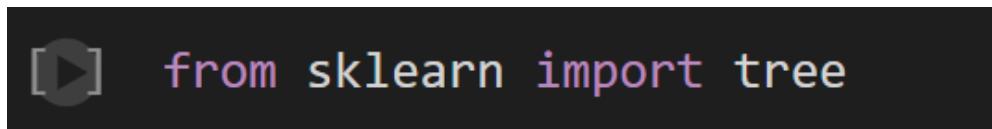
# Lấy dữ liệu từ các thuộc tính quan trọng nhất
X_fs = data[top_features].values

# Chia dữ liệu thành tập train và test
X_train, X_test, y_train, y_test = train_test_split(X_fs, y, test_size=0.2, random_state=42)
```

Hình 4. 17. Lấy dữ liệu top 5 thuộc tính quan trọng và chia tập dữ liệu.

4.2.2 Xây dựng mô hình Decision Tree

Bước 1: Import thêm các thư viện cần thiết.



```
[ ] from sklearn import tree
```

Hình 4. 18. Import thêm thư viện

Bước 2: Huấn luyện mô hình.

```
[ ] clf= tree.DecisionTreeClassifier(criterion="gini", random_state=0)
#Train Decision Tree Classifier
clf.fit(X_train, y_train)

→ /usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning
  and should_run_async(code)
    ▾ DecisionTreeClassifier ⓘ ⓘ
DecisionTreeClassifier(random_state=0)
```

Hình 4. 20. Huấn luyện mô hình Decision Tree

Bước 3: Tính độ chính xác.

```
▶ #Predict the response for test dataset
tree_pred = clf.predict(X_test)

#Model Accuracy, how often is the classifier correct?
tree_score = metrics.accuracy_score(y_test, tree_pred)
print("Accuracy:", tree_score)
print("Report:", metrics.classification_report(y_test, tree_pred))

