

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THÔNG THÔNG TIN



BÁO CÁO ĐỒ ÁN
KHO DỮ LIỆU VÀ OLAP

ĐỀ TÀI: GIÁ NHÀ Ở BẮC KINH

Lớp: IS217.O22.HTCL

GVHD: Đỗ Thị Minh Phụng

Thành viên nhóm:

- Ché Duy Khang – 21522187
- Trịnh Đức – 21521969

TP. Hồ Chí Minh, tháng 3 năm 2024

LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến toàn thể giảng viên trường Đại học Công nghệ thông tin – Đại học Quốc gia TP.HCM cũng như là nhà trường vì đã giúp nhóm em có những kiến thức cơ bản làm nền tảng để thực hiện đồ án môn học này.

Chúng em xin bày tỏ lòng biết ơn sâu sắc đến tất cả những người đã hỗ trợ và đóng góp vào việc hoàn thành báo cáo này. Sự giúp đỡ ấy đã đóng vai trò quan trọng trong quá trình nghiên cứu và thực hiện dự án.

Đặc biệt trong đó chúng em xin gửi lời cảm ơn chân thành đến cô Đỗ Thị Minh Phụng (Giảng viên lý thuyết Kho dữ liệu và OLAP). Thời gian vừa qua, cô đã trực tiếp giảng dạy, truyền đạt kinh nghiệm, cung cấp sự hướng dẫn, kiến thức chuyên sâu và hỗ trợ nhóm chúng em trong suốt quá trình nghiên cứu, giúp nhóm chúng em hoàn thành tốt đồ án môn học của mình. Chúc cô sẽ luôn dồi dào sức khỏe, tràn đầy nhiệt huyết để có thể tiếp tục giảng dạy, dù dắt những thế hệ sinh viên tiếp theo.

Chúng em cũng gửi lời cảm ơn đến tập thể lớp IS217.O22.HTCL vì thời gian qua đã đồng hành cùng nhau. Cùng nhau học tập, tranh luận một cách sôi nổi để xây dựng bài học một cách tốt nhất. Cảm ơn các bạn đã cùng thảo luận, đánh giá và đóng góp ý kiến, cùng học hỏi nghiên cứu để thực hiện đồ án một cách tốt nhất có thể.

Đồng thời, chúng em muốn gửi lời cảm ơn sâu sắc đến mỗi thành viên trong nhóm, hai người chúng em đã đồng hành và chung tay thực hiện đồ án môn học. Thời gian gắn bó cùng nhau đã giúp chúng em rút ra được nhiều bài học và kinh nghiệm quý báu. Bài báo cáo này không thể hoàn thành mà không có sự góp sức của các bạn.

Mặc dù đã vận dụng tối đa những gì đã học được nhưng chúng em vẫn khó có thể tránh khỏi những sai sót. Chính vì vậy, nhóm em rất mong nhận được sự góp ý từ phía cô để có thể hoàn thiện một cách tốt nhất có thể. Qua đó cũng tích lũy và học hỏi kinh nghiệm để làm hành trang cho tương lai.

Chúng em xin hết và xin một lần nữa gửi lời cảm ơn chân thành đến với cô!

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

....., ngày ... tháng ... năm 2024
Người nhận xét
(Ký và ghi rõ họ tên)

MỤC LỤC

CHƯƠNG I: GIỚI THIỆU VỀ TỔNG QUAN DỮ LIỆU	21
1. Phát biểu về dữ liệu.....	21
1.1. Mô tả về dữ liệu	21
1.2. Thuộc tính của kho dữ liệu	21
1.3. Kho dữ liệu đã xử lý	22
2. XÂY DỰNG KHO DỮ LIỆU	24
2.1. Sơ đồ bông tuyết minh hoạ	24
2.2. DIM_LOCATION.....	24
2.3. DIM_BUILDING	24
2.4. DIM_HOUSE	25
2.5. DIM_DATE.....	25
2.6. FACT_TRANSACTION	25
3. NỘI DUNG 16 CÂU HỎI TRUY VẤN	26
CHƯƠNG II: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)	28
1. Tạo Project SSIS	28
2. Chuẩn bị cơ sở dữ liệu	30
3. Tạo mới project SSIS	30
4. Tạo bảng Dim và bảng Fact	33
4.1. Chuẩn bị dữ liệu từ file .csv	33
4.2. Tạo bảng DIM_DATE.....	39
4.3. Bảng DIM_LOCATION	49
4.4. Bảng DIM_BUILDING	53
4.5. Bảng DIM_HOUSE_RAW:.....	57
4.6. Merge DIM_HOUSE_RAW và DIM_LOCATION vào DIM_HOUSE1.....	62
4.7. Merge DIM_HOUSE1 và DIM_BUILDING vào DIM_HOUSE	71
4.8. Bảng FACT_RAW	83
4.8. Merge FACT_RAW và DIM_LOCATION vào FACT1	88
4.9. Merge FACT1 và DIM_BUILDING vào FACT2.....	98
4.10. Merge FACT2 và DIM_HOUSE vào FACT3.....	106

4.11. Merge FACT3 và DIM_DATE vào FACT_TRANSACTION.....	115
4.12. Tạo khóa ngoại từ bảng Fact đến các Dimension	123
4.13. Chạy dự án SSIS	125
4.14. Kiểm tra dữ liệu	130
CHƯƠNG III: PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS)	133
1. Tạo Project SSAS	133
2. Xác định dữ liệu nguồn (Data Sources)	135
3. Xác định khung nhìn dữ liệu nguồn (Data Source Views)	141
4. Xây dựng các khối (Cubes) và xác định các độ đo (Measures).....	145
5. Xác định các độ đo (Measures).....	170
5.1. Phân cấp bảng Dim_Time.....	175
6. Thực hiện 16 câu truy vấn – Quá trình phân tích dữ liệu bằng thao tác tay trên các khối CUBE, Pivot table trong Excel, ngôn ngữ MDX.....	186
6.1. Câu truy vấn 1	186
6.2. Câu truy vấn 2	189
6.3. Câu truy vấn 3	192
6.4. Câu truy vấn 4	194
6.5. Câu truy vấn 5	197
6.6. Câu truy vấn 6	199
6.7. Câu truy vấn 7	203
6.8. Câu truy vấn 8	206
6.9. Câu truy vấn 9	211
6.10. Câu truy vấn 10	213
6.11. Câu truy vấn 11	218
6.12. Câu truy vấn 12	219
6.13. Câu truy vấn 13	222
6.14. Câu truy vấn 14.....	226
6.15. Câu truy vấn 15	228
6.16. Câu truy vấn 16	232
CHƯƠNG IV: QUÁ TRÌNH LẬP BÁO BIÊU (SSRS)	236
1. Cài đặt các công cụ cần thiết.....	236

2. Report Designer (Reporting Services)	252
2.1. Tạo Report đơn giản	252
2.1.1. Mục đích, câu truy vấn 1.....	252
2.1.2. Tạo một Report Server Project mới	252
2.1.3. Tạo Dataset cho Report.....	257
2.1.5. Thêm dữ liệu vào Report	260
2.1.6. Định dạng Report.....	264
2.1.7. Xuất báo cáo	267
2.1.7.1. Cấu hình Report Server.....	267
2.1.7.2. Xuất báo cáo	267
2.2. Tạo report có Total và Grouping	270
2.2.1. Mục đích, câu truy vấn 2.....	270
2.2.2. Tạo các dữ liệu cần thiết cho Report.....	270
2.2.3. Group dữ liệu	272
2.2.4. Total và Grand Total.....	274
2.3. Tạo report 3	290
2.3.1. Mục đích, câu truy vấn 3.....	290
2.3.2 Tạo các dữ liệu cần thiết cho Report.....	290
2.3.3. Group dữ liệu	292
3. Quá trình lập báo biểu bằng công cụ Power BI	304
3.1. Cài đặt Power BI và kết nối với SQL Server	304
3.2. Báo biểu 1: Thống kê tổng số tiền giao dịch qua từng năm từ 2011 đến 2017.....	308
3.3. Báo biểu 2: Thống kê diện tích bất động sản của từng khu vực được giao dịch qua từng năm, chỉ tính những mảnh đất có loại hình kiến trúc là “Bungalow” và “Tower”	313
3.4. Báo biểu 3: Thống kê số lượng và tổng số tiền giao dịch đạt chuẩn chất lượng cao qua từng năm	317
3.5. Cách kiểm tra và xem lại report đã được Publish	321
CHƯƠNG V: DATA MINING.....	324
1. Tạo Mining Structures	324
2. Nhận xét kết quả sau khi Mining dữ liệu.....	339
2.1. Thuật toán Microsoft Decision Trees.....	339

2.2. Thuật toán Microsoft Clustering	343
2.3. Thuật toán Microsoft Naive Bayes	346
2.4. Đánh giá các thuật toán bằng Mining Accuracy Chart	350

DANH MỤC HÌNH

Hình 1: Bảng chi tiết có trong file csv	22
Hình 2: Sơ đồ bông tuyết minh họa	24
Hình 3: Giao diện sau khi mở Visual Studio	28
Hình 4: Chọn “Extensions” -> “Manage Extensions”	28
Hình 5: SQL Server Intergration Services Projects	29
Hình 6: Thông báo cài SSIS thành công	29
Hình 7: Kết nối Database Engine.....	30
Hình 8: Tạo cơ sở dữ liệu HOUSE_PRICE_BEIJING.....	30
Hình 9: Create a new project	31
Hình 10: Chọn Intergration Services Project	31
Hình 11: Đặt tên và thiết lập đường dẫn	32
Hình 12: Thêm Data Flow Task	33
Hình 13: Chọn New để tạo một Flat File Connection Manager	34
Hình 14: Chọn nút Browse.. ..	35
Hình 15: Chọn file dữ liệu .csv	36
Hình 16: Xem lại các cột dữ liệu trong file dữ liệu	37
Hình 17: Click chọn OK	38
Hình 18: Tạo Multicast	39
Hình 19: Tạo mới một Sort có tên là Sort_DIM_DATE.....	39
Hình 20: Đặt tên cho Data Conversion này là DIM_DATE Change type string to datetime	40
Hình 21: Click chuột phải vào Data conversion này và chọn Edit	40
Hình 22: Đặt lại Output alias cho tradeTime	41
Hình 23: Thêm thành phần Derived Column.....	41
Hình 24: Chọn phương thức DATEPART.....	42
Hình 25: Chia TRADE_DATE thành các cột	43
Hình 26: Khi hoàn thành, nhấn nút Configure Error Output... ..	44
Hình 27: Tạo một connection mới đến MS SQL Server.....	45
Hình 28: Tạo một connection mới	45
Hình 29: Test Connection để kiểm tra kết nối	46
Hình 30: Chọn connection vừa tạo đến MS SQL Server và nhấn OK.....	47
Hình 31: Tạo mới bảng	47
Hình 32: Tạo mới bảng(2)	48
Hình 33: Xem xét việc ánh xạ các cột dữ liệu	49
Hình 34: Chọn một Sort để tạo ra Sort_DIM_LOCATION cho DIM_LOCATION	49
Hình 35: Chọn các cột để đổ dữ liệu vào Sort_DIM_LOCATION	50
Hình 36: Đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu.....	50
Hình 37: Chọn New... để tạo bảng DIM_LOCATION	51
Hình 38: Câu lệnh SQL tạo bảng DIM_LOCATION	52

Hình 39: Chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu.....	53
Hình 40: Chọn một Sort để tạo ra Sort_DIM_BUILDING	53
Hình 41: Tick chọn Remove rows with duplicate sort values	54
Hình 42: Tạo mới một OLE DB Destination.....	55
Hình 43: New... để tạo bảng DIM_BUILDING.....	55
Hình 44: Câu lệnh SQL tạo bảng DIM_BUILDING	56
Hình 45: Chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu.....	57
Hình 46: Chọn một Sort để tạo ra Sort_DIM_HOUSE_RAW	58
Hình 47: Chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau	59
Hình 48: Tạo mới một OLE DB Destination để đỗ dữ liệu gốc sau khi đã được xử lý vào trong bảng DIM_HOUSE_RAW.....	59
Hình 49: Câu lệnh SQL tạo bảng DIM_HOUSE_RA	60
Hình 50: Xem xét việc ánh xạ các cột dữ liệu	61
Hình 51: Tạo hai Data Flow Task và đổi tên	62
Hình 52: Tạo 2 OLE DB Source và đổi tên	62
Hình 53: Làm data source cho bảng DIM_HOUSE_RAW mới	63
Hình 54: Xem xét các cột được ánh xạ	64
Hình 55: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (1).....	65
Hình 56: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (2).....	65
Hình 57: Tạo 2 Sort tương ứng với mỗi Source	66
Hình 58: Chuẩn bị cho quá trình merge (1)	66
Hình 59: Chuẩn bị cho quá trình merge (2)	67
Hình 60: Nối Sort với Merge Join	67
Hình 61: Nối Sort1 với Merge Join	68
Hình 62: Tick chọn tất cả các cột của Sort nhưng không lấy 4 thuộc tính LNG, LAT, DISTRICT và CID	69
Hình 63: Câu lệnh SQL tạo bảng DIM_HOUSE1	70
Hình 64: Xem xét việc ánh xạ các cột dữ liệu	71
Hình 65: Tạo thêm một Data Flow Task và đổi tên	72
Hình 66: Tạo 2 OLE DB Source và đổi tên	72
Hình 67: Chọn bảng DIM_HOUSE1 đã được tạo khi merge	73
Hình 68: Xem xét các cột được ánh xạ	74
Hình 69: Thực hiện chọn ánh xạ các cột cho DIM_BUILDING.....	75
Hình 70: Xem xét các cột được ánh xạ	76
Hình 71: Tạo 2 Sort tương ứng với mỗi Source	76
Hình 72: Chọn các cột theo thứ tự giống với bảng DIM_BUILDING để chuẩn bị cho quá trình merge(1).....	77
Hình 73: : Chọn các cột theo thứ tự giống với bảng DIM_BUILDING để chuẩn bị cho quá trình merge(2).....	78
Hình 74: Chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng DIM_HOUSE1 ..	78

Hình 75: Nối Sort1 với Merge Join	79
Hình 76: Kết quả sau khi merge bảng DIM_HOUSE1.....	80
Hình 77: Câu lệnh SQL tạo bảng DIM_HOUSE.....	81
Hình 78: Xem xét việc ánh xạ các cột dữ liệu	82
Hình 79: Kết quả.....	82
Hình 80: Tạo một Data Conversion	83
Hình 81: Chọn lại kiểu dữ liệu cho cột này date[DT_DATE]	84
Hình 82: Tạo bảng Fact.....	84
Hình 83: Câu lệnh SQL tạo bảng FACT_RAW	85
Hình 84: Ánh xạ cột id trong Input Column vào cột ID_TRANS của bảng FACT_RAW.....	87
Hình 85: Kết quả.....	87
Hình 86: Tạo hai Data Flow Task và đổi tên	88
Hình 87: Tạo 2 OLE DB Source và đổi tên	88
Hình 88: Chọn bảng FACT_RAW đã tạo trước đó làm data source cho bảng FACT_RAW	89
Hình 89: Xem xét các cột được ánh xạ	90
Hình 90: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (1).....	91
Hình 91: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (2).....	91
Hình 92: Tạo 2 Sort là Sort và Sort1 tương ứng với mỗi Source	92
Hình 93: Chọn các cột LNG, LAT, CID, DISTRICT	92
Hình 94: Tương tự ta chọn cột LNG, LAT, CID, DISTRICT	93
Hình 95: Chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng FACT_RAW	94
Hình 96: Kết quả sau khi merge	95
Hình 97: Câu lệnh SQL tạo bảng FACT1	96
Hình 98: Xem xét việc ánh xạ các cột dữ liệu	97
Hình 99: Tạo thêm một Data Flow Task và đổi tên	98
Hình 100: Tạo 2 OLE DB Source và đổi tên	99
Hình 101: Chọn bảng FACT1 đã được tạo khi merge	99
Hình 102: Xem xét các cột được ánh xạ	100
Hình 103: Tạo 2 Sort tương ứng với mỗi Source	100
Hình 104: Chọn các cột để chuẩn bị cho quá trình merge	101
Hình 105: Tương tự ta chọn cột cho Sort1	102
Hình 106: Nối Sort1 với Merge Join	103
Hình 107: Kết quả sau khi merge	104
Hình 108: Tạo bảng FACT2 để chứa tất cả những gì đã merge.....	104
Hình 109: Câu lệnh SQL tạo bảng FACT2	105
Hình 110: Xem xét việc ánh xạ các cột dữ liệu	106
Hình 111: Tạo thêm một Data Flow Task và đổi tên	107
Hình 112: Tạo 2 OLE DB Source và đổi tên	107
Hình 113: Chọn bảng FACT2 đã được tạo khi merge	108
Hình 114: Xem xét các cột được ánh xạ	109

Hình 115: Tạo 2 Sort tương ứng với mỗi Source.....	109
Hình 116: Chọn các cột để chuẩn bị cho quá trình merge	110
Hình 117: Chọn cột cho Sort1.....	111
Hình 118: Nối Sort1 với Merge Join.....	112
Hình 119: Kết quả sau khi merge.....	113
Hình 120: Tạo bảng FACT3.....	113
Hình 121: Câu lệnh SQL tạo bảng FACT3	114
Hình 122: Xem xét việc ánh xạ các cột dữ liệu	115
Hình 123: Tạo thêm một Data Flow Task và đổi tên	116
Hình 124: Tạo 2 OLE DB Source và đổi tên	116
Hình 125: Chọn bảng FACT3 đã được tạo khi merge	117
Hình 126: Xem xét các cột được ánh xạ.....	118
Hình 127: Tạo 2 Sort tương ứng với mỗi Source	118
Hình 128: Chọn cột để chuẩn bị cho quá trình merge	119
Hình 129: Chọn cột TRADE_DATE cho Sort1	120
Hình 130: Nối Sort1 với Merge Join	121
Hình 131: Kết quả sau khi merge	121
Hình 132: Tạo bảng FACT_TRANSACTION để chứa tất cả những gì đã merge	122
Hình 133: Câu lệnh SQL tạo bảng FACT_TRANSACTION.....	122
Hình 134: Xem xét việc ánh xạ các cột dữ liệu	123
Hình 135: Tạo một Execute SQL Task	124
Hình 136: Chọn connection đã thiết lập đến data warehouse	124
Hình 137: Thêm các câu truy vấn SQL thực hiện tạo các khóa ngoại từ các Dimension đến bảng Fact.....	125
Hình 138: Thêm vào một Execute SQL Task	126
Hình 139: Chọn connection đã thiết lập đến data warehouse	126
Hình 140: Thêm các câu truy vấn SQL thực hiện xóa dữ liệu cũ trong các bảng	127
Hình 141: Nút Start.....	127
Hình 142: Kết quả chạy project (1).....	128
Hình 143: Kết quả chạy project (2).....	128
Hình 144: Thêm các lệnh SQL để xóa và thêm lại các khóa ngoại	129
Hình 145: Chạy lại project (1)	129
Hình 146: Chạy lại project (2)	130
Hình 147: Kiểm tra dữ liệu bảng DIM_LOCATION.....	130
Hình 148: Kiểm tra dữ liệu bảng DIM_BUILDING	131
Hình 149: Kiểm tra dữ liệu bảng DIM_HOUSE	131
Hình 150: Kiểm tra dữ liệu bảng DIM_DATE	132
Hình 151: Kiểm tra dữ liệu bảng FACT_TRANSACTION	132
Hình 152: Tải extention “Microsoft Analysis Services Project”	133
Hình 153: Chọn “Create a new project”	134

Hình 154: Chọn Analysis Services Multidimensional Project.....	134
Hình 155: Đặt tên và thiết lập đường dẫn cho Project.....	135
Hình 156: Chọn New Data Source.....	135
Hình 157: Chọn Next để tiếp tục	136
Hình 158: Tạo kết nối với cơ sở dữ liệu đã được tạo từ quá trình SSIS	137
Hình 159: Ta nhập Server name và chọn cơ sở dữ liệu	138
Hình 160: Chọn data source vừa tạo	139
Hình 161: Chọn “Use the service account”	140
Hình 162: Chọn Finish để hoàn tất	140
Hình 163: Chọn New Data Source View	141
Hình 164: Chọn Next để tiếp tục	142
Hình 165: Chọn data source vừa tạo	142
Hình 166: Thêm bảng Fact vào data source.....	143
Hình 167: Thêm tất cả các bảng Dim vào data source view.	144
Hình 168: Chọn Finish.....	144
Hình 169: Tất cả các bảng đã được chọn và thể hiện sự liên kết giữa chúng	145
Hình 170: Chọn New Cube.....	146
Hình 171: Chọn Next để tiếp tục	146
Hình 172: Chọn use existing tables	147
Hình 173: Chọn Fact để phân chia các measure group.....	148
Hình 174: Chọn những độ đo để xuất	149
Hình 175: Chọn danh sách các bảng Dimension	150
Hình 176: Chọn Finish để hoàn tất quy trình.....	151
Hình 177: Kết quả như hình.....	152
Hình 178: Lần lượt double-click vào từng bảng dimension	153
Hình 179: Sau đó kéo thả các thuộc tính từ Data Source View vào Attributes	153
Hình 180: Kéo thả các thuộc tính từ Data Source View vào Attributes.....	154
Hình 181: Chọn Properties.....	155
Hình 182: OrderBy "Key"	156
Hình 183: Kết quả khi deploy thành công	157
Hình 184: Xuất hiện lỗi.....	158
Hình 185: Đường dẫn “SQL Server 2022 Configuration Manager”	158
Hình 186: Giao diện SQL Server 2022 Configuration Manager	158
Hình 187: Vào trong Properties	159
Hình 188: Chọn Browse.....	159
Hình 189: Chọn Advanced -> Find Now	160
Hình 190: Chọn một tài khoản có đủ quyền	161
Hình 191: Connect	162
Hình 192: Mục Logins của server.....	162
Hình 193: Kiểm tra Windows authentication.....	163

Hình 194: Tích chọn public và sysadmin	164
Hình 195: Tích chọn như hình	165
Hình 196: Chọn Properties.....	166
Hình 197: Điền Server sẽ Deploy	167
Hình 198: Deploy dự án.....	168
Hình 199: Chọn Run	169
Hình 200: Thành công	170
Hình 201: Chọn Show Measures Grid	171
Hình 202: Chi tiết các độ đo	171
Hình 203: Thêm các Measures tương ứng	172
Hình 204: Nhấn Ok	172
Hình 205: Các Measures	173
Hình 206: Thông báo xuất hiện yêu cầu ta phải có một time dimension.....	173
Hình 207: Kiểu bảng từ Regular sang Time.	174
Hình 208: Các độ đo	175
Hình 209: Kéo những thuộc tính cần phân cấp qua cửa sổ Hierarchies	175
Hình 210: Phân cấp theo thứ tự từ trái sang phải.....	175
Hình 211: Chọn Rigid	176
Hình 212: Chọn KeyColumns.....	177
Hình 213: Thêm các thuộc tính cấp cao hơn vào KeyColumns.....	178
Hình 214: Chọn Name Column và chọn tên thuộc tính là Start Time Minute.....	179
Hình 215: Chọn KeyColumns.....	180
Hình 216: Thêm các những thuộc tính cấp cao hơn vào.....	181
Hình 217: Chọn Name Column và chọn tên thuộc tính là Start Time Hour.....	182
Hình 218: Chọn KeyColumns.....	183
Hình 219: Thêm các những thuộc tính cấp cao hơn vào.....	184
Hình 220: Chọn Name Column và chọn tên thuộc tính là Start Time Day	185
Hình 221: Deploy thành công	185
Hình 222: Kéo thả thuộc tính và độ đo	186
Hình 223: Thực thi câu lệnh và được kết quả	187
Hình 224: Kết quả Pivot table trong Excel	188
Hình 225: Kết quả	189
Hình 226: Tạo NameSet.....	189
Hình 227: Kéo thuộc tính và độ đo vào cửa sổ thực thi.....	190
Hình 228: Được kết quả như hình.....	190
Hình 229: Kết quả trong Excel	191
Hình 230: Kết quả MDX.....	192
Hình 231: Tạo NameSet.....	192
Hình 232: Kéo vào thuộc tính	193
Hình 233: Được kết quả như hình.....	193

Hình 234: Kết quả bằng Excel	194
Hình 235: Kết quả bằng MDX.....	194
Hình 236: Kéo thuộc tính.....	195
Hình 237: Kết quả trên Cube	195
Hình 238: Kết quả trên Excel.....	196
Hình 239: Kết quả bằng MDX.....	197
Hình 240: Két thả thuộc tính.....	197
Hình 241: Kết quả trên Cube	198
Hình 242: Kết quả bằng Excel.....	198
Hình 243: Kết quả bằng MDX.....	198
Hình 244: Chọn thuộc tính.....	199
Hình 245: Kết quả bằng Cube.....	200
Hình 246: Sử dụng Filter	201
Hình 247: Kết quả bằng Excel	202
Hình 248: Kết quả bằng MDX.....	203
Hình 249: Chọn thuộc tính.....	204
Hình 250: Kết quả bằng Cube.....	204
Hình 251: Chọn các năm tương ứng	205
Hình 252: Kết quả bằng Excel	205
Hình 253: Kết quả bằng MDX.....	206
Hình 254: Chọn thuộc tính và độ đo	207
Hình 255: Kết quả bằng Cube.....	207
Hình 256: Chọn 3 và 4	208
Hình 257: Kết quả bằng Excel	208
Hình 258: : Kết quả bằng MDX.....	210
Hình 259: Chọn NameSet với biểu thức như hình.....	211
Hình 260: Chọn thuộc tính và độ đo	211
Hình 261: Kết quả bằng Cube.....	211
Hình 262: Chọn Filter	212
Hình 263: Chọn Top 10 Filter	212
Hình 264: Kết quả bằng Cube.....	212
Hình 265: Kết quả bằng MDX.....	213
Hình 266: Filter Custom cho thuộc tính BUILDINGTYPE	214
Hình 267: Chọn thuộc tính.....	214
Hình 268: Kết quả Cube	215
Hình 269: Sử dụng Filter	215
Hình 270: Kết quả bằng Excel	216
Hình 271: Kết quả bằng MDX.....	217
Hình 272: Tạo NameSet.....	218
Hình 273: Kết quả trên Cube	218

Hình 274: Tạo Filter Top 5.....	218
Hình 275: Kết quả trên Excel.....	219
Hình 276: Kết quả trên MDX	219
Hình 277: Chọn thuộc tính và độ đo.....	220
Hình 278: Kết quả trên Cube	220
Hình 279: Tạo Filter between	221
Hình 280: Kết quả bằng Excel	221
Hình 281: Kết quả bằng MDX.....	222
Hình 282: Chọn thuộc tính và độ đo.....	223
Hình 283: Kết quả bằng Cube.....	223
Hình 284: Chọn Filter	224
Hình 285: Kết quả bằng Excel	224
Hình 286: Kết quả bằng MDX.....	225
Hình 287: Chọn thuộc tính và độ đo	226
Hình 288: Kết quả bằng Cube.....	226
Hình 289: Chọn trường dữ liệu và filter	227
Hình 290: Kết quả bằng Excel	227
Hình 291: Kết quả bằng MDX.....	228
Hình 292: Chọn các thuộc tính và độ đo.....	229
Hình 293: Kết quả bằng Cube.....	230
Hình 294: Chọn các trường dữ liệu và Filter	230
Hình 295: Kết quả bằng Excel	231
Hình 296: Kết quả bằng MDX.....	232
Hình 297: Chọn các thuộc tính và độ đo.....	233
Hình 298: Kết quả bằng Cube.....	233
Hình 299: Chọn các trường dữ liệu và Filter	234
Hình 300: Kết quả bằng Excel	234
Hình 301: Kết quả bằng MDX.....	235
Hình 302: Cài đặt công cụ Microsoft SQL Server 2022 Reporting Services	236
Hình 303: Chọn choose a free edition.....	237
Hình 304: Chọn I accept the license terms	238
Hình 305: Nhấn Install.....	238
Hình 306: Chọn Configure Report Server	239
Hình 307: Cấu hình Report Server Configuration Manager	240
Hình 308: Giao diện tổng thể.....	241
Hình 309: Thu được kết quả như sau	241
Hình 310: Chọn Change Database	242
Hình 311: Chọn Create a new report server database	242
Hình 312: Điền Server Name phù hợp.....	243
Hình 313: Đặt tên cho Database	243

Hình 314: Tiếp tục nhấn Next	244
Hình 315: Tiếp tục	244
Hình 316: Đang tiến hành tạo Database	245
Hình 317: Tạo Database thành công	245
Hình 318: Sửa tên Virtual Directory thành tên ReportServer1	246
Hình 319: Chọn Apply	247
Hình 320: Giao diện	248
Hình 321: Add Data Source	248
Hình 322: Giao diện Data Source Properties	249
Hình 323: Chọn Use a connection embedded in my report	250
Hình 324: Kết nối với Database	251
Hình 325: Kết nối thành công	251
Hình 326: Kiểm tra lại	251
Hình 327: Tạo Report Server Project	252
Hình 328: Đặt tên Project	253
Hình 329: Chọn New Item	253
Hình 330: Report trống được tạo	254
Hình 331: Kết nối với Cube dữ liệu SSAS	255
Hình 332: Nhập localhost vào Server name và nhập tên project SSAS	256
Hình 333: Chọn Use Windows Authentication	257
Hình 334: Chọn Use a dataset embedded in my report	258
Hình 335: Chọn Query	259
Hình 336: Chọn MDX	259
Hình 337: Kết quả Query	260
Hình 338: SSRS Toolbox	261
Hình 339: Kéo vào một Table mới	261
Hình 340: Kéo lần lượt các trường vào bảng	262
Hình 341: Xem báo cáo	263
Hình 342: Hiển thị báo cáo giống với trang in	263
Hình 343: Chọn FontWeight Bold	265
Hình 344: Sau khi định dạng lại	266
Hình 345: Print Layout của Report	267
Hình 346: Giao diện trình duyệt	268
Hình 347: Thư mục SSRS_HOUSE_PRICE_BEIJING	268
Hình 348: Trans_DistrictRoom	269
Hình 349: Report mới	270
Hình 350: Các Datasets	271
Hình 351: Report thô	271
Hình 352: Sau khi được định dạng lại	272
Hình 353: Đặt trường trên bên trên vị trí của trường mặc định (1)	273

Hình 354: Đặt trường trên bên trên vị trí của trường mặc định (2)	273
Hình 355: Kết quả sau khi group	274
Hình 356: Thêm các trường cần thiết	275
Hình 357: Sau khi được thêm Background Color.....	275
Hình 358: Kết quả sau khi thêm Total (1).....	276
Hình 359: Kết quả sau khi thêm Total (2).....	277
Hình 360: Kết quả trên trình duyệt	278
Hình 361: Chọn thuộc tính cần thiết.....	279
Hình 362: Insert Table.....	279
Hình 363: Các thuộc tính được kéo vào	280
Hình 364: Chart Wizard	280
Hình 365: Chọn lược đồ dạng Column.....	281
Hình 366: Tạo Chart mới	282
Hình 367: Biểu đồ.....	283
Hình 368: Báo cáo sau khi định dạng	284
Hình 369: Report trong PrintLayout (1)	285
Hình 370: Report trong PrintLayout (2)	286
Hình 371: Report trong PrintLayout (3)	287
Hình 372: Xuất file pdf	288
Hình 373: Deploy lên Report Server	288
Hình 374: In report 2 trên Report Server	289
Hình 375: Add Dataset.....	290
Hình 376: Chọn Query	290
Hình 377: Chọn những thuộc tính cần thiết cho báo biểu	291
Hình 378: Kéo những thuộc tính vào bảng table như hình	291
Hình 379: Tạo Grouping, Add Group để gom nhóm dữ liệu.....	292
Hình 380: Add Total để tính tổng các giá trị	293
Hình 381: Thêm Title	293
Hình 382: Chọn lược đồ các dạng hình tròn, cột và đường	294
Hình 383: Kéo thả độ đo vào ô thích hợp	295
Hình 384: Biểu đồ thu được.....	295
Hình 385: Chọn Yes khi gấp thông báo	296
Hình 386: Hiển thị được thay đổi	297
Hình 387: Tạo biểu đồ đường	297
Hình 388: Chọn độ đo thích hợp.....	298
Hình 389: Kết quả trả về	299
Hình 390: Kết quả Report	300
Hình 391: Kết quả trong PrintLayout	301
Hình 392: Trang 2 của kết quả.....	302
Hình 393: Xuất file PDF	303

Hình 394: Đặt tên và chọn đường dẫn	303
Hình 395: In report 3 trên Report Server	304
Hình 396: Bấm Install	305
Hình 397: Bấm nút Sign in	305
Hình 398: Giao diện Power BI	306
Hình 399: Nhập tên server Sql và tên Database tương ứng	306
Hình 400: Chọn các bảng Fact và Dim để import	307
Hình 401: Giao diện Model View	308
Hình 402: Add Textbox	308
Hình 403: Chọn Clustered column chart và kéo thuộc tính tương ứng	309
Hình 404: Thêm điều kiện	310
Hình 405: Kết quả biểu đồ	311
Hình 406: Kết quả sau khi thêm Visualizations Area Chart và Multi-row Card	311
Hình 407: Hiển thị chi tiết hơn	312
Hình 408: Publish thành công	312
Hình 409: Tạo Textbox	313
Hình 410: Tạo mới 1 Matrix	314
Hình 411: Sử dụng filter	315
Hình 412: Kết quả sau khi dùng filter	315
Hình 413: Kết quả sau khi thêm Visualizations Donut Chart và Stacked Bar Chart	316
Hình 414: Publish thành công	316
Hình 415: Tạo Textbox	317
Hình 416: Tạo mới Matrix và thêm hàng cột tương ứng	318
Hình 417: Sử dụng filter	319
Hình 418: Kết quả sau khi áp Filter	320
Hình 419: Thêm Visualizations Donut Chart, Clustered Bar Chart và Stacked Bar Chart	320
Hình 420: Publish thành công	321
Hình 421: Giao diện web	322
Hình 422: Nhấn Open report	322
Hình 423: Đang xem lại Báo biểu 3	323
Hình 424: Chọn New Mining Structure	324
Hình 425: Chọn Next	325
Hình 426: Tích From existing relational or data warehouse	326
Hình 427: Chọn thuật toán Decision Trees	327
Hình 428: Chọn Data source	328
Hình 429: Chọn DIM_HOUSE làm case để chạy thuật toán mining	329
Hình 430: Chọn thuộc tính key và thuộc tính Predict	330
Hình 431: Xem các thuộc tính được suggest	331
Hình 432: Chọn các thuộc tính để làm input	332
Hình 433: Detect các trường dữ liệu	333

Hình 434: Chọn các thông số thích hợp.....	334
Hình 435: Nhấn Finish.....	335
Hình 436: Thêm Mining Model	336
Hình 437: Thêm thuật toán Microsoft Clustering	336
Hình 438: Sau khi thêm thành công.....	337
Hình 439: Thêm thuật toán Microsoft Naive Bayes	337
Hình 440: Sau khi thêm thành công.....	338
Hình 441: Tiến hành Mining.....	339
Hình 442: Xem Mining Model Viewer của thuật toán Microsoft Decision Tree với thuộc tính dự đoán là ELEVATOR.....	340
Hình 443: Tỷ lệ cao 76.02% là những nhà có 2 phòng tắm và 2 phòng khách	340
Hình 444: Tỷ lệ cao 75.96% là những nhà có 1 phòng tắm, có số phòng vă lớn hơn hoặc bằng 2 và có ít hơn 2 phòng khách	341
Hình 445: Tỷ lệ cao 98.61% là những nhà có số phòng tắm lớn hơn hoặc bằng 3, có số phòng vă bé hơn 2 và có 2 phòng khách.....	341
Hình 446: Pependency Network của thuật toán (1)	342
Hình 447: Pependency Network của thuật toán (2)	342
Hình 448: Pependency Network của thuật toán (3)	343
Hình 449: Pependency Network của thuật toán (4)	343
Hình 450: Cluster Diagram của thuật toán Shading Variable là ELEVATOR	344
Hình 451: Tập lục của những nhà có thang máy chiếm 76.1%	344
Hình 452: Tập lục của những nhà không có thang máy chiếm 61.4%	345
Hình 453: Cluster Profiles của Cluster Low của tập lục trên (1).....	346
Hình 454: Cluster Profiles của Cluster Low của tập lục trên (2).....	346
Hình 455: Dependency Network của thuật toán Microsoft Naive Bayes (1)	347
Hình 456: Dependency Network của thuật toán Microsoft Naive Bayes (2)	347
Hình 457: Dependency Network của thuật toán Microsoft Naive Bayes (3)	348
Hình 458: Dependency Network của thuật toán Microsoft Naive Bayes (4)	348
Hình 459: Nhà có số phòng tắm từ 1-2 khả năng cao sẽ không có thang máy	349
Hình 460: Nhà có số phòng tắm từ 2-3 khả năng cao sẽ có thang máy	349
Hình 461: Nhà có số phòng vă lớn hơn hoặc bằng 2 khả năng cao sẽ có thang máy	350
Hình 462: Thiết lập thông số Input Selection với các nhà có thang máy	350
Hình 463: Tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà có thang máy.....	351
Hình 464: Thiết lập Input Selection với các nhà không có thang máy	351
Hình 465: Tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà không có thang máy...	352

DANH MỤC BẢNG

Bảng 1: Bảng thuộc tính của kho dữ liệu.....	22
Bảng 2: Thuộc tính dữ liệu đã được xử lý	23
Bảng 3: DIM_LOCATION	24
Bảng 4: DIM_BUILDING.....	25
Bảng 5: DIM_HOUSE.....	25
Bảng 6: DIM_DATE.....	25
Bảng 7: FACT_TRANSACTION.....	26

CHƯƠNG I: GIỚI THIỆU VỀ TỔNG QUAN DỮ LIỆU

1. Phát biểu về dữ liệu

1.1. Mô tả về dữ liệu

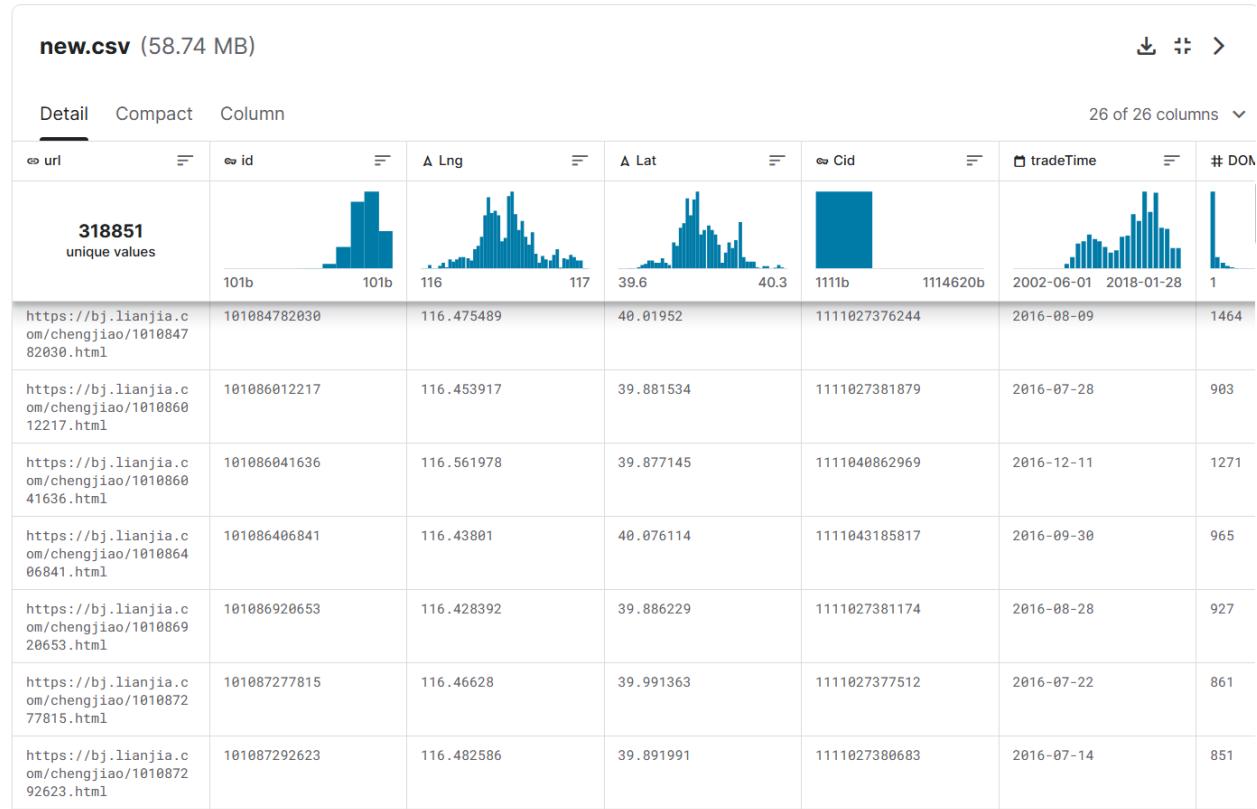
- Kho dữ liệu “House price in Beijing” là một kho dữ liệu được thu thập từ năm 2011 đến 2017 về giá nhà ở Beijing, dữ liệu được lấy từ trang Lianjia.com
- Thông qua dữ liệu người dùng có thể biết được định danh của giao dịch, url, tọa độ Lat(giao thức BD09), định danh cộng đồng, thời gian giao dịch, số ngày hoạt động trên thị trường, số người theo dõi giao dịch, tổng giá, giá theo mét vuông, số lượng các phòng có trong nhà, ...
- Kho dữ liệu gồm 318851 dòng và 26 thuộc tính.
- Link gốc: [Housing price in Beijing \(kaggle.com\)](https://www.kaggle.com/c/house-prices-advanced-regression-techniques)

1.2. Thuộc tính của kho dữ liệu

STT	Tên thuộc tính	Kiểu	Ý nghĩa của thuộc tính
1	url	String	Url lấy dữ liệu
2	id	String	Định danh của giao dịch
3	Lng	Float	Kinh độ (sử dụng giao thức BD09)
4	Lat	Float	Vĩ độ (sử dụng giao thức BD09)
5	Cid	BigInt	Định danh cộng đồng
6	tradeTime	Date	Thời gian giao dịch
7	DOM	Int	Số ngày hoạt động trên thị trường
8	followers	Int	Số người theo dõi
9	totalPrice	Float	Tổng giá
10	price	Float	Giá trung bình trên từng mét vuông
11	square	Float	Diện tích nhà
12	livingRoom	Int	Số lượng phòng khách
13	drawingRoom	Int	Số lượng phòng vẽ
14	kitchen	Int	Số lượng nhà bếp
15	bathroom	Int	Số lượng nhà tắm
16	floor	Int	Số tầng
17	buildingType	Int	Loại công trình bao gồm (1) tháp, (2) nhà cấp bốn, (3) kết hợp tắm và tháp, (4) tắm
18	constructionTime	Int	Năm xây dựng
19	renovationCondition	Int	Điều kiện cải tạo bao gồm (1) khác, (2) thô, (3) đơn giản, (4) cứng cáp
20	buildingStructure	Int	Kết cấu xây dựng bao gồm (1) chua biệt, (2) hỗn hợp, (3) gạch và gỗ, (4)

			gạch và bê tông, (5) thép, (6) thép và bê tông
21	ladderRatio	Float	Tỉ lệ thang máy, đây là tỷ lệ giữa số cư dân trên cùng một tầng và số thang máy hoặc thang cuốn. Nó mô tả bao nhiêu thang mỗi cư dân trung bình có trên một tòa nhà.
22	elevator	Boolean	Cho biết tòa nhà có thang máy hay không
23	fiveYearsProperty	Boolean	Cho biết nếu chủ sở hữu sở hữu tài sản dưới 5 năm
24	Subway	Boolean	Cho biết có gầm tàu điện ngầm hay không
25	District	Int	Năm ở khu vực nào
26	communityAverage	Int	Giá trung bình trên mét vuông cộng đồng gần đó

Bảng 1: Bảng thuộc tính của kho dữ liệu



Hình 1: Bảng chi tiết có trong file csv

1.3. Kho dữ liệu đã xử lý

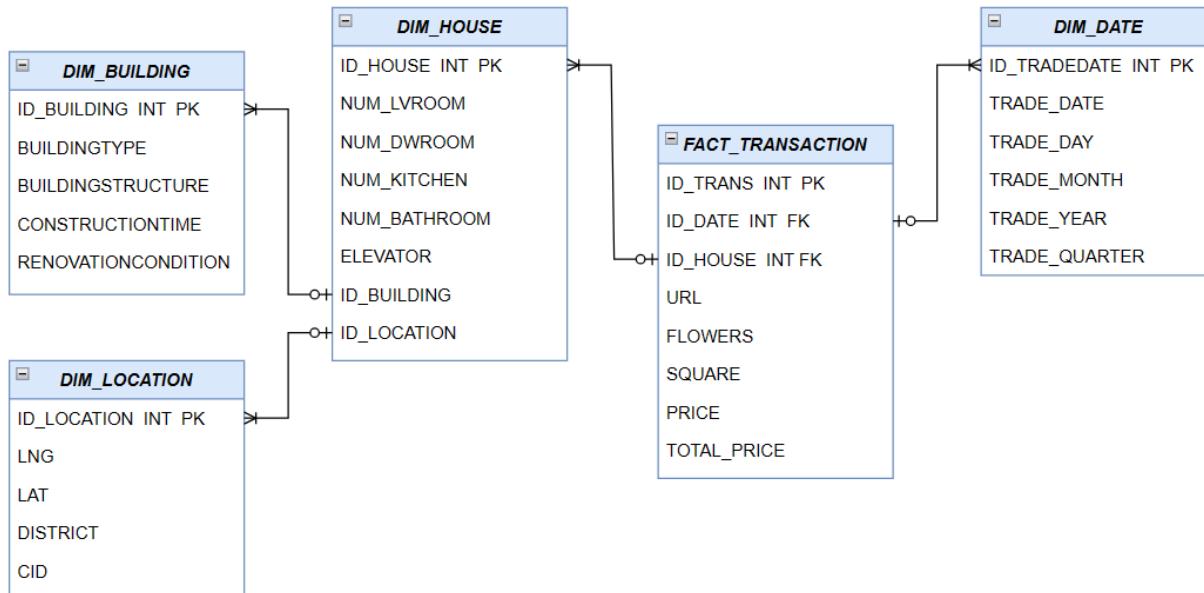
Sau khi lọc dữ liệu ta được 318851 và 20 thuộc tính để sử dụng việc phân tích cho đề tài.

STT	Tên thuộc tính	Kiểu	Ý nghĩa của thuộc tính
1	url	String	Url lấy dữ liệu
2	Id	String	Định danh của giao dịch
3	Lng	Float	Kinh độ (sử dụng giao thức BD09)
4	Lat	Float	Vĩ độ (sử dụng giao thức BD09)
5	Cid	BigInt	Định danh cộng đồng
6	tradeTime	Date	Thời gian giao dịch
7	DOM	Int	Số ngày hoạt động trên thị trường
8	followers	Int	Số người theo dõi
9	totalPrice	Float	Tổng giá
10	Price	Float	Giá trung bình trên từng mét vuông
11	square	Float	Diện tích nhà
12	livingRoom	Int	Số lượng phòng khách
13	drawingRoom	Int	Số lượng phòng vẽ
14	kitchen	Int	Số lượng nhà bếp
15	bathroom	Int	Số lượng nhà tắm
16	buildingType	Int	Loại công trình bao gồm (1) tháp, (2) nhà cấp bốn, (3) kết hợp tâm và tháp, (4) tâm
17	constructionTime	Int	Năm xây dựng
18	buildingStructure	Int	Kết cấu xây dựng bao gồm (1) chưa biệt, (2) hỗn hợp, (3) gạch và gỗ, (4) gạch và bê tông, (5) thép, (6) thép và bê tông
19	elevator	Boolean	Cho biết tòa nhà có thang máy hay không
20	District	Boolean	Năm ở khu vực nào
21	renovationCondition	Int	Điều kiện cải tạo bao gồm (1) khác, (2) thô, (3) đơn giản, (4) cứng cáp

Bảng 2: Thuộc tính dữ liệu đã được xử lý

2. XÂY DỰNG KHO DỮ LIỆU

2.1. Sơ đồ bông tuyết minh họa



Hình 2: Sơ đồ bông tuyết minh họa

2.2. DIM_LOCATION

Gồm có 318851 dòng dữ liệu, 5 thuộc tính chứa tất cả các địa điểm ..., có khoá chính là ID_LOCATION

Khoá	Tên thuộc tính	NULL	Kiểu dữ liệu	Mô tả thuộc tính
PK	ID_LOCATION		Int	Mã địa điểm
	LNG		Float	Kinh độ của nhà
	LAT		Float	Vĩ độ của nhà
	DISTRICT		Int	Nằm ở khu vực nào
	CID		BigInt	Định danh cộng đồng khu vực đó

Bảng 3: DIM_LOCATION

2.3. DIM_BUILDING

Gồm 5 thuộc tính chứa thông tin loại công trình, kết cấu xây dựng của nhà

Khoá	Tên thuộc tính	NULL	Kiểu dữ liệu	Mô tả thuộc tính
PK	ID_BUILDING		Int	Mã thiết kế
	BUILDINGTYPE		Int	Loại công trình

	BUILDINGSTRUCTURE		Int	Kết cấu xây dựng
	CONSTRUCTIONTIME		Int	Thời gian xây dựng
	RENOVATIONCONDITION		Int	Điều kiện cải tạo bao gồm (1) khác, (2) thô, (3) đơn giản, (4) cứng cáp

Bảng 4: DIM_BUILDING

2.4. DIM_HOUSE

Gồm 8 thuộc tính chứa thông tin của nhà

Khoá	Tên thuộc tính	NULL	Kiểu dữ liệu	Mô tả thuộc tính
PK	ID HOUSE		Int	Mã nhà
	NUM LVROOM		Int	Số phòng khách
	NUM DWROOM		Int	Số phòng vệ
	NUM KITCHEN		Int	Số phòng bếp
	NUM BATHROOM		Int	Số phòng tắm
	ELEVATOR		Bool	Có thang máy hay không
	ID BUILDING		Int	Mã thiết kế
	ID LOCATION		Int	Mã địa điểm

Bảng 5: DIM_HOUSE

2.5. DIM_DATE

Gồm 6 thuộc tính chứa thông tin ngày giao dịch

Khoá	Tên thuộc tính	NULL	Kiểu dữ liệu	Mô tả thuộc tính
PK	ID TRADEDATE		Int	Mã ngày
	TRADE DATE		Date	Thời gian giao dịch
	TRADE DAY		Int	Ngày giao dịch
	TRADE MONTH		Int	Tháng giao dịch
	TRADE YEAR		Int	Năm giao dịch
	TRADE QUARTER		Int	Quý giao dịch

Bảng 6: DIM_DATE

2.6. FACT_TRANSACTION

Gồm 8 thuộc tính, là bảng sự kiện của kho dữ liệu

Khoá	Tên thuộc tính	NULL	Kiểu dữ liệu	Mô tả thuộc tính
PK	ID TRANS		String	Mã giao dịch
FK	ID DATE		Int	Mã ngày giao dịch

FK	ID HOUSE		Int	Mã nhà
	URL		String	Url lấy dữ liệu
	FLOWERS		Int	Số người theo dõi
	SQUARE		Float	Diện tích nhà
	PRICE		Float	Giá nhà trên một m ²
	TOTAL PRICE		Float	Tổng giá nhà

Bảng 7: FACT_TRANSACTION

3. NỘI DUNG 16 CÂU HỎI TRUY VẤN

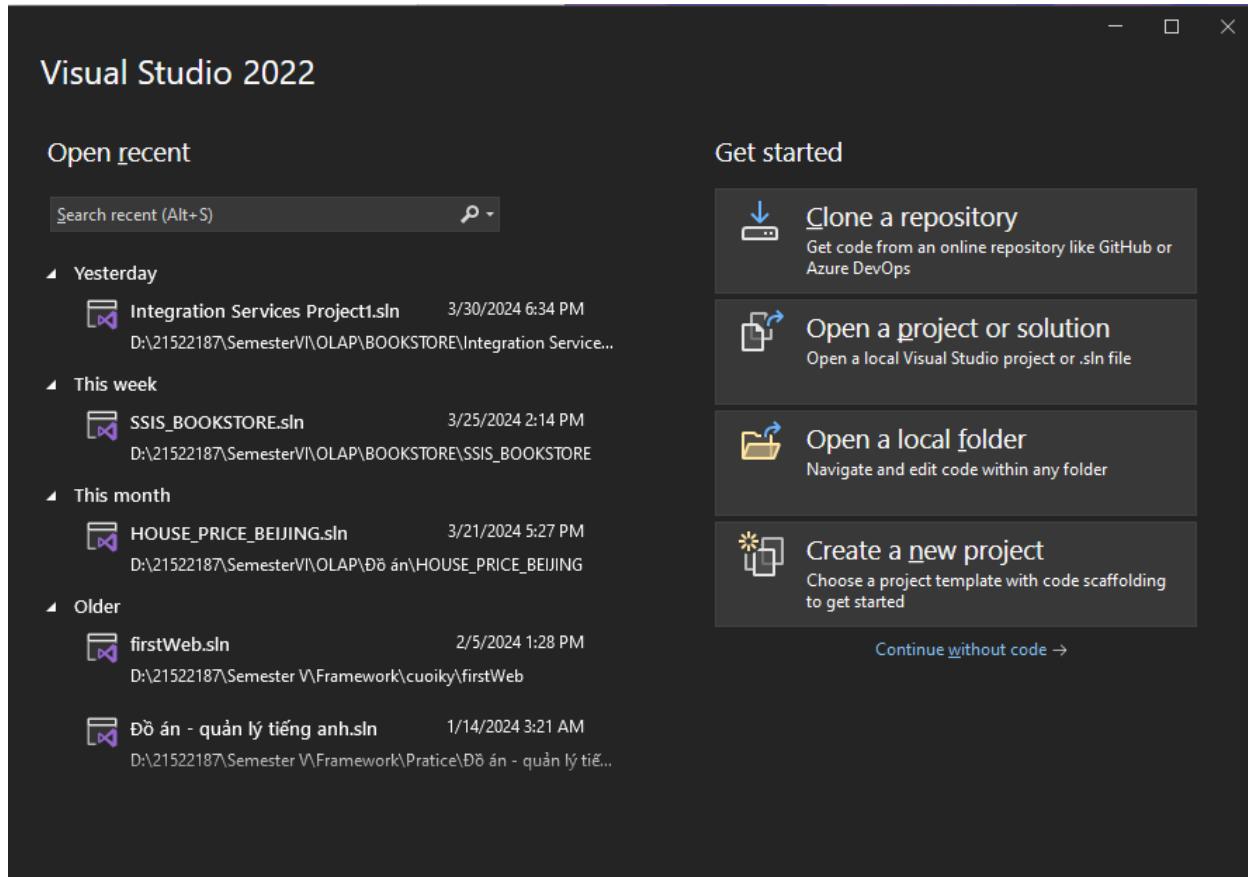
- a) Với mỗi khu vực (**DISTRICT**) đưa ra nhà có diện tích lớn nhất.
- b) Liệt kê có bao nhiêu giao dịch với có giá trên một m² lớn hơn 100000 theo từng năm (Sử dụng bảng phân cấp [**TRANS TIME**], roll up từ **DAY** lên **YEAR**)
- c) Liệt kê **TOP 3** giao dịch (**ID_TRANS**, **ID_DATE**, **ID_HOUSE**) có tổng giá thấp nhất.
- d) Thống kê vào quý 3 theo các năm để có bao nhiêu nhà có số phòng khách (**NUM_LVROOM**) lớn hơn hoặc bằng 2.
- e) Trong năm 2016, cho biết trung bình số người theo dõi (**FOLLOWERS**), của nhà là loại công trình là nhà cấp bốn (**BUILDINGTYPE=2**) (roll up **TRANS TIME** từ **TRADE DAY** sang **TRADE YEAR**)
- f) Cho biết tổng số tiền các nhà có số phòng vẽ bằng 2 (**DRAWINGROOM**) theo từng tháng, quý, năm.
- g) Cho biết khu vực nào có giá tiền trên một m² (**PRICE**) cao nhất từ năm 2016 đến năm 2018.
- h) Với kết cấu xây dựng là gạch và bê tông (**BUILDINGSTRUCTURE=4**) hoặc gạch và gỗ (**BUILDINGSTRUCTURE=3**), cho biết trung bình của tổng giá (**TOTALPRICE**) cho mỗi nhà là bao nhiêu, theo từng năm, tháng.
- i) Cho biết trong quý một năm 2014 (**QUARTER=1**), TOP 2 khu vực nào bán được ít nhà nhất và bán được bao nhiêu.
- j) Thống kê theo năm, quý, tháng, trung bình tổng tiền giao dịch của nhà có loại (**BUILDINGTYPE**) nhiều nhất tại từng thời điểm.
- k) Top 5 cộng đồng (**CID**) có tổng số lượng giao dịch thấp nhất, và cho biết tổng tiền giao dịch của các cộng đồng đó.
- l) Kiến trúc hiện đại là phong cách thống trị giới kiến trúc phương Tây từ những năm 1933 đến những năm 1960, cho biết trung bình giá, trung bình diện tích của những ngôi nhà được xây dựng trong khoảng thời gian này (**CONSTRUCTIONTIME**) theo từng năm.

- m) Liệt kê số lượng và tổng số tiền giao dịch đạt chuẩn chất lượng cao qua từng quý của các năm. Được biết giao dịch đạt chuẩn chất lượng cao chỉ bàn giao những ngôi nhà có tình trạng cải tạo đạt mức “Hardcover” (**RENOVATIONCONDITION**), được làm bởi các vật liệu có chất lượng cao kết hợp thép và bê tông (**BUILDINGSTRUCTURE**)
- n) Cho biết số giao dịch, trung bình m², của những nhà có thang máy (**ELEVATOR**) giao dịch trong năm 2018, xếp theo thứ tự trung bình m² **giảm dần**.
- o) Thống kê cho biết giá trên một m² (**PRICE**) lớn nhất qua các năm của những nhà có thể loại công trình (**BUILDINGTYPE**) là nhà cấp 4 và có điều kiện cải tạo (**RENOVATIONCONDITION**) không phải là đơn giản, xếp theo thứ tự giá trên một m² (**PRICE**) **tăng dần**.
- p) Cho biết thống kê theo từng năm ứng với mỗi huyện (**DISTRICT**) có bao nhiêu đường link giao dịch (**URL**) có kết cấu xây dựng (**BUILDINGSTRUCTURE**) là thép

CHƯƠNG II: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

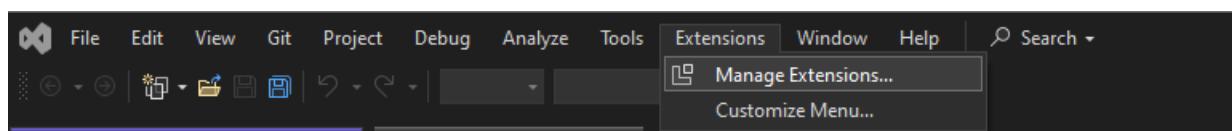
1. Tạo Project SSIS

- Mở **Visual Studio**, bấm nút **Esc** để về giao diện chính, lúc này giao diện chính rỗng.



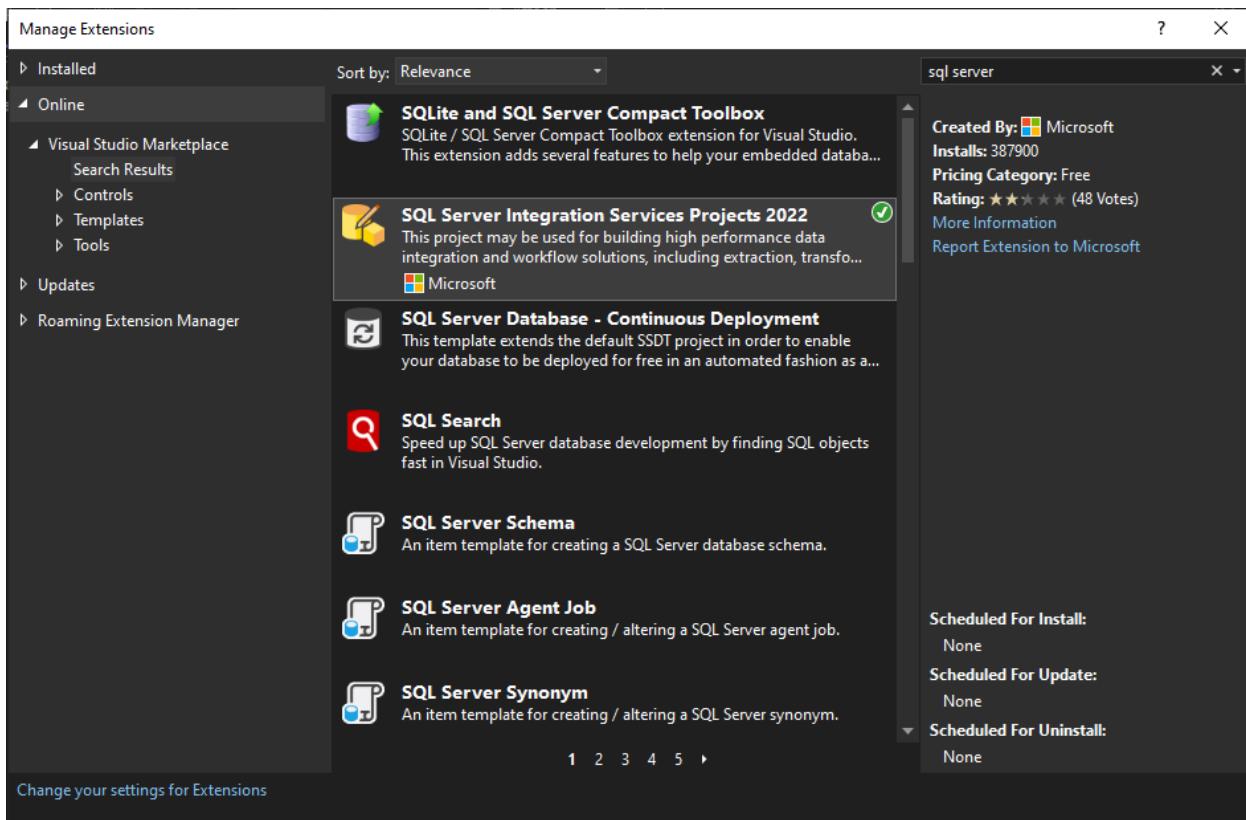
Hình 3: Giao diện sau khi mở Visual Studio

- Trong giao diện chính, click chọn “Extensions” -> “Manage Extensions”



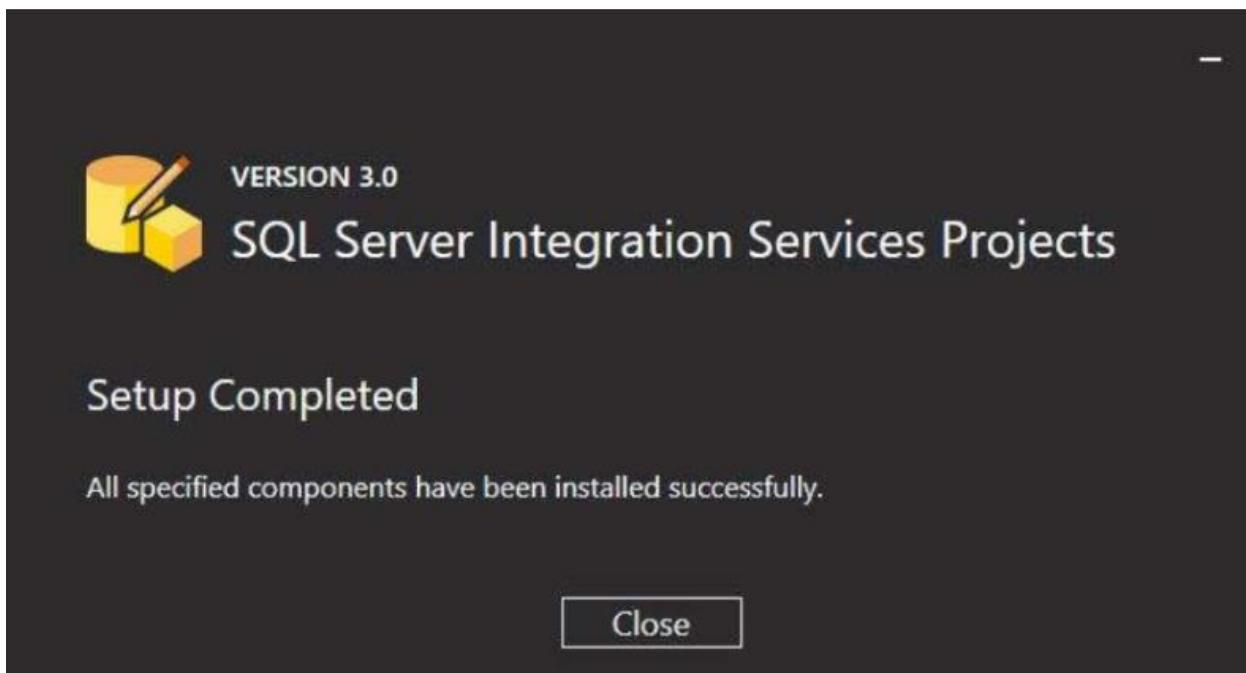
Hình 4: Chọn “Extensions” -> “Manage Extensions”

- Tìm và tải về công cụ “SQL Server Integration Services Projects”



Hình 5: SQL Server Intergration Services Projects

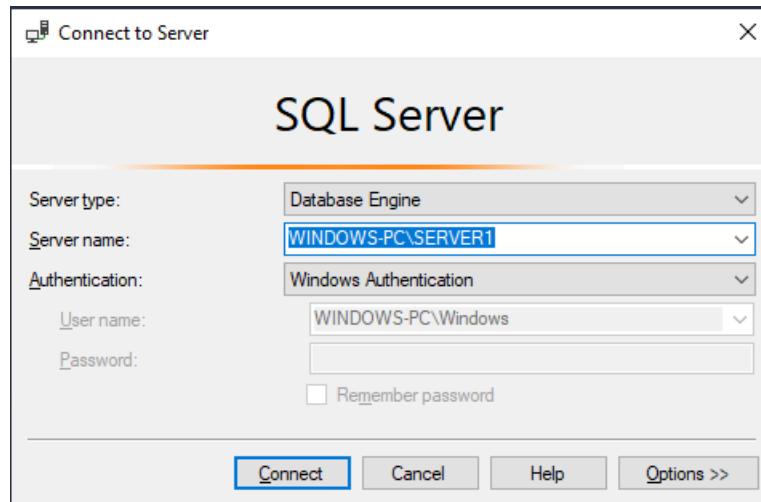
- Sau khi cài đặt thành công sẽ nhận được thông báo:



Hình 6: Thông báo cài SSIS thành công

2. Chuẩn bị cơ sở dữ liệu

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



Hình 7: Kết nối Database Engine

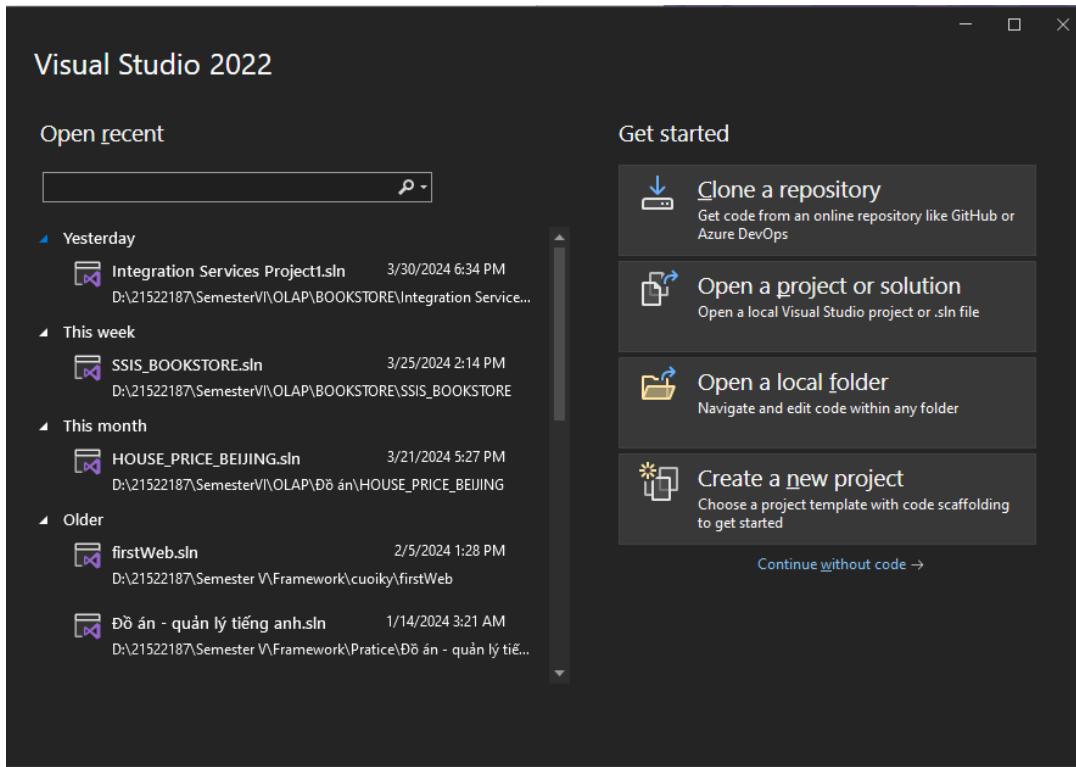
- Tạo cơ sở dữ liệu có tên **HOUSE_PRICE_BEIJING**, đây là nơi lưu các bảng **Dim** và bảng **Fact** cùng dữ liệu của các bảng đó

```
SQLQuery1.sql - W...S-PC\Windows (58)*
CREATE DATABASE HOUSE_PRICE_BEIJING
```

Hình 8: Tạo cơ sở dữ liệu HOUSE_PRICE_BEIJING

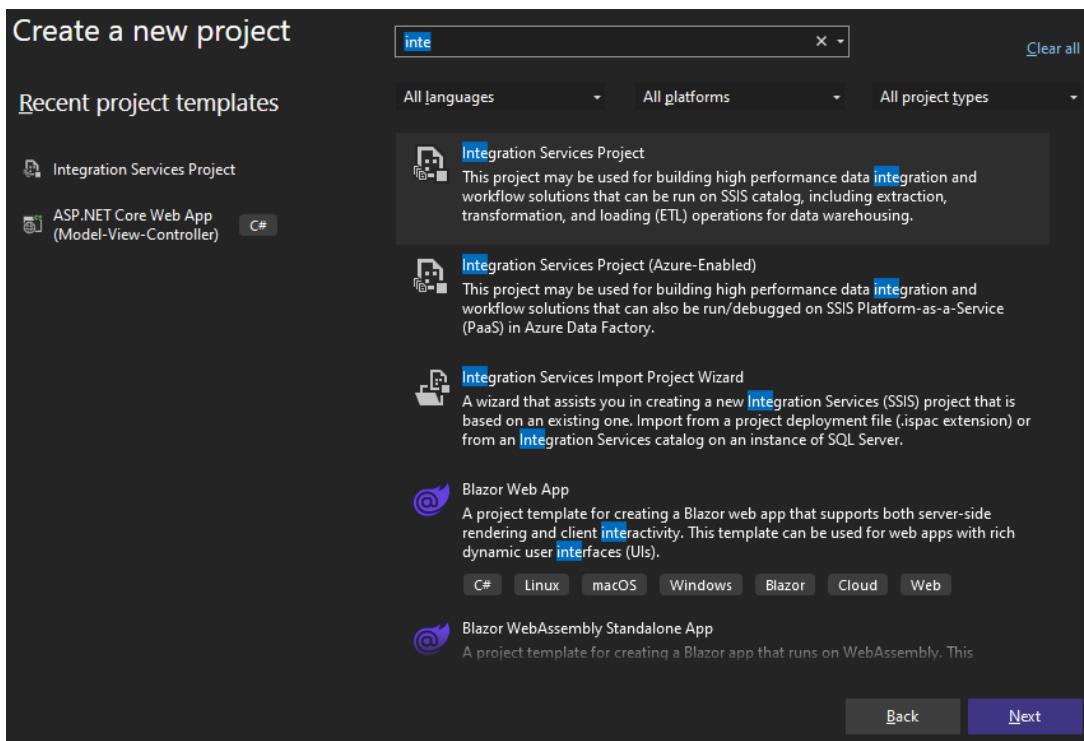
3. Tạo mới project SSIS

- Mở **Visual Studio** và chọn “Create a new project”.



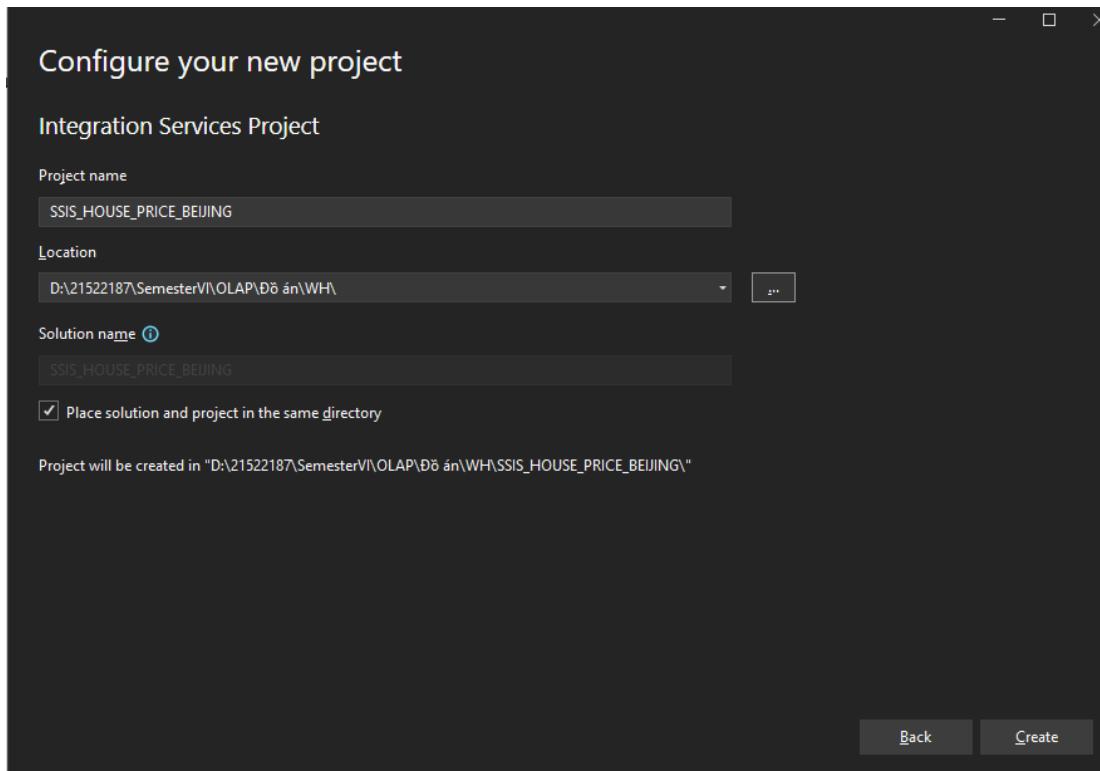
Hình 9: Create a new project

- Chọn **Integration Services Project** và chọn **Next**



Hình 10: Chọn Integration Services Project

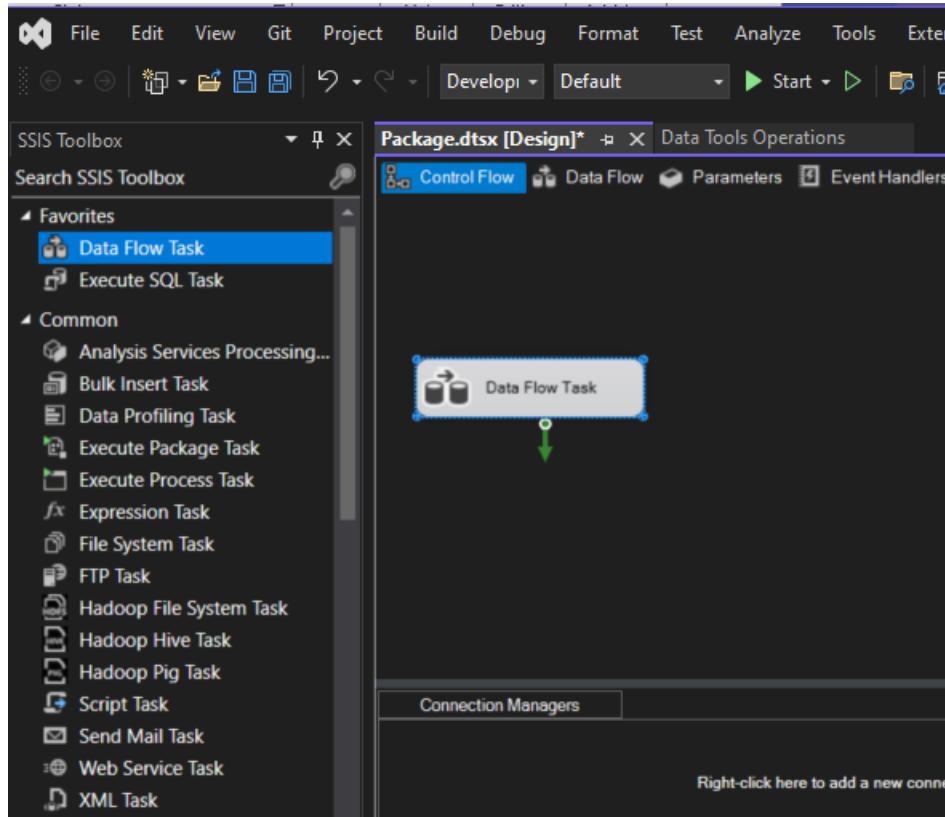
- Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn **Create**



Hình 11: Đặt tên và thiết lập đường dẫn

4. Tạo bảng Dim và bảng Fact

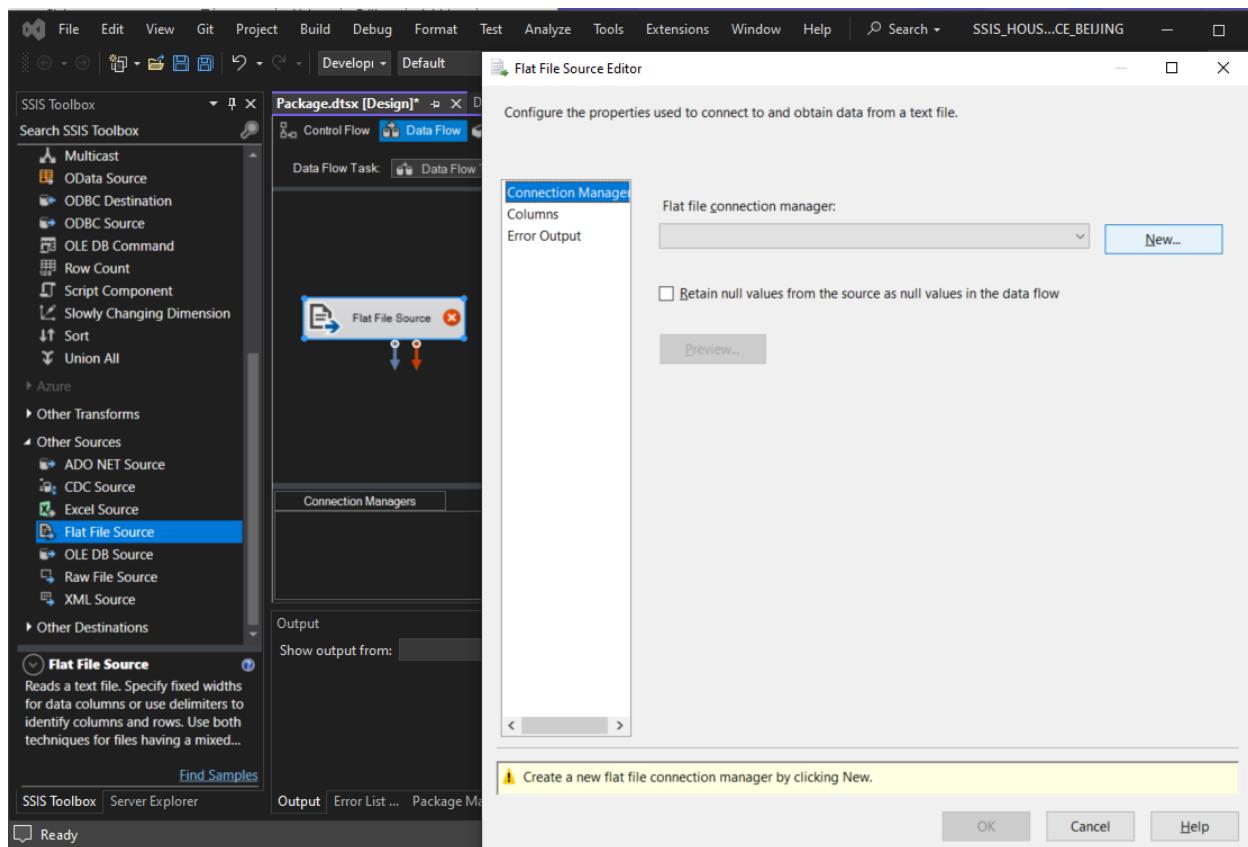
4.1. Chuẩn bị dữ liệu từ file .csv



Hình 12: Thêm Data Flow Task

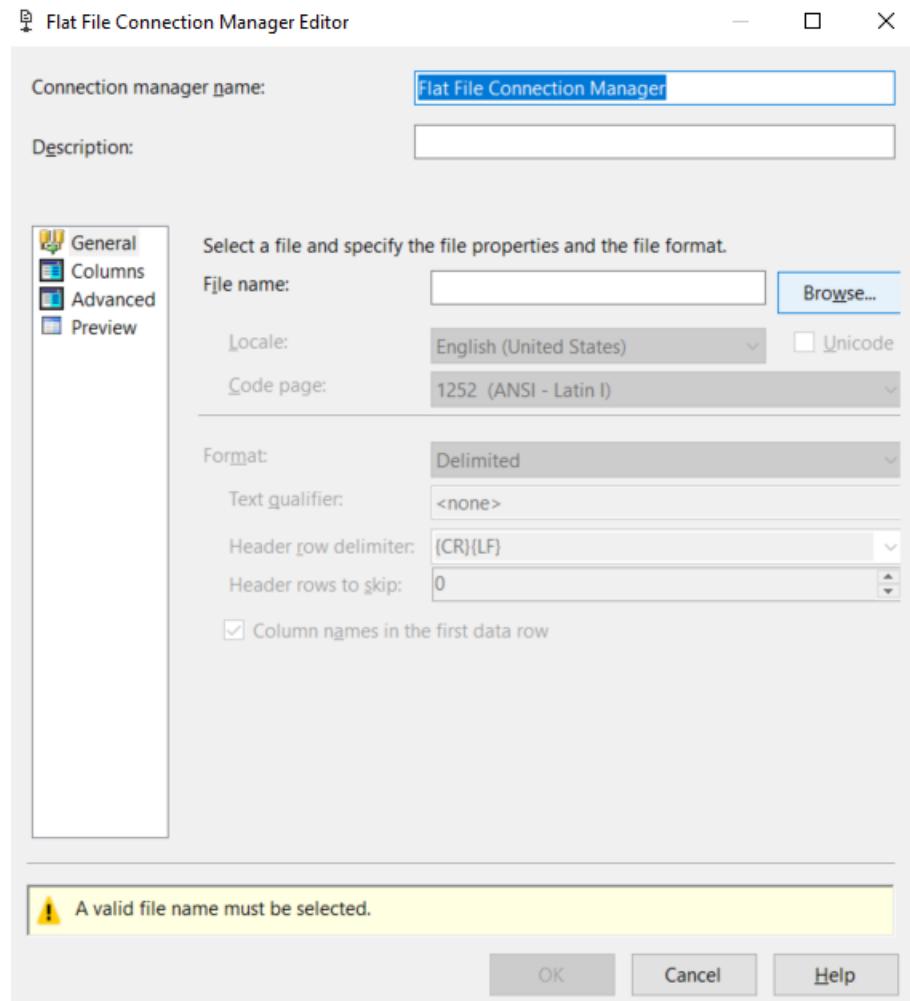
Trước khi tiến hành chia Dimension và bảng Fact, ta cần load dữ liệu gốc từ file .csv vào Data Flow:

- Trong **Data Flow**, tạo một đối tượng **Flat File Source** để lấy dữ liệu gốc từ file .csv. Chọn **New** để tạo một **Flat File Connection Manager**



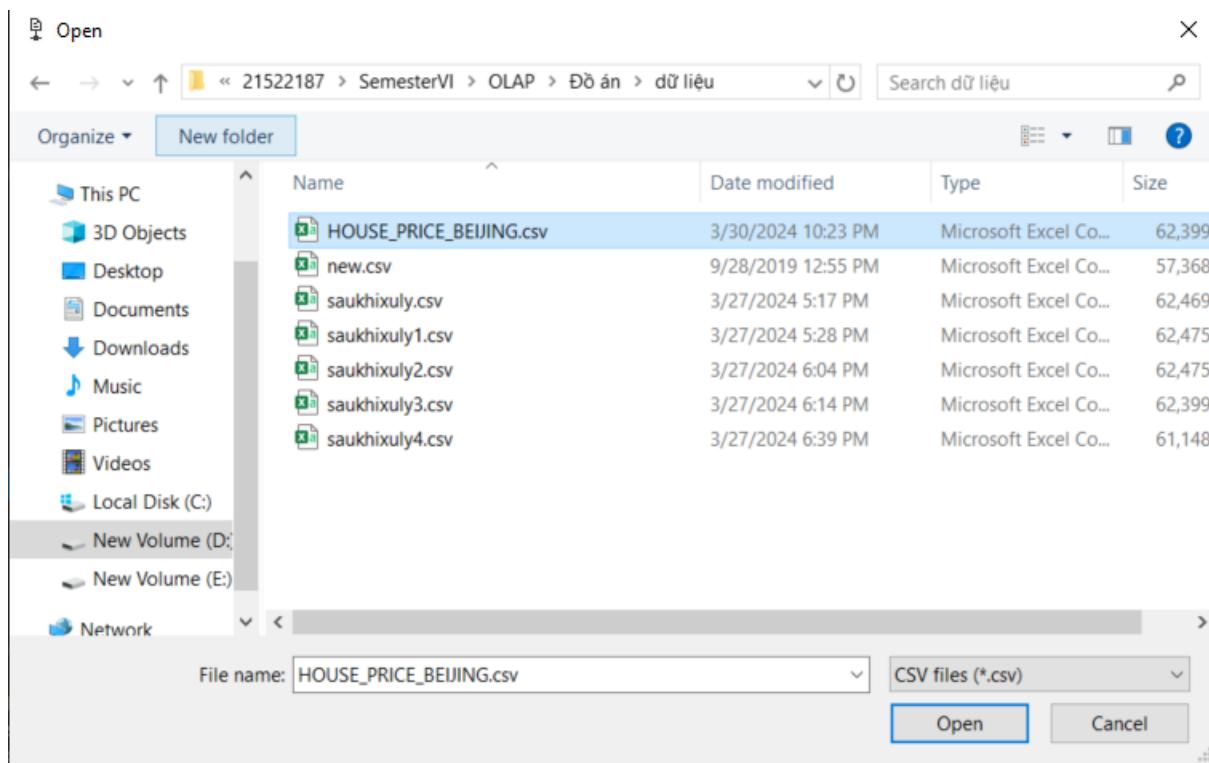
Hình 13: Chọn New để tạo một Flat File Connection Manager

- Chọn nút **Browse..** để tải lên file dữ liệu gốc lưu trong máy



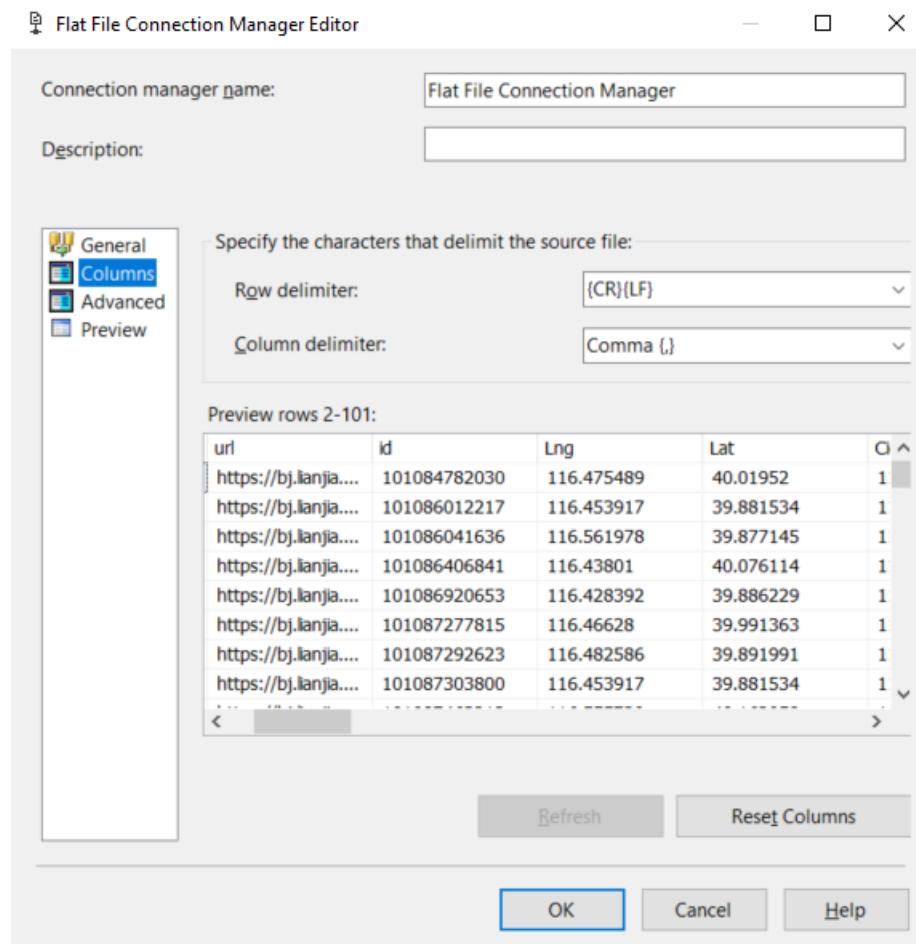
Hình 14: Chọn nút Browse..

- Chọn file dữ liệu .csv và chọn **Open**



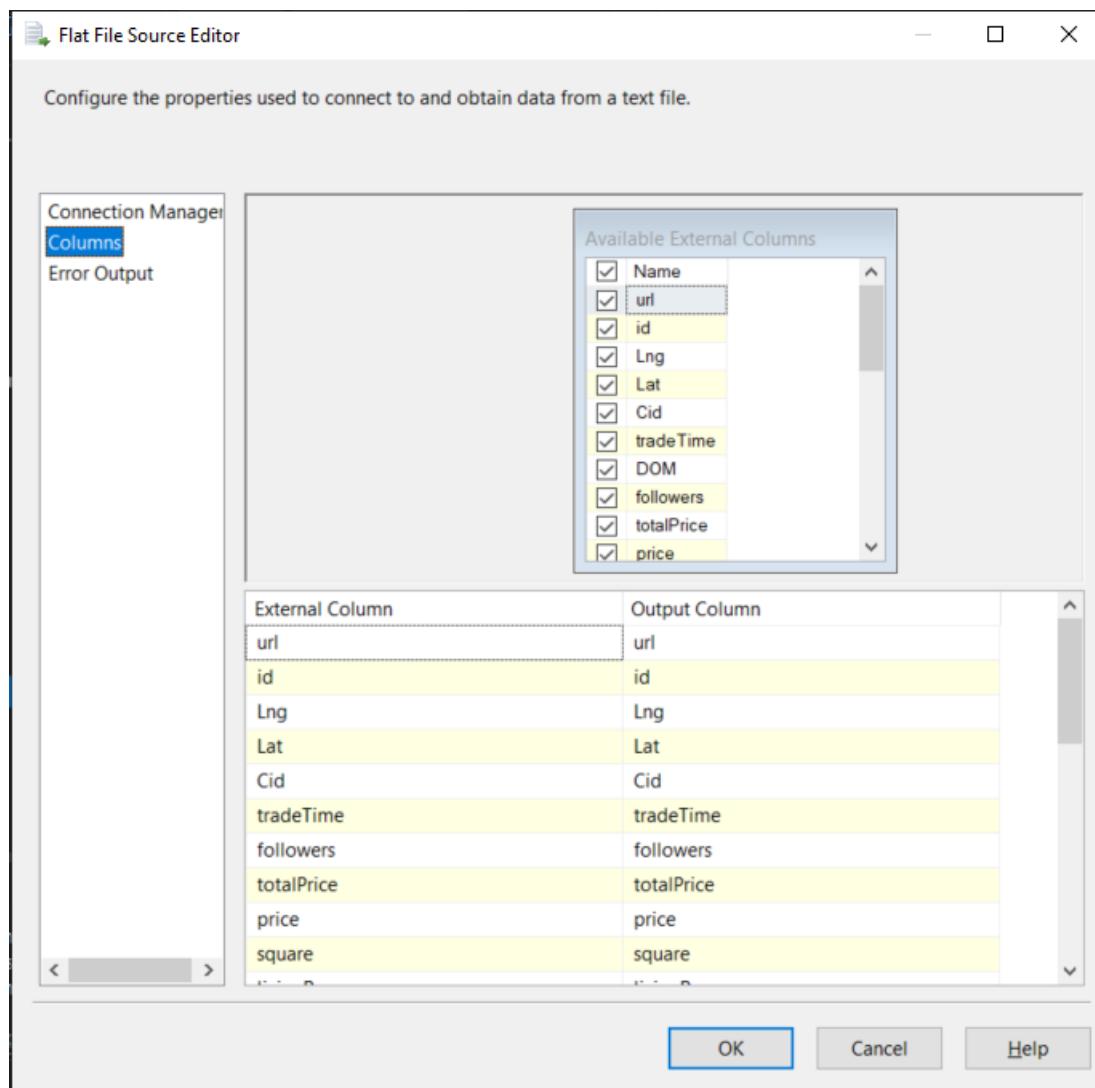
Hình 15: Chọn file dữ liệu .csv

- Xem lại các cột dữ liệu trong file dữ liệu đã được tải lên ở menu Columns



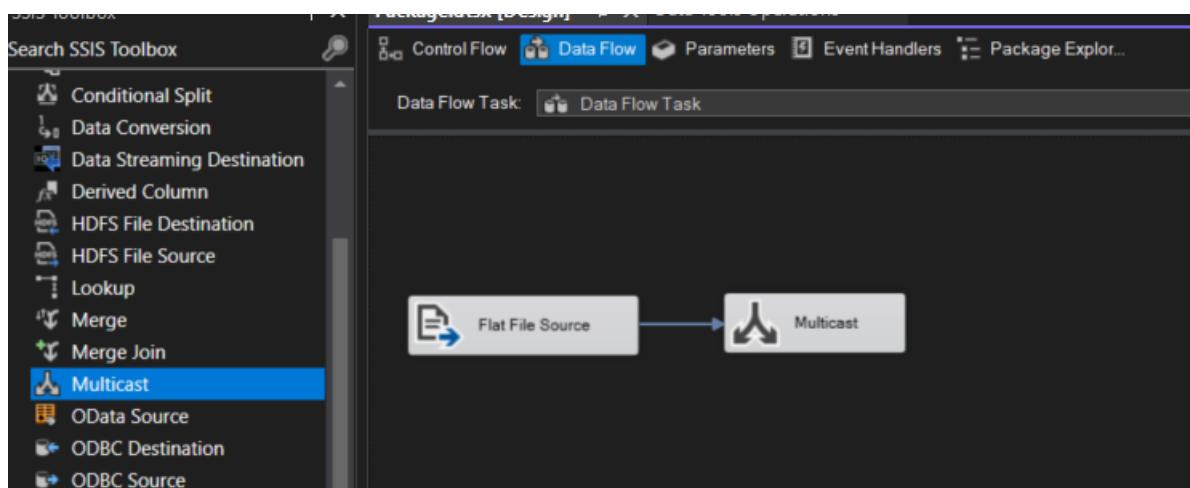
Hình 16: Xem lại các cột dữ liệu trong file dữ liệu

- Click chọn **OK** và kiểm tra lần nữa các cột dữ liệu ở dạng danh sách. Nhấn **OK** lần nữa để tiến hành hoàn tất quá trình load dữ liệu vào **Flat File Source**



Hình 17: Click chọn OK

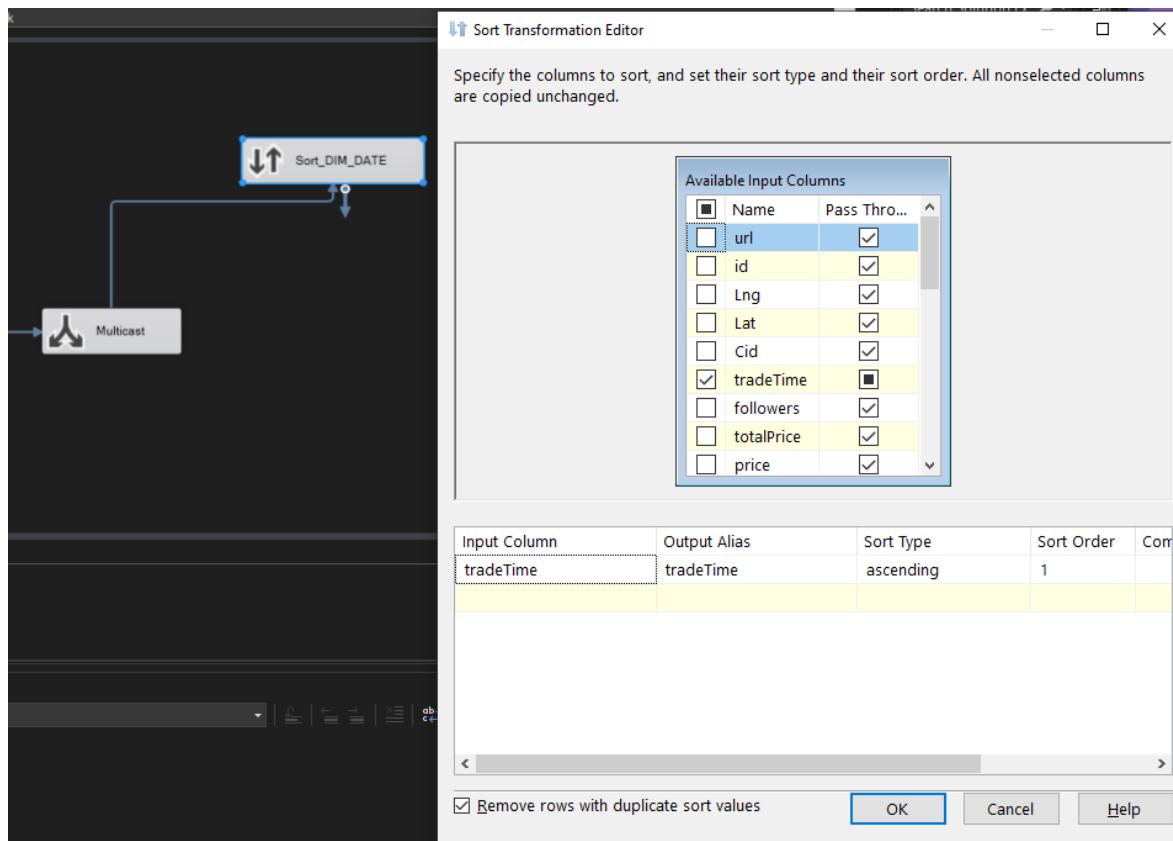
- Tạo **Multicast** để phân tán dữ liệu từ **Flat File Source** đến các **Dimension**. Tiến hành kết nối **Flat File Source** và **Multicast**.