→ Accuracy: 0.349
Report:
precision      recall   f1-score   support
          0       0.36      0.36      0.36      346
          1       0.34      0.34      0.34      316
          2       0.35      0.35      0.35      338
accuracy           0.35      0.35      0.35     1000
macro avg       0.35      0.35      0.35     1000
weighted avg     0.35      0.35      0.35     1000
```

Hình 4. 21. Độ chính xác của mô hình Decision Tree

Nhân xét: Kết quả cho thấy mô hình đạt độ chính xác thấp trên tập dữ liệu. Điều này hoàn toàn đúng, bởi vì các thuộc tính trên tập dữ liệu có độ tương quan rất thấp với nhau dẫn đến không có sự ảnh hưởng mạnh đến nhãn cần dự đoán.

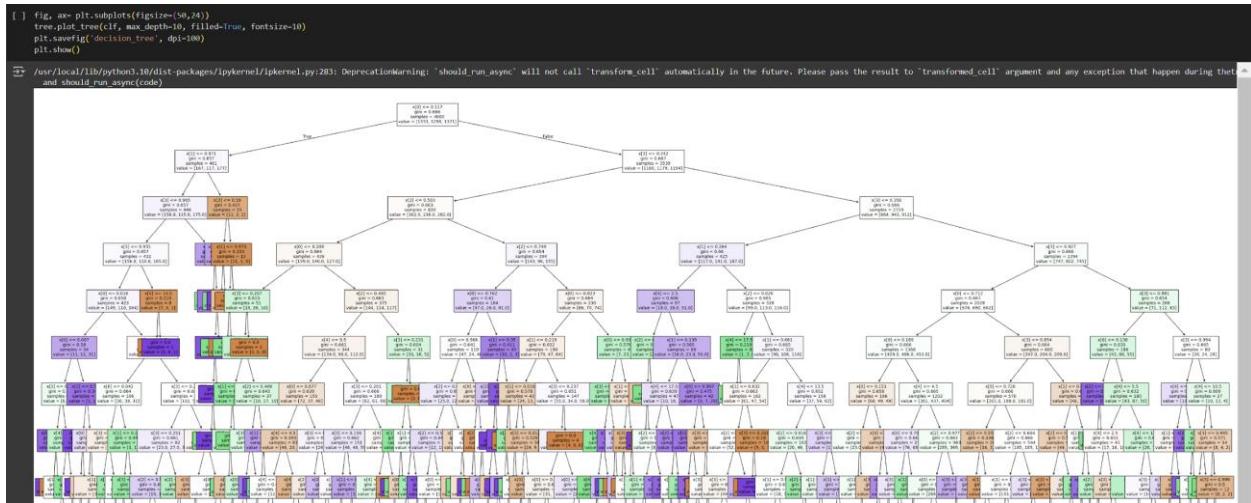
Bước 4: Tính toán ma trận nhầm lẫn.



Hình 4. 22. Ma trận nhầm lẫn của mô hình Decision Tree

Nhận xét: Mô hình sai phần lớn các nhãn, trong 3 nhãn thì nhãn 0 được dự đoán đúng cao nhất.

Bước 5: Vẽ cây CART:



Hình 4. 23. Cây CART

⇒ Từ cây CART ở trên ta có thể rút ra một tập luật mẫu như sau:

- Với Water_Usage_Liters ≤ 0.12 , Carbon_Footprint_MT ≤ 0.97 , Average_Price_USD ≤ 0.96 , Average_Price_USD ≤ 0.93 , Water_Usage_Liters ≤ 0.02 , Water_Usage_Liters ≤ 0.01 , Average_Price_USD ≤ 0.24 , Water_Usage_Liters ≤ 0.01 thuộc class: 2

Bước 6: In tập luật

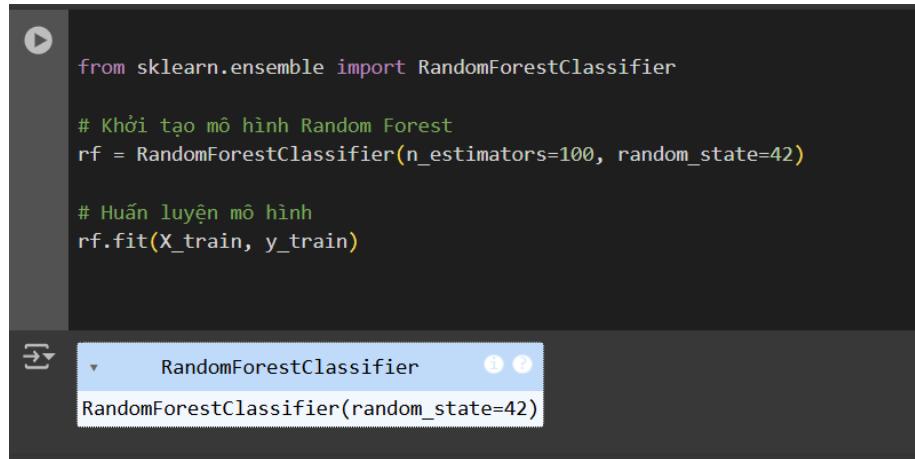
```
from sklearn.tree import export_text
# Giả sử 'clf' là mô hình Decision Tree đã được huấn luyện
tree_rules = export_text(clf, feature_names=top_features)
print(tree_rules)