Hình 18: Tạo Multicast

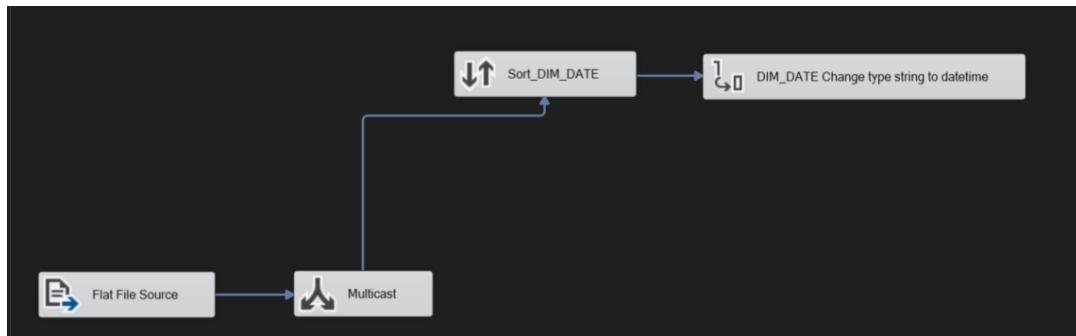
4.2. Tạo bảng DIM_DATE

- Tạo mới một Sort có tên là **Sort_DIM_DATE** để lấy ra các cột dữ liệu cần thiết cho **DIM_DATE**. Nhấn đúp chuột vào để Edit, chọn **tradeTime** làm cột dữ liệu cho **Sort_DIM_DATE** (Tick chọn **Remove rows with duplicate sort values** xoá các dòng dữ liệu trùng nhau, sau đó chọn **OK**.)



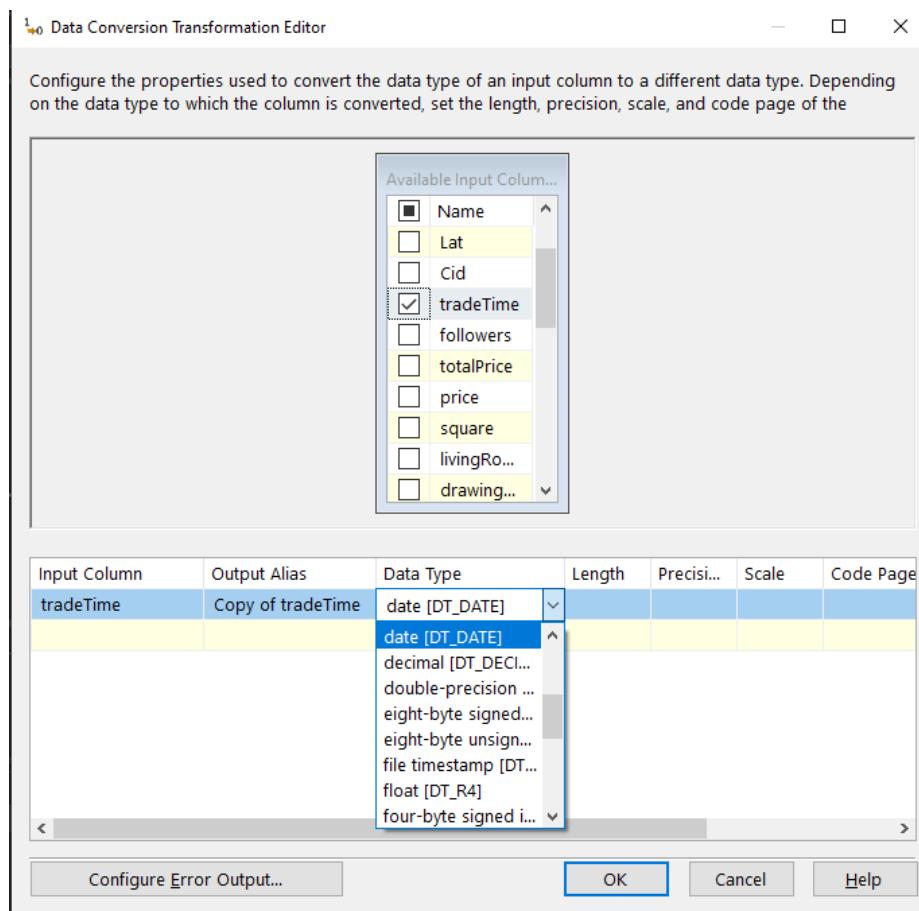
Hình 19: Tạo mới một Sort có tên là Sort_DIM_DATE

- Kiểu dữ liệu DateTime khi lấy từ dữ liệu gốc sẽ được mặc định là kiểu string. Ta dùng **Data Conversion** để chuyển đổi kiểu string về dạng DateTime. Đặt tên cho **Data Conversion** này là **DIM_DATE Change type string to datetime**



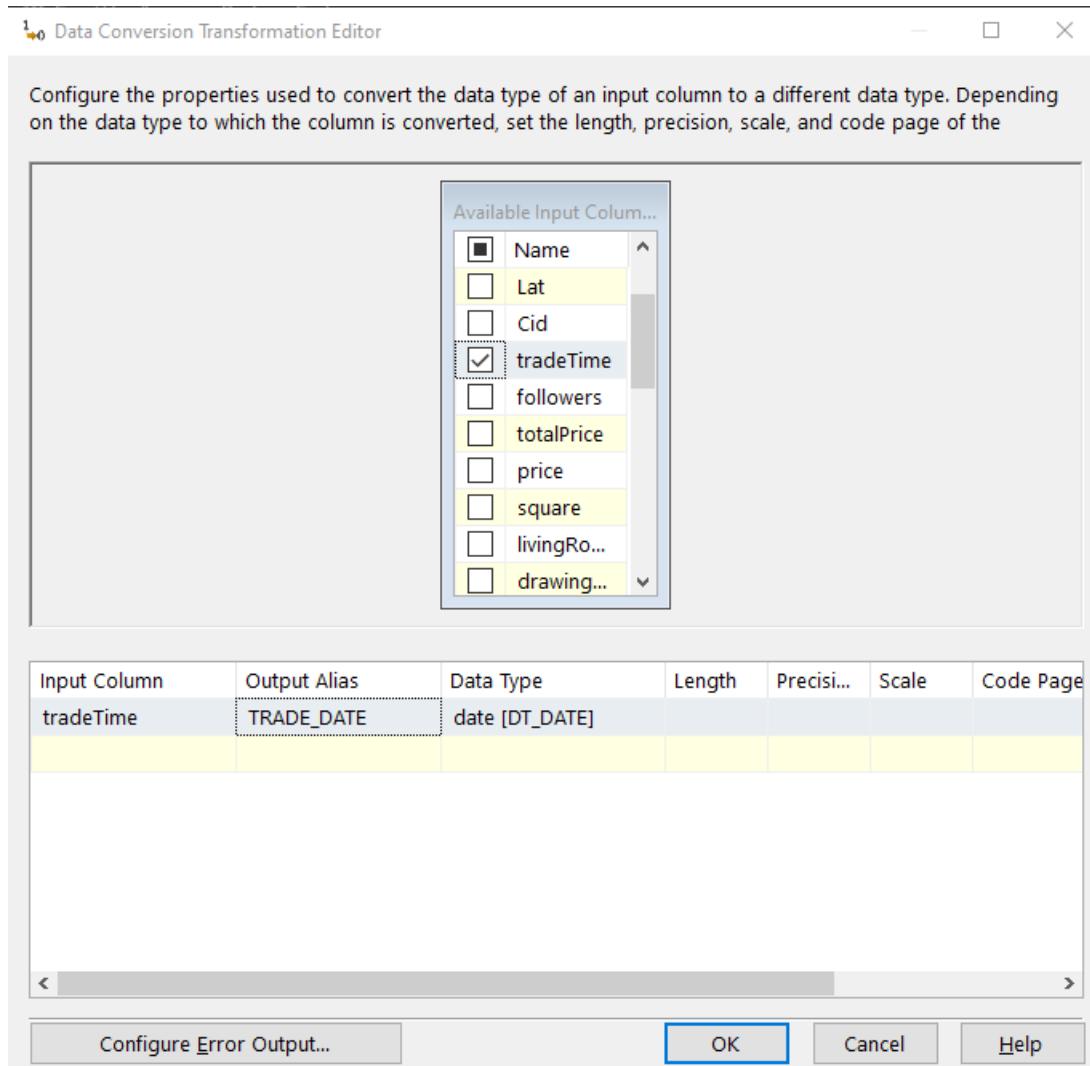
Hình 20: Đặt tên cho Data Conversion này là DIM_DATE Change type string to datetime

- Click chuột phải vào **Data conversion** này và chọn **Edit**, click chọn cột **tradeTime**, ở cột Data Type ta thấy mặc định kiểu dữ liệu là string, ta chọn lại kiểu dữ liệu cho cột này **date[DT_DATE]**



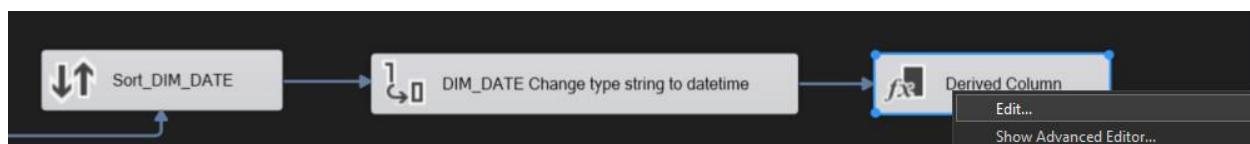
Hình 21: Click chuột phải vào Data conversion này và chọn Edit

- Đặt lại Output alias cho tradeTime từ **Copy of tradeTime** thành **TRADE_DATE** rồi nhấn **OK**.



Hình 22: Đặt lại Output alias cho tradeTime

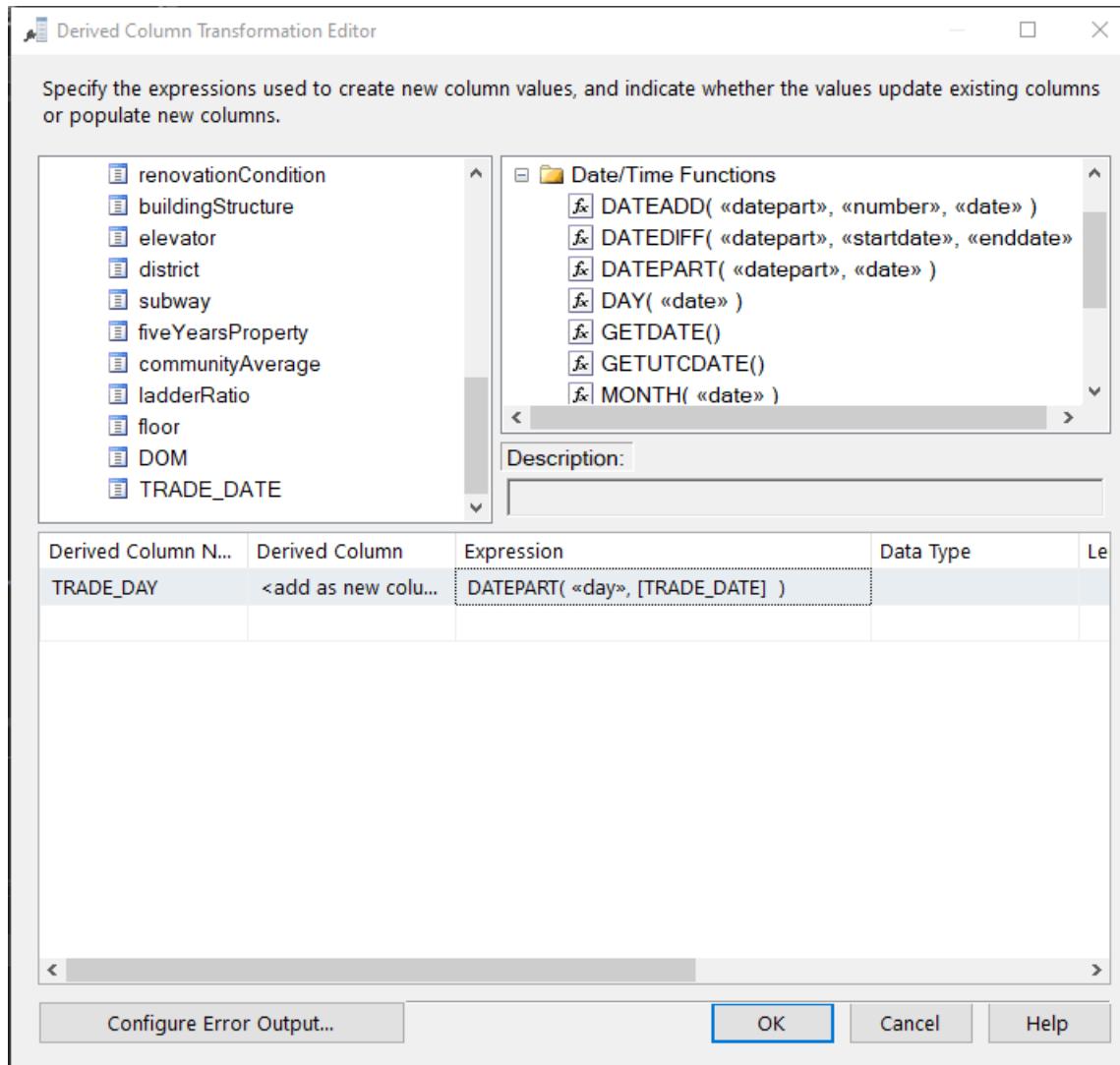
- Thêm thành phần **Derived Column** để chia cột dữ liệu.



Hình 23: Thêm thành phần Derived Column

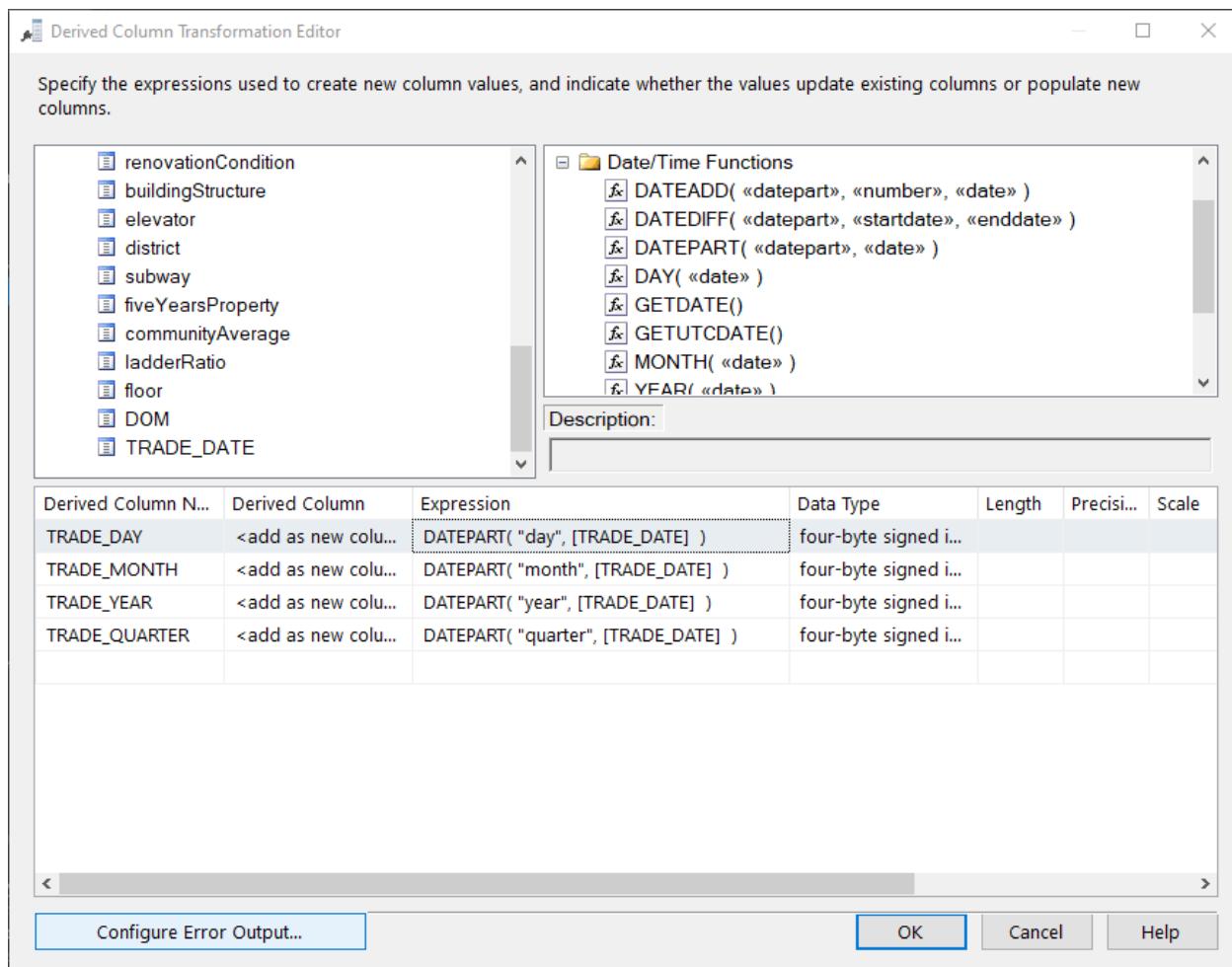
- Nhấn đúp chuột để edit, chúng ta sẽ chia dữ liệu từ cột **TRADE_DATE** có kiểu dữ liệu **mm/dd/yyyy** thành các cột **TRADE_DAY**, **TRADE_MONTH**, **TRADE_YEAR**, **TRADE_QUARTER**

- Chọn phương thức **DATEPART**(<kiểu dữ liệu thời gian>, cột dữ liệu). Bắt đầu với **TRADE_DATE** thành cột **TRADE_DAY**, cài đặt theo cú pháp: **DATEPART("day", [Start_Time_Date])**.



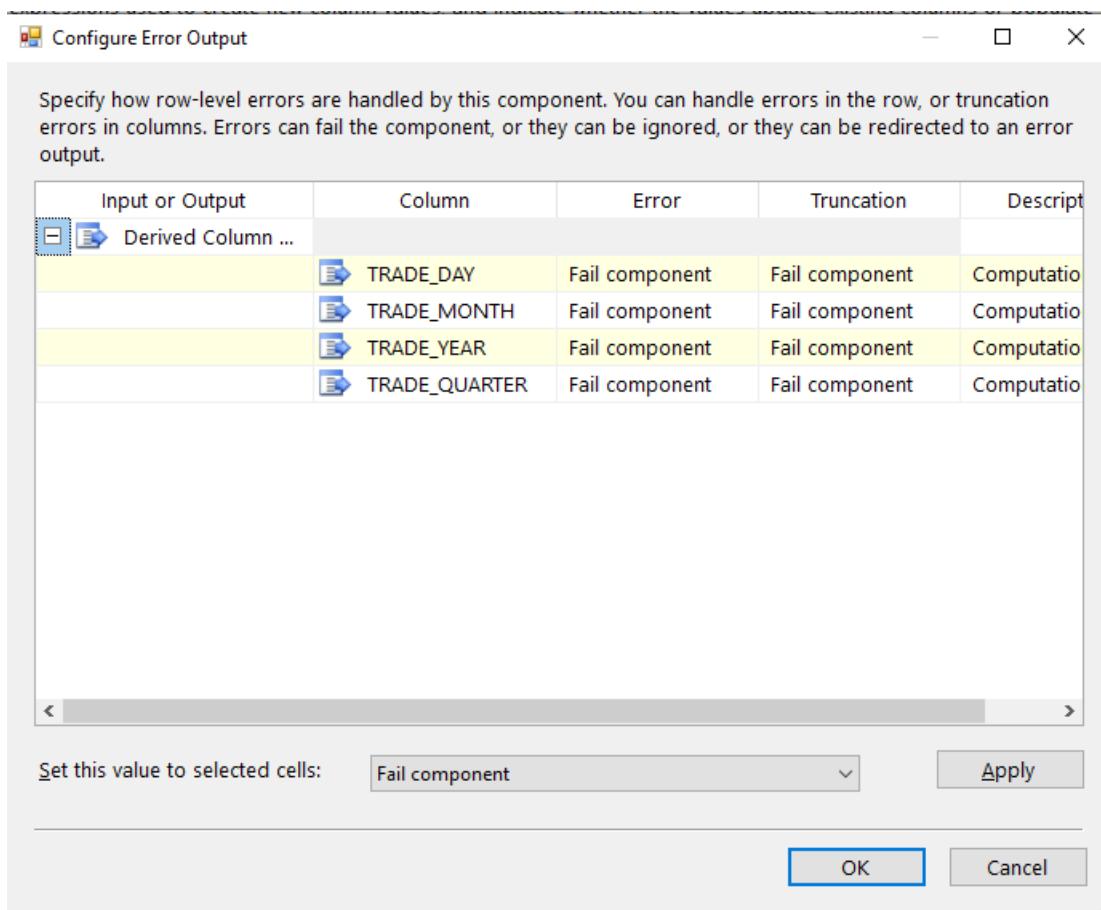
Hình 24: Chọn phương thức DATEPART

- Làm tương tự để chia **TRADE_DATE** thành các cột **TRADE_MONTH**, **TRADE_YEAR**, **TRADE_QUARTER**



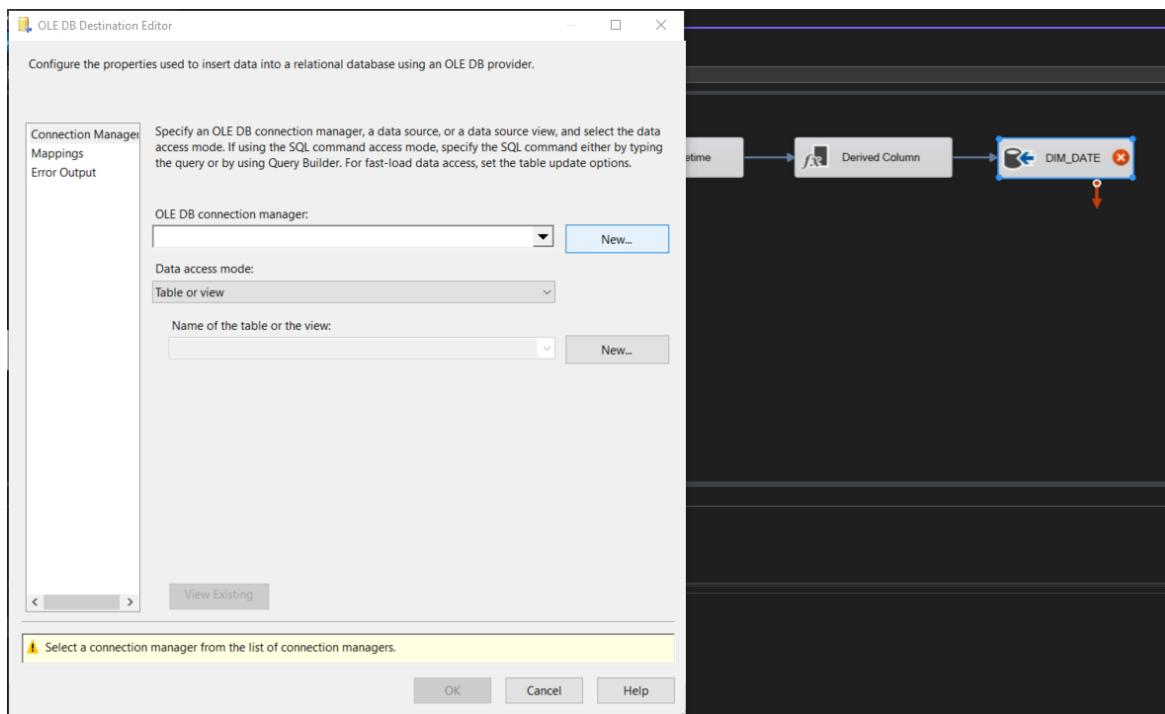
Hình 25: Chia TRADE_DATE thành các cột

- Khi hoàn thành, nhấn nút **Configure Error Output...**. Ta thấy từ 1 cột TRADE_DATE được chia thành 4 cột ứng với lược đồ dữ liệu bảng **DIM_DATE**. Nhấn **OK** để hoàn tất quá trình chia cột.



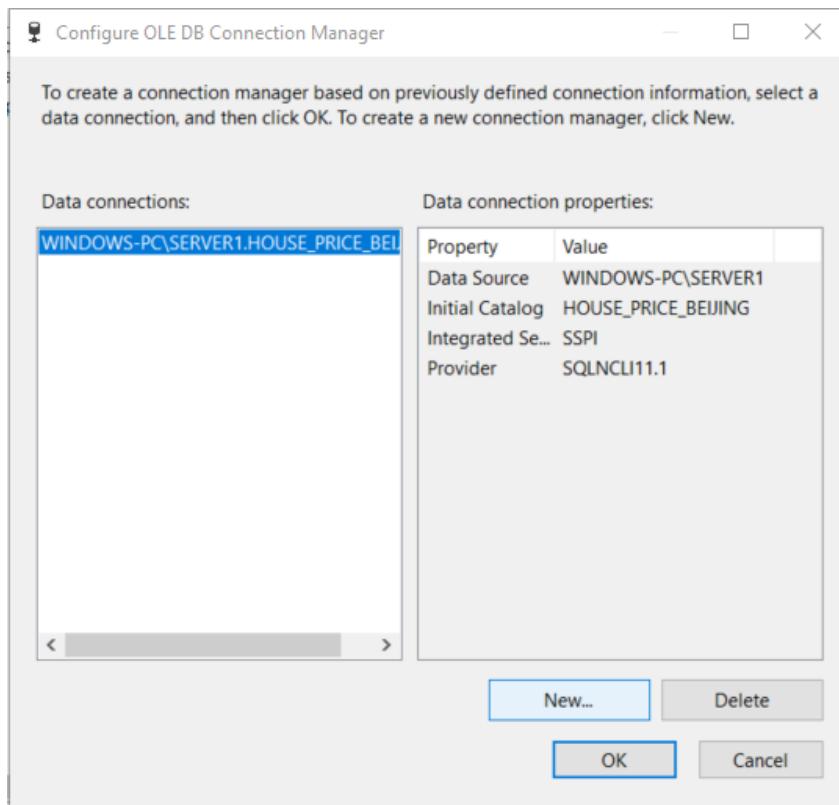
Hình 26: Khi hoàn thành, nhấn nút Configure Error Output...

- Tạo **DIM_DATE** từ một **OLE DB Destination**. Double click vào **OLE DB Destination** này để tạo một connection mới đến MS SQL Server.



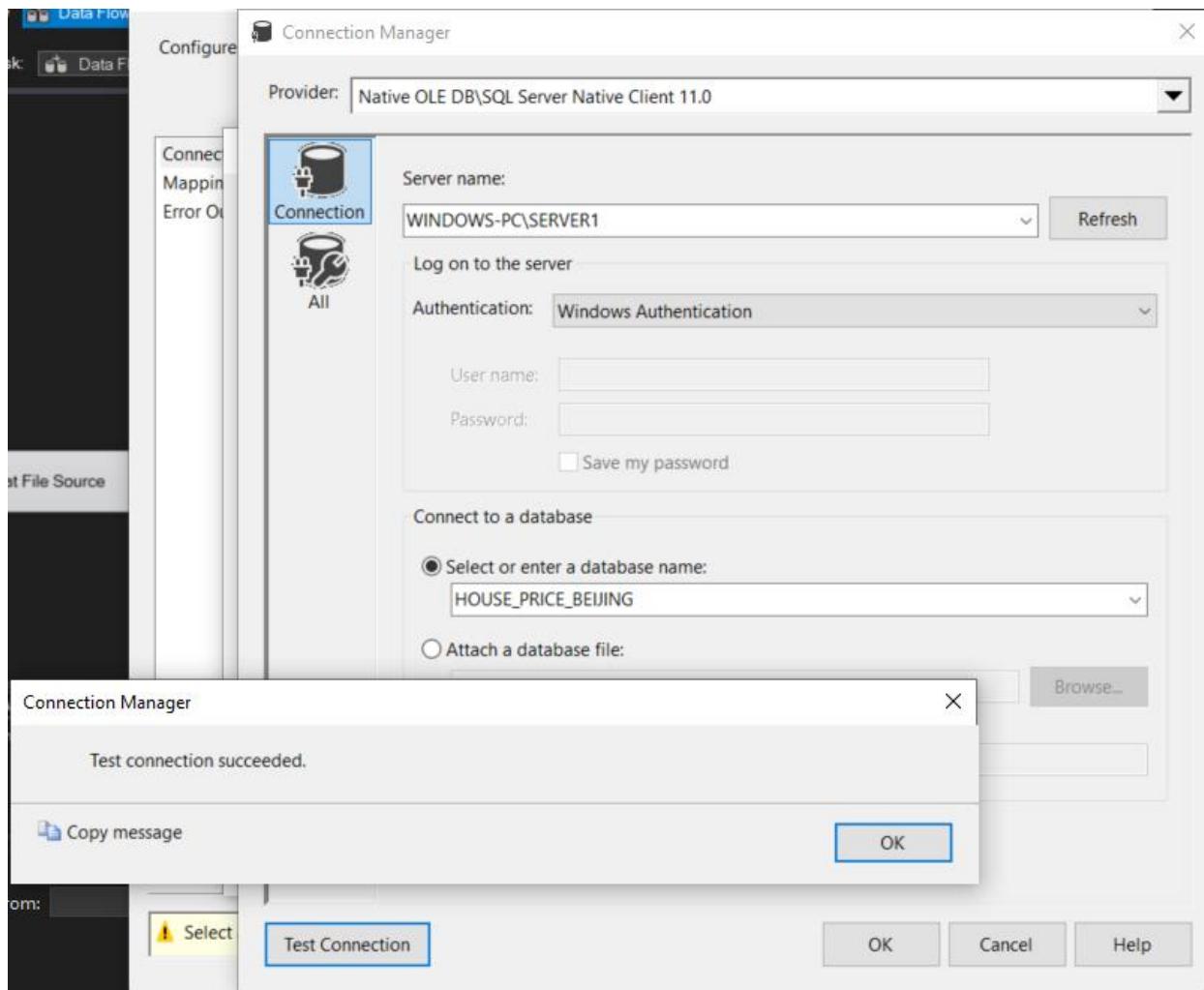
Hình 27: Tạo một connection mới đến MS SQL Server

- Tiếp tục chọn **New...** để tạo một connection mới:



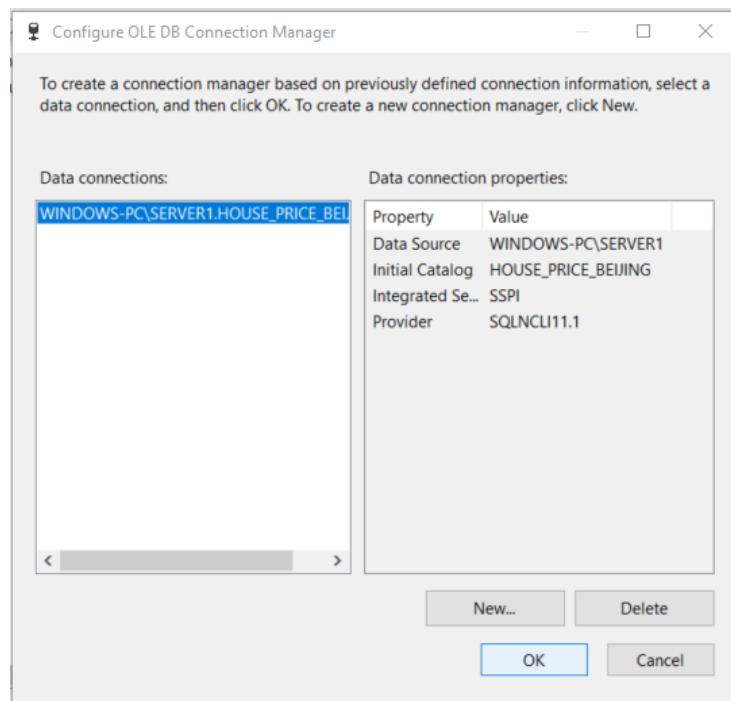
Hình 28: Tạo một connection mới

- Chọn tên server name trùng với server name MS SQL Server để ta có thể kết nối đến datawarehouse **HOUSE_PRICE_BEIJING** vừa tạo. Kết nối đến server bằng tài khoản window mặc định (Windows Authentication)
- Nhấn **Test Connection** để kiểm tra kết nối



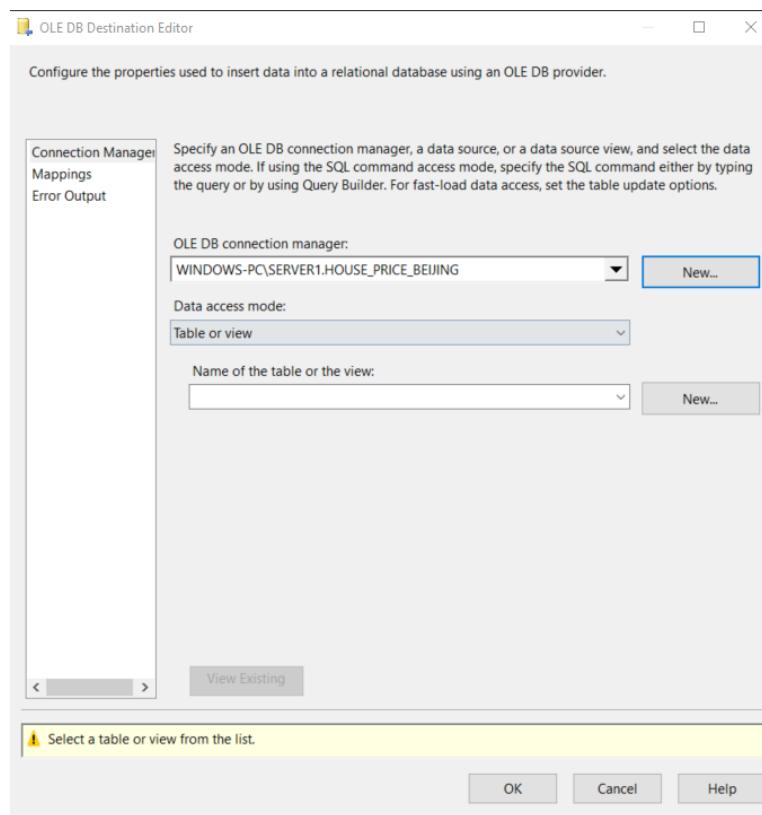
Hình 29: Test Connection để kiểm tra kết nối

- Tiến hành tạo bảng DIM_DATE ở MS SQL Server
- Chọn connection vừa tạo đến MS SQL Server và nhấn **OK**.

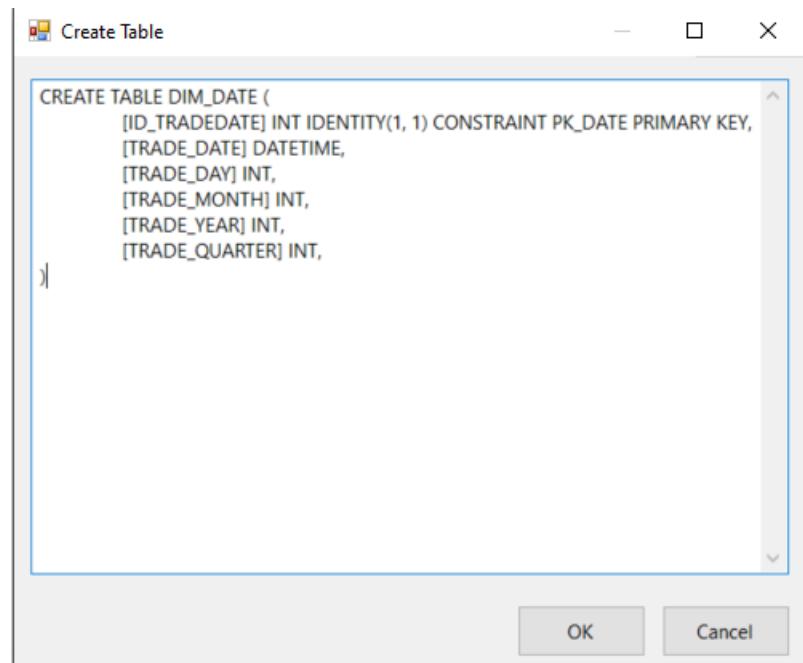


Hình 30: Chọn connection vừa tạo đến MS SQL Server và nhấn OK

Chọn New.. để tạo mới bảng



Hình 31: Tạo mới bảng



Hình 32: Tạo mới bảng(2)

*Nội dung câu lệnh SQL tạo bảng DIM_DATE như sau:

`CREATE TABLE DIM_DATE (`

`[ID_TRADEDATE] INT IDENTITY(1, 1) CONSTRAINT PK_DATE PRIMARY
 KEY,`

`[TRADE_DATE] DATETIME,`

`[TRADE_DAY] INT,`

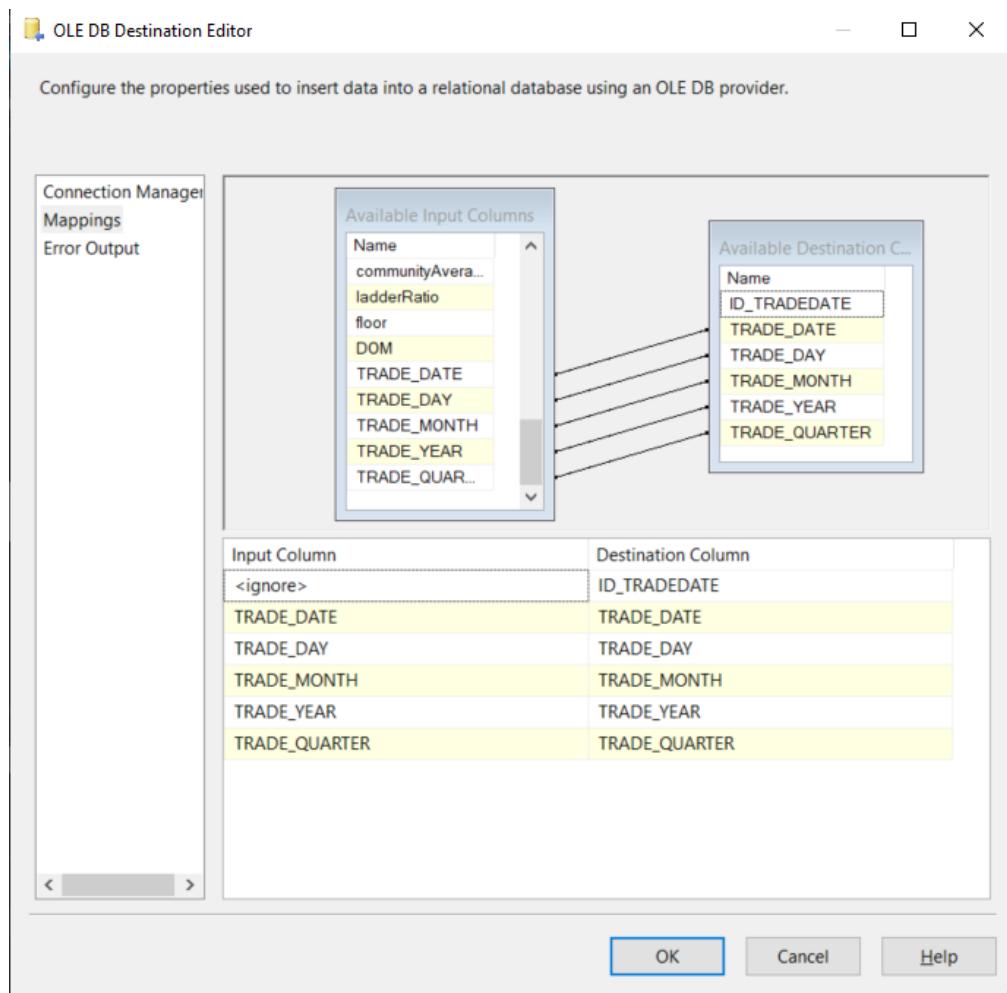
`[TRADE_MONTH] INT,`

`[TRADE_YEAR] INT,`

`[TRADE_QUARTER] INT,`

`)`

- Tiếp đến ta cần chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

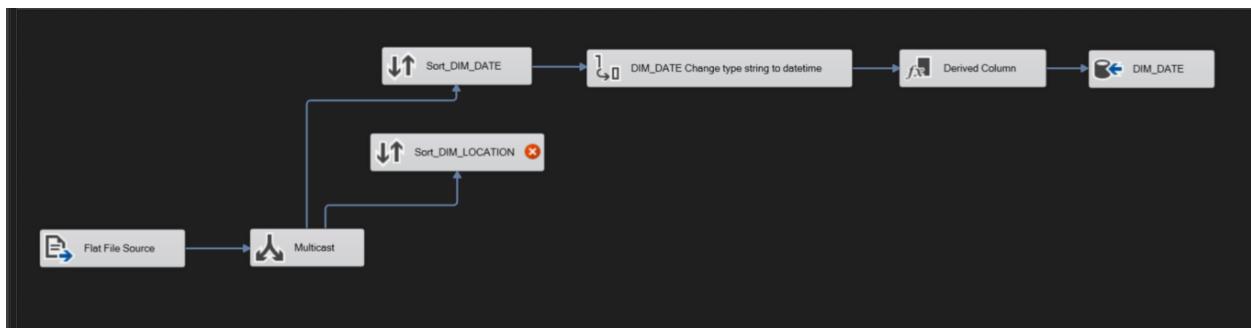


Hình 33: Xem xét việc ánh xạ các cột dữ liệu

- Chọn **OK** để hoàn tất thiết lập.

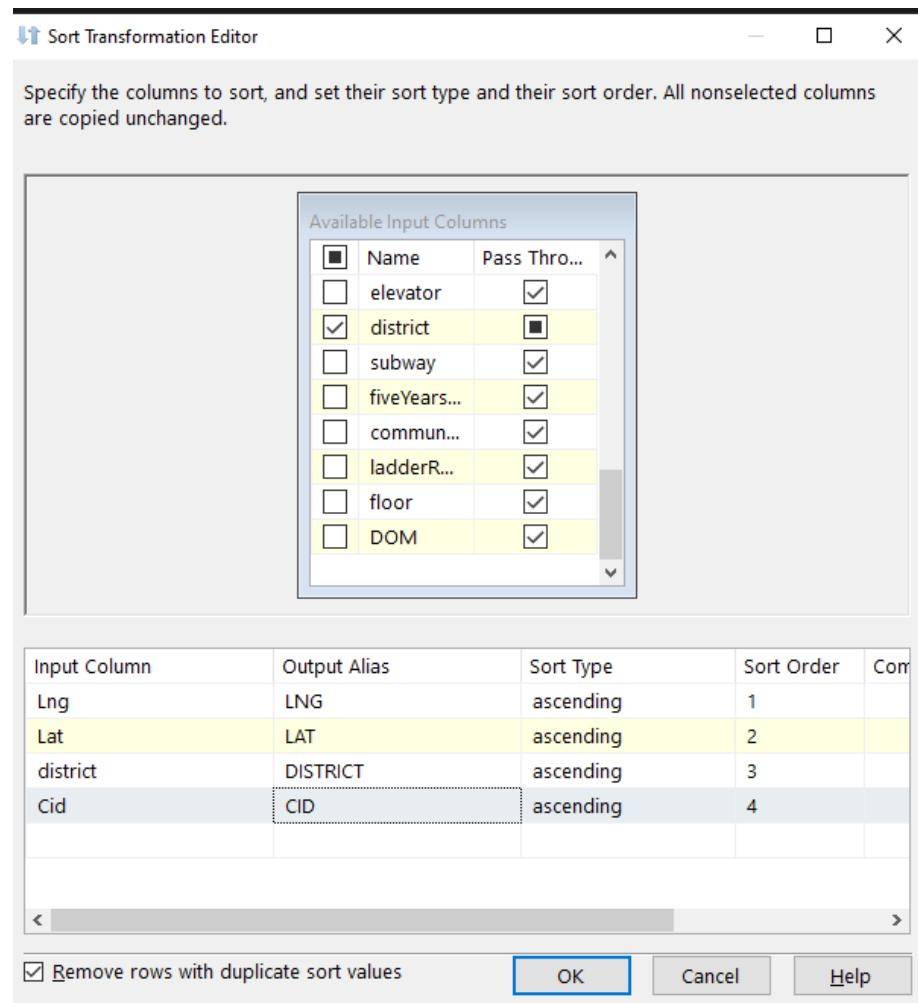
4.3. Bảng DIM_LOCATION

Bước 1. Chọn một Sort để tạo ra Sort_DIM_LOCATION cho DIM_LOCATION



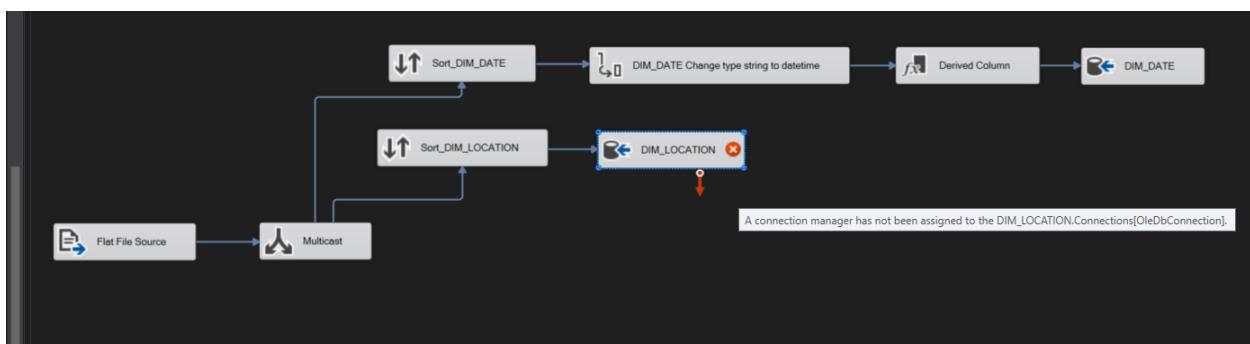
Hình 34: Chọn một Sort để tạo ra Sort_DIM_LOCATION cho DIM_LOCATION

Bước 2. Click chuột phải vào Sort_DIM_LOCATION, chọn **Edit:** lần lượt chọn các cột LNG, LAT, DISTRICT và CID làm các cột để đổ dữ liệu vào Sort_DIM_LOCATION



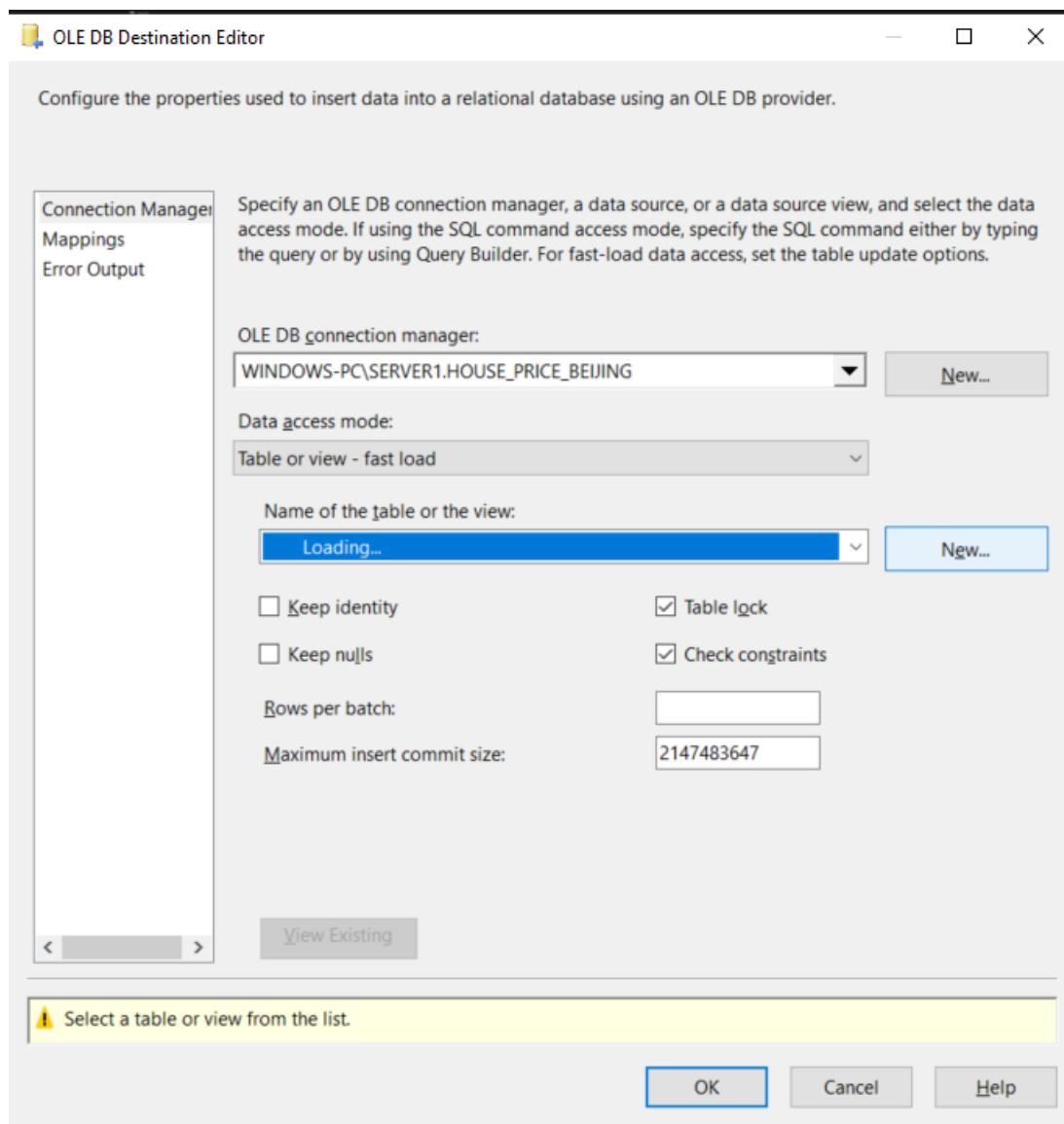
Hình 35: Chọn các cột để đổ dữ liệu vào Sort_DIM_LOCATION

Bước 3. Tạo mới một **OLE DB Destination** để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu HOUSE_PRICE_BEIJING.

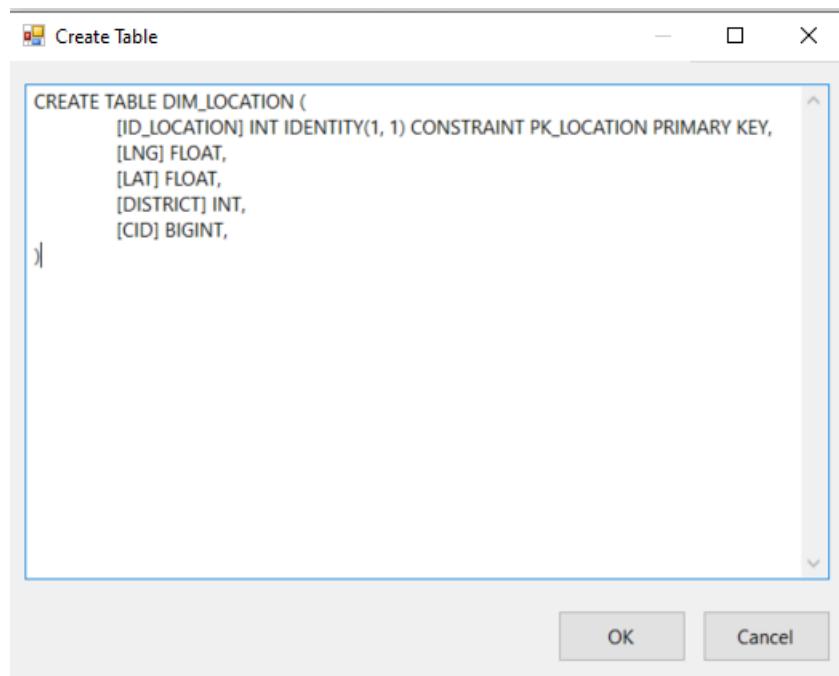


Hình 36: Đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu

Bước 4. Connect đến kho dữ liệu đã được tạo khi tạo DIM_DATE, vì vậy chỉ cần chọn **New...** để tạo bảng DIM_LOCATION



Hình 37: Chọn New... để tạo bảng DIM_LOCATION



Hình 38: Câu lệnh SQL tạo bảng DIM_LOCATION

*Nội dung câu lệnh SQL tạo bảng DIM_LOCATION như sau:

CREATE TABLE DIM_LOCATION (

 [ID_LOCATION] INT IDENTITY(1, 1) CONSTRAINT PK_LOCATION
 PRIMARY KEY,

 [LNG] FLOAT,

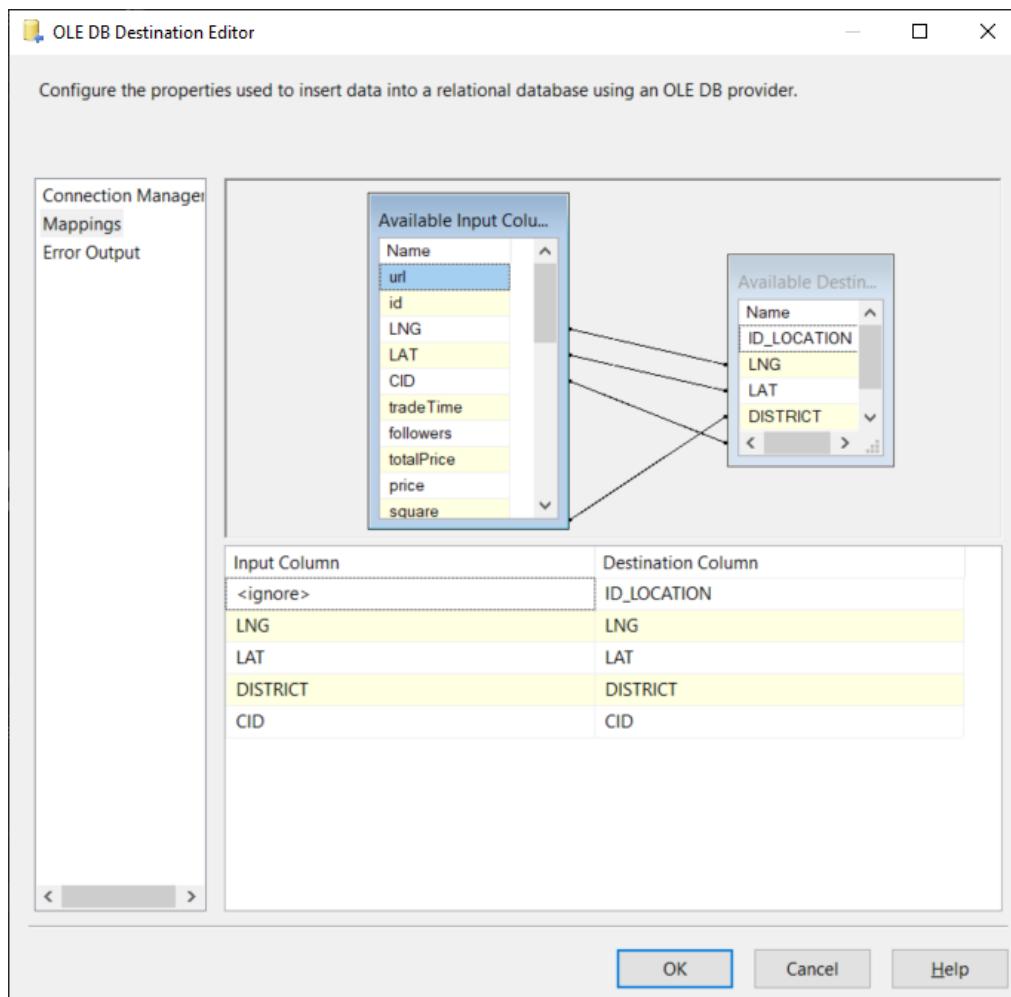
 [LAT] FLOAT,

 [DISTRICT] INT,

 [CID] BIGINT,

)

Bước 5. Tiếp đến ta cần chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

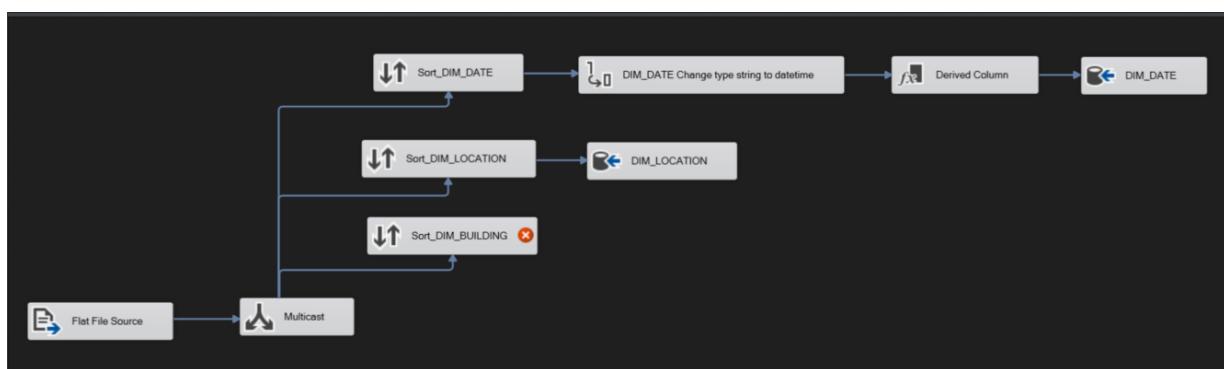


Hình 39: Chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu

- Chọn **OK** để hoàn tất thiết lập.

4.4. Bảng DIM_BUILDING

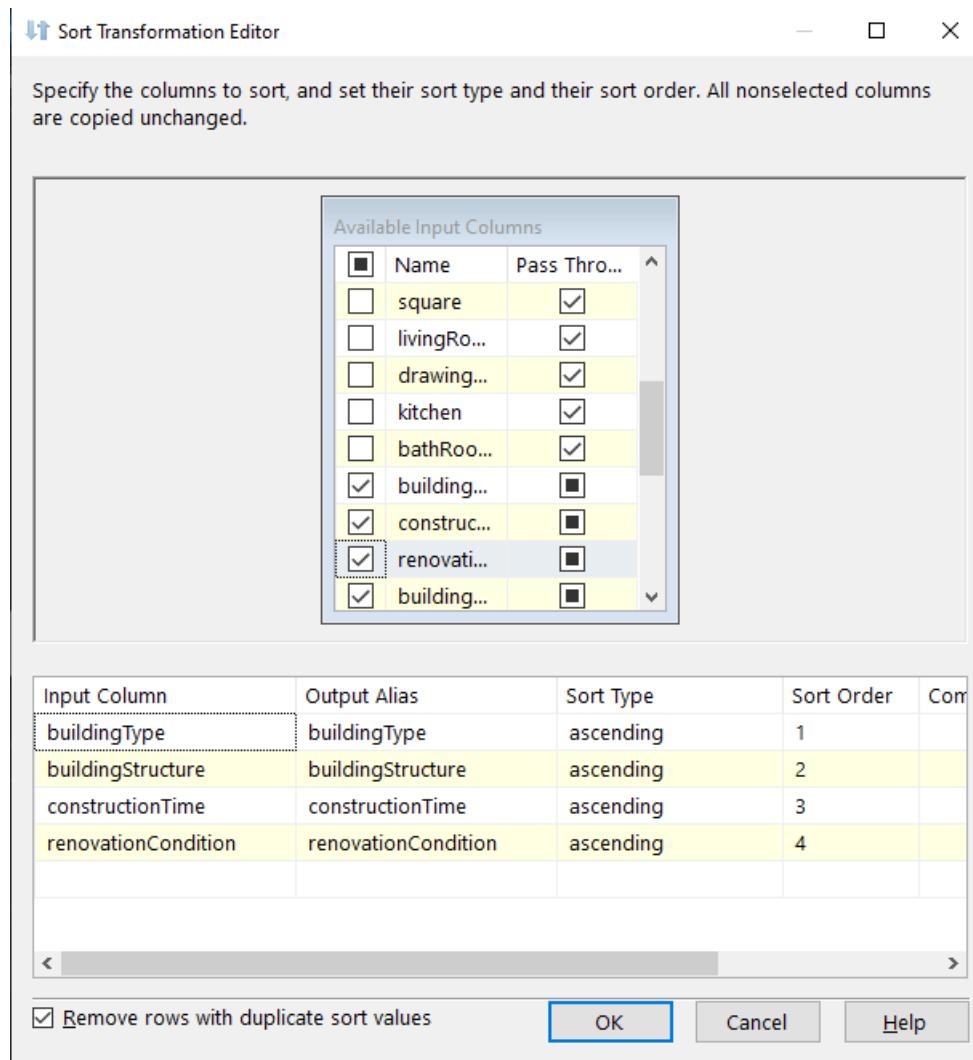
Bước 1. Chọn một Sort để tạo ra Sort_DIM_BUILDING cho DIM_BUILDING



Hình 40: Chọn một Sort để tạo ra Sort_DIM_BUILDING

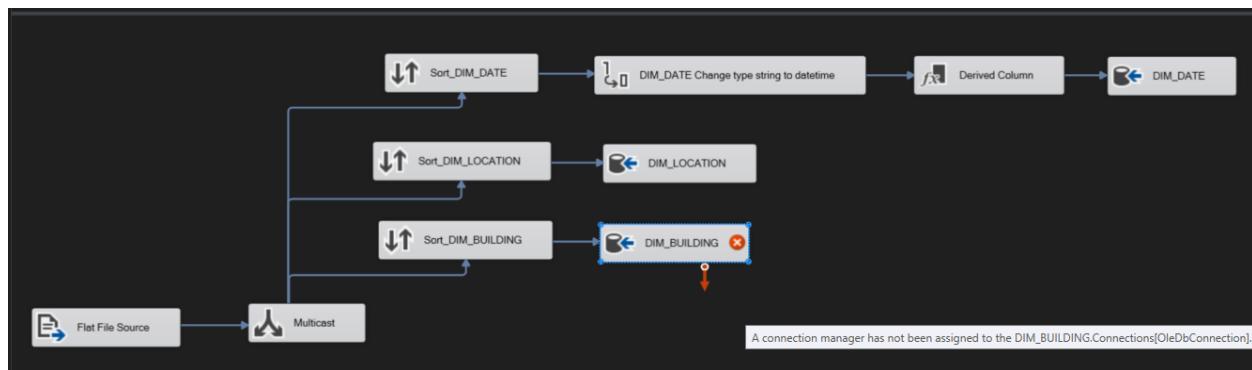
Bước 2. Click chuột phải vào Sort_DIM_BUILDING, chọn **Edit:** lần lượt chọn các cột BUILDINGTYPE, BUILDINGSTRUCTURE, CONSTRUCTIONTIME, RENOVATIONCONDITION làm các cột để đổ dữ liệu vào Sort_DIM_BUILDING

- Tick chọn **Remove rows with duplicate sort values** xóa đi các dòng dữ liệu trùng nhau và sau đó chọn **OK**.



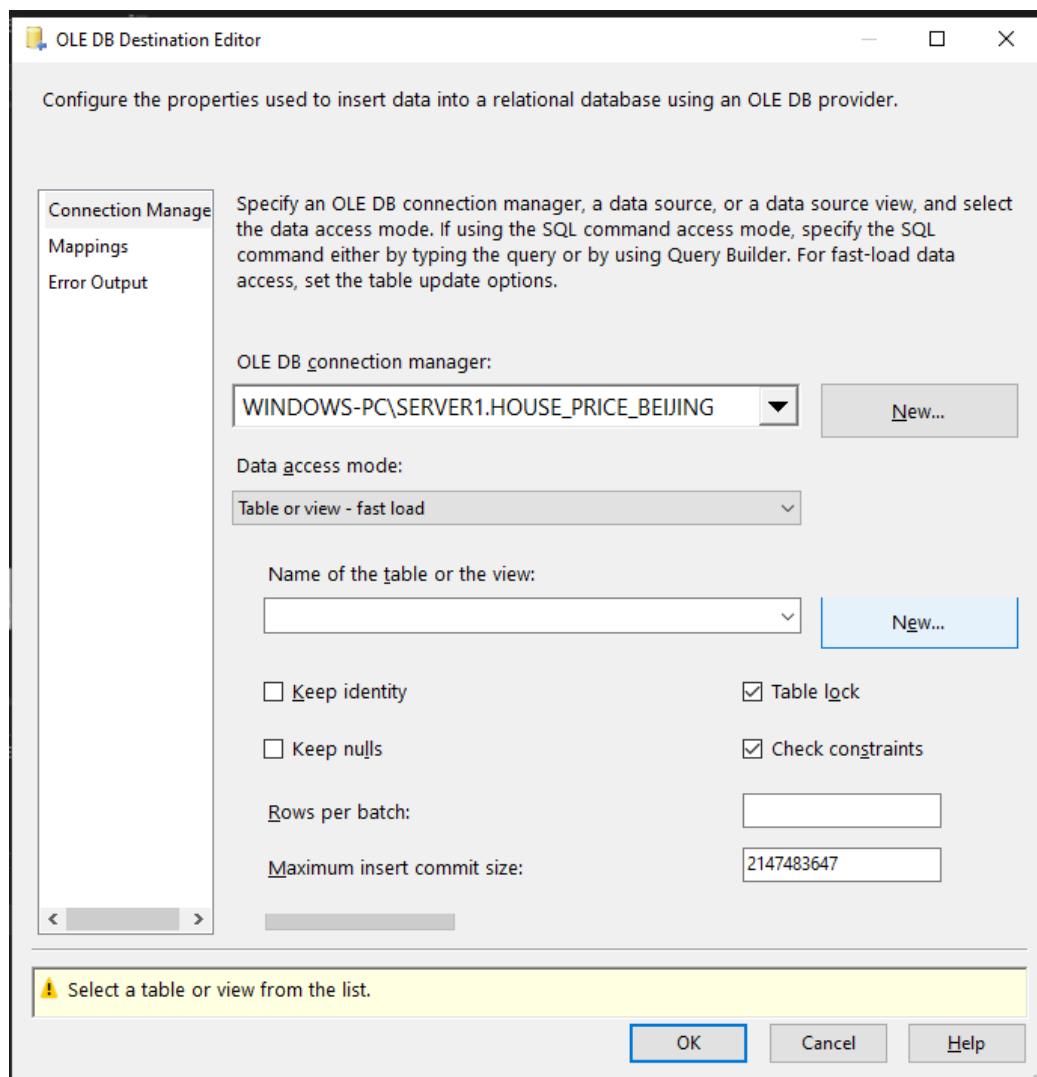
Hình 41: Tick chọn Remove rows with duplicate sort values

Bước 3. Tạo mới một **OLE DB Destination** để đổ dữ liệu gốc sau khi đã được xử lý vào trong bảng DIM_BUILDING kho dữ liệu HOUSE_PRICE_BEIJING.

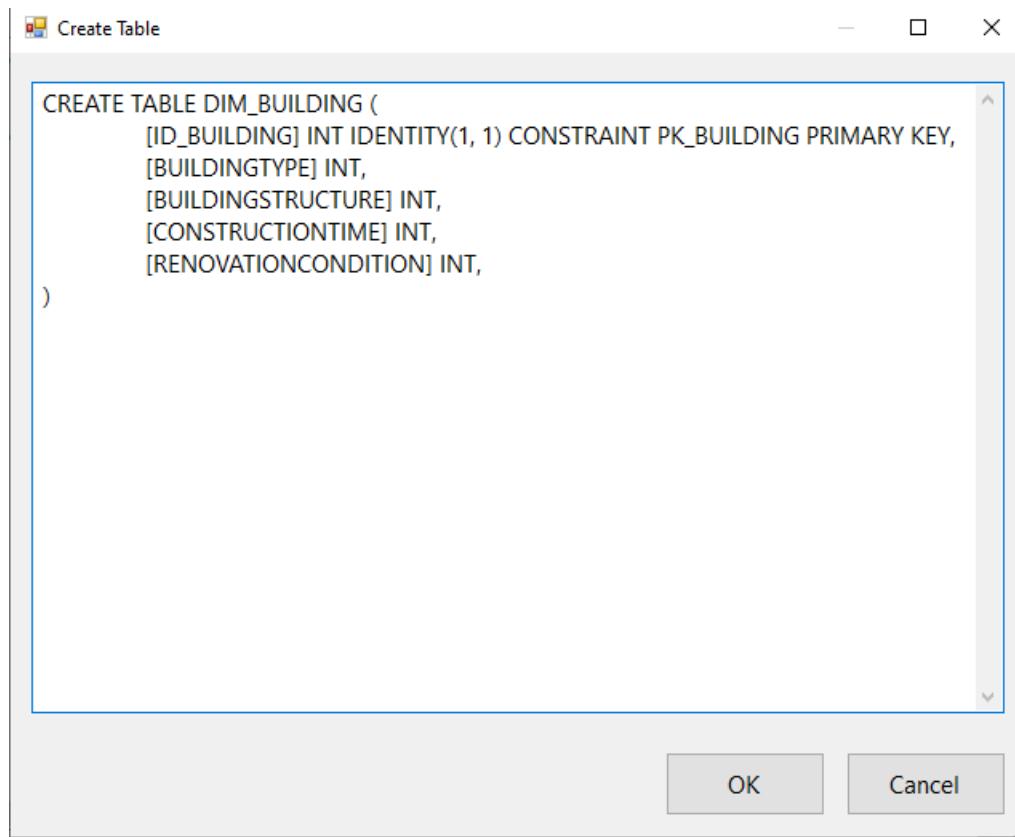


Hình 42: Tạo mới một OLE DB Destination

Bước 4. Connection đến kho dữ liệu đã được tạo khi tạo DIM_DATE, vì vậy ta chỉ cần chọn **New...** để tạo bảng DIM_BUILDING



Hình 43: New... để tạo bảng DIM_BUILDING

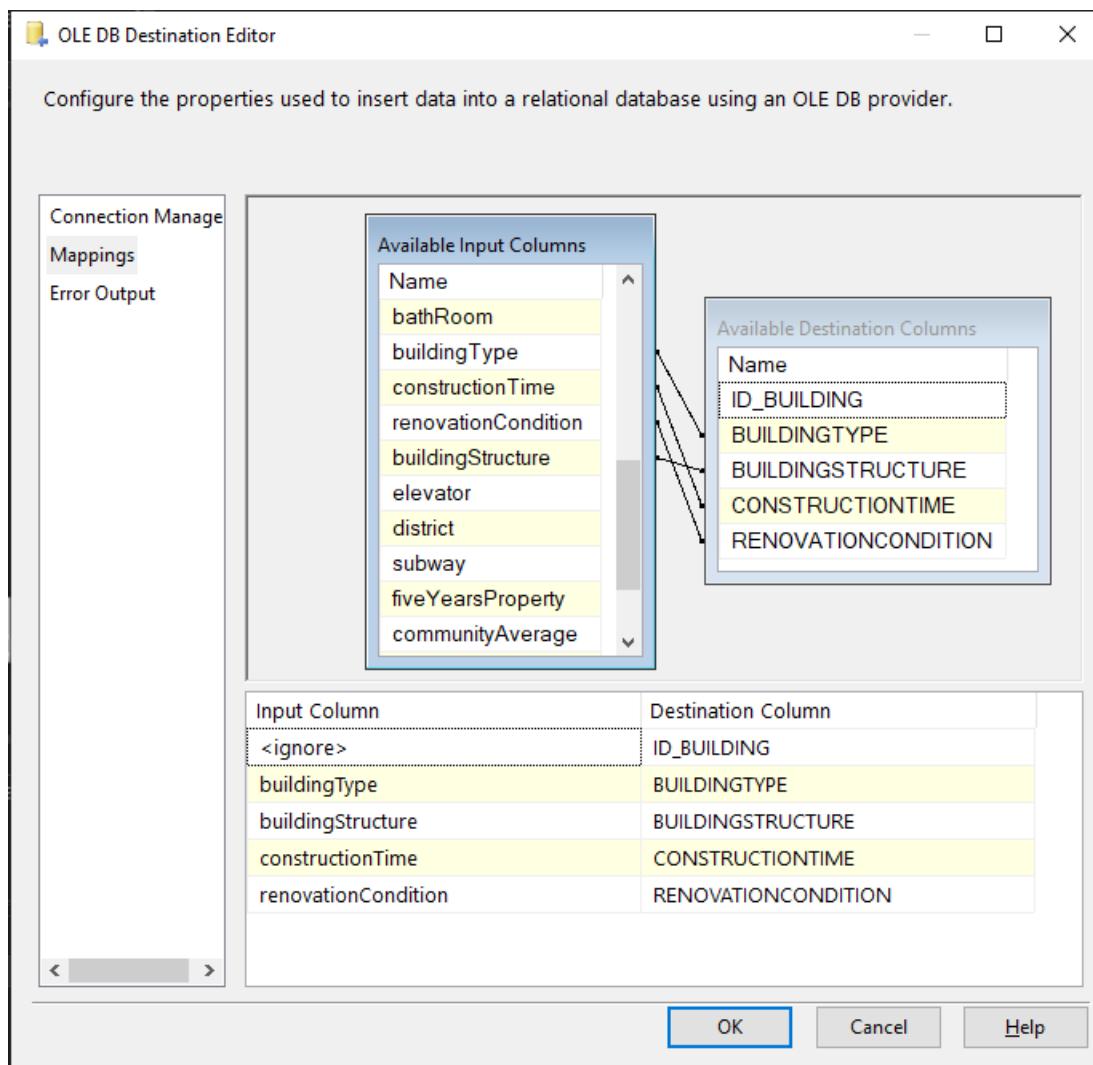


Hình 44: Câu lệnh SQL tạo bảng DIM_BUILDING

*Nội dung câu lệnh SQL tạo bảng DIM_BUILDING như sau:

```
CREATE TABLE DIM_BUILDING (
    [ID_BUILDING] INT IDENTITY(1, 1) CONSTRAINT PK_BUILDING
    PRIMARY KEY,
    [BUILDINGTYPE] INT,
    [BUILDINGSTRUCTURE] INT,
    [CONSTRUCTIONTIME] INT,
    [RENOVATIONCONDITION] INT,
)
```

Bước 5. Tiếp đến ta cần chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu



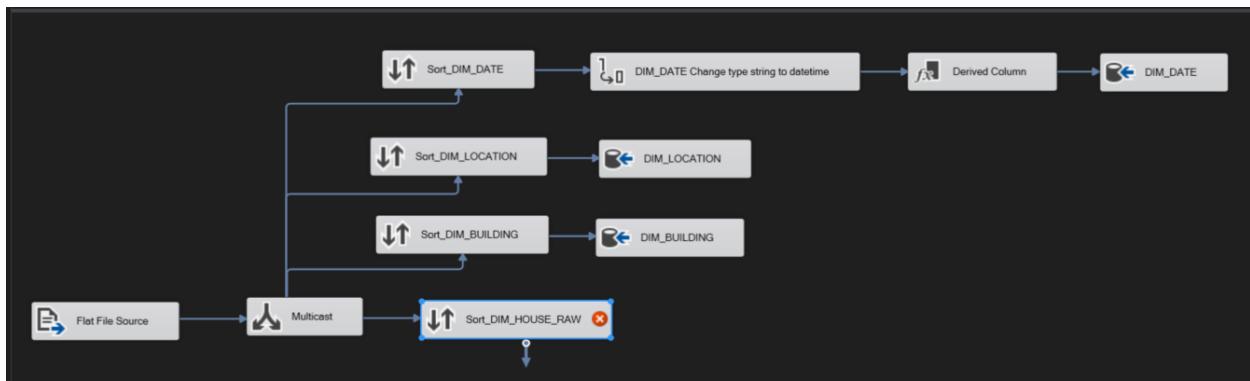
Hình 45: Chọn mục Mappings để xem xét việc ánh xạ các cột dữ liệu

- Chọn **OK** để hoàn tất thiết lập.

4.5. Bảng DIM_HOUSE_RAW:

Vì chiều **DIM_HOUSE** có các thuộc tính: ID_HOUSE, ID_LOCATION, để dễ dàng mappings các thuộc tính đúng với loại (ID_LOCATION) đã tạo ở **DIM_LOCATION** ta sẽ tạo một chiều mang tên là **DIM_HOUSE_RAW** dùng để mappings.

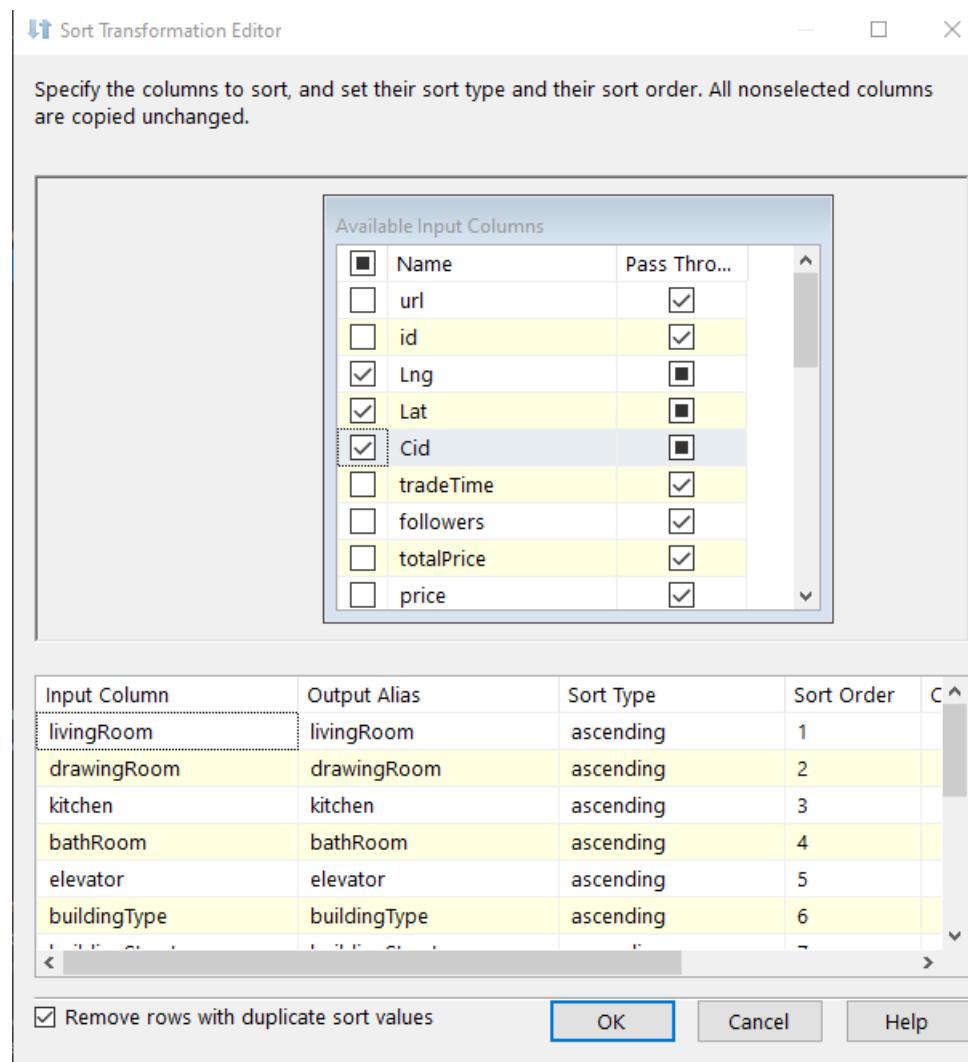
Bước 1. Chọn một Sort để tạo ra Sort_DIM_HOUSE_RAW cho DIM_HOUSE_RAW.



Hình 46: Chọn một Sort để tạo ra Sort_DIM_HOUSE_RAW

Bước 2. Click chuột phải vào **Sort_DIM_HOUSE_RAW**, chọn **Edit**: lần lượt chọn các cột làm cột dữ liệu để đổ dữ liệu vào **Sort_DIM_HOUSE_RAW**.

- Tick chọn **Remove rows with duplicate sort values** xóa đi các dòng dữ liệu trùng nhau và sau đó chọn **OK**.



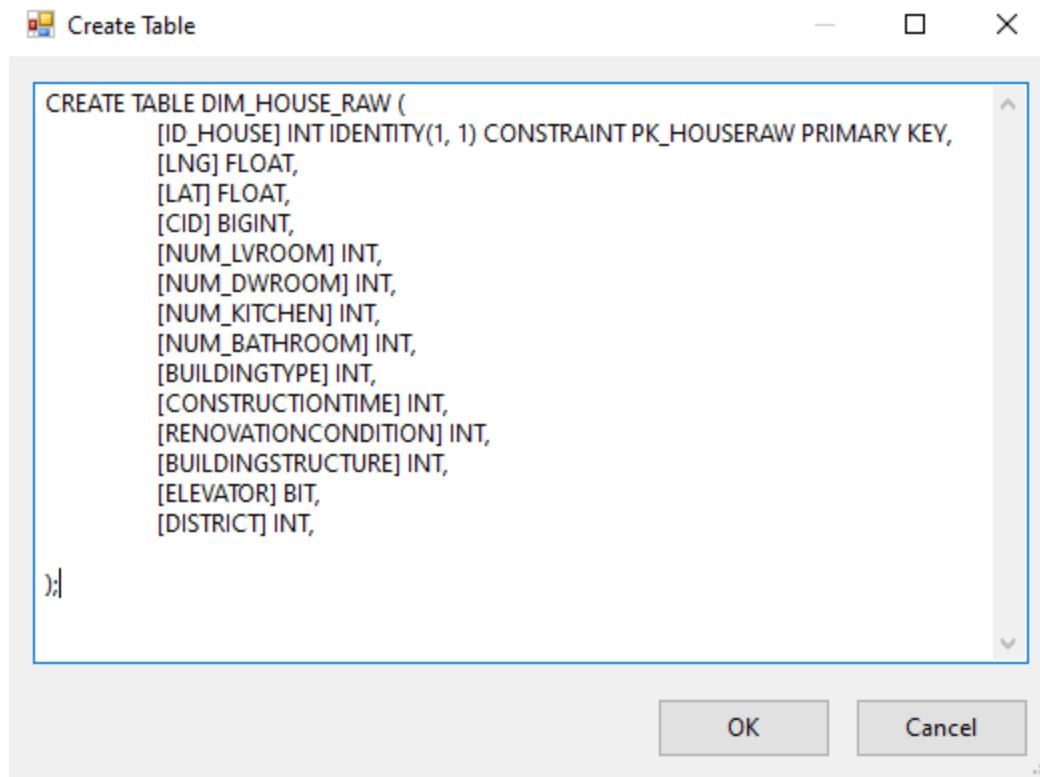
Hình 47: Chọn Remove rows with duplicate sort values xóa đi các dòng dữ liệu trùng nhau

Bước 3. Tạo mới một **OLE DB Destination** để đổ dữ liệu gốc sau khi đã được xử lý vào trong bảng **DIM_HOUSE_RAW** kho dữ liệu **HOUSE_PRICE_BEIJING**.



Hình 48: Tạo mới một OLE DB Destination để đổ dữ liệu gốc sau khi đã được xử lý vào trong bảng DIM_HOUSE_RAW

Bước 4. Click chuột phải và chọn **Edit** để tạo bảng **DIM_HOUSE_RAW** có các cột là tất cả các cột từ dữ liệu của **LOCATION** và **BUILDING** và **HOUSE**



Hình 49: Câu lệnh SQL tạo bảng DIM_HOUSE_RA

*Nội dung câu lệnh SQL tạo bảng DIM_HOUSE_RAW như sau:

CREATE TABLE DIM_HOUSE_RAW (

```

    [ID_HOUSE] INT IDENTITY(1, 1) CONSTRAINT PK_Houseraw
    PRIMARY KEY,
    [LNG] FLOAT,
    [LAT] FLOAT,
    [CID] BIGINT,
    [NUM_LVROOM] INT,
    [NUM_DWROOM] INT,
    [NUM_KITCHEN] INT,
    [NUM_BATHROOM] INT,
    [BUILDINGTYPE] INT,
    [CONSTRUCTIONTIME] INT,

```

```
[RENOVATIONCONDITION] INT,  

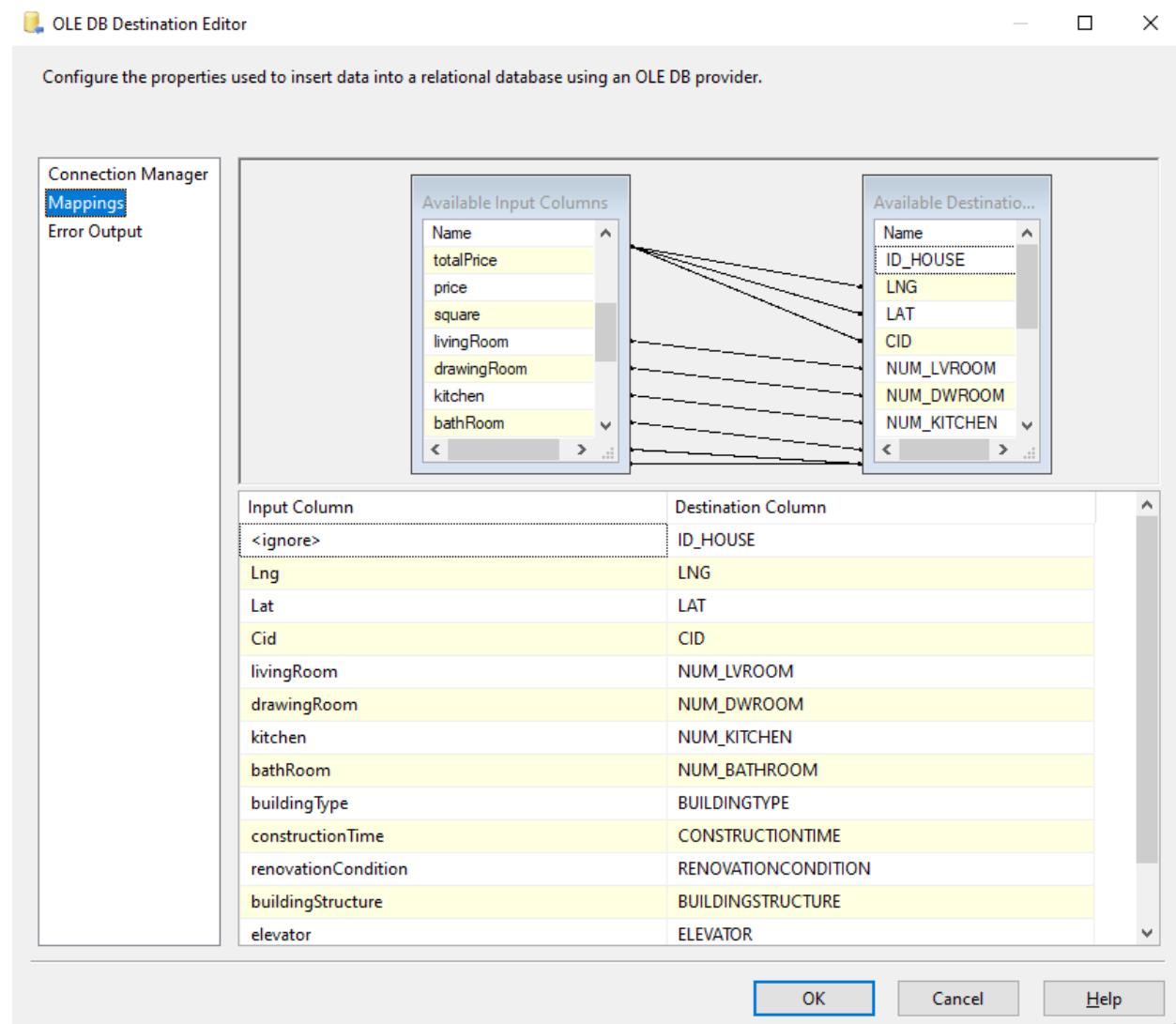
[BUILDINGSTRUCTURE] INT,  

[ELEVATOR] BIT,  

[DISTRICT] INT,  

);
```

Bước 5. Tiếp đến ta cần chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

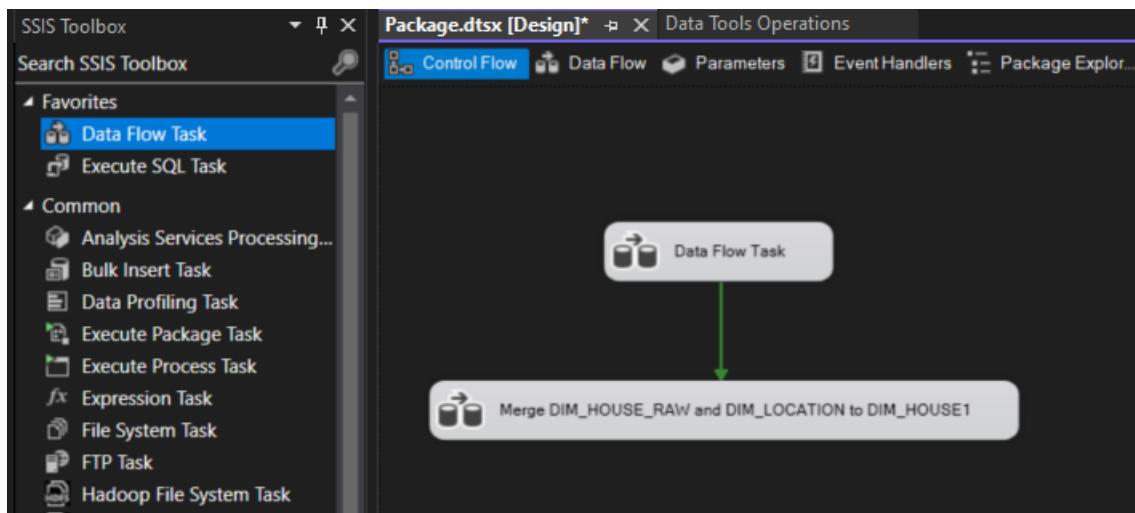


Hình 50: Xem xét việc ánh xạ các cột dữ liệu

- Tiếp theo đây ta sẽ thực hiện quá trình lần lượt loại bỏ các cột dữ liệu trùng của bảng DIM_HOUSE_RAW với các Dimension

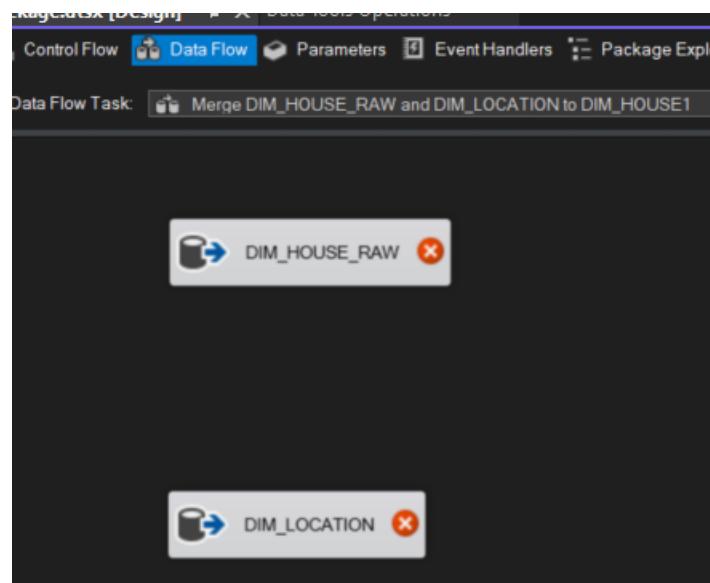
4.6. Merge DIM_HOUSE_RAW và DIM_LOCATION vào DIM_HOUSE1

Bước 1. Ở tab Control Flow, tạo hai Data Flow Task và đổi tên Data FlowTask thứ 2 là “Merge DIM_HOUSE_RAW and DIM_LOCATION to DIM_HOUSE1”



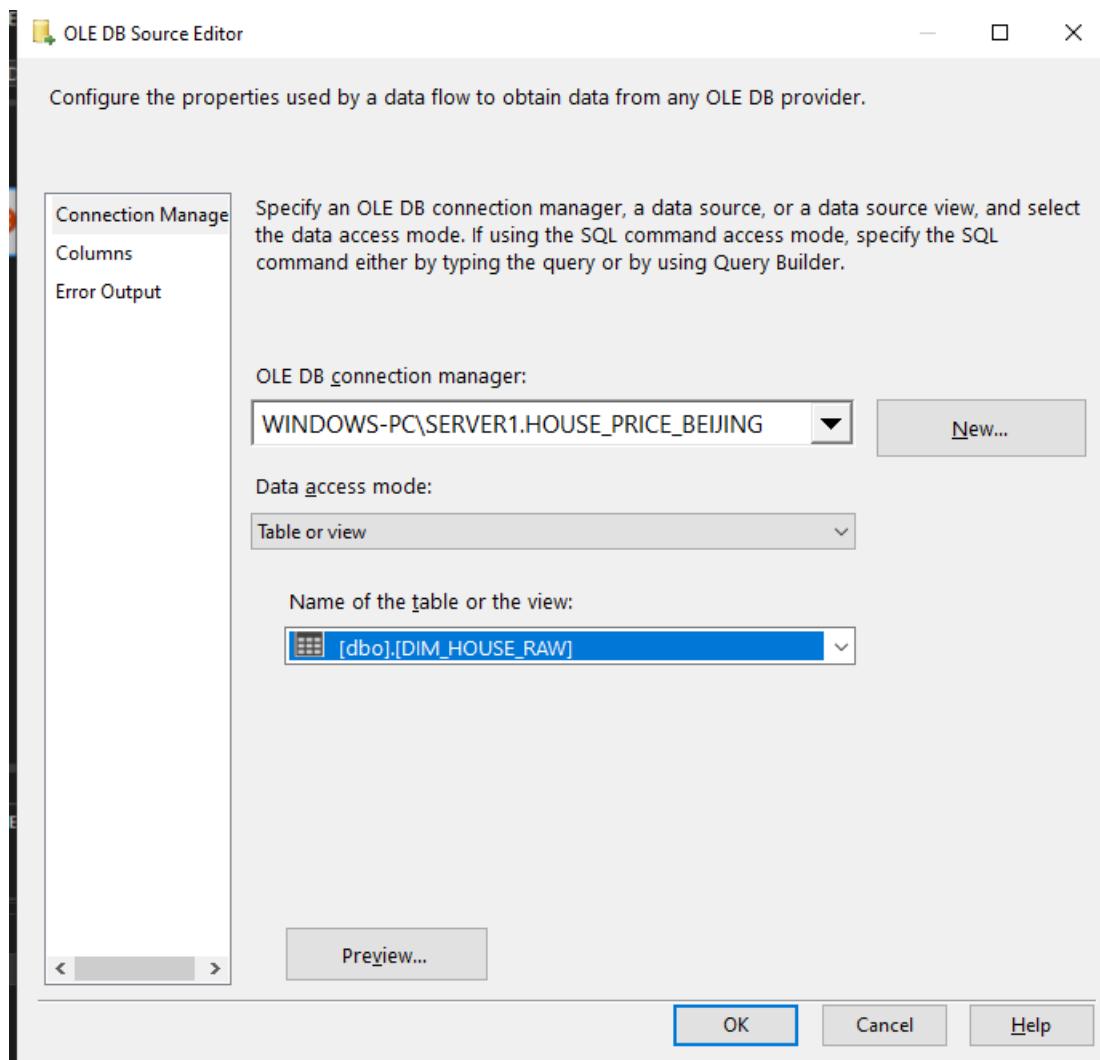
Hình 51: Tạo hai Data Flow Task và đổi tên

Bước 2. Click chuột phải vào Data Flow Task nói trên và chọn Edit, trong tab Data Flow ta tạo 2 OLE DB Source và đổi tên DIM_HOUSE_RAW và DIM_LOCATION



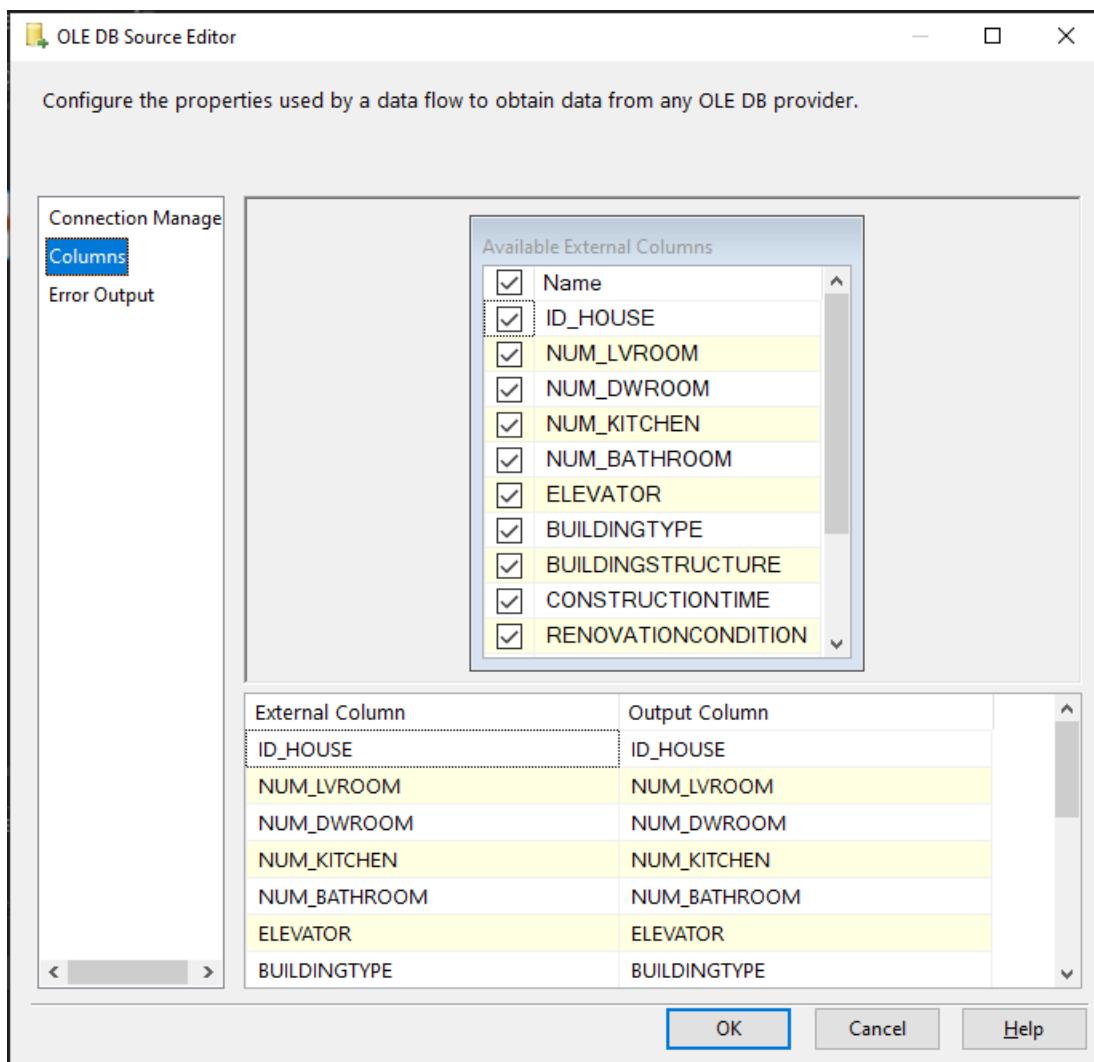
Hình 52: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải chọn **Edit**, sau đó chọn bảng **DIM_HOUSE_RAW** đã tạo trước đó làm data source cho bảng **DIM_HOUSE_RAW** mới này.