--- Water_Usage_Liters <= 0.12
|--- Carbon_Footprint_MT <= 0.97
|   |--- Average_Price_USD <= 0.96
|   |   |--- Water_Usage_Liters <= 0.02
|   |   |   |--- Water_Usage_Liters <= 0.01
|   |   |   |   |--- Average_Price_USD <= 0.24
|   |   |   |   |   |--- Water_Usage_Liters <= 0.01
|   |   |   |   |   |--- class: 2
|   |   |   |   |   |--- Water_Usage_Liters > 0.01
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- Average_Price_USD > 0.24
|   |   |   |   |--- Product_Lines <= 12.50
|   |   |   |   |   |--- Carbon_Footprint_MT <= 0.44
|   |   |   |   |   |   |--- Waste_Production_KG <= 0.59
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- Waste_Production_KG > 0.59
|   |   |   |   |   |   |--- class: 2
|   |   |   |   |   |--- Carbon_Footprint_MT > 0.44
|   |   |   |   |   |--- Product_Lines <= 6.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Product_Lines > 6.50
|   |   |   |   |   |   |--- class: 2
|   |   |   |   |--- Product_Lines > 12.50
|   |   |   |   |   |--- Water_Usage_Liters <= 0.00
|   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Water_Usage_Liters > 0.00
|   |   |   |   |   |--- class: 1
|   |   |   |--- Water_Usage_liters > 0.01
|   |   |   |--- Carbon_Footprint_MT <= 0.60
|   |   |   |   |--- Average_Price_USD <= 0.18
|   |   |   |   |   |--- Carbon_Footprint_MT <= 0.12
|   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Carbon_Footprint_MT > 0.12
|   |   |   |   |   |--- class: 2
|   |   |   |   |--- Average_Price_USD > 0.18
|   |   |   |   |   |--- class: 2
|   |   |   |   |--- Carbon_Footprint_MT > 0.60
|   |   |   |   |   |--- Water_Usage_Liters <= 0.01
|   |   |   |   |   |--- Product_Lines <= 12.00
|   |   |   |   |   |   |--- Water_Usage_Liters <= 0.01
|   |   |   |   |   |   |--- class: 2
|   |   |   |   |   |   |--- Water_Usage_Liters > 0.01
```

Hình 4. 24. Tập luật từ cây CART

4.2.3 Xây dựng mô hình Random Forest.

Bước 1: Import các thư viện cần thiết. Sau đó tiến hành huấn luyện mô hình Random Forest.



```
from sklearn.ensemble import RandomForestClassifier

# Khởi tạo mô hình Random Forest
rf = RandomForestClassifier(n_estimators=100, random_state=42)

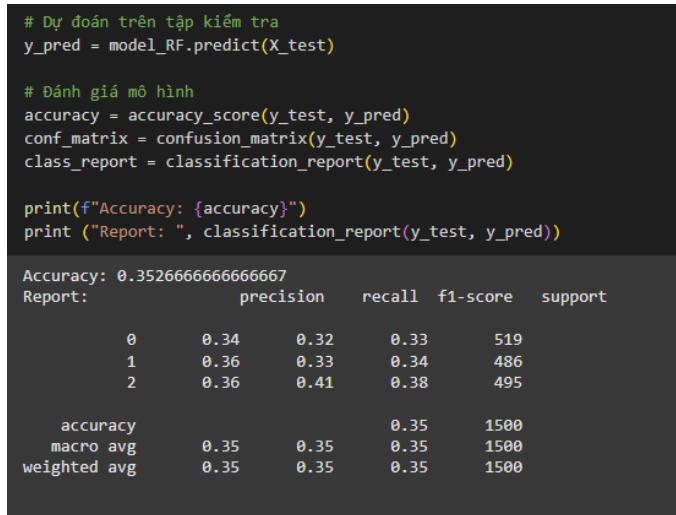
# Huấn luyện mô hình
rf.fit(X_train, y_train)
```

RandomForestClassifier

```
RandomForestClassifier(random_state=42)
```

Hình 4. 25. Import thư viện và huấn luyện mô hình Random forest

Bước 2: Dự đoán và tính độ chính xác.



```
# Dự đoán trên tập kiểm tra
y_pred = model_RF.predict(X_test)

# Đánh giá mô hình
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)

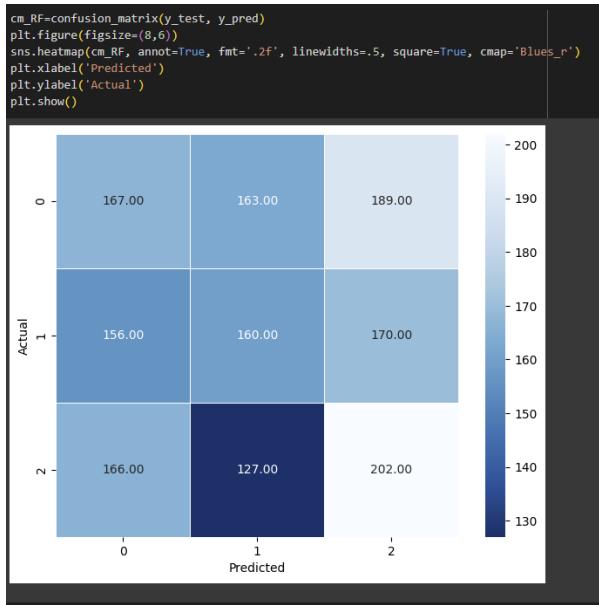
print(f"Accuracy: {accuracy}")
print ("Report: ", classification_report(y_test, y_pred))

Accuracy: 0.3526666666666667
Report:
precision    recall  f1-score   support
          0       0.34      0.32      0.33      519
          1       0.36      0.33      0.34      486
          2       0.36      0.41      0.38      495