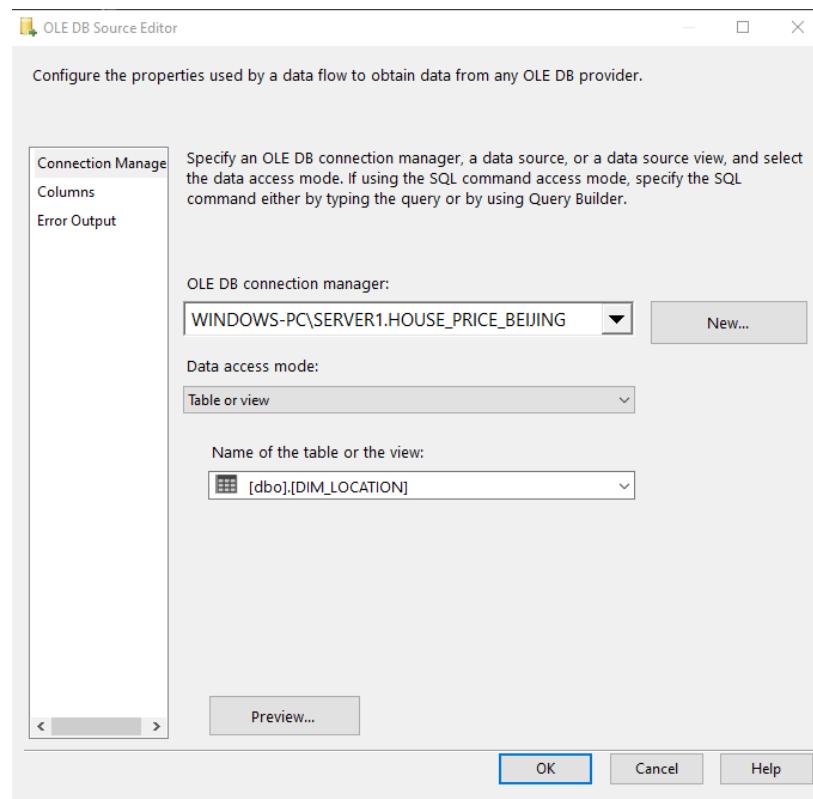
Hình 53: Làm data source cho bảng DIM_HOUSE_RAW mới

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.

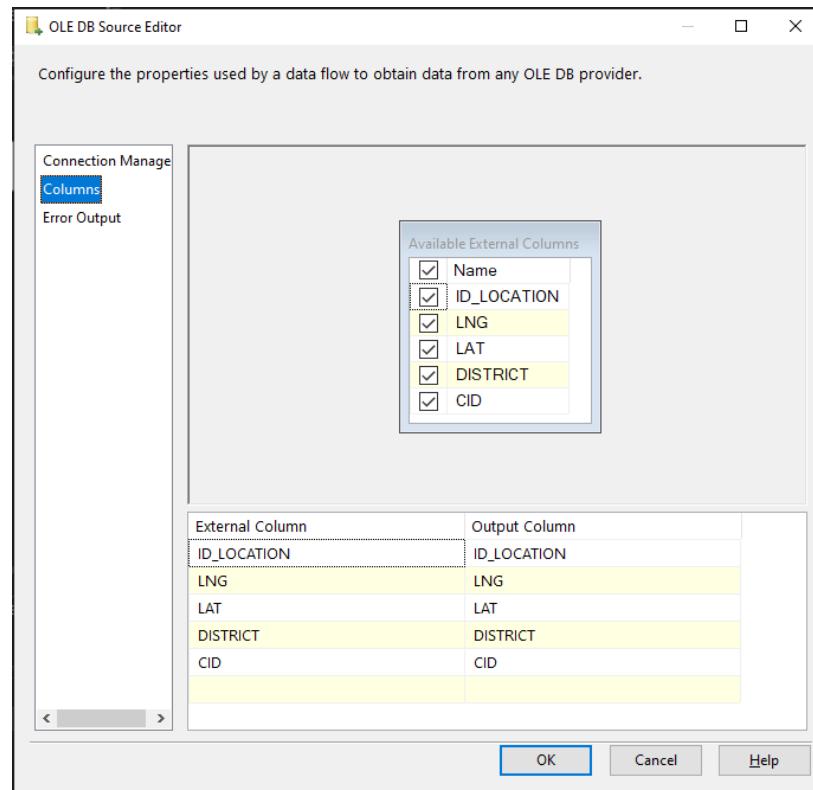


Hình 54: Xem xét các cột được ánh xạ

Bước 5. Tương tự thực hiện chọn ánh xạ cột cho **DIM_LOCATION**

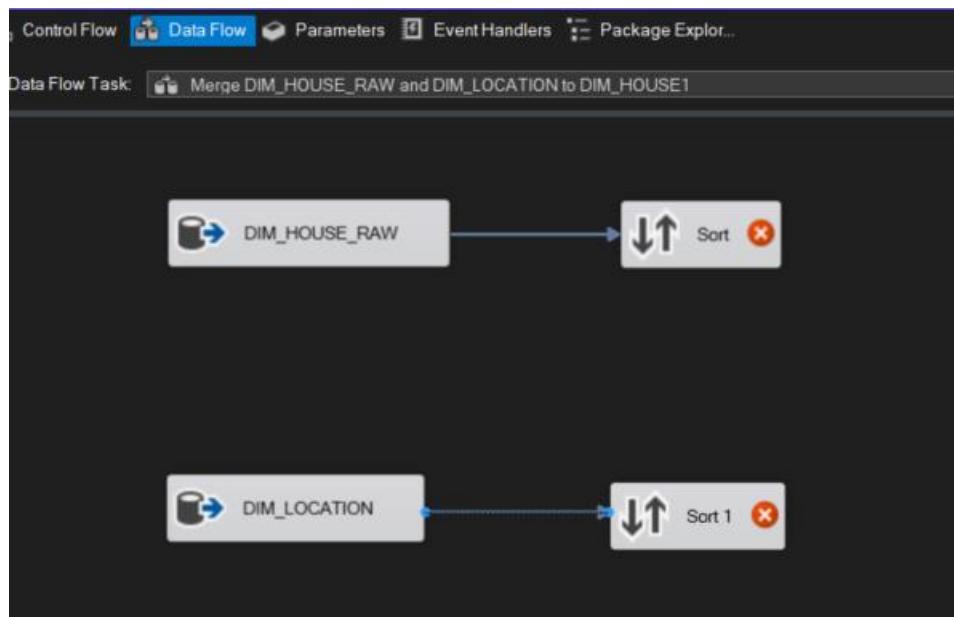


Hình 55: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (1)



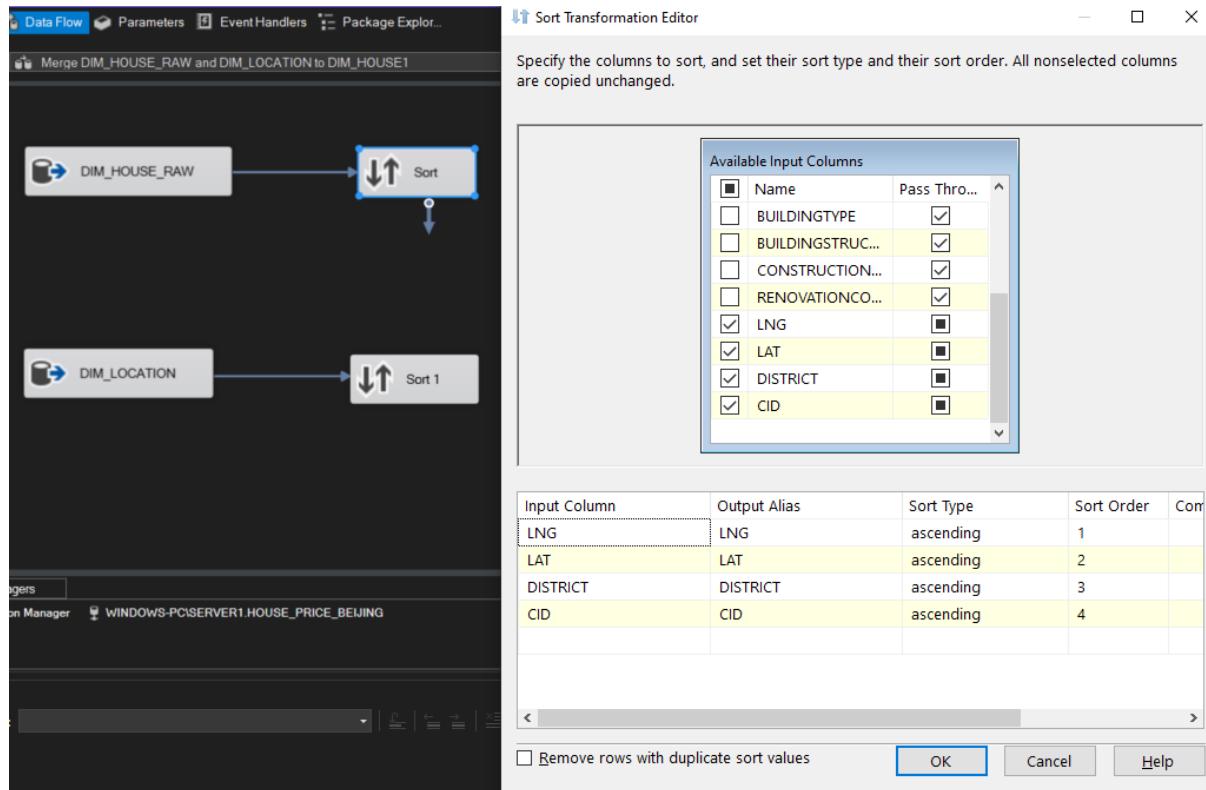
Hình 56: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (2)

Bước 6. Tạo 2 Sort là Sort và Sort1 tương ứng với mỗi Source.

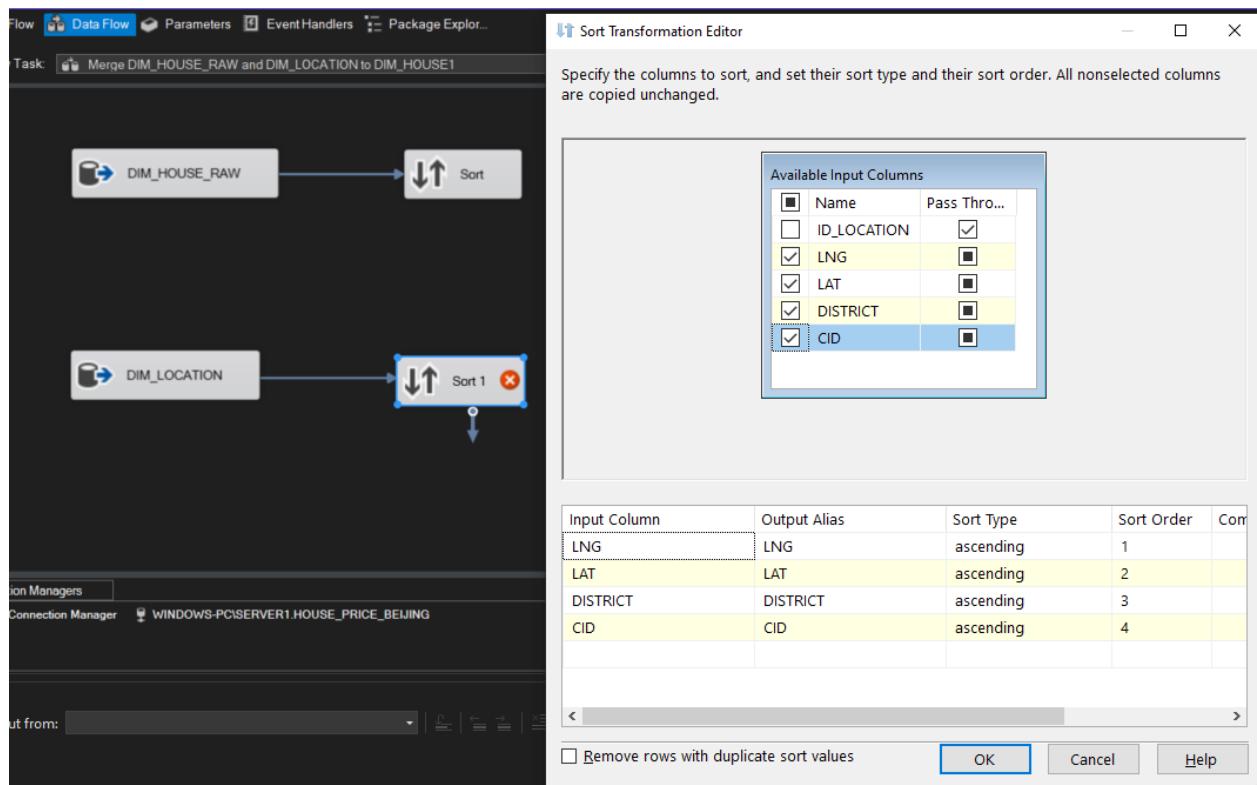


Hình 57: Tạo 2 Sort tương ứng với mỗi Source

Bước 7. Ở Sort, click chuột phải chọn Edit và chọn các cột LNG, LAT, DISTRICT, CID theo thứ tự giống với bảng **DIM_LOCATION** để chuẩn bị cho quá trình merge.



Hình 58: Chuẩn bị cho quá trình merge (I)



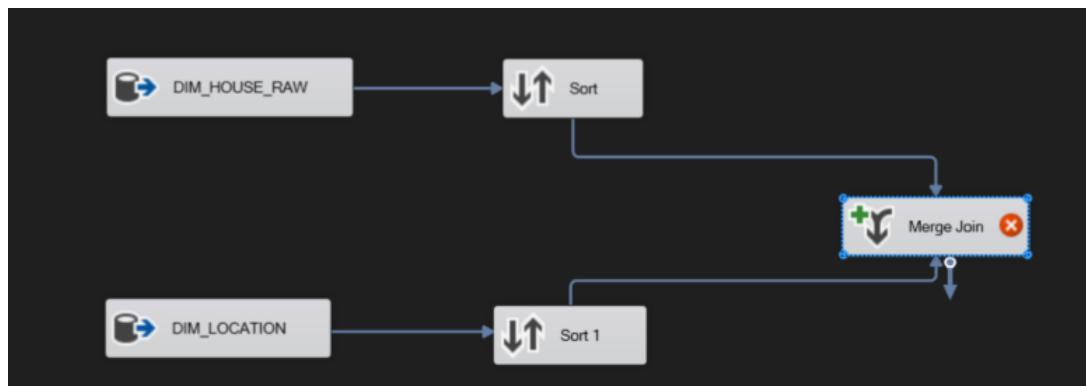
Hình 59: Chuẩn bị cho quá trình merge (2)

Bước 8. Tạo một **Merge Join** và nối với **Sort**, tiếp theo ta chọn **Merge Join Left** Input để giữ lại toàn bộ các dòng trong bảng **DIM_HOUSE_RAW** bất kể có kết quả khi thực hiện phép kết trái với cột ID của bảng **DIM_LOCATION** hay không.



Hình 60: Nối Sort với Merge Join

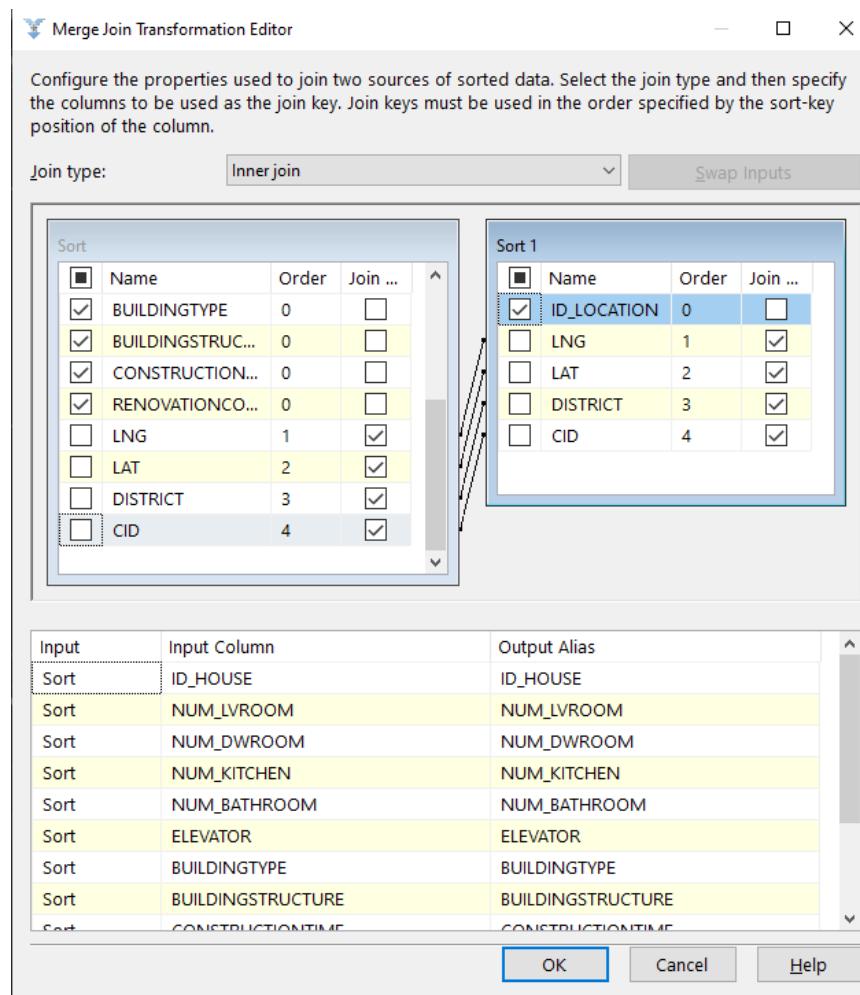
- Nối **Sort1** với **Merge Join**



Hình 61: Nối Sort1 với Merge Join

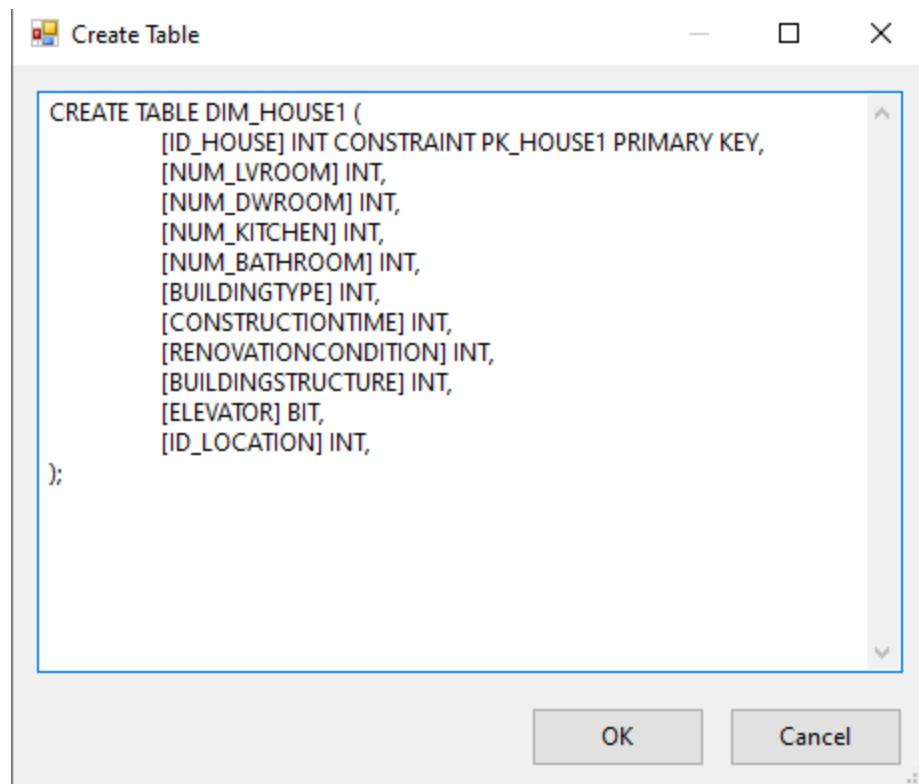
Bước 9. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy 4 thuộc tính **LNG, LAT, DISTRICT** và **CID**.

- Tiếp theo ta chọn **ID_LOCATION** ở **Sort1** để merge vào **DIM_HOME_RAW**
- Kết quả sau khi merge là bảng **DIM_HOME_RAW** không còn 4 thuộc tính **LNG, LAT, DISTRICT** và **CID** và có thêm 1 thuộc tính mới là **ID_LOCATION**



Hình 62: Tick chọn tất cả các cột của Sort nhưng không lấy 4 thuộc tính LNG, LAT, DISTRICT và CID

Bước 10. Tạo bảng **DIM_HOUSE1** từ một **OLE DB Destination** để chứa tất cả những gì đã merge



Hình 63: Câu lệnh SQL tạo bảng DIM_HOUSE1

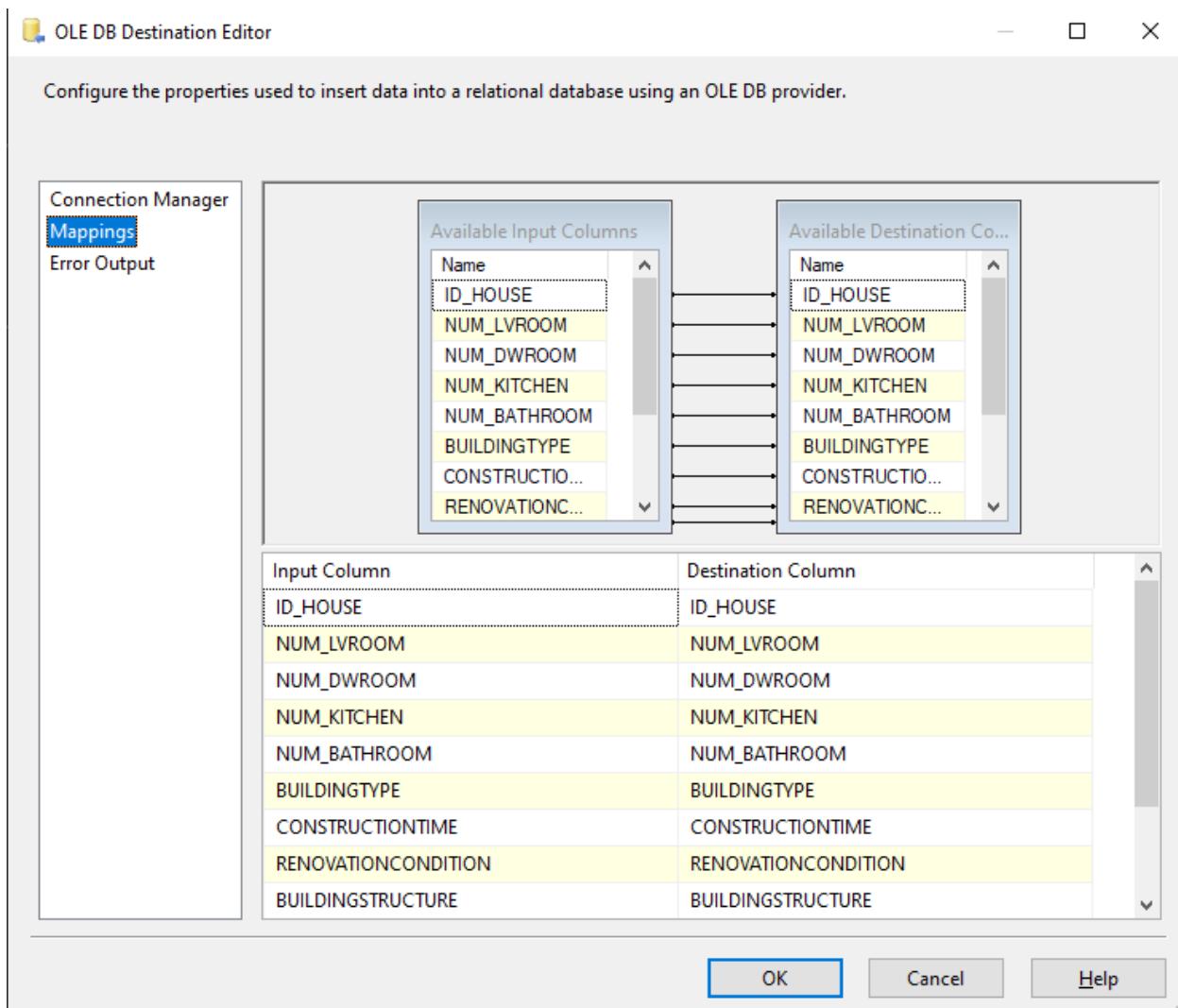
*Nội dung câu lệnh SQL tạo bảng DIM_HOUSE1 như sau:

CREATE TABLE DIM_HOUSE1 (

```
[ID_HOUSE] INT CONSTRAINT PK_HOUSE1 PRIMARY KEY,  
[NUM_LVROOM] INT,  
[NUM_DWROOM] INT,  
[NUM_KITCHEN] INT,  
[NUM_BATHROOM] INT,  
[BUILDINGTYPE] INT,  
[CONSTRUCTIONTIME] INT,  
[RENOVATIONCONDITION] INT,  
[BUILDINGSTRUCTURE] INT,  
[ELEVATOR] BIT,  
[ID_LOCATION] INT,
```

);

Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

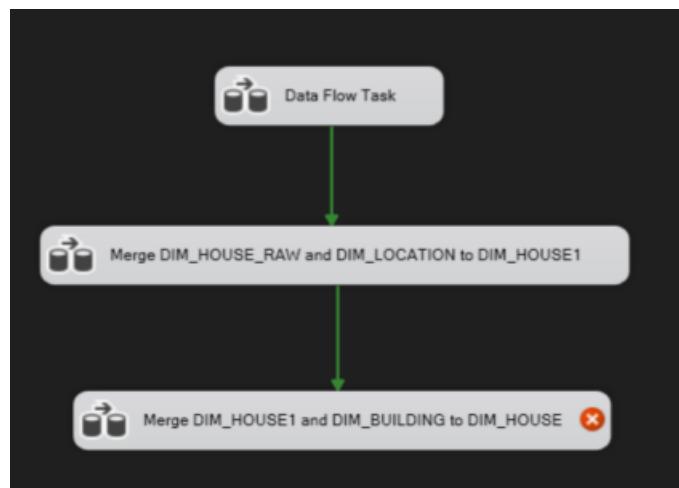


Hình 64: Xem xét việc ánh xạ các cột dữ liệu

- Tiếp tục quá trình **merge** bảng **DIM_HOUSE1** với các **Dimension** còn lại để có 1 bảng **DIM_HOUSE** hoàn chỉnh.

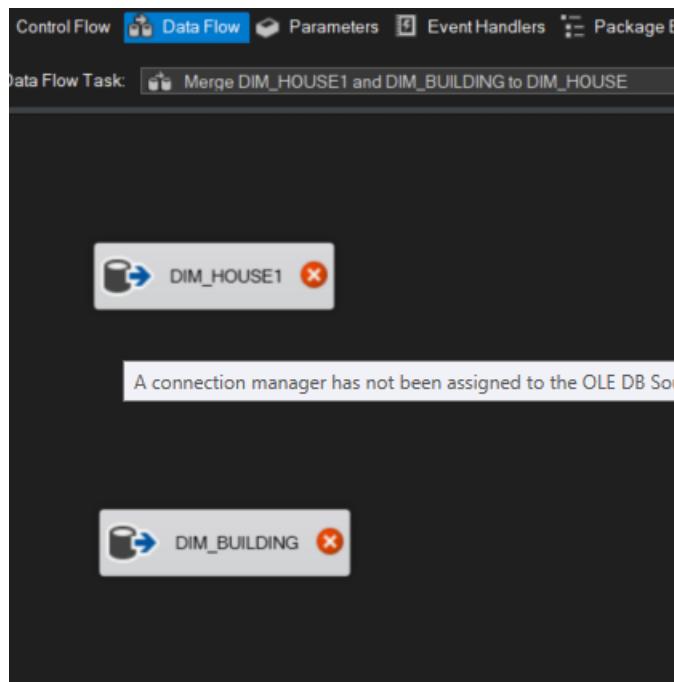
4.7. Merge DIM_HOUSE1 và DIM_BUILDING vào DIM_HOUSE

Bước 1. Ở tab **Control Flow**, tạo thêm một **Data Flow Task** và đổi tên **Data Flow Task** này là “**Merge DIM_HOUSE1 and DIM_BUILDING to DIM_HOUSE**”



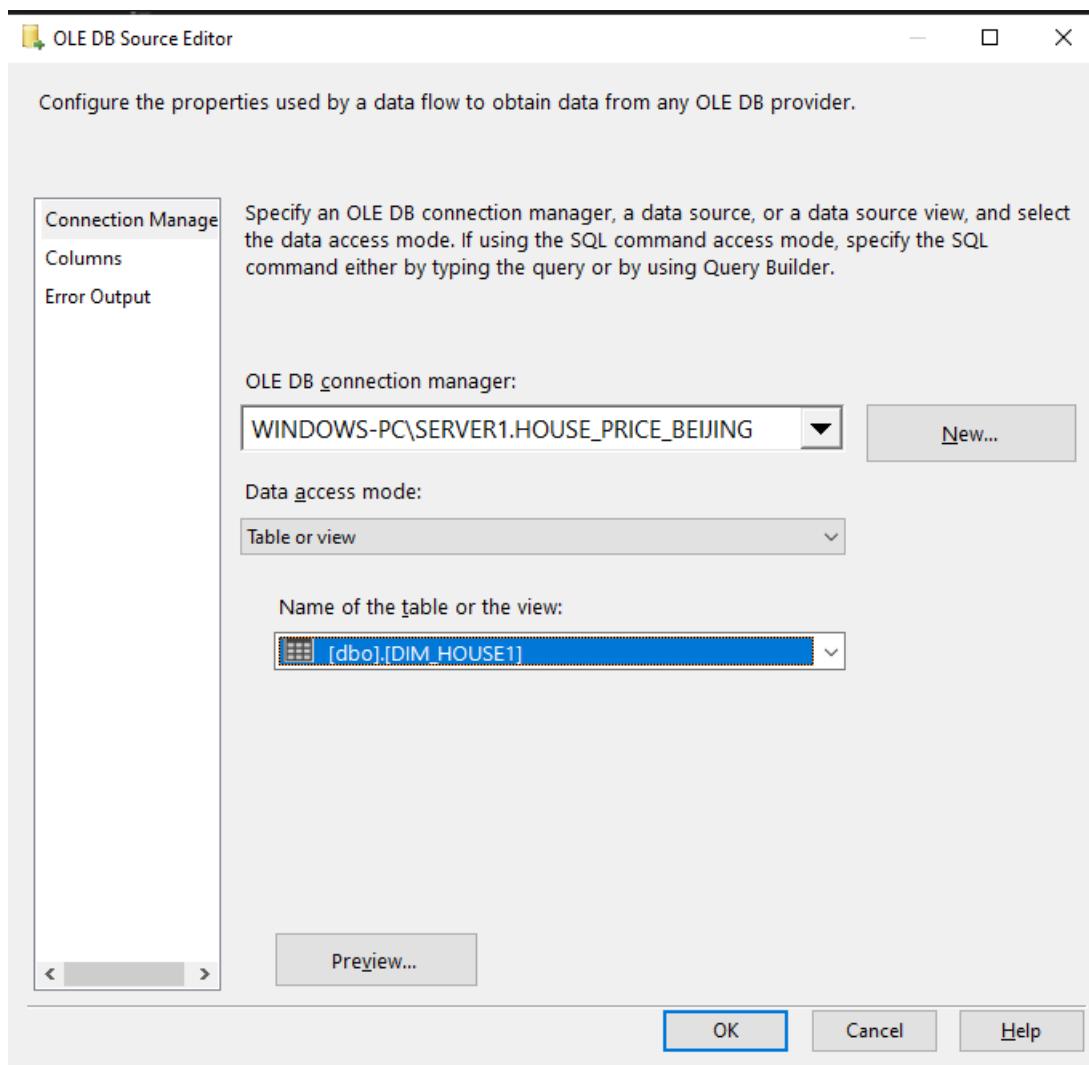
Hình 65: Tạo thêm một Data Flow Task và đổi tên

Bước 2. Click chuột phải vào **Data Flow Task** nói trên và chọn **Edit**, trong tab **Data Flow** ta tạo 2 **OLE DB Source** và đổi tên **DIM_HOME1** và **DIM_BUILDING**



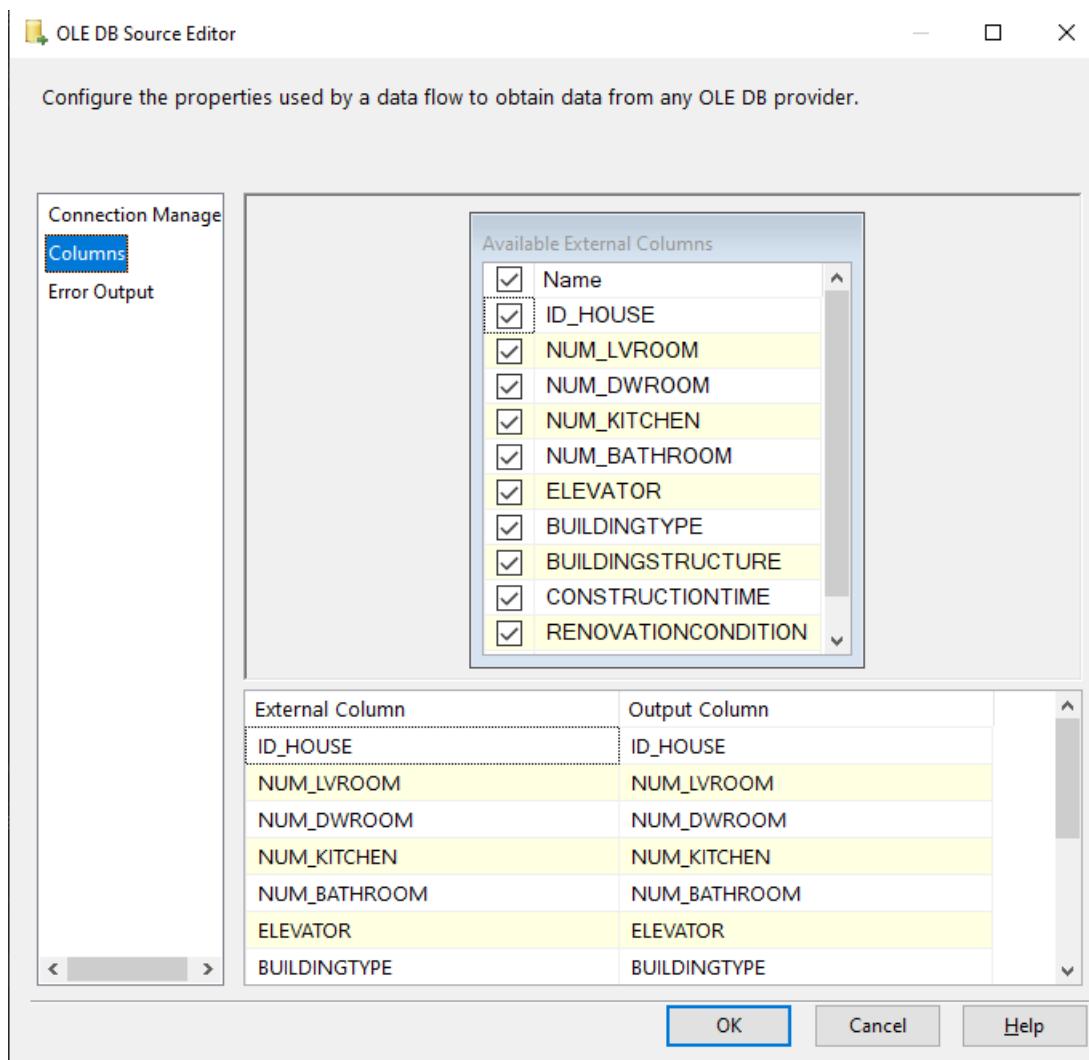
Hình 66: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải vào **DIM_HOME1** chọn **Edit**, sau đó chọn bảng **DIM_HOME1** đã được tạo khi merge **DIM_HOME_RAW** và **DIM_LOCATION** làm **data source**.



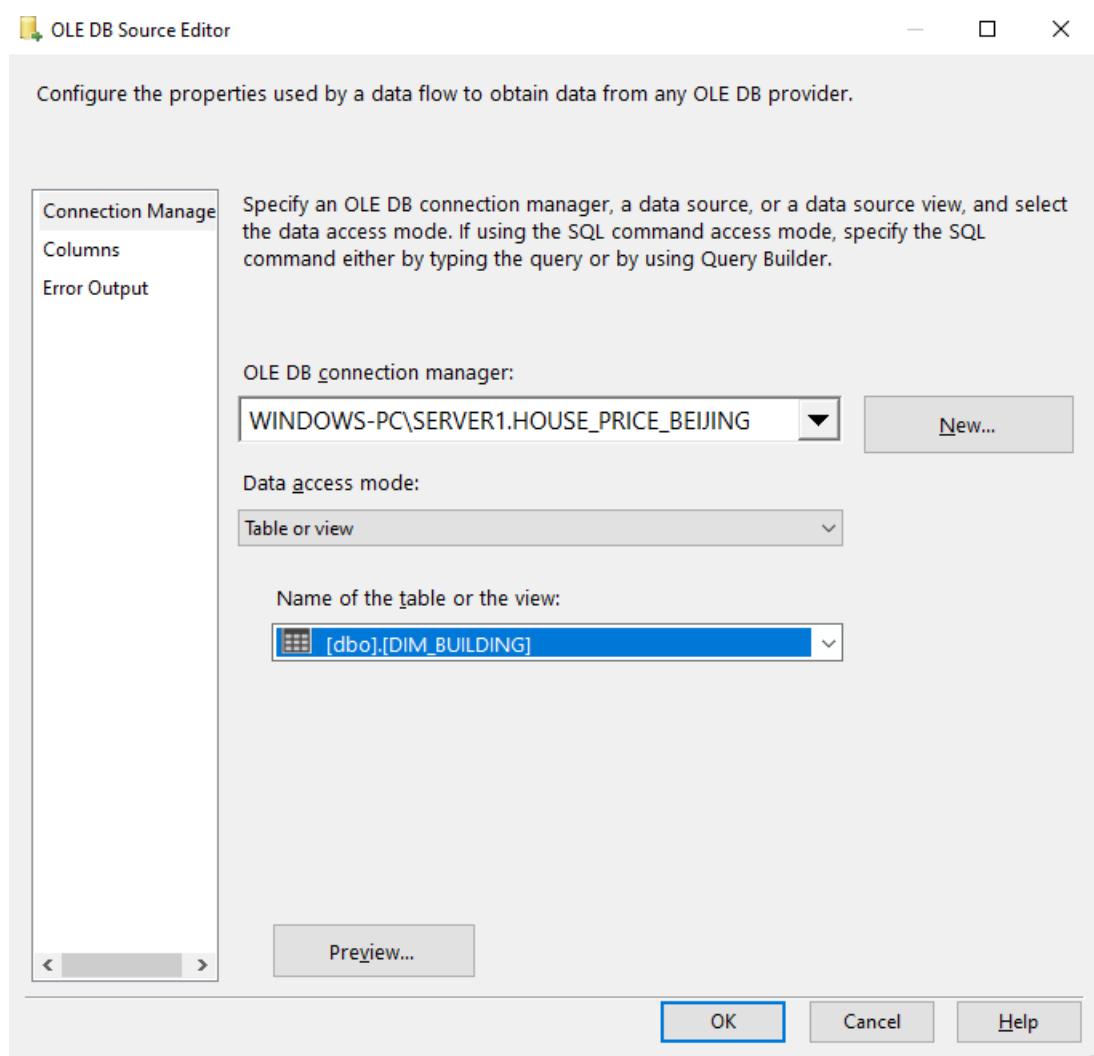
Hình 67: Chọn bảng DIM_HOUSE1 đã được tạo khi merge

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.



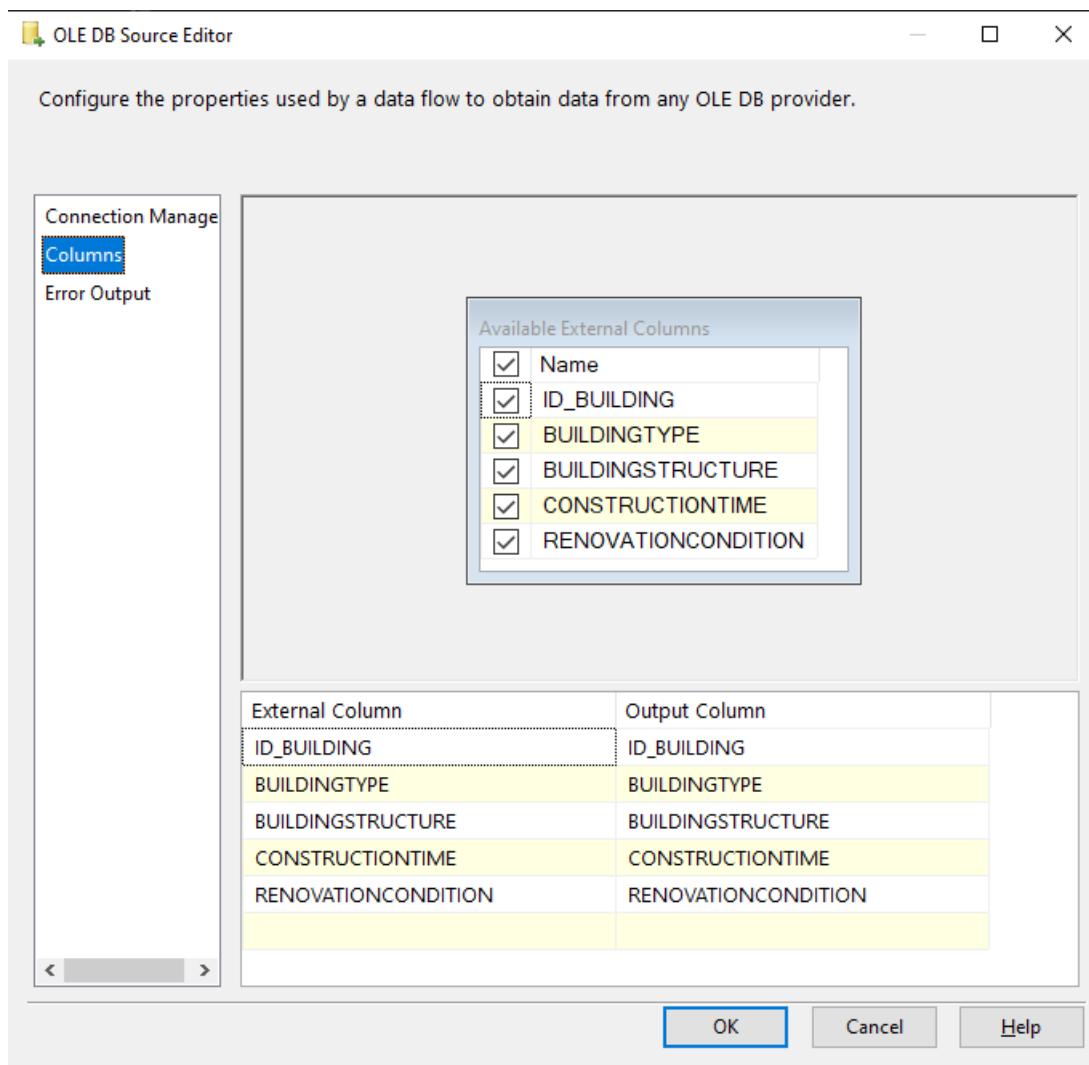
Hình 68: Xem xét các cột được ánh xạ

Bước 5. Thực hiện chọn ánh xạ các cột cho **DIM_BUILDING**



Hình 69: Thực hiện chọn ánh xạ các cột cho DIM_BUILDING

- Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.