   accuracy                           0.35      1500
  macro avg       0.35      0.35      0.35      1500
weighted avg       0.35      0.35      0.35      1500
```

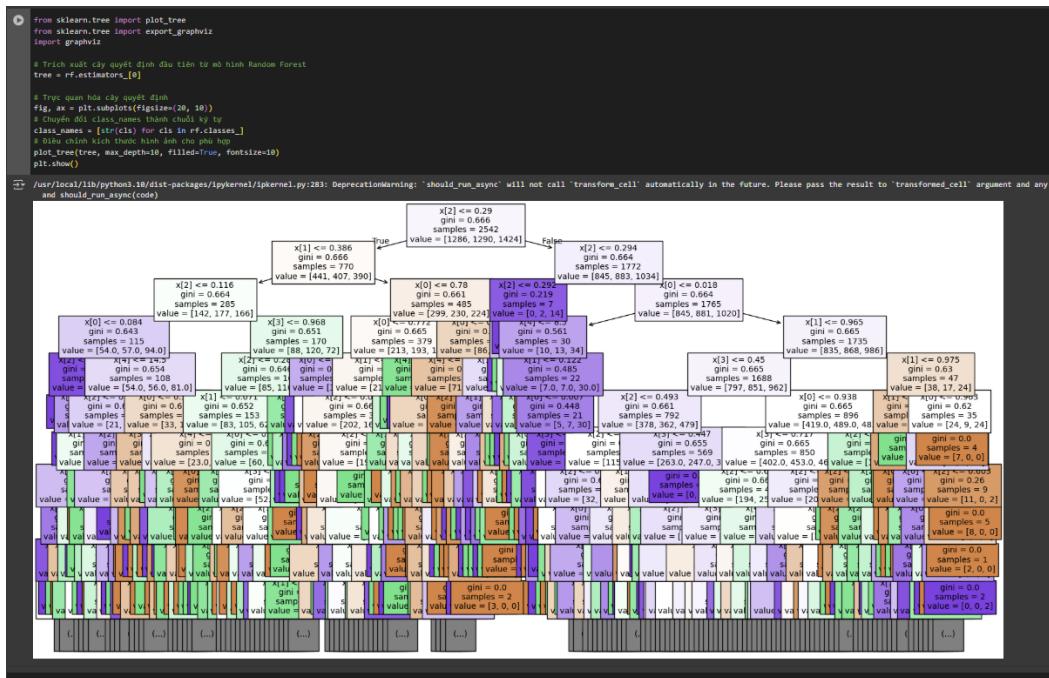
Hình 4. 26. Độ chính xác của mô hình Random forest

Bước 3: Tính toán ma trận nhầm lẫn



Hình 4. 27. Ma trận nhầm lẫn của mô hình Random forest

Bước 4: In cây quyết định đầu tiên của mô hình Random forest.

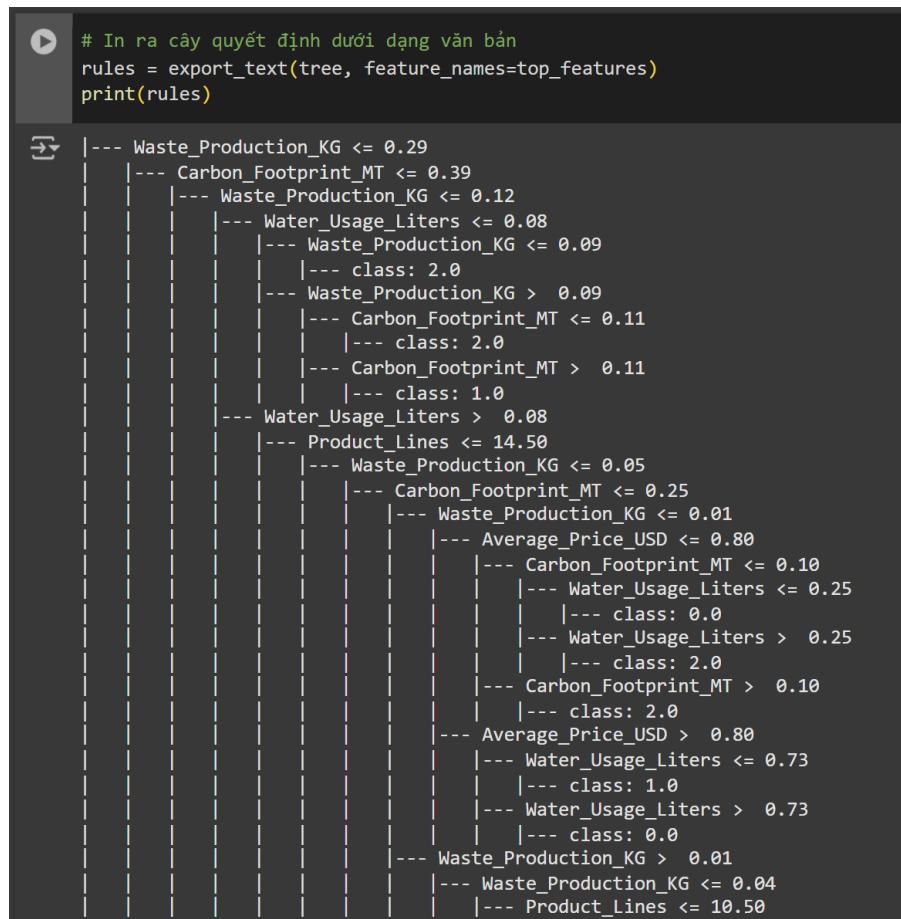


Hình 4. 28. Cây quyết định đầu tiên của mô hình Random forest

⇒ Từ cây ở trên ta có thể rút ra một tập luật mẫu như sau:

- Waste_Production_KG <= 0.29, Carbon_Footprint_MT <= 0.39
 Water_Usage_Liters > 0.08, Product_Lines <= 14.50, Waste_Production_KG <= 0.05, Carbon_Footprint_MT <= 0.25, Waste_Production_KG <= 0.01, Average_Price_USD <= 0.80, Carbon_Footprint_MT <= 0.10, Water_Usage_Liters <= 0.25 thuộc class: 0.0

Bước 5: In ra tập luật



```
# In ra cây quyết định dưới dạng văn bản
rules = export_text(tree, feature_names=top_features)
print(rules)
```

```

--- Waste_Production_KG <= 0.29
|--- Carbon_Footprint_MT <= 0.39
|   |--- Waste_Production_KG <= 0.12
|   |   |--- Water_Usage_Liters <= 0.08
|   |   |--- Waste_Production_KG <= 0.09
|   |   |   |--- class: 2.0
|   |   |--- Waste_Production_KG > 0.09
|   |   |   |--- Carbon_Footprint_MT <= 0.11
|   |   |   |   |--- class: 2.0
|   |   |   |--- Carbon_Footprint_MT > 0.11
|   |   |   |   |--- class: 1.0
|   |--- Water_Usage_Liters > 0.08
|   |--- Product_Lines <= 14.50
|   |   |--- Waste_Production_KG <= 0.05
|   |   |   |--- Carbon_Footprint_MT <= 0.25
|   |   |   |   |--- Waste_Production_KG <= 0.01
|   |   |   |   |   |--- Average_Price_USD <= 0.80
|   |   |   |   |   |--- Carbon_Footprint_MT <= 0.10
|   |   |   |   |   |   |--- Water_Usage_Liters <= 0.25
|   |   |   |   |   |   |--- class: 0.0
|   |   |   |   |   |--- Water_Usage_Liters > 0.25
|   |   |   |   |   |--- class: 2.0
|   |   |   |   |--- Carbon_Footprint_MT > 0.10
|   |   |   |   |   |--- class: 2.0
|   |   |--- Average_Price_USD > 0.80
|   |   |   |--- Water_Usage_Liters <= 0.73
|   |   |   |   |--- class: 1.0
|   |   |   |   |--- Water_Usage_Liters > 0.73
|   |   |   |   |--- class: 0.0
|   |   |--- Waste_Production_KG > 0.01
|   |   |   |--- Waste_Production_KG <= 0.04
|   |   |   |   |--- Product_Lines <= 10.50

```

Hình 4. 29. Tập luật từ cây quyết định đầu tiên của Random Forest

5. TỔNG KẾT

5.1 Kết quả đạt được

- Biết cách thiết kế kho dữ liệu sử dụng lược đồ hình sao.
- Thực hiện thành công quá trình SSIS, SSAS, và Data Mining.
- Sửa được các lỗi phát sinh trong quá trình chạy project.
- Biết cách viết các câu lệnh MDX.
- Dựa vào kết quả của các quá trình để có cái nhìn tổng quan về tình hình cũng như là xu hướng phát triển thời trang bền vững trên thế giới nói chung và từng quốc gia nói riêng.

5.2 Thuận lợi

- Công việc hoàn thành nhanh chóng, phân chia thời gian hợp lý.
- Thao tác ít phát sinh lỗi nên dễ dàng hơn.

5.3 Khó khăn

- Do chưa có nhiều kinh nghiệm nên phải chạy project nhiều lần để đạt được kết quả cuối cùng.
- Khó khăn trong việc tìm tập luật, pivot table vì tập luật quá phức tạp.

BẢNG PHÂN CÔNG CÔNG VIỆC

Tên công việc	Thành viên	
	Võ Thanh Nhàn (Leader)	Huỳnh Ngọc Diễm Phúc
Lựa chọn đề tài và tìm dataset	x	x
Mô tả dữ liệu và làm sạch dữ liệu	x	x
Thiết kế kho dữ liệu	x	x
SSIS	Đỗ dữ liệu vào các bảng Dim	x
	Merge dữ liệu vào bảng Fact	x
SSAS	Chạy dự án SSAS	x
	Cube	x
	PowerBI	x
	Excel	x
Data Mining	Tiền xử lý dữ liệu	x
	Thuật toán Decision Tree	x
	Thuật toán Random Forest	x
Báo cáo	Tạo khung báo cáo	x
	Viết báo cáo	x
Đánh giá hoàn thành công việc		100%
		100%