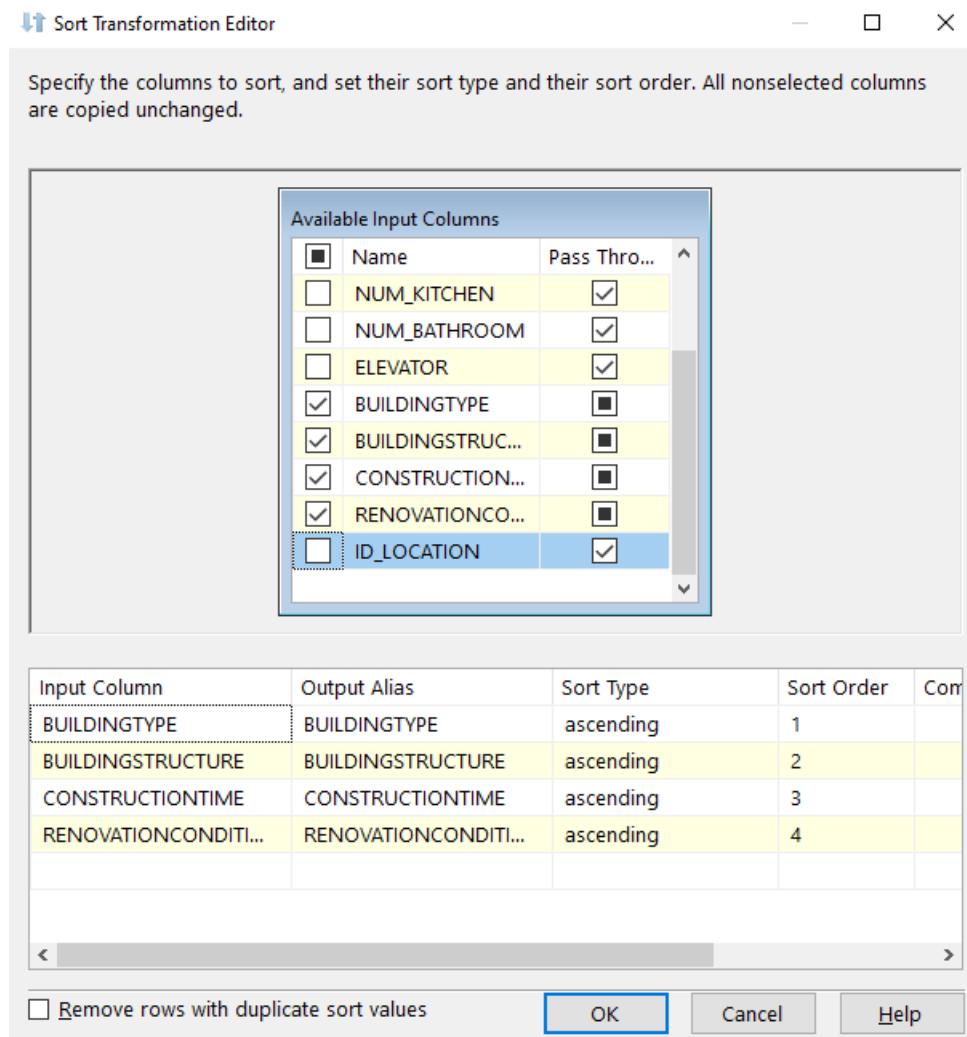
Hình 70: Xem xét các cột được ánh xạ

Bước 6. Tạo 2 Sort tương ứng với mỗi Source

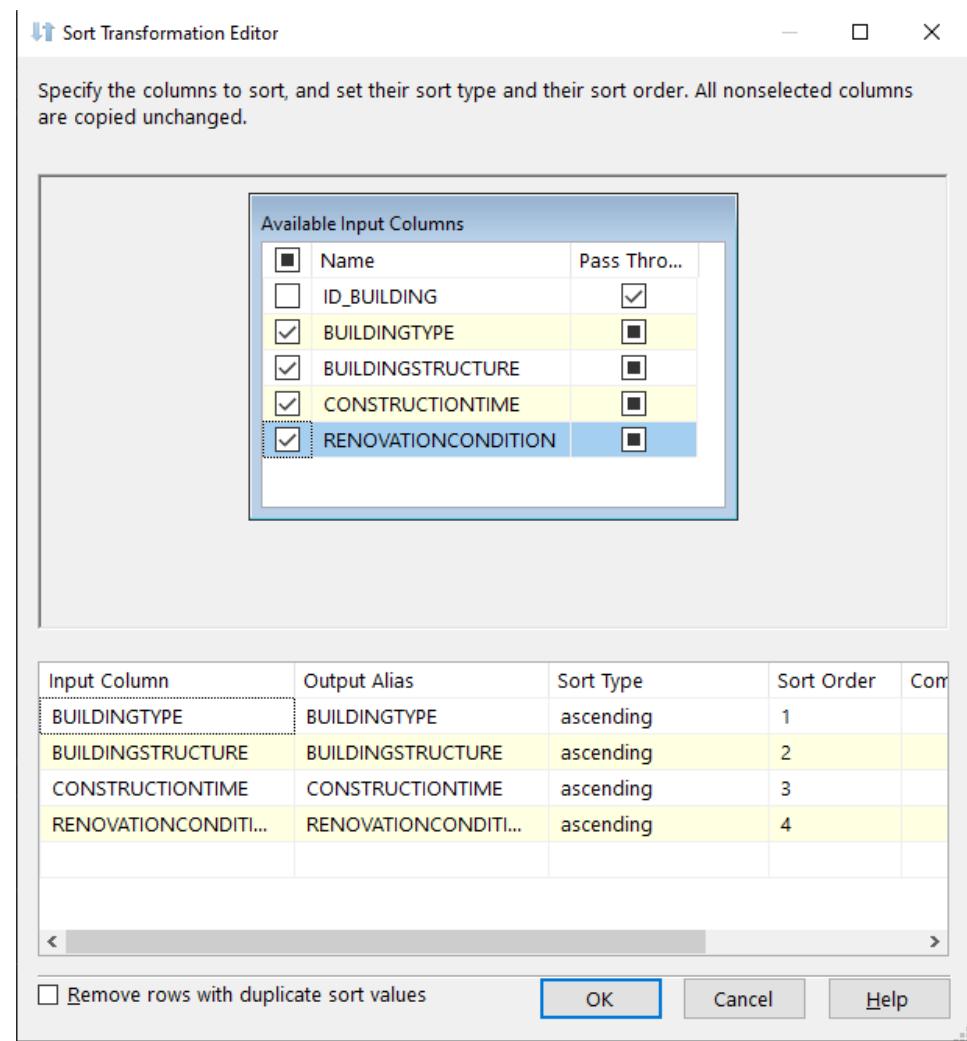


Hình 71: Tạo 2 Sort tương ứng với mỗi Source

Bước 7. Tại **Sort**, click chuột phải chọn **Edit** và chọn các cột **BUILDINGTYPE**, **BUILDINGSTRUCTURE**, **CONSTRUCTIONTIME**, **RENOVATIONCONDITION** theo thứ tự giống với bảng **DIM_BUILDING** để chuẩn bị cho quá trình merge

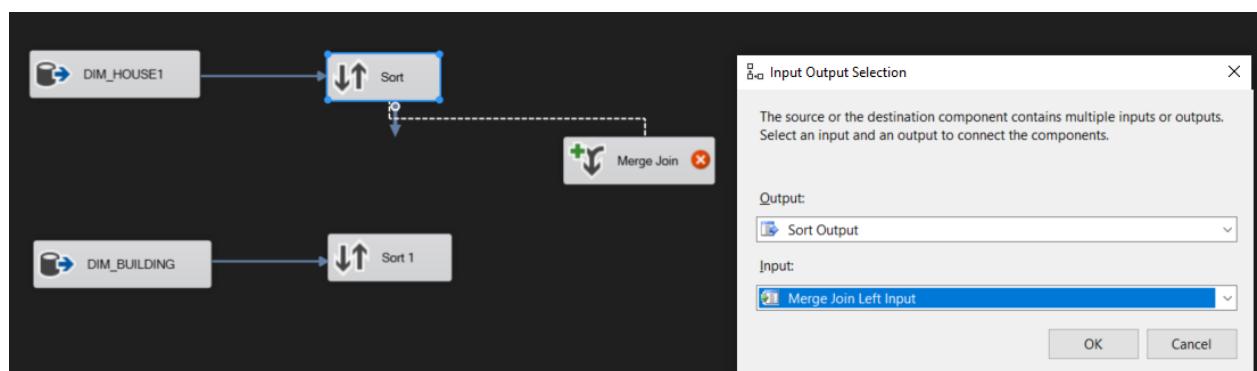


Hình 72: Chọn các cột theo thứ tự giống với bảng DIM_BUILDING để chuẩn bị cho quá trình merge(1)



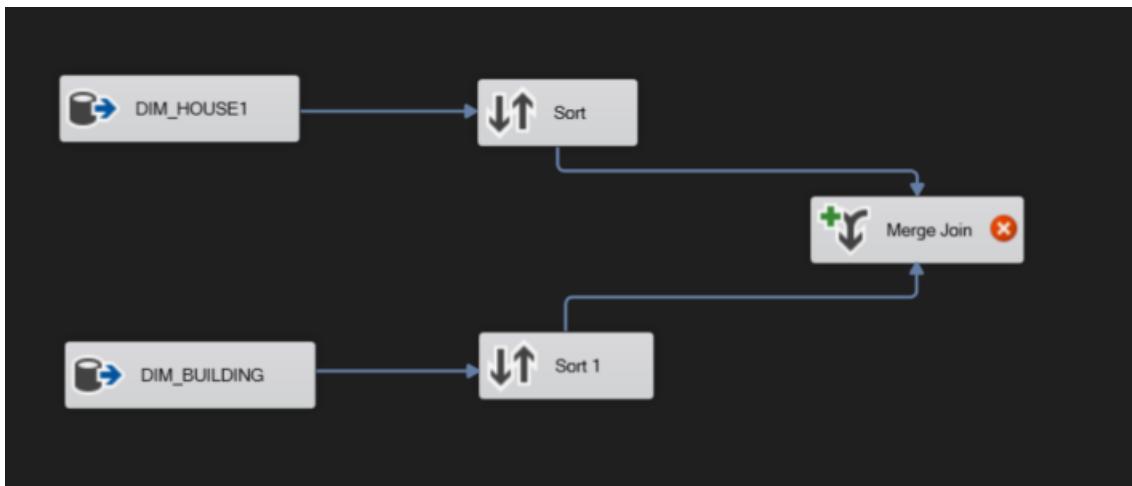
Hình 73: : Chọn các cột theo thứ tự giống với bảng DIM_BUILDING để chuẩn bị cho quá trình merge(2)

Bước 8. Tạo một Merge Join và nối với Sort, tiếp theo chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng DIM_HOUSE1 bất kể có kết quả khi thực hiện phép kết trái với cột ID của bảng DIM_BUILDING hay không.



Hình 74: Chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng DIM_HOUSE1

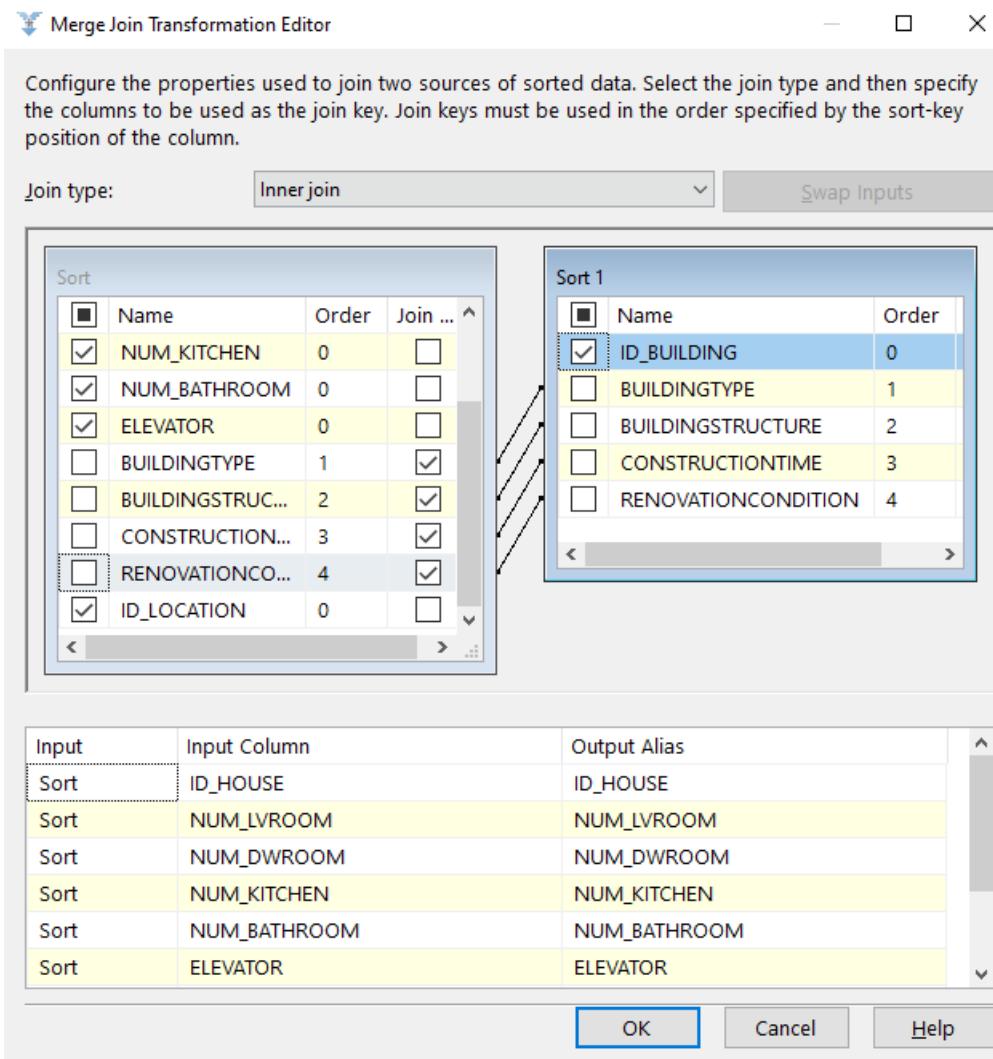
- Nối Sort1 với Merge Join



Hình 75: Nối Sort1 với Merge Join

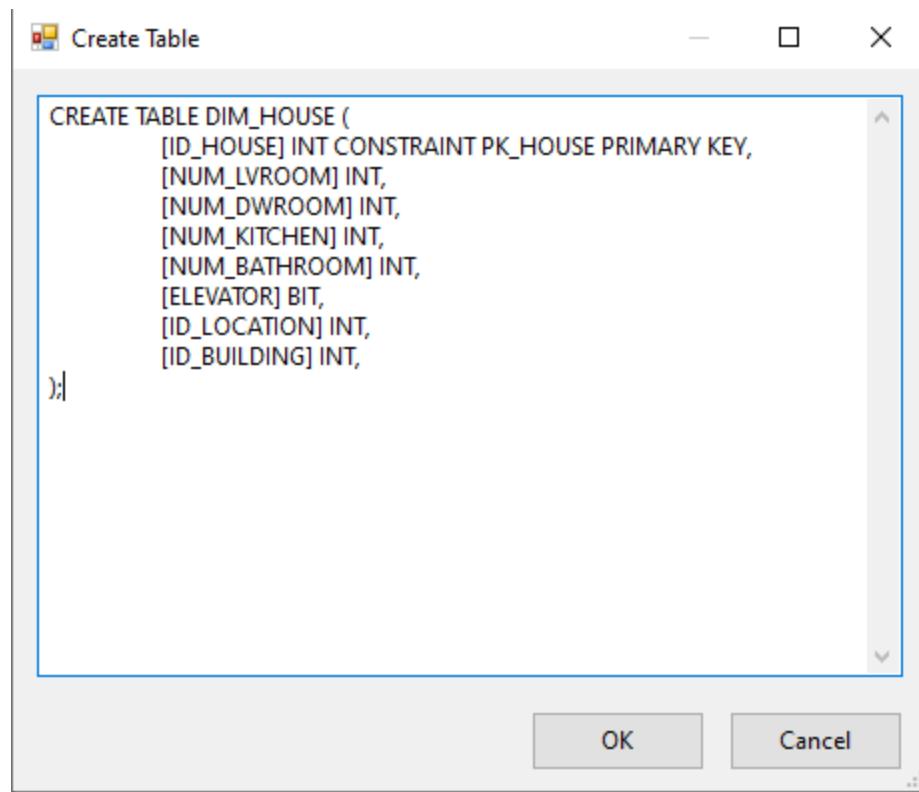
Bước 9. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy 4 thuộc tính **BUILDINGTYPE**, **BUILDINGSTRUCTURE**, **CONSTRUCTIONTIME**, **RENOVATIONCONDITION**.

- Tiếp theo ta chọn **ID_BUILDING** ở **Sort1** để merge vào **DIM_HOUSE1**
- Kết quả sau khi merge là bảng **DIM_HOUSE1** không còn 4 thuộc tính **BUILDINGTYPE**, **BUILDINGSTRUCTURE**, **CONSTRUCTIONTIME**, **RENOVATIONCONDITION** và có thêm 1 thuộc tính mới là **ID_BUILDING**



Hình 76: Kết quả sau khi merge bảng DIM_HOUSE1

Bước 10. Tạo bảng **DIM_HOUSE** từ một **OLE DB Destination** để chứa tất cả những gì đã **merge**



Hình 77: Câu lệnh SQL tạo bảng DIM_HOUSE

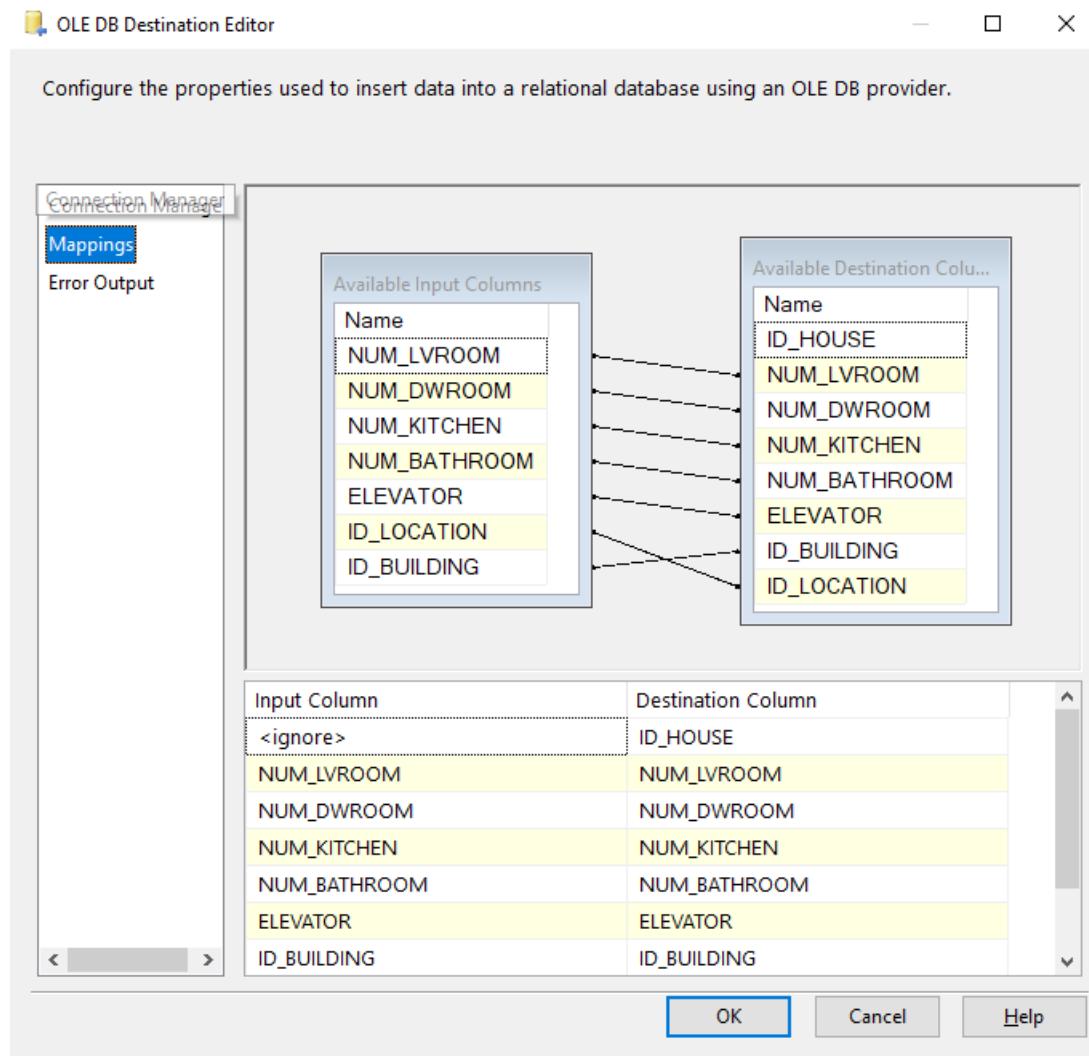
*Nội dung câu lệnh SQL tạo bảng DIM_HOUSE như sau:

CREATE TABLE DIM_HOUSE (

```
[ID_HOUSE] INT CONSTRAINT PK_HOUSE PRIMARY KEY,
[NUM_LVROOM] INT,
[NUM_DWROOM] INT,
[NUM_KITCHEN] INT,
[NUM_BATHROOM] INT,
[ELEVATOR] BIT,
[ID_LOCATION] INT,
[ID_BUILDING] INT,
```

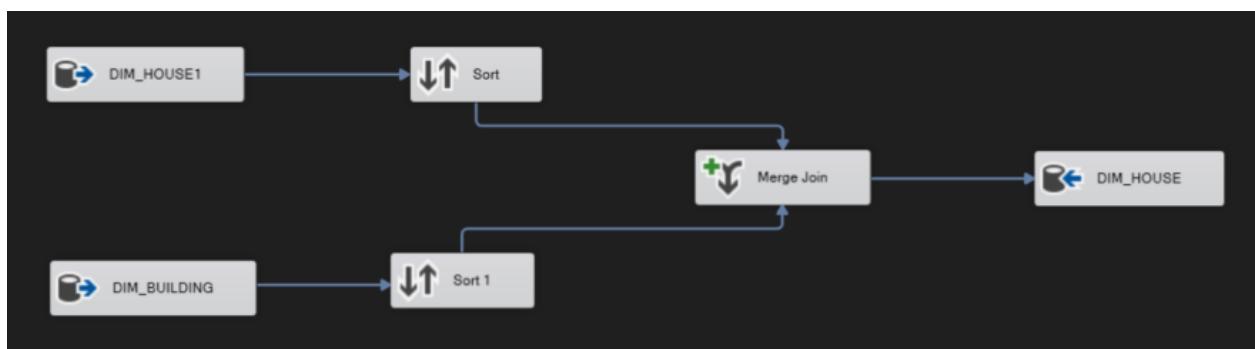
);

- Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu



Hình 78: Xem xét việc ánh xạ các cột dữ liệu

- Nhấn OK:

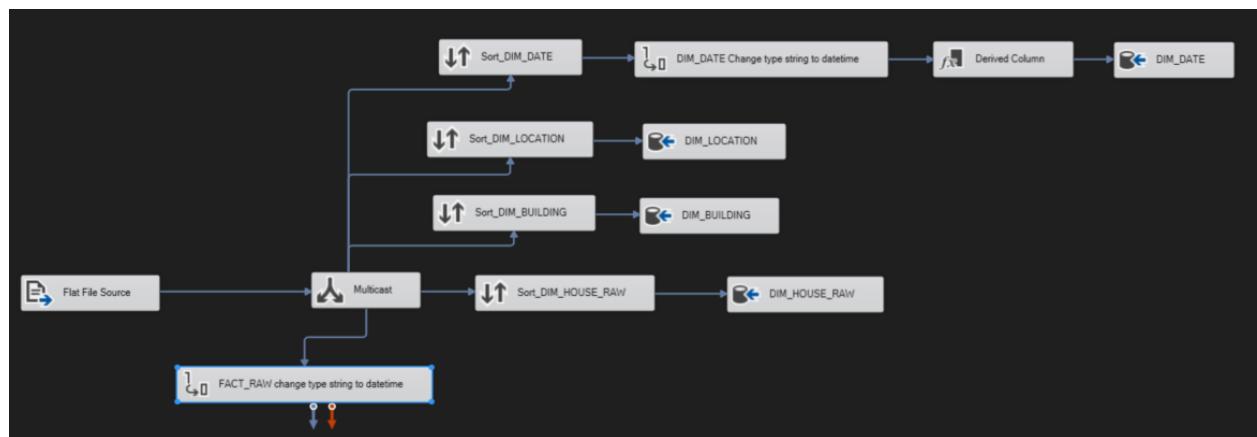


Hình 79: Kết quả

4.8. Bảng FACT_RAW

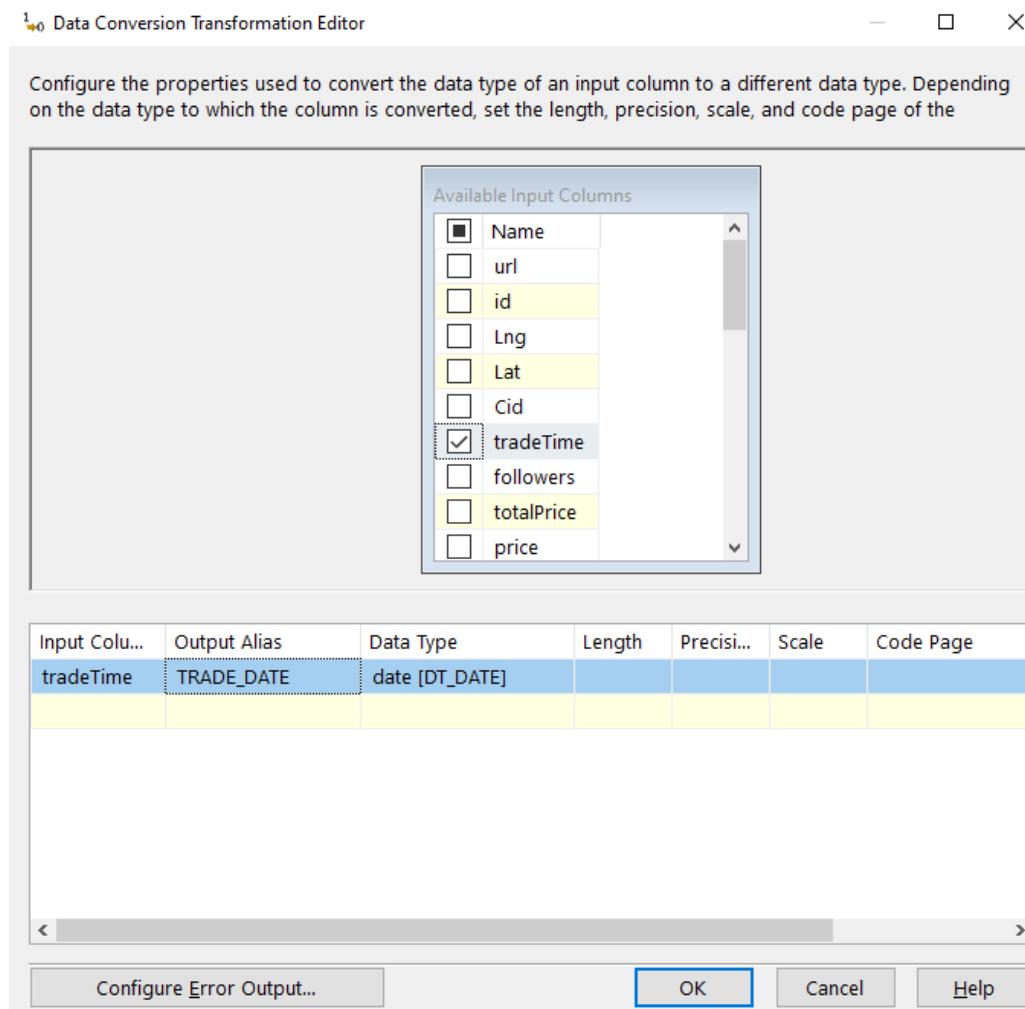
Ta thấy bảng **Fact** có một cột **TRADE_DATE** chứa kiểu dữ liệu DateTime, tuy nhiên kiểu dữ liệu DateTime khi lấy từ dữ liệu gốc (file .csv) sẽ được mặc định là kiểu string. Ta dùng **Data Conversion** để chuyển đổi kiểu string về dạng DateTime.

Bước 1: Tạo một **Data Conversion** và đặt tên là “**FACT_RAW change type string to datetime**”



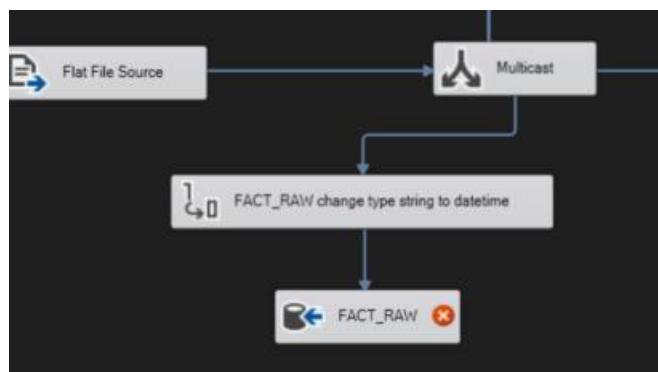
Hình 80: Tạo một Data Conversion

Bước 2: Click phải chuột vào **Data Conversion** trên và chọn **Edit**, tick chọn cột **tradeTime**, ở cột **Data Type** ta thấy mặc định kiểu dữ liệu là string, ta chọn lại kiểu dữ liệu cho cột này **date[DT_DATE]** sau đó nhấn nút **OK**.



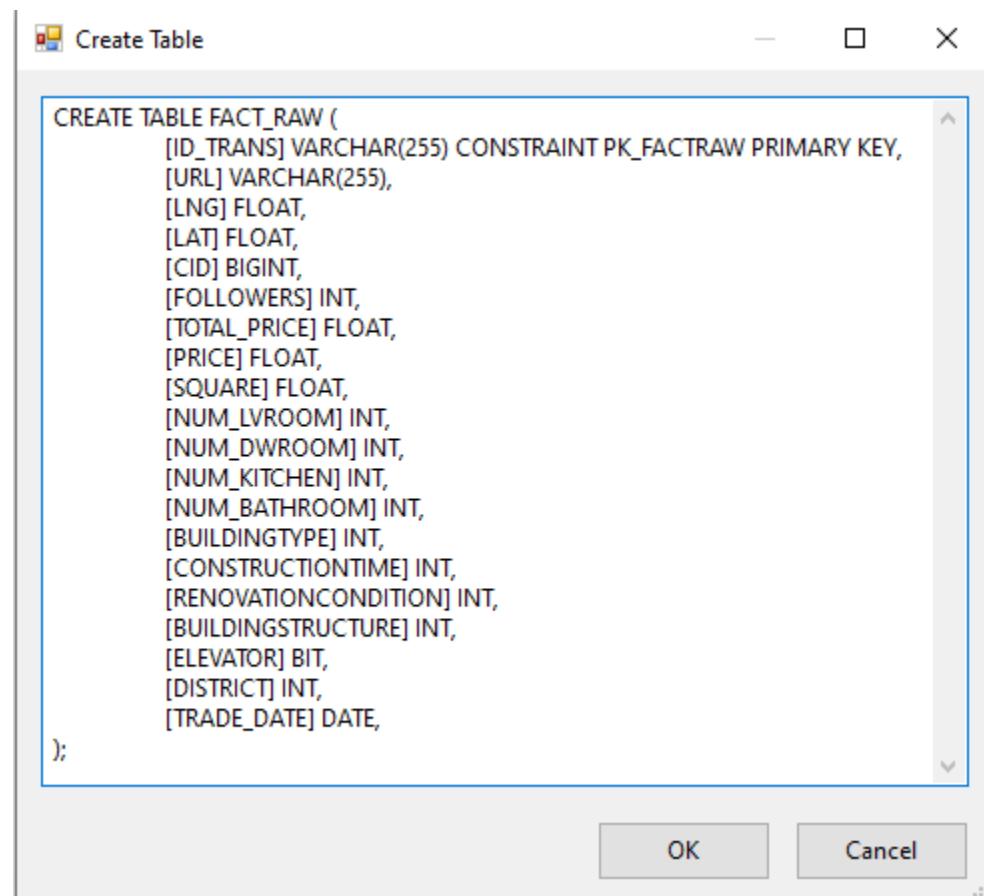
Hình 81: Chọn lại kiểu dữ liệu cho cột này date[DT_DATE]

Bước 3. Tiến hành tạo bảng **Fact** và đặt tên là **FACT_RAW** từ một OLE DB Destination



Hình 82: Tạo bảng Fact

Bước 4. Click chuột phải và chọn **Edit** để tạo bảng **FACT_RAW** có các cột là tất cả các cột từ dữ liệu gốc và chứa tất cả các dòng dữ liệu



Hình 83: Câu lệnh SQL tạo bảng FACT_RAW

*Nội dung câu lệnh SQL tạo bảng FACT_RAW như sau:

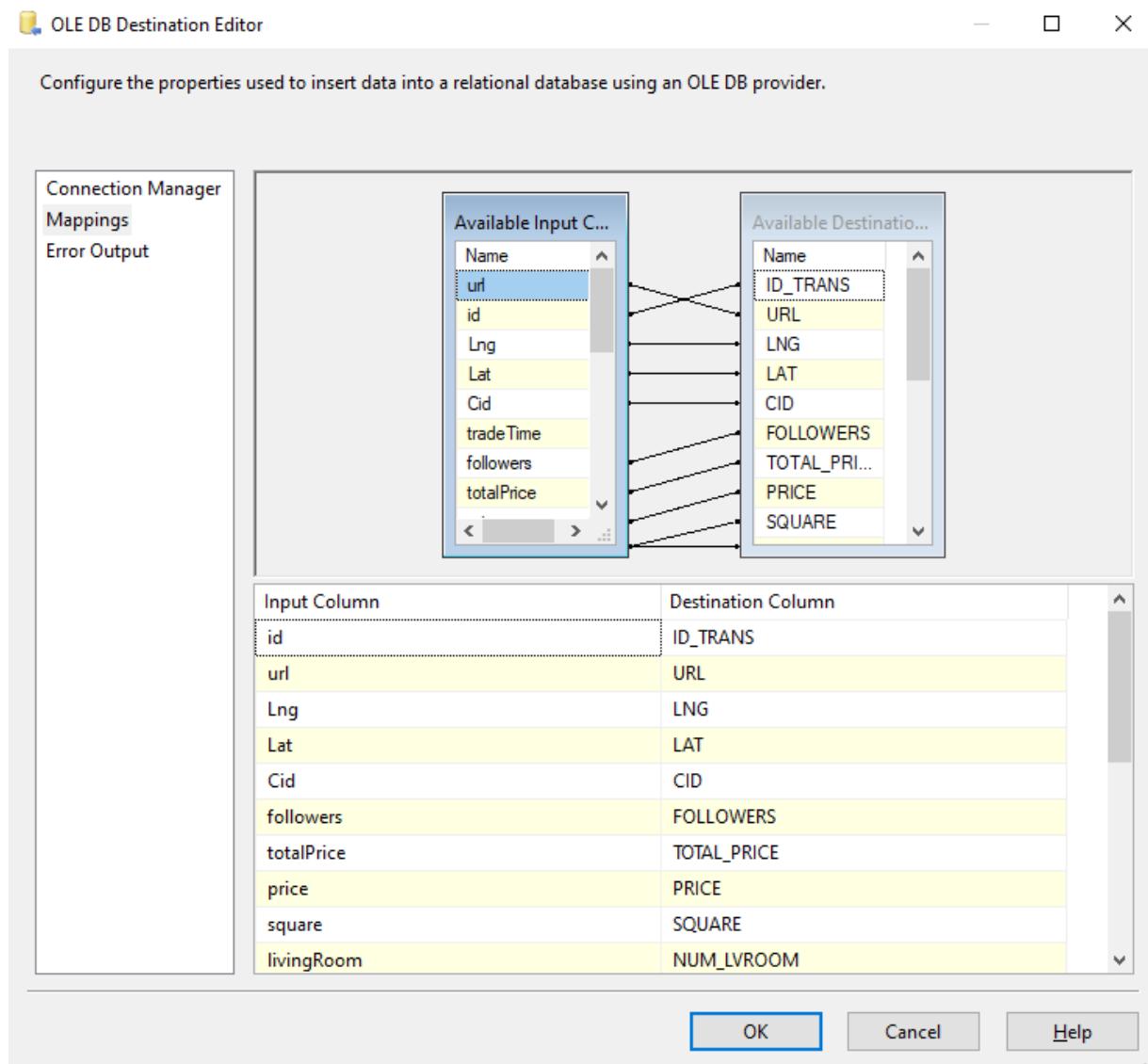
CREATE TABLE FACT_RAW (

```
[ID_TRANS] VARCHAR(255) CONSTRAINT PK_FACTRAW PRIMARY KEY,
[URL] VARCHAR(255),
[LNG] FLOAT,
[LAT] FLOAT,
[CID] BIGINT,
[FOLLOWERS] INT,
[TOTAL_PRICE] FLOAT,
```

```
[PRICE] FLOAT,  
[SQUARE] FLOAT,  
[NUM_LVROOM] INT,  
[NUM_DWROOM] INT,  
[NUM_KITCHEN] INT,  
[NUM_BATHROOM] INT,  
[BUILDINGTYPE] INT,  
[CONSTRUCTIONTIME] INT,  
[RENOVATIONCONDITION] INT,  
[BUILDINGSTRUCTURE] INT,  
[ELEVATOR] BIT,  
[DISTRICT] INT,  
[TRADE_DATE] DATE,  
);
```

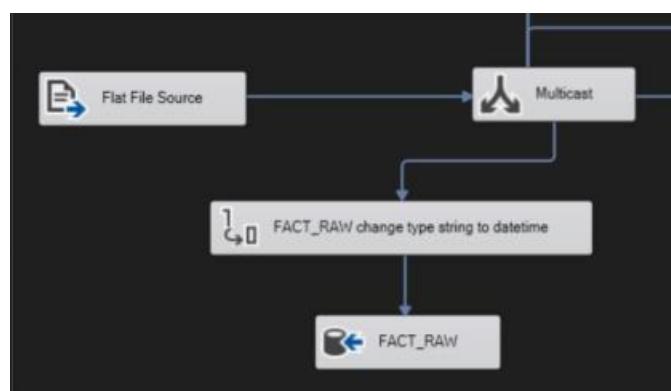
Bước 5. Tiếp đến ta cần chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu. Ta thấy **ID** của bảng **FACT_RAW** cũng chính là id của tập dữ liệu nên ta chọn ánh xạ cột id trong **Input Column** vào cột **ID_TRANS** của bảng **FACT_RAW**

- Cuối cùng nhấn nút **OK** để hoàn tất quá trình tạo bảng.



Hình 84: Ánh xạ cột id trong Input Column vào cột ID_TRANS của bảng FACT_RAW

- Nhận OK:

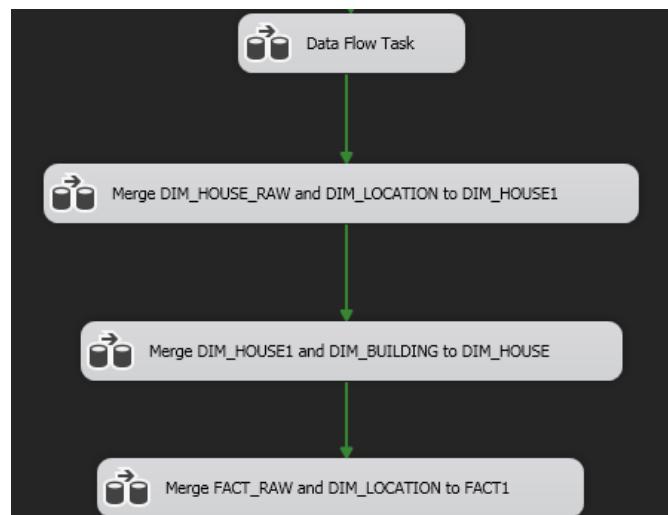


Hình 85: Kết quả

- Tiếp theo đây ta sẽ thực hiện quá trình lần lượt loại bỏ các cột dữ liệu trùng của bảng **Fact** với các **Dimension**, thực hiện thêm khóa ngoại vào bảng **Fact** nhằm thu gọn bảng **Fact**, tối ưu hóa quá trình phân tích dữ liệu.

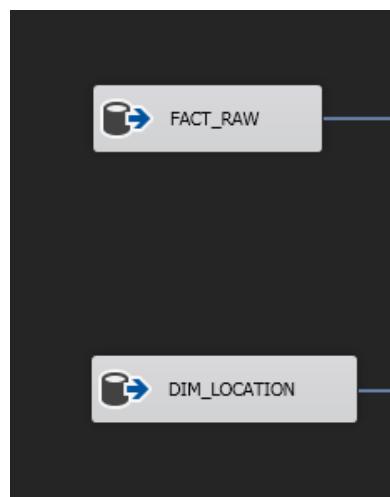
4.8. Merge FACT_RAW và DIM_LOCATION vào FACT1

Bước 1. Ở tab **Control Flow**, tạo hai **Data Flow Task** và đổi tên **Data Flow Task** thứ 2 là “**Merge FACT_RAW and DIM_LOCATION to FACT1**”



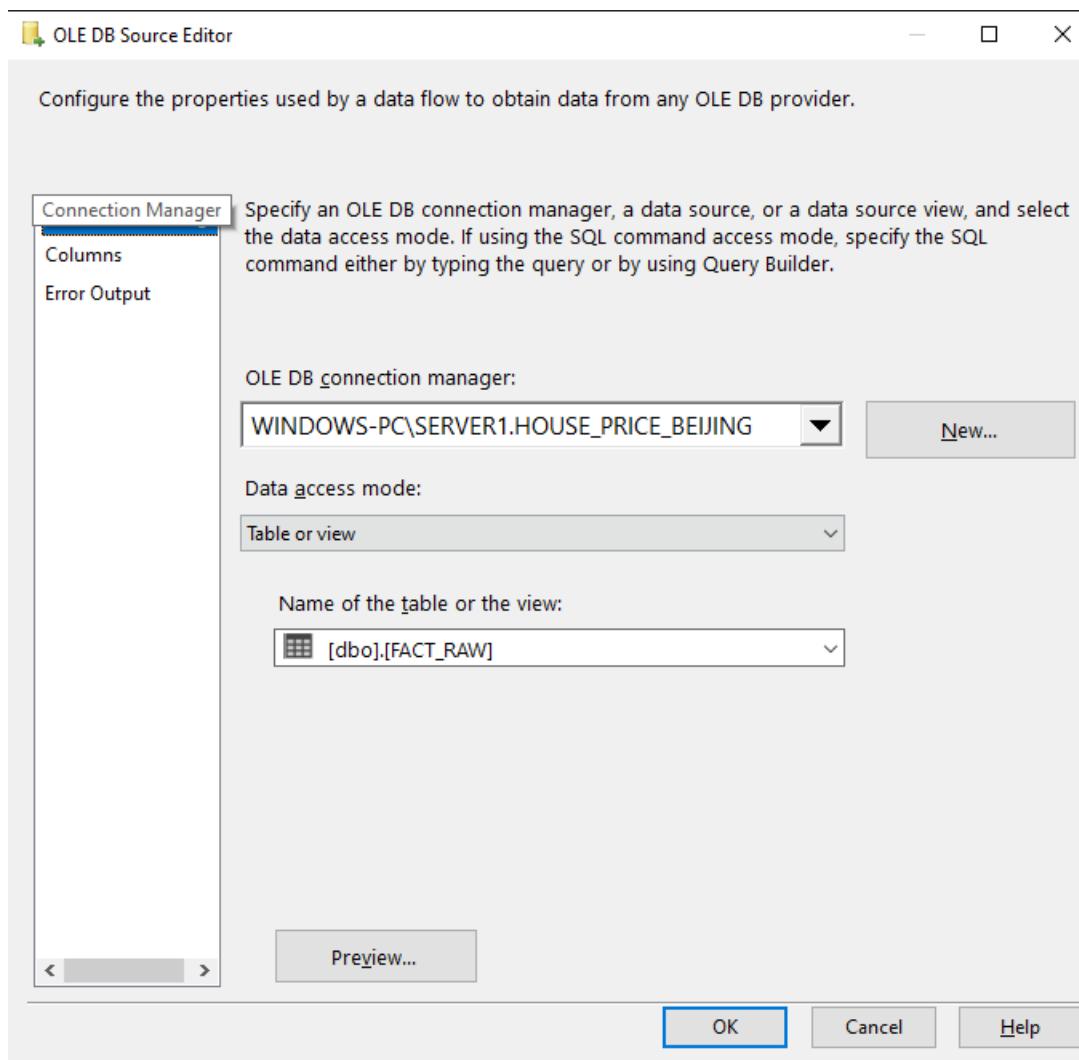
Hình 86: Tạo hai Data Flow Task và đổi tên

Bước 2. Click chuột phải vào **Data Flow Task** nói trên và chọn **Edit**, trong tab **Data Flow** ta tạo 2 **OLE DB Source** và đổi tên **FACT_RAW** và **DIM_LOCATION**



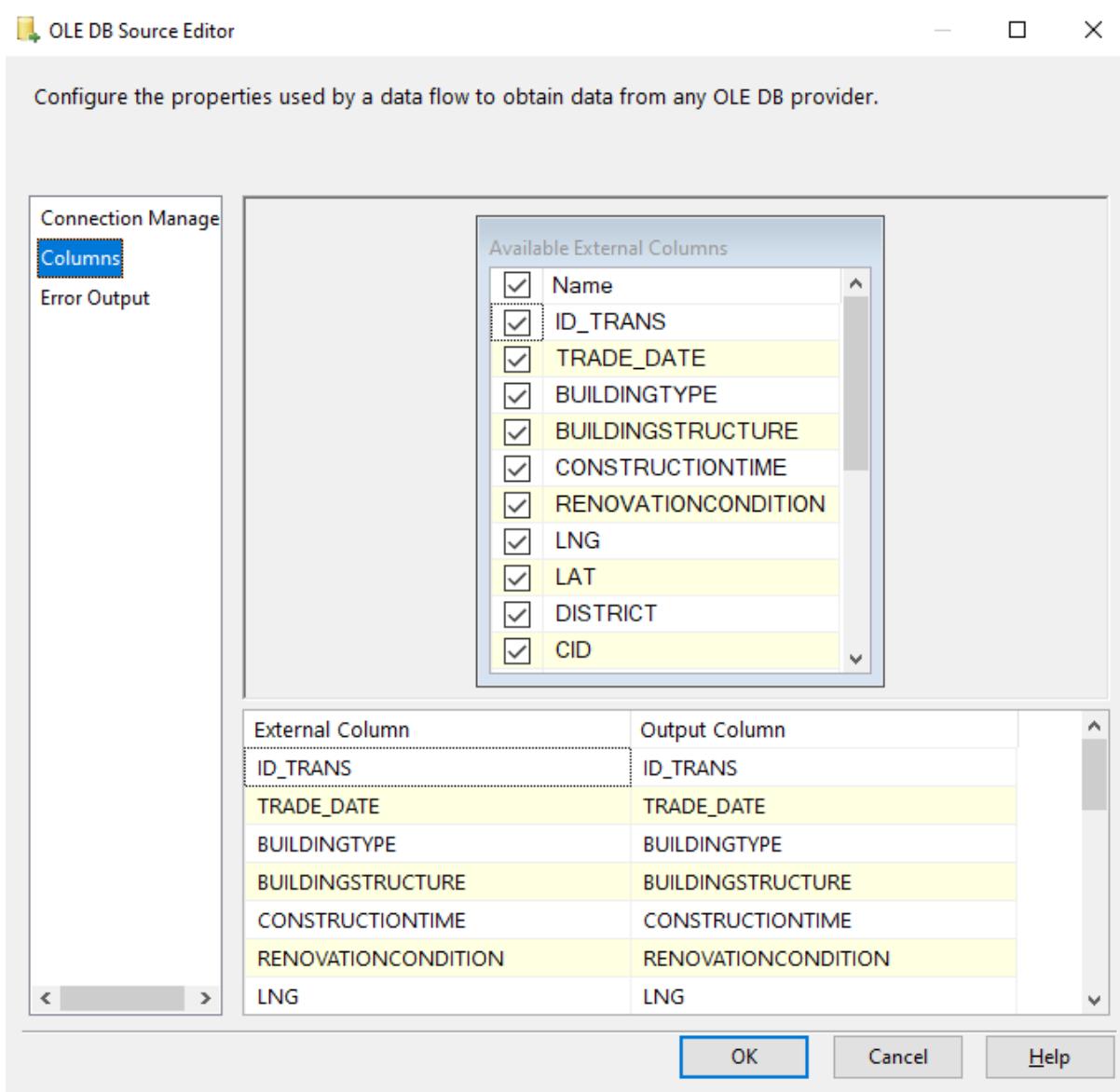
Hình 87: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải chọn **Edit**, sau đó chọn bảng **FACT_RAW** đã tạo trước đó làm data source cho bảng **FACT_RAW** mới này.



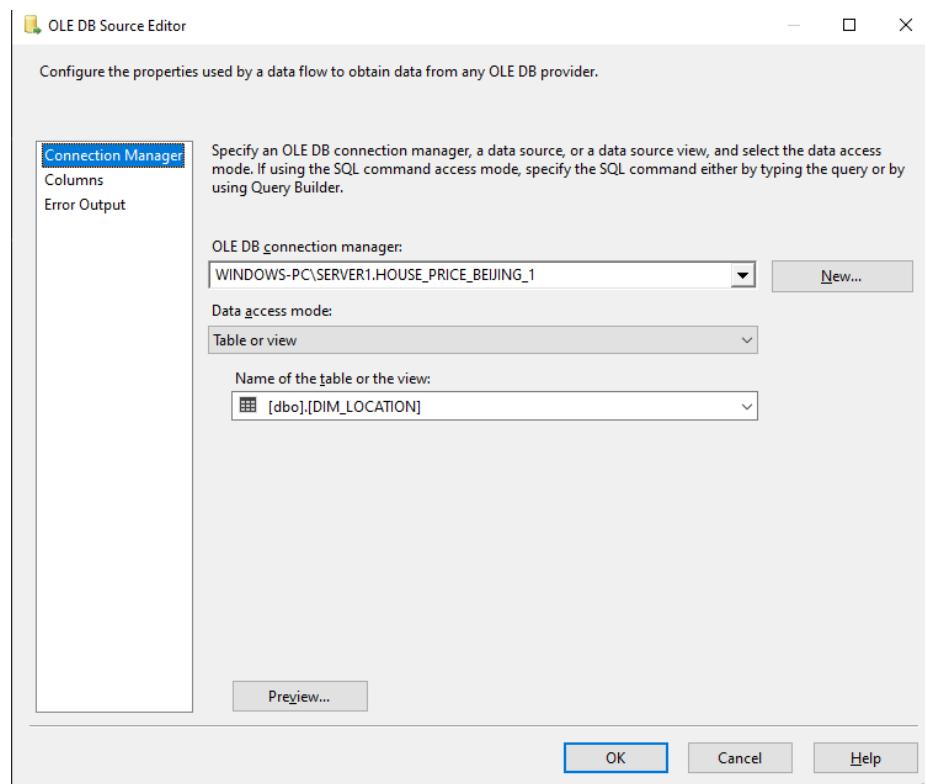
Hình 88: Chọn bảng FACT_RAW đã tạo trước đó làm data source cho bảng FACT_RAW

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.

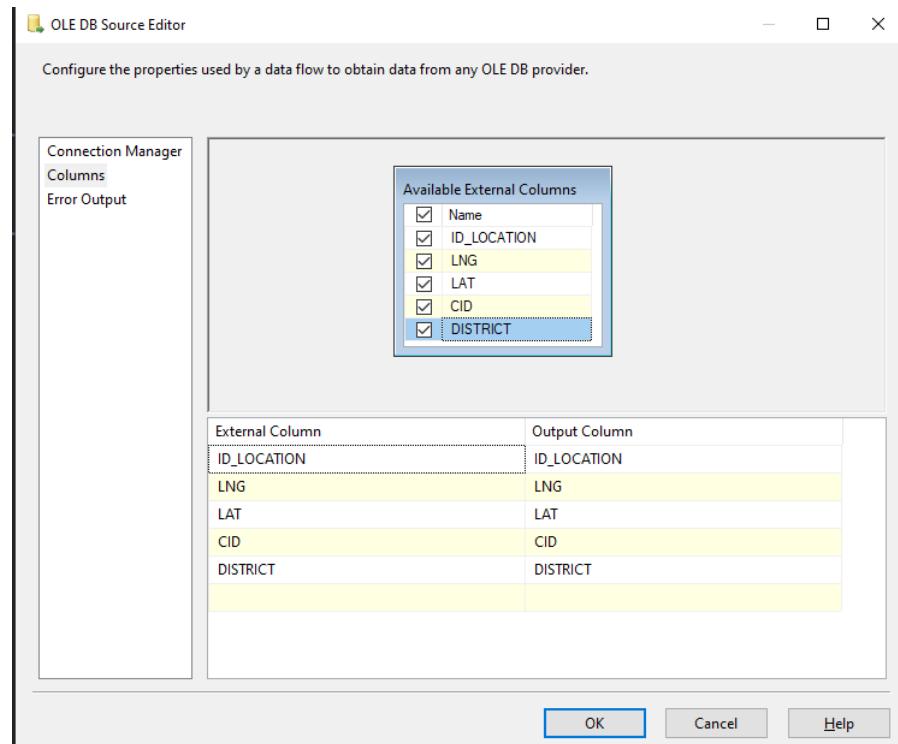


Hình 89: Xem xét các cột được ánh xạ

Bước 5. Tương tự thực hiện chọn ánh xạ cột cho **DIM_LOCATION**

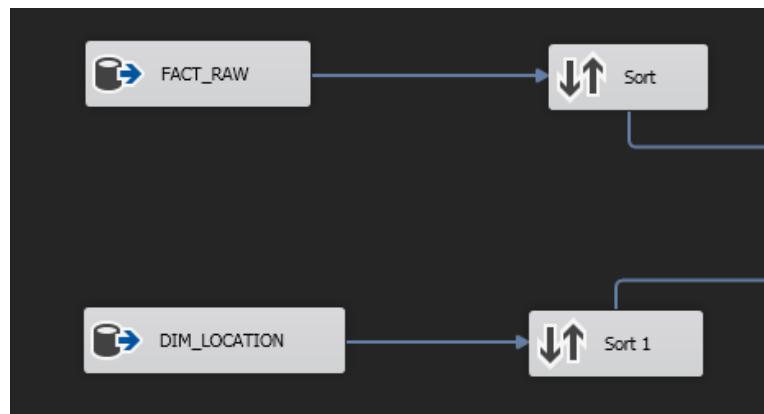


Hình 90: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (1)



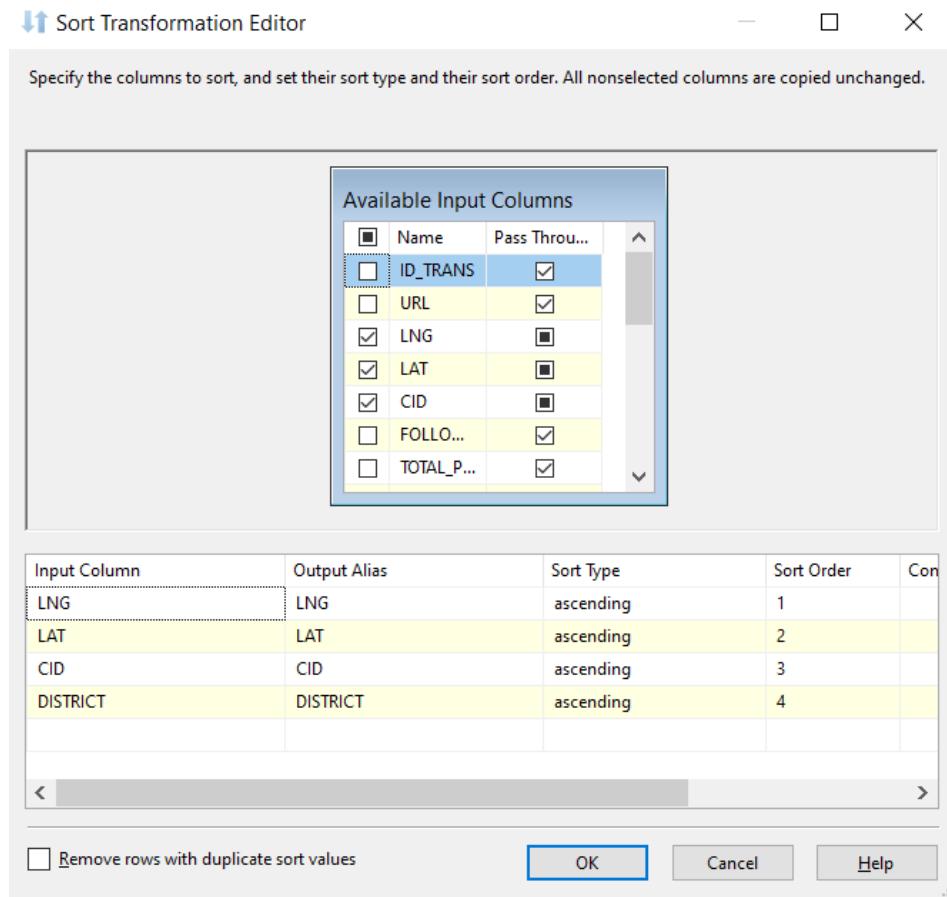
Hình 91: Thực hiện chọn ánh xạ cột cho DIM_LOCATION (2)

Bước 6. Tạo 2 Sort là Sort và Sort1 tương ứng với mỗi Source.



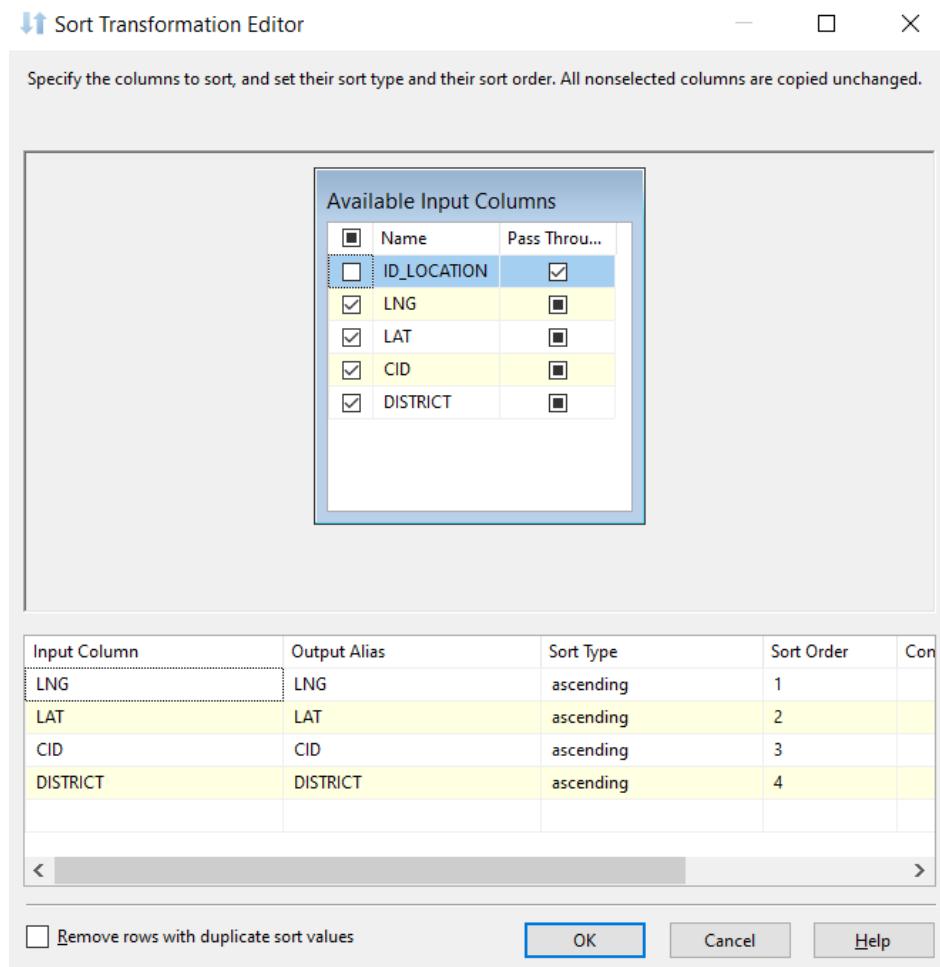
Hình 92: Tạo 2 Sort là Sort và Sort1 tương ứng với mỗi Source

Bước 7. Ở Sort, click chuột phải chọn **Edit** và chọn các cột LNG, LAT, CID, DISTRICT theo thứ tự giống với bảng **DIM_LOCATION** để chuẩn bị cho quá trình merge.



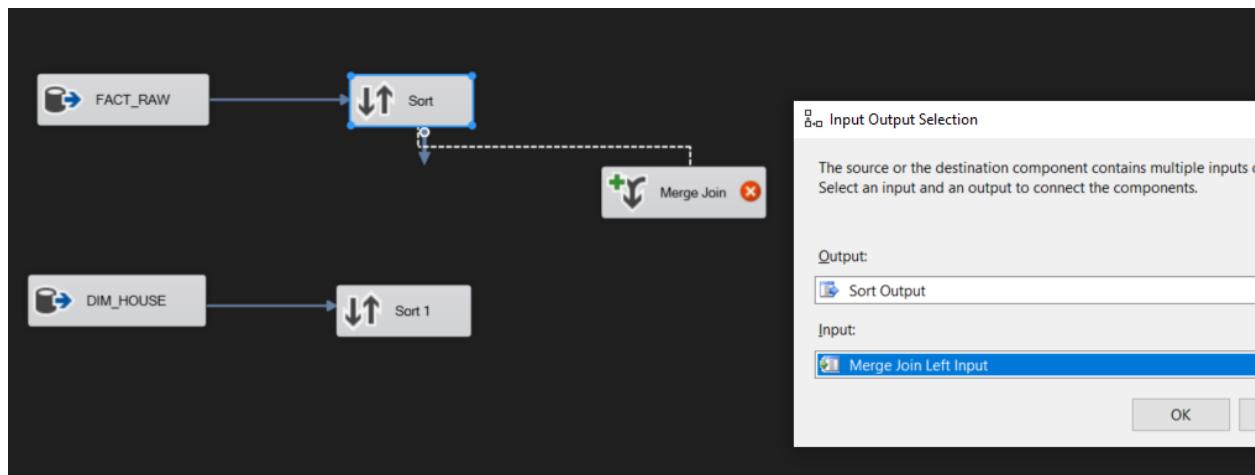
Hình 93: Chọn các cột LNG, LAT, CID, DISTRICT

- Tương tự ta chọn cột **LNG, LAT, CID, DISTRICT** cho **Sort1**



Hình 94: Tương tự ta chọn cột LNG, LAT, CID, DISTRICT

Bước 8. Tạo một **Merge Join** và nối với **Sort**, tiếp theo ta chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng **FACT_RAW** bất kể có kết quả khi thực hiện phép kết trái với cột **ID** của bảng **DIM_LOCATION** hay không.

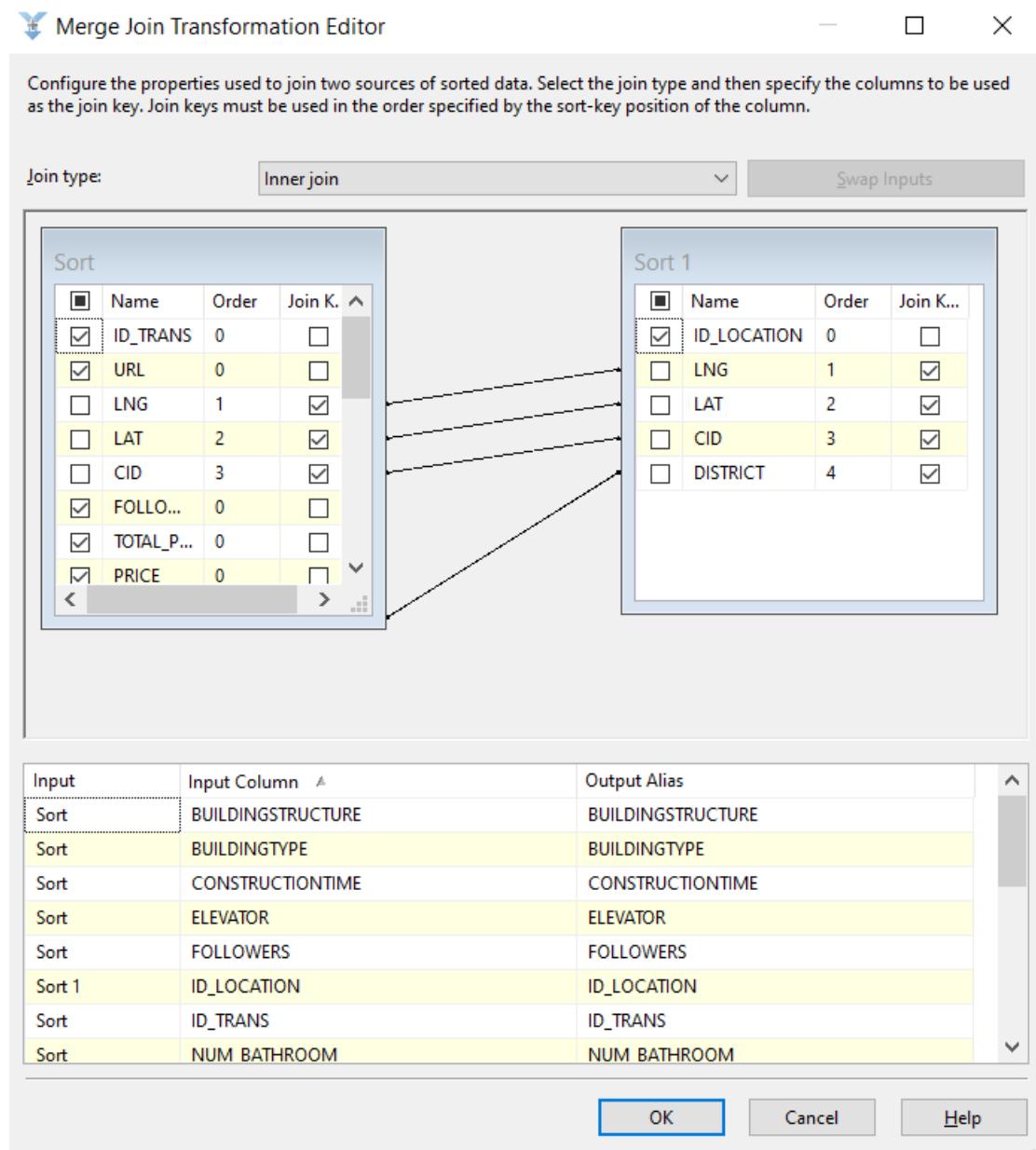


Hình 95: Chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng FACT_RAW

Bước 10. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy thuộc tính **LNG, LAT, CID, DISTRICT**

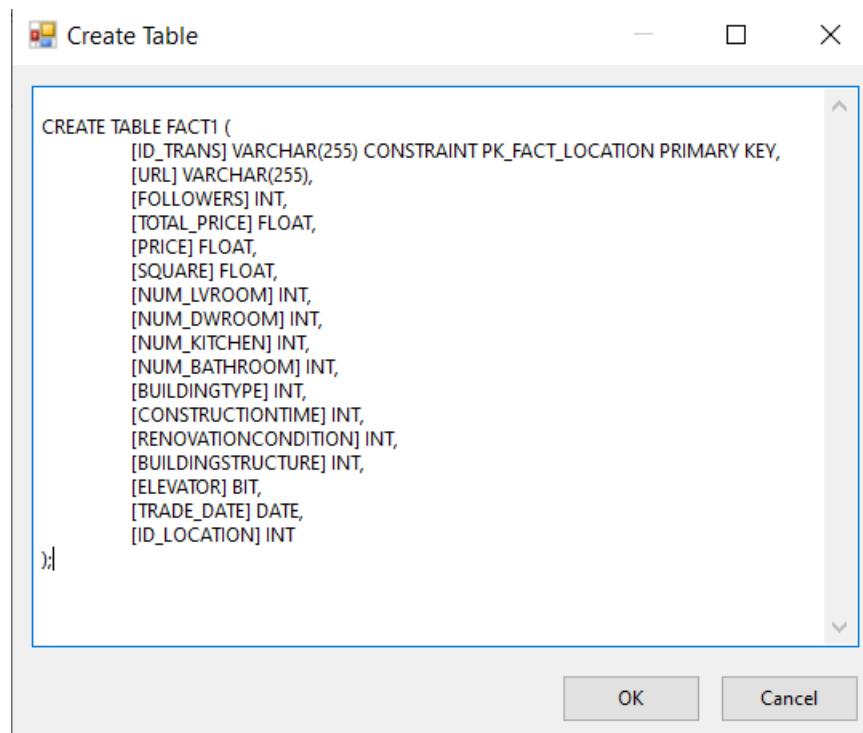
- Tiếp theo ta chọn **ID_LOCATION** ở **Sort1** để merge vào **FACT_RAW**

Kết quả sau khi merge là bảng **FACT_RAW** không còn thuộc tính **LNG, LAT, CID, DISTRICT** và có thêm thuộc tính mới là **ID_LOCATION**



Hình 96: Kết quả sau khi merge

Bước 11. Tạo bảng FACT1 từ một OLE DB Destination để chứa tất cả những gì đã merge



Hình 97: Câu lệnh SQL tạo bảng FACT1

*Nội dung câu lệnh SQL tạo bảng FACT1 như sau:

```
CREATE TABLE FACT1 (
```

```
    [ID_TRANS] VARCHAR(255) CONSTRAINT PK_FACT_LOCATION  
    PRIMARY KEY,
```

```
    [URL] VARCHAR(255),
```

```
    [FOLLOWERS] INT,
```

```
    [TOTAL_PRICE] FLOAT,
```

```
    [PRICE] FLOAT,
```

```
    [SQUARE] FLOAT,
```

```
    [NUM_LVROOM] INT,
```

```
    [NUM_DWROOM] INT,
```

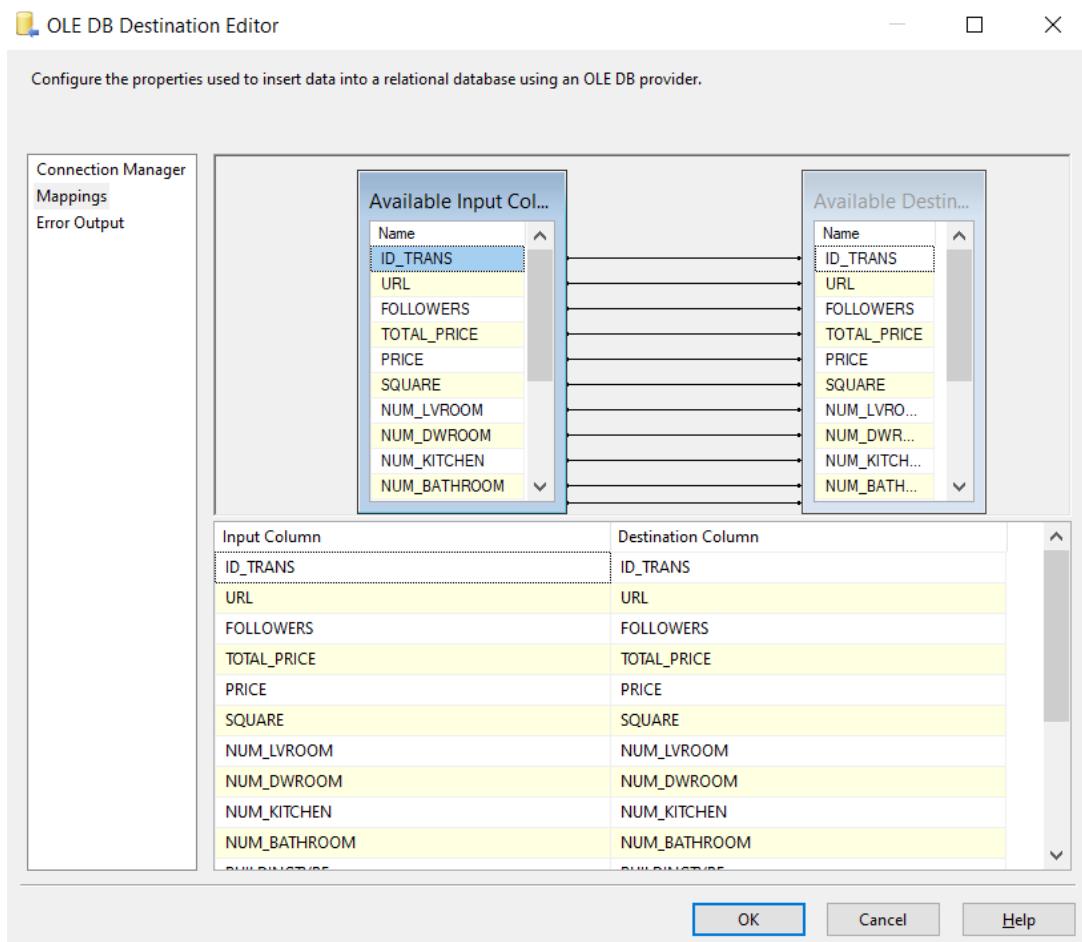
```
    [NUM_KITCHEN] INT,
```

```
    [NUM_BATHROOM] INT,
```

```
    [BUILDINGTYPE] INT,
```

[CONSTRUCTIONTIME] INT,
 [RENOVATIONCONDITION] INT,
 [BUILDINGSTRUCTURE] INT,
 [ELEVATOR] BIT,
 [TRADE_DATE] DATE,
 [ID_LOCATION] INT
);

- Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

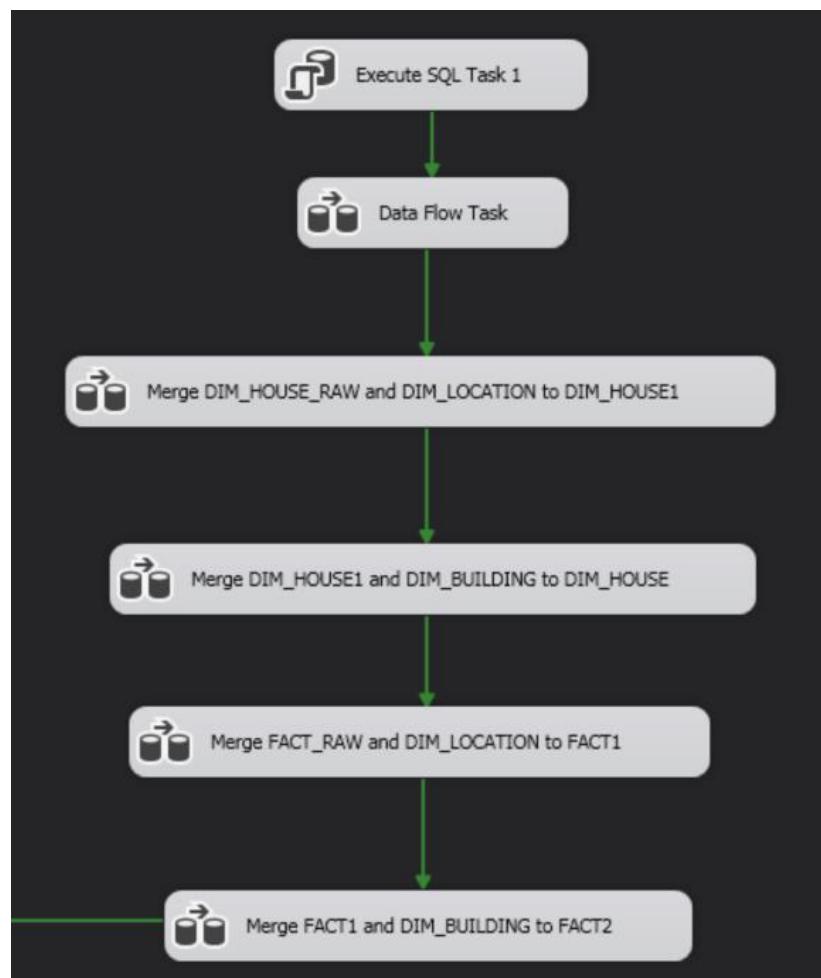


Hình 98: Xem xét việc ánh xạ các cột dữ liệu

- Nhấn **OK**

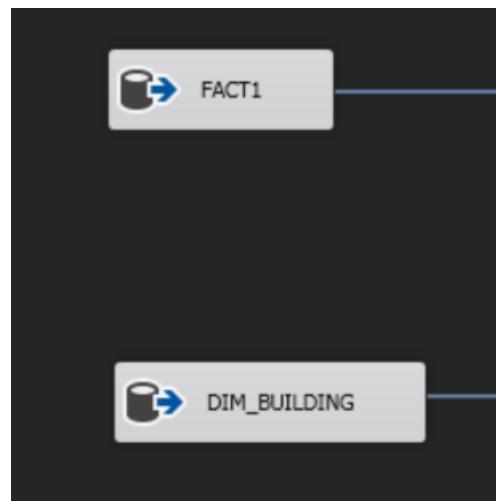
4.9. Merge FACT1 và DIM_BUILDING vào FACT2

Bước 1. Ở tab Control Flow, tạo thêm một Data Flow Task và đổi tên Data Flow Task này là “Merge FACT1 and DIM_BUILDING to FACT2”



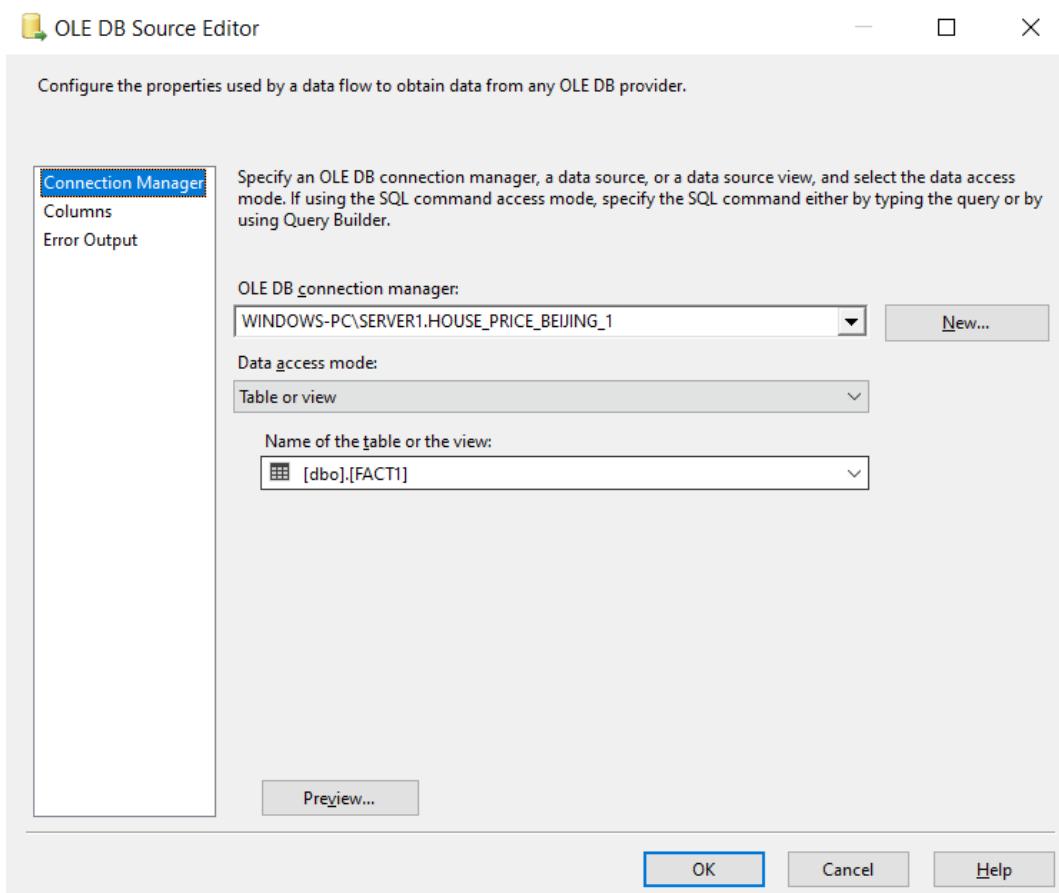
Hình 99: Tạo thêm một Data Flow Task và đổi tên

Bước 2. Click chuột phải vào **Data Flow Task** nói trên và chọn **Edit**, trong tab **Data Flow** ta tạo 2 **OLE DB Source** và đổi tên **FACT1** và **DIM_BUILDING**



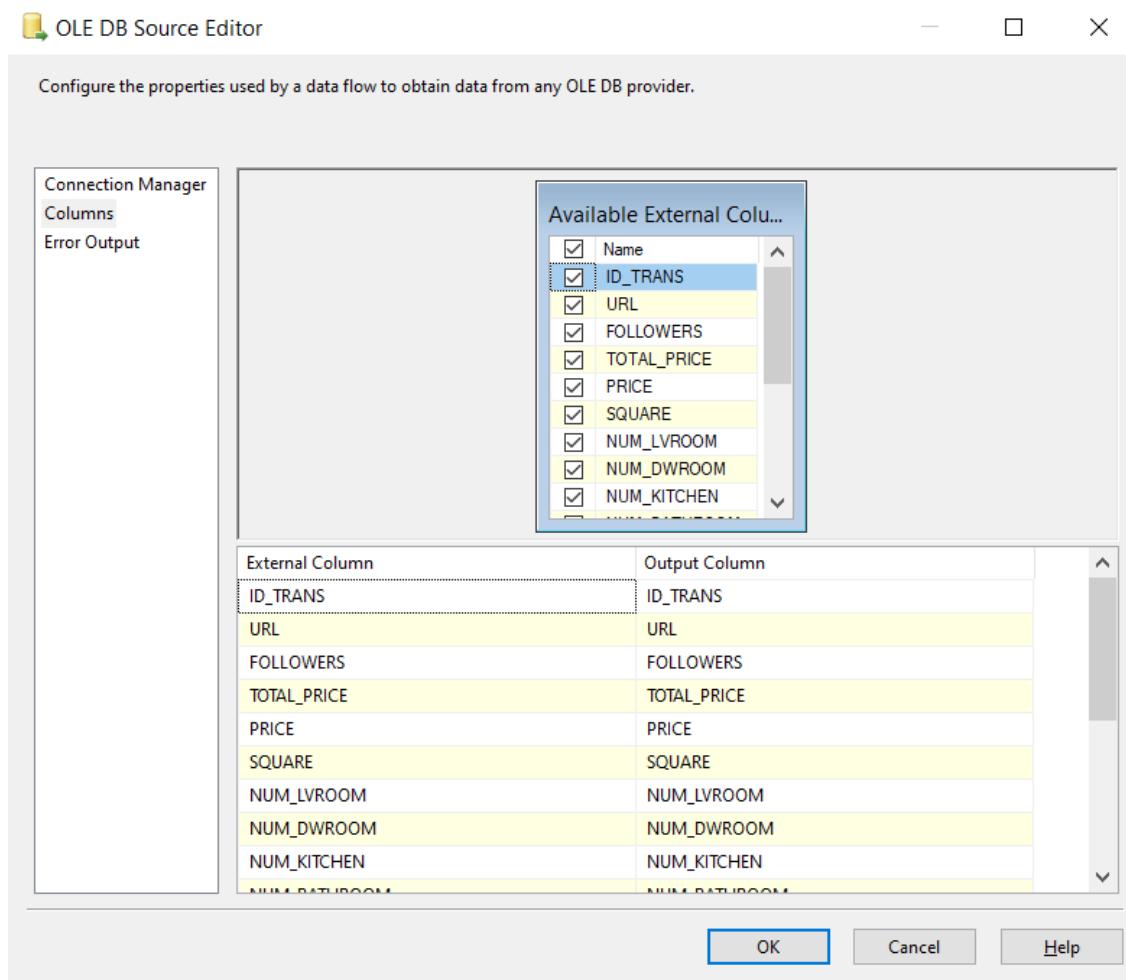
Hình 100: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải vào FACT1 chọn Edit, sau đó chọn bảng FACT1 đã được tạo khi merge FACT_RAW và DIM_LOCATION làm data source.



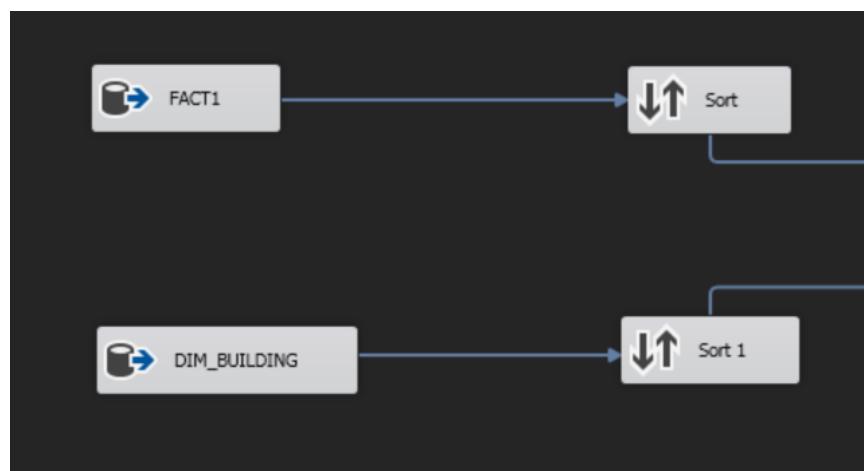
Hình 101: Chọn bảng FACT1 đã được tạo khi merge

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.



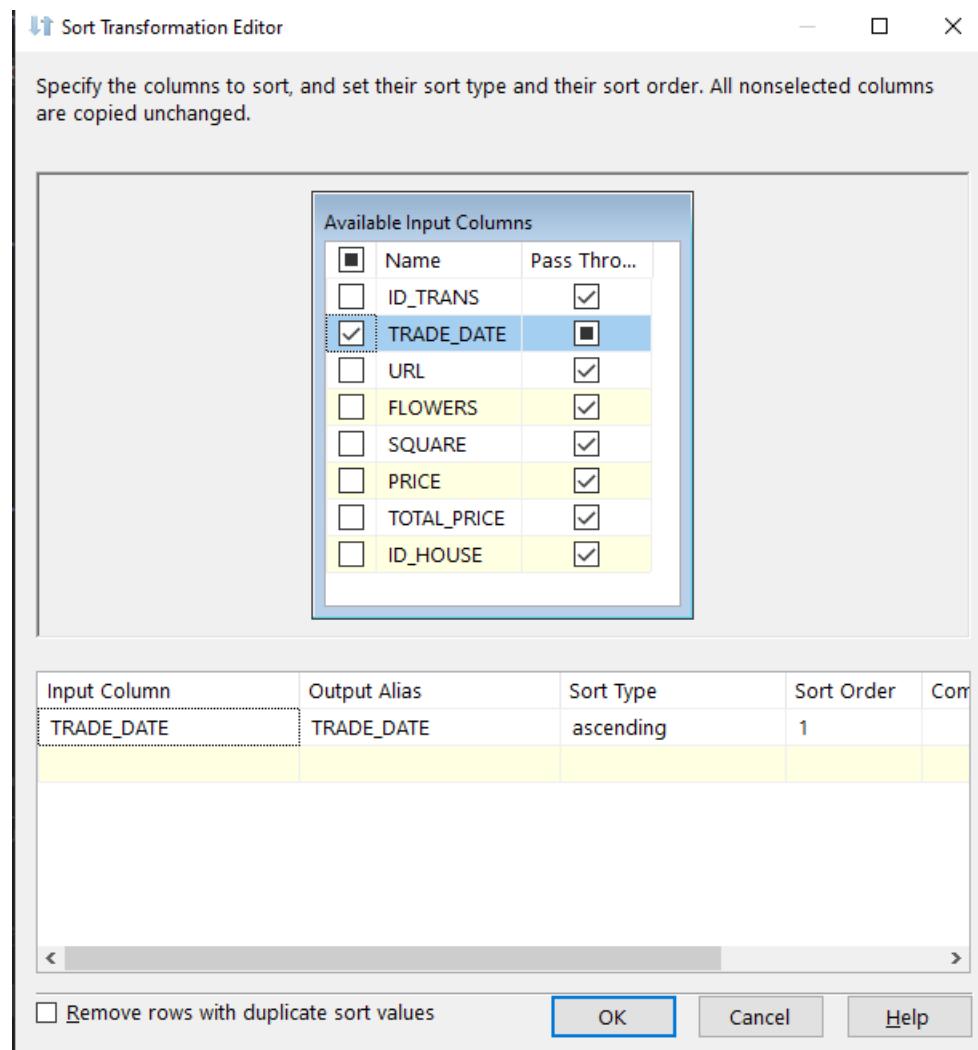
Hình 102: Xem xét các cột được ánh xạ

Bước 5. Tạo 2 Sort tương ứng với mỗi Source



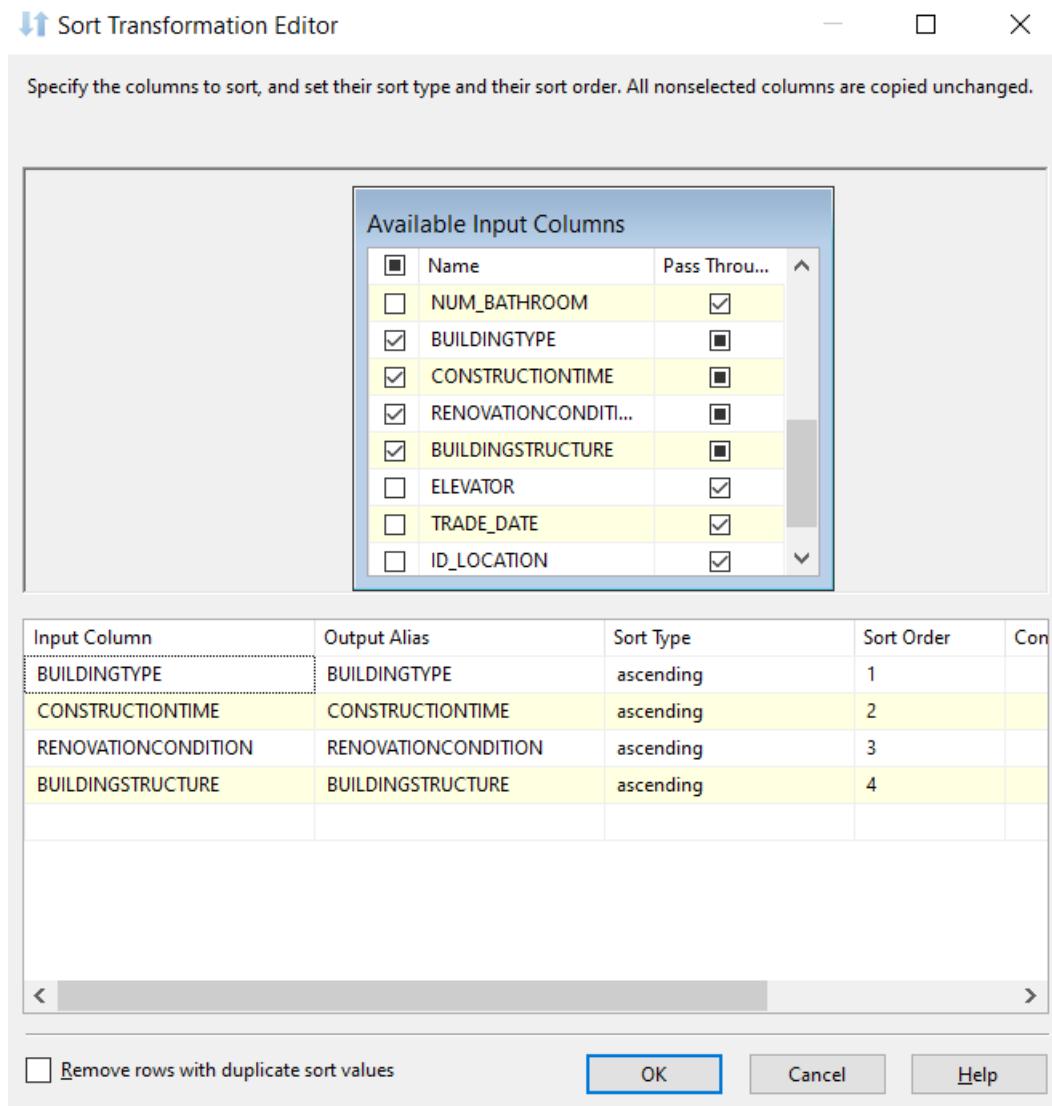
Hình 103: Tạo 2 Sort tương ứng với mỗi Source

Bước 6. Tại **Sort**, click chuột phải chọn **Edit** và chọn các cột **BUILDINGTYPE**, **CONSTRUCTIONTIME**, **RENOVATIONCONDITION**, **BUILDINGSTRUCTURE** theo thứ tự giống với bảng **DIM_BUILDING** để chuẩn bị cho quá trình **merge**



Hình 104: Chọn các cột để chuẩn bị cho quá trình merge

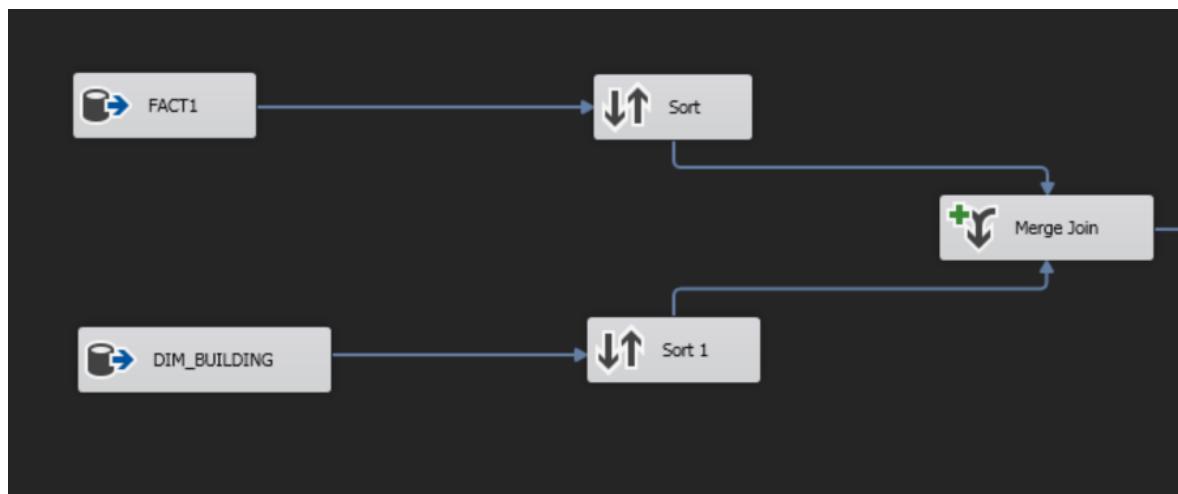
Bước 7. Tương tự ta chọn cột **BUILDINGTYPE**, **CONSTRUCTIONTIME**, **RENOVATIONCONDITION**, **BUILDINGSTRUCTURE** cho **Sort1**



Hình 105: Tương tự ta chọn cột cho Sort1

Bước 8. Tạo một **Merge Join** và nối với **Sort**, tiếp theo chọn **Merge Join Left Input** để giữ lại toàn bộ các dòng trong bảng **FACT1** bất kể có kết quả khi thực hiện phép kết trái với cột **ID** của bảng **DIM_BUILDING** hay không.

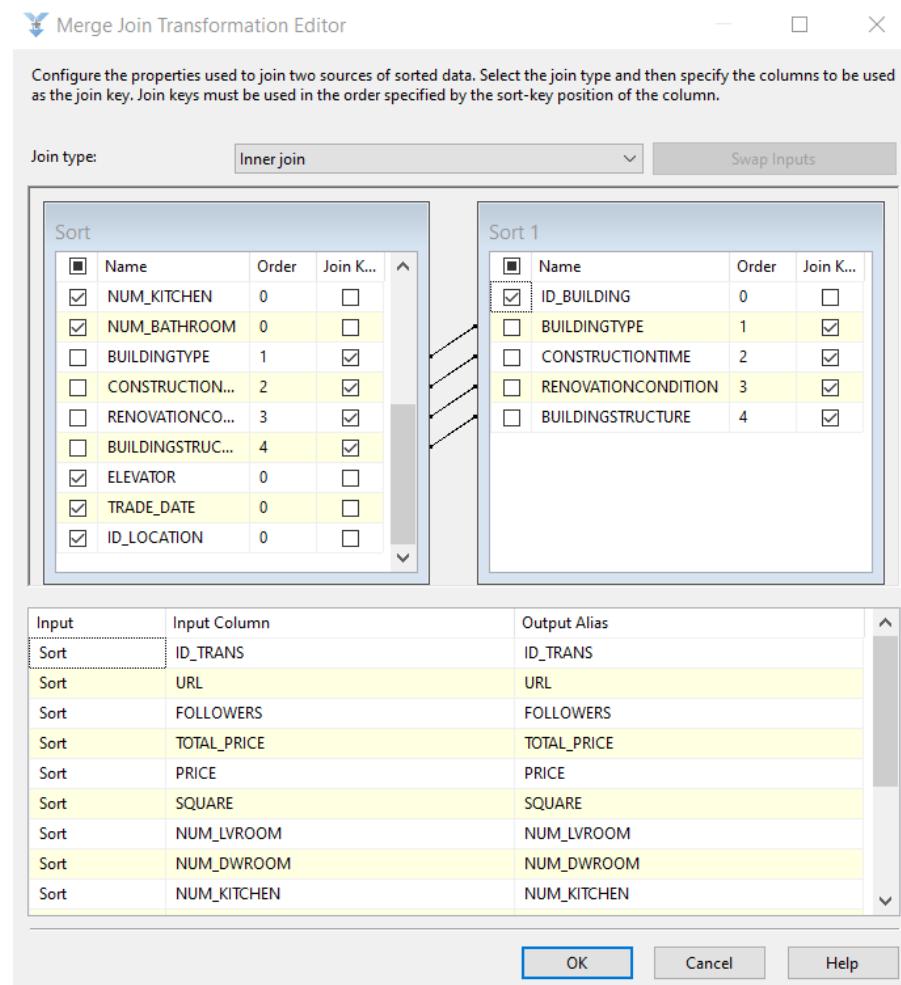
- Nối **Sort1** với **Merge Join**



Hình 106: Nối Sort1 với Merge Join

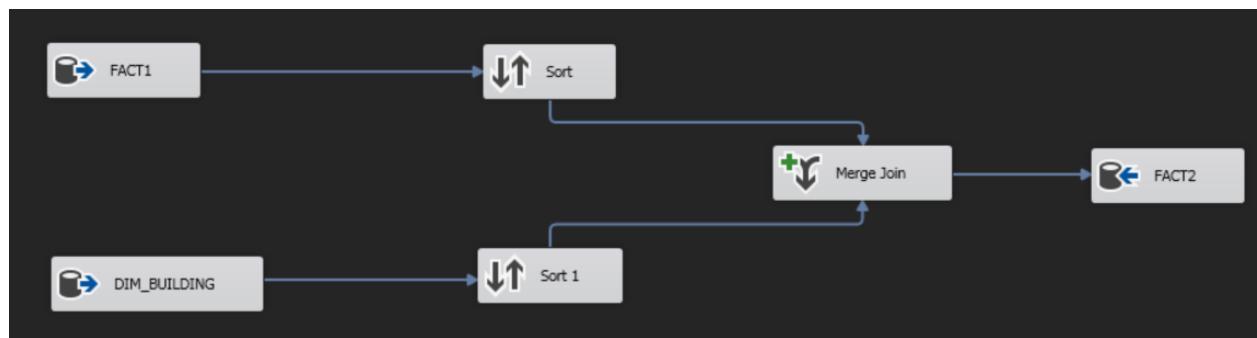
Bước 9. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy thuộc tính **BUILDINGTYPE, CONSTRUCTIONTIME, RENOVATIONCONDITION, BUILDINGSTRUCTURE**.

- Tiếp theo ta chọn **ID_BUILDING** ở **Sort1** để merge vào **FACT1**
- Kết quả sau khi merge là bảng **FACT1** không còn thuộc tính **BUILDINGTYPE, CONSTRUCTIONTIME, RENOVATIONCONDITION, BUILDINGSTRUCTURE** và có thêm 1 thuộc tính mới là **ID_BUILDING**

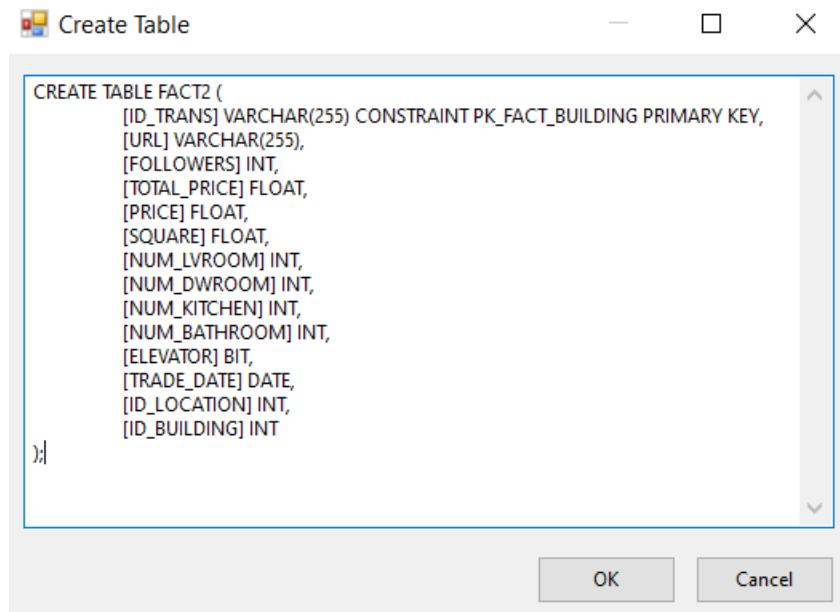


Hình 107: Kết quả sau khi merge

Bước 10. Tạo bảng FACT2 từ một OLE DB Destination để chứa tất cả những gì đã merge



Hình 108: Tạo bảng FACT2 để chứa tất cả những gì đã merge



Hình 109: Câu lệnh SQL tạo bảng FACT2

*Nội dung câu lệnh SQL tạo bảng FACT2 như sau:

CREATE TABLE FACT2 (

[ID_TRANS] VARCHAR(255) CONSTRAINT PK_FACT_BUILDING
PRIMARY KEY,

[URL] VARCHAR(255),

[FOLLOWERS] INT,

[TOTAL_PRICE] FLOAT,

[PRICE] FLOAT,

[SQUARE] FLOAT,

[NUM_LVROOM] INT,

[NUM_DWROOM] INT,

[NUM_KITCHEN] INT,

[NUM_BATHROOM] INT,

[ELEVATOR] BIT,

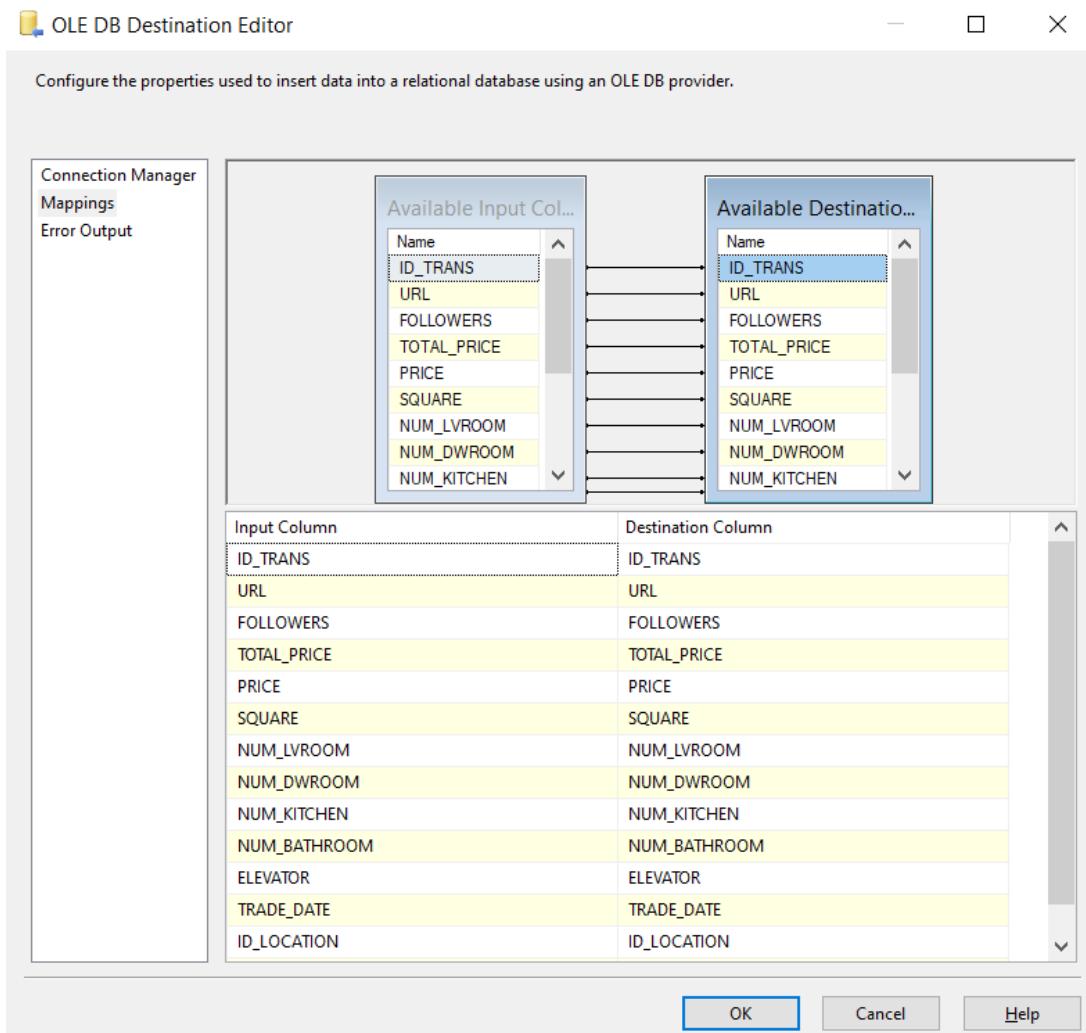
[TRADE_DATE] DATE,

[ID_LOCATION] INT,

[ID_BUILDING] INT

);

- Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu



Hình 110: Xem xét việc ánh xạ các cột dữ liệu

- Nhấn **OK**

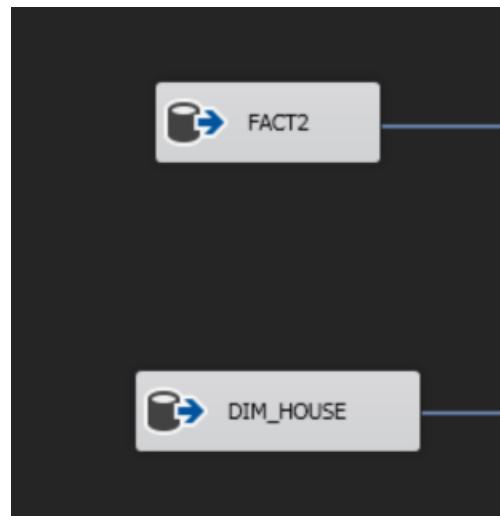
4.10. Merge FACT2 và DIM_HOUSE vào FACT3

Bước 1. Ở tab Control Flow, tạo thêm một Data Flow Task và đổi tên Data Flow Task này là “Merge FACT2 and DIM_HOUSE to FACT3”



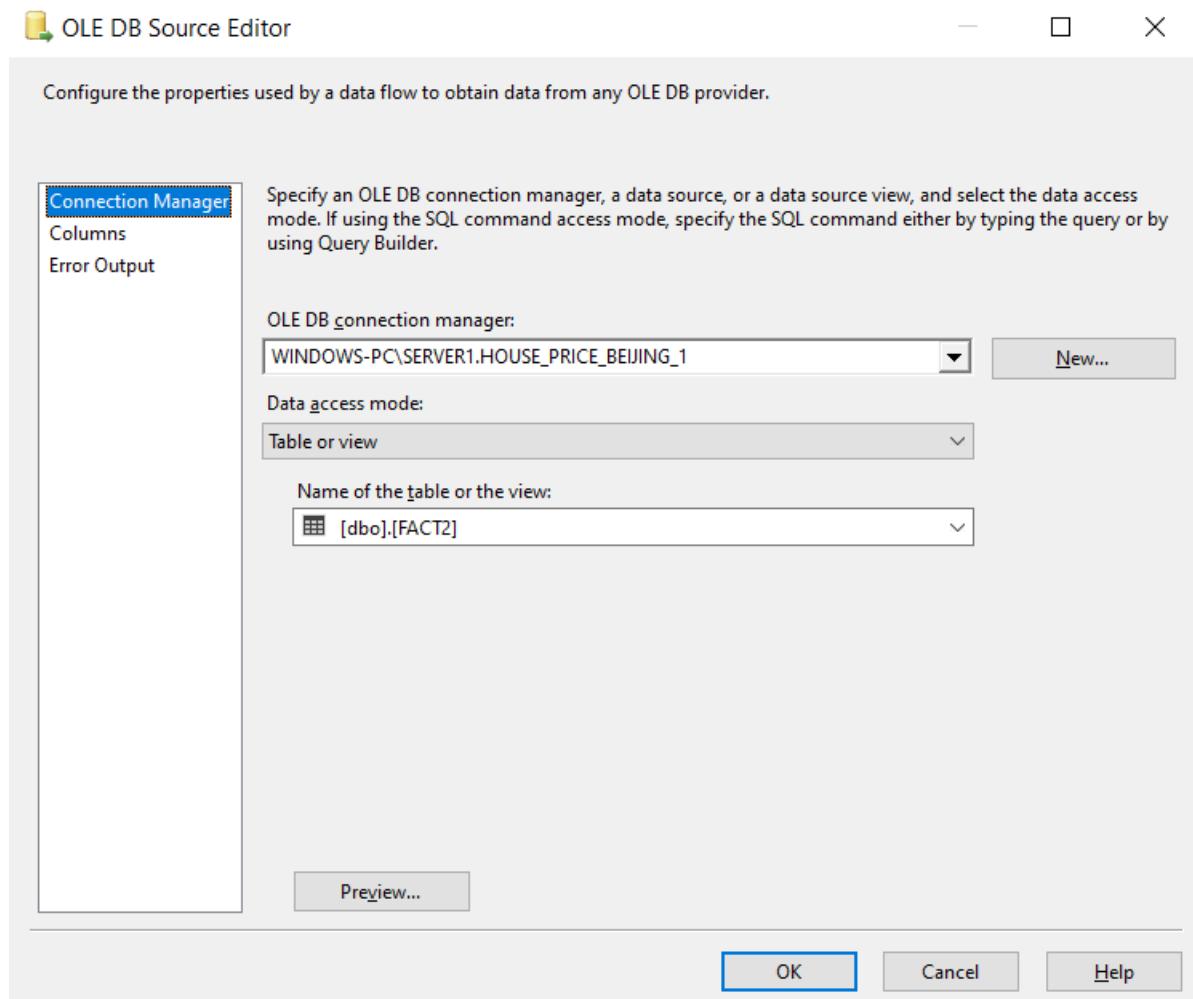
Hình 111: Tạo thêm một Data Flow Task và đổi tên

Bước 2. Click chuột phải vào **Data Flow Task** nói trên và chọn **Edit**, trong tab **Data Flow** ta tạo 2 **OLE DB Source** và đổi tên **FACT2** và **DIM_HOME**



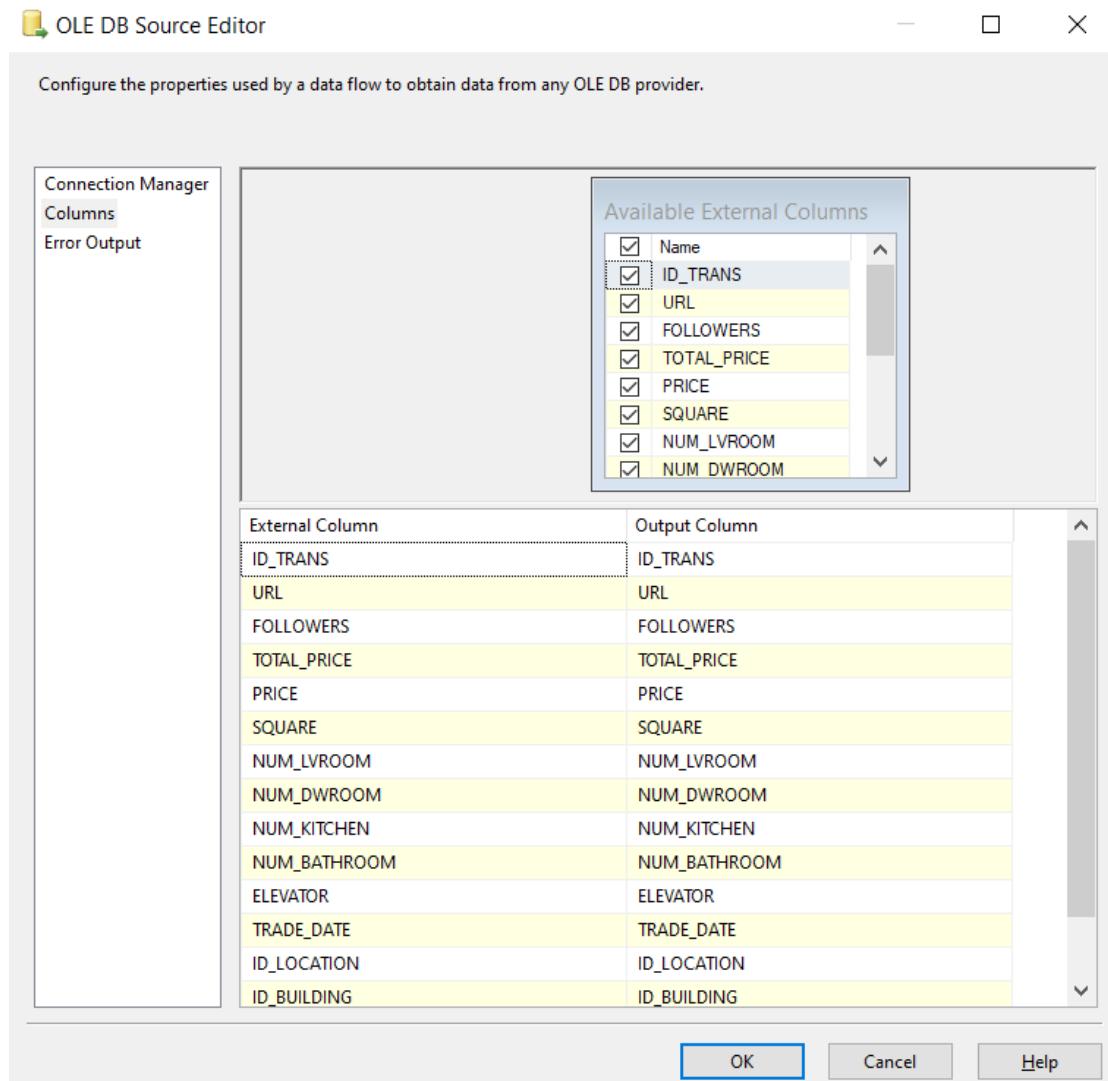
Hình 112: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải vào **FACT2** chọn **Edit**, sau đó chọn bảng **FACT2** đã được tạo khi merge **FACT1** và **DIM_BUILDING** làm data source.



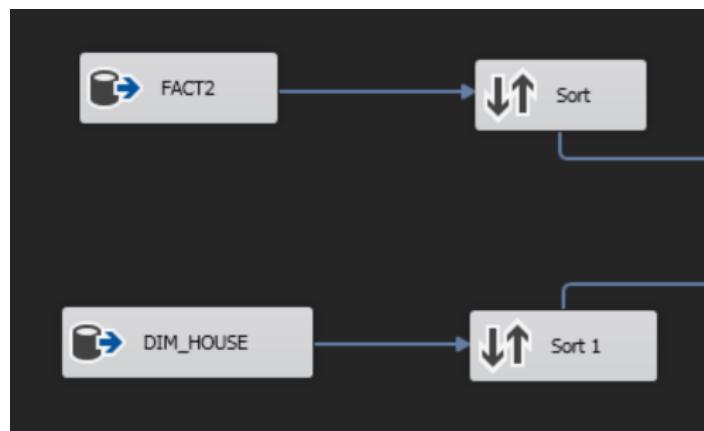
Hình 113: Chọn bảng FACT2 đã được tạo khi merge

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.



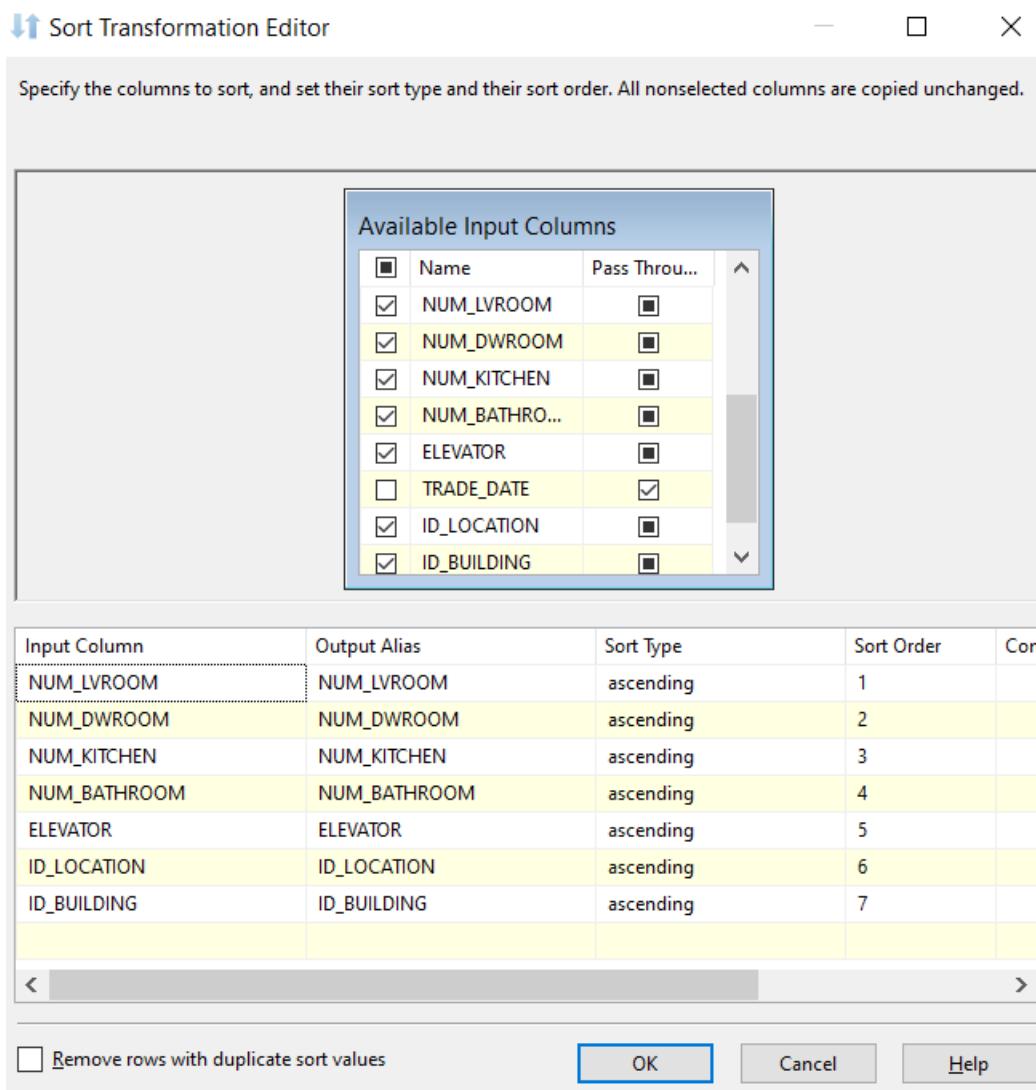
Hình 114: Xem xét các cột được ánh xạ

Bước 5. Tạo 2 Sort tương ứng với mỗi Source



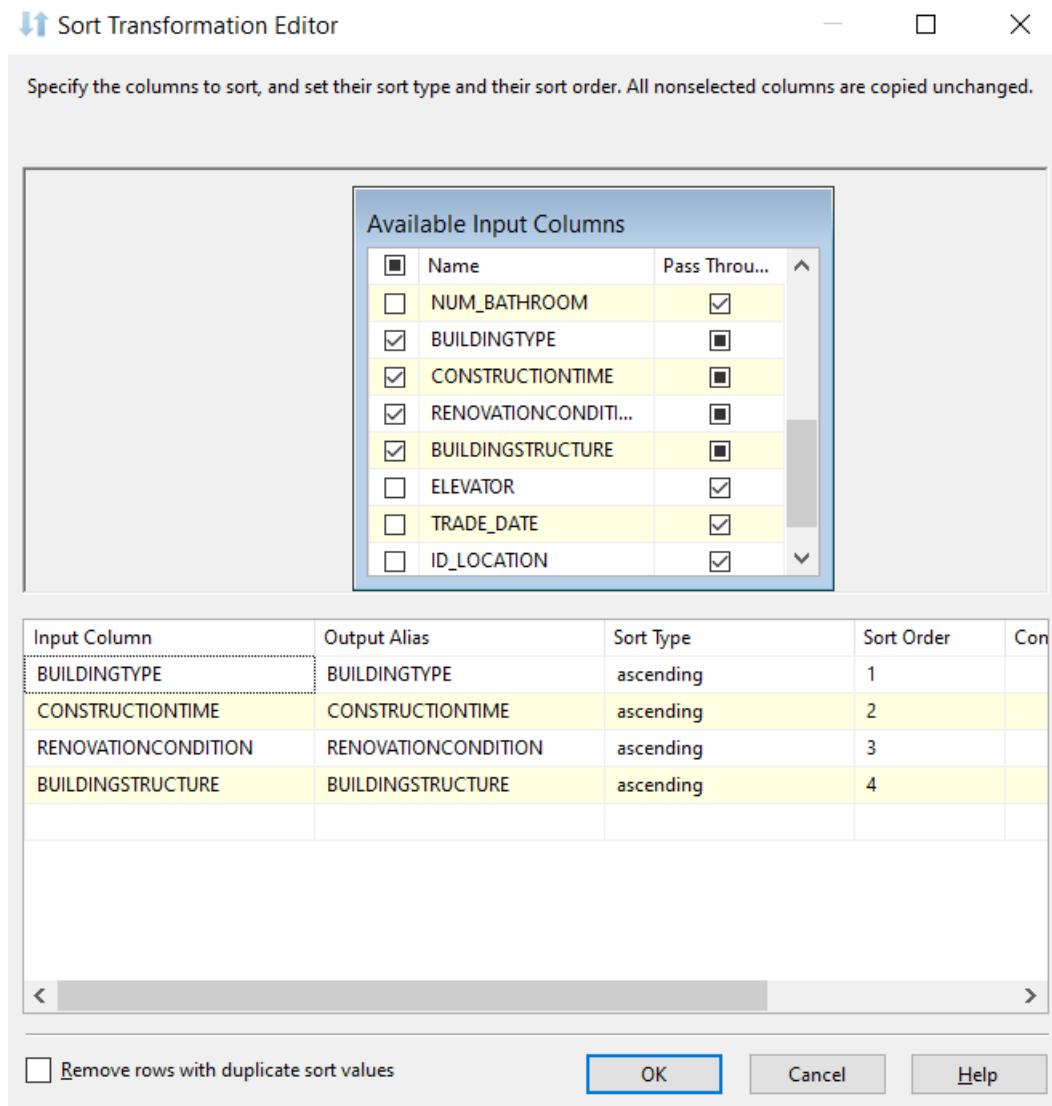
Hình 115: Tạo 2 Sort tương ứng với mỗi Source

Bước 6. Tại **Sort**, click chuột phải chọn **Edit** và chọn các cột **NUM_LVROOM**, **NUM_DWROOM**, **NUM_KITCHEN**, **NUM_BATHROOM**, **ELEVATOR**, **ID_LOCATION**, **ID_BUILDING** theo thứ tự giống với bảng **DIM_HOUSE** để chuẩn bị cho quá trình merge



Hình 116: Chọn các cột để chuẩn bị cho quá trình merge

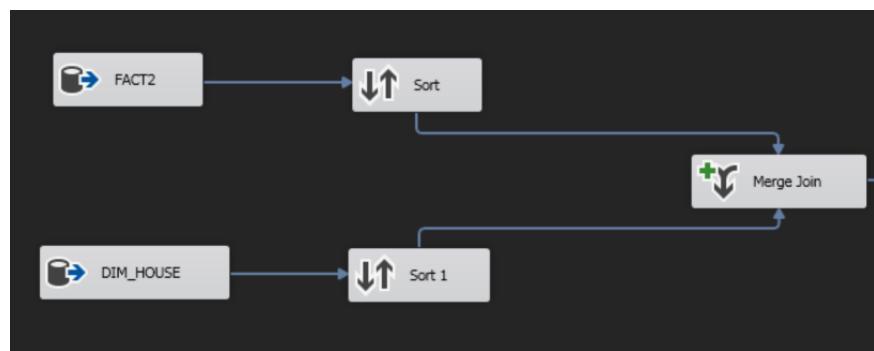
Bước 7. Tương tự ta chọn cột **NUM_LVROOM**, **NUM_DWROOM**, **NUM_KITCHEN**, **NUM_BATHROOM**, **ELEVATOR**, **ID_LOCATION**, **ID_BUILDING** cho **Sort1**



Hình 117: Chọn cột cho Sort1

Bước 8. Tạo một **Merge Join** và nối với **Sort**, tiếp theo chọn **Merge Join Left Input** để giữ lại toàn bộ các dòng trong bảng **FACT2** bất kể có kết quả khi thực hiện phép kết trái với cột **ID** của bảng **DIM_HOUSE** hay không.

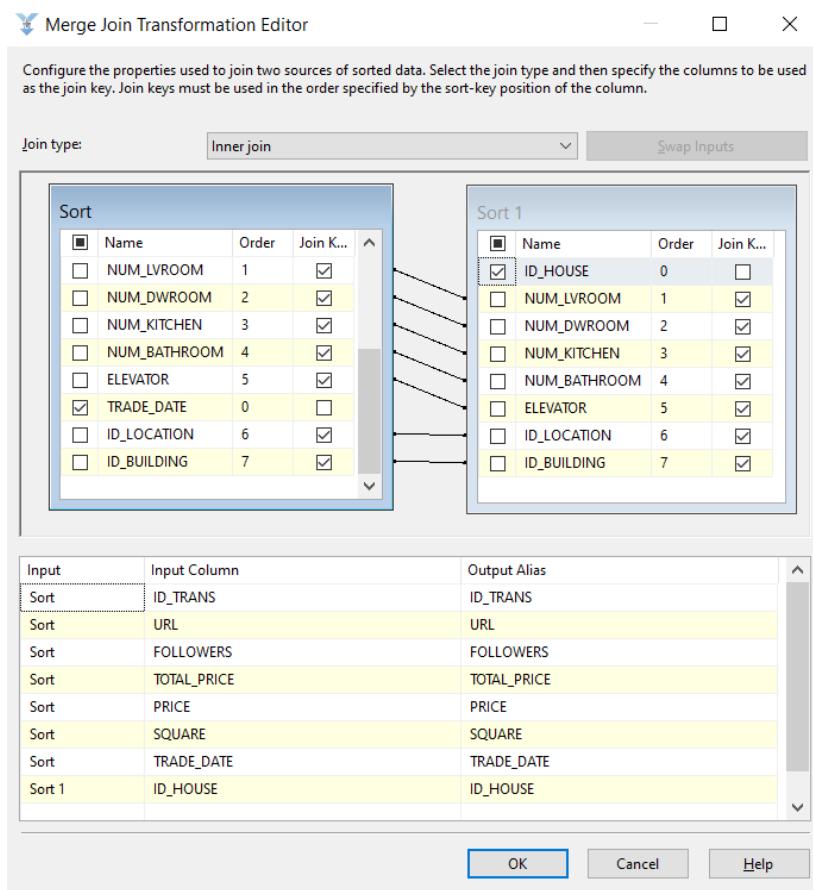
- Nối **Sort1** với **Merge Join**



Hình 118: Nối Sort1 với Merge Join

Bước 9. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy thuộc tính **NUM_LVROOM**, **NUM_DWROOM**, **NUM_KITCHEN**, **NUM_BATHROOM**, **ELEVATOR**, **ID_LOCATION**, **ID_BUILDING**.

- Tiếp theo ta chọn **ID_HOUSE** ở **Sort1** để merge vào **FACT2**
- Kết quả sau khi merge là bảng **FACT2** không còn thuộc **NUM_LVROOM**, **NUM_DWROOM**, **NUM_KITCHEN**, **NUM_BATHROOM**, **ELEVATOR**, **ID_LOCATION**, **ID_BUILDING** và có thêm 1 thuộc tính mới là **ID_HOUSE**

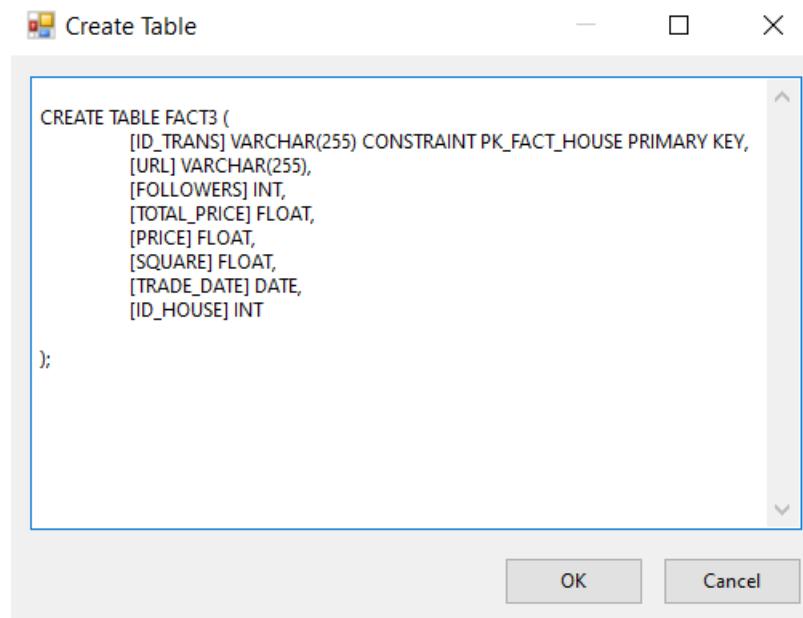


Hình 119: Kết quả sau khi merge

Bước 10. Tạo bảng FACT3 từ một **OLE DB Destination** để chứa tất cả những gì đã merge



Hình 120: Tạo bảng FACT3



Hình 121: Câu lệnh SQL tạo bảng FACT3

*Nội dung câu lệnh SQL tạo bảng FACT3 như sau:

CREATE TABLE FACT3 (

 [ID_TRANS] VARCHAR(255) CONSTRAINT PK_FACT_HOUSE PRIMARY
 KEY,

 [URL] VARCHAR(255),

 [FOLLOWERS] INT,

 [TOTAL_PRICE] FLOAT,

 [PRICE] FLOAT,

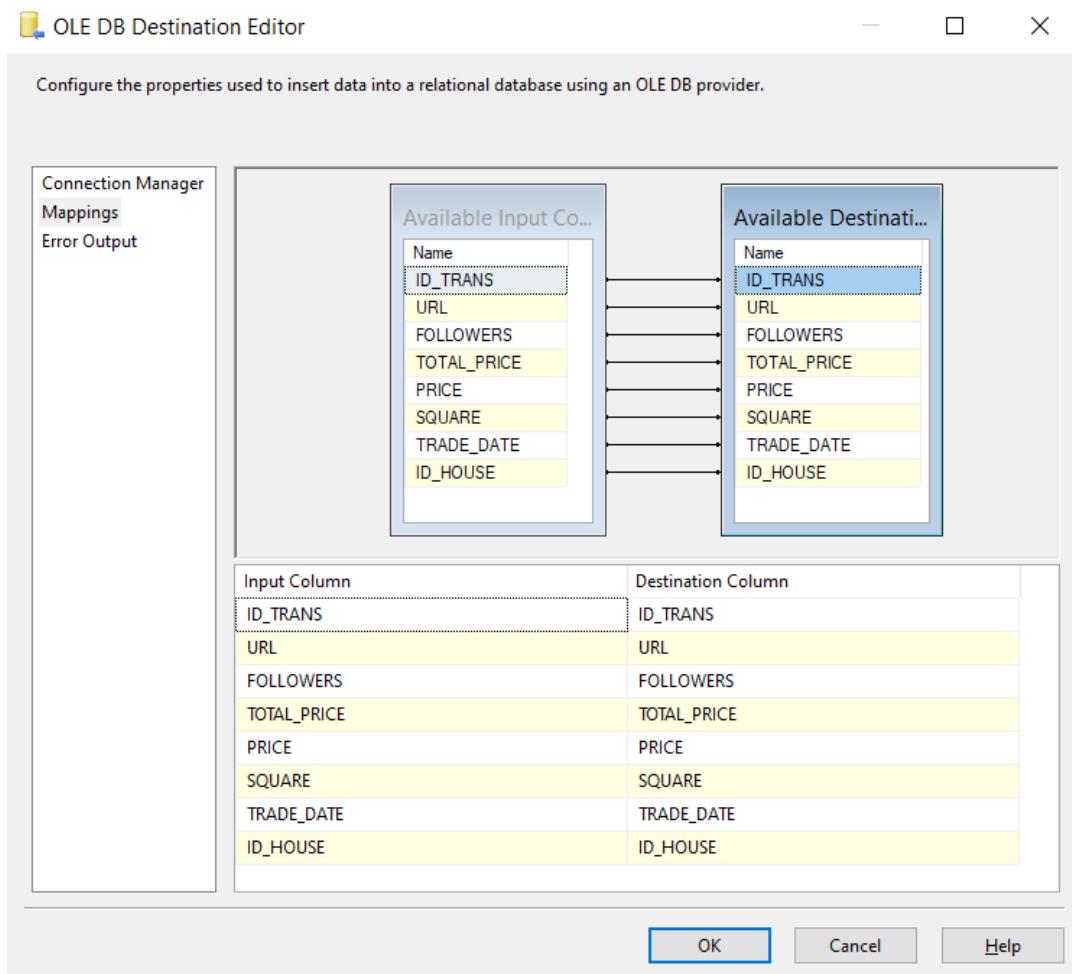
 [SQUARE] FLOAT,

 [TRADE_DATE] DATE,

 [ID_HOUSE] INT

);

- Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu

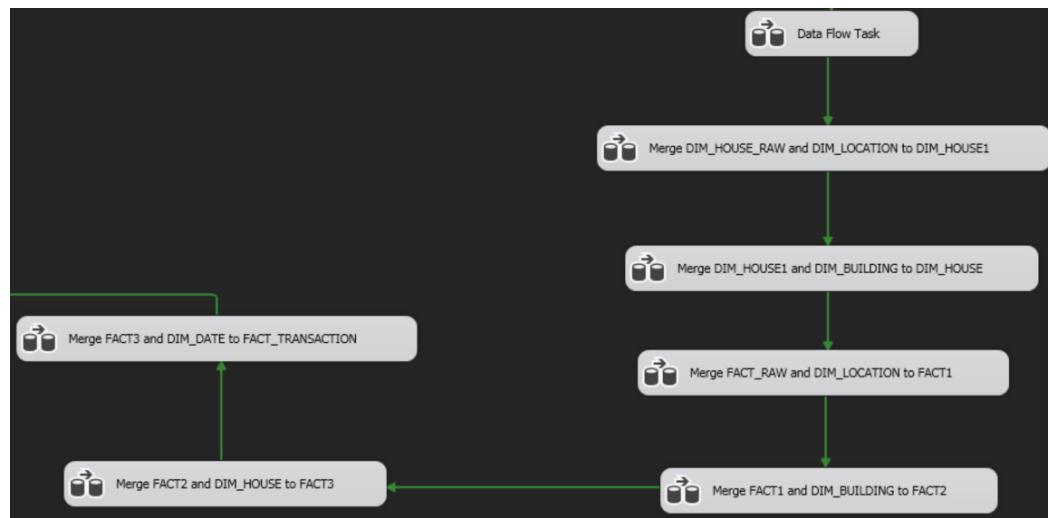


Hình 122: Xem xét việc ánh xạ các cột dữ liệu

- Nhấn **OK**

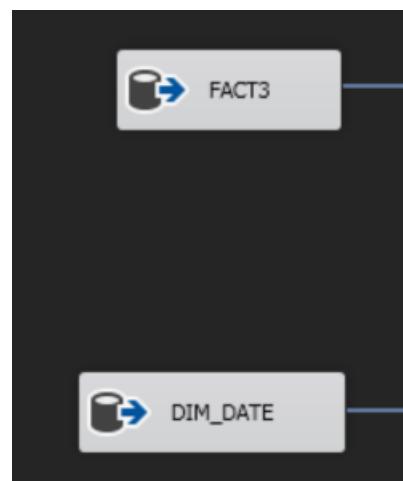
4.11. Merge FACT3 và DIM_DATE vào FACT_TRANSACTION

Bước 1. Ở tab Control Flow, tạo thêm một Data Flow Task và đổi tên Data Flow Task này là “Merge FACT3 and DIM_DATE to FACT_TRANSACTION”



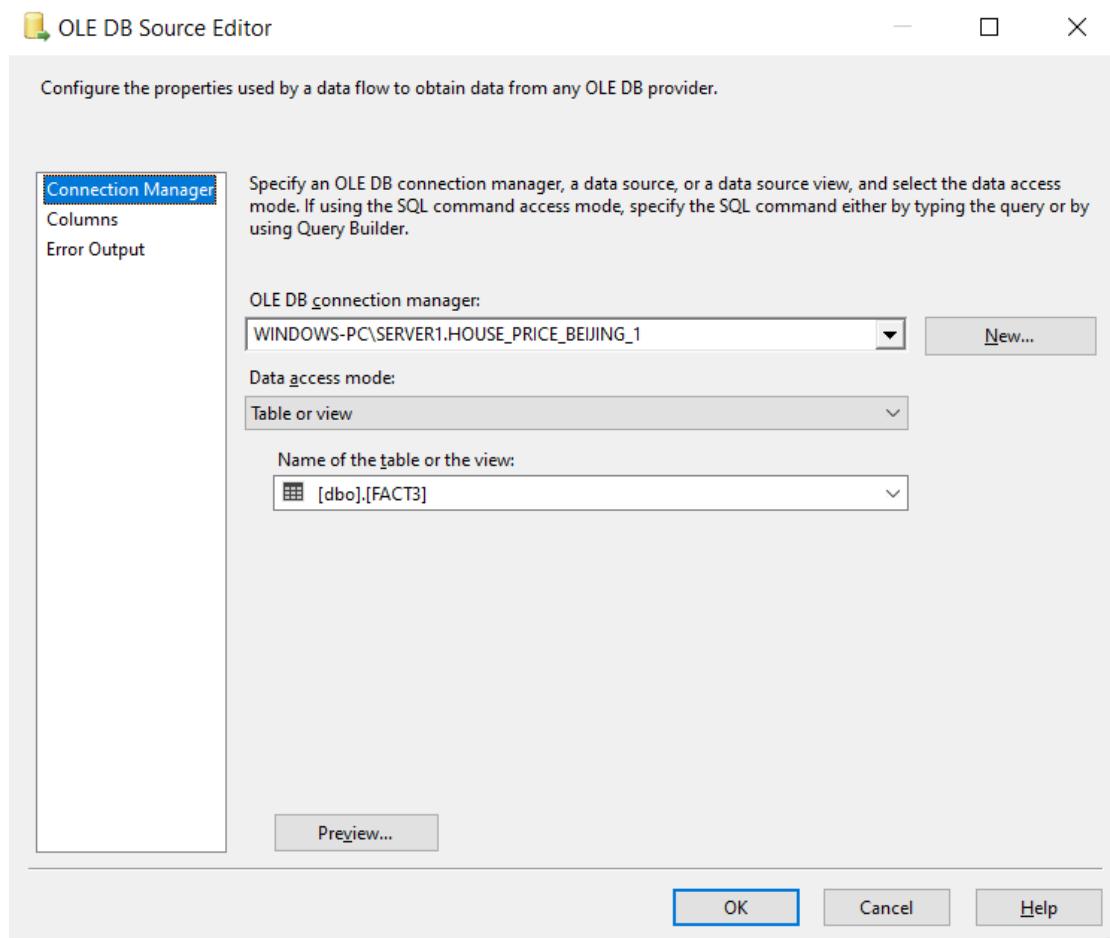
Hình 123: Tạo thêm một Data Flow Task và đổi tên

Bước 2. Click chuột phải vào **Data Flow Task** nói trên và chọn **Edit**, trong tab **Data Flow** ta tạo 2 **OLE DB Source** và đổi tên **FACT3** và **DIM_DATE**



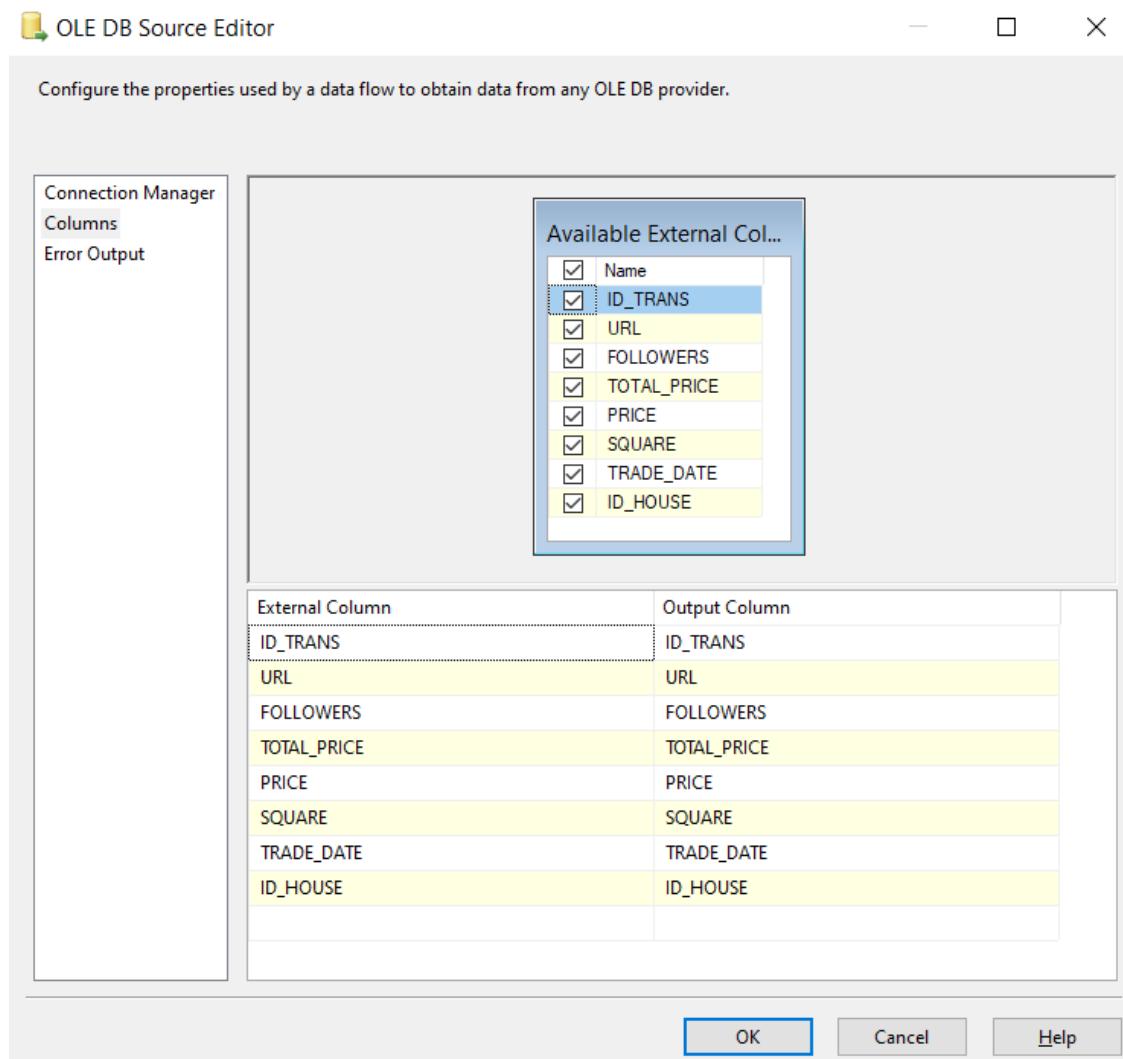
Hình 124: Tạo 2 OLE DB Source và đổi tên

Bước 3. Click chuột phải vào **FACT3** chọn **Edit**, sau đó chọn bảng **FACT3** đã được tạo khi merge **FACT2** và **DIM_HOME** làm data source.



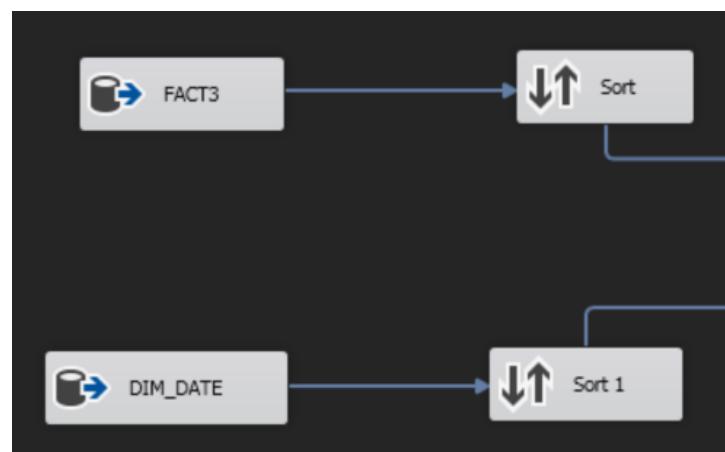
Hình 125: Chọn bảng FACT3 đã được tạo khi merge

Bước 4. Chọn mục **Columns** để xem xét các cột được ánh xạ. Nhấn **OK**.



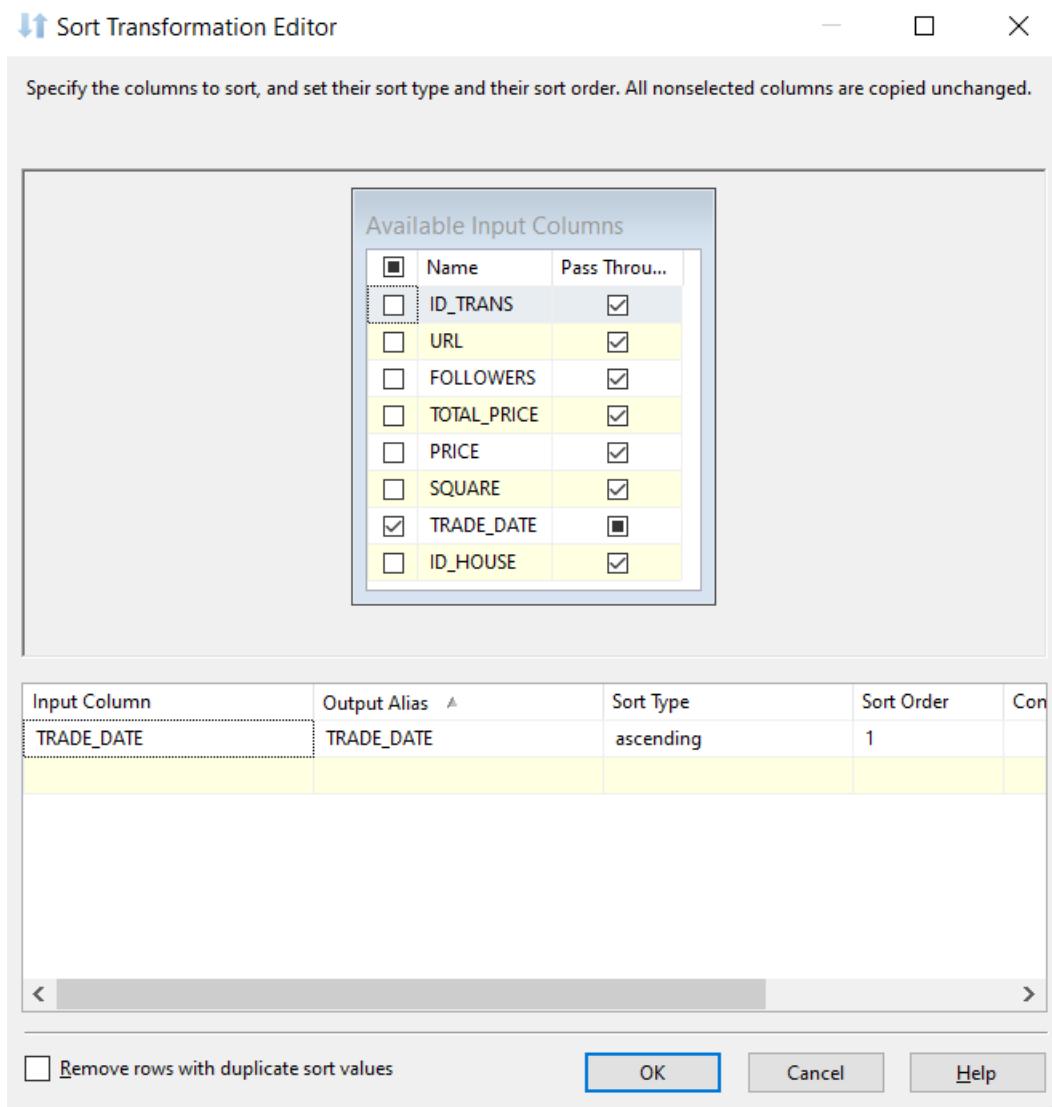
Hình 126: Xem xét các cột được ánh xạ

Bước 5. Tạo 2 Sort tương ứng với mỗi Source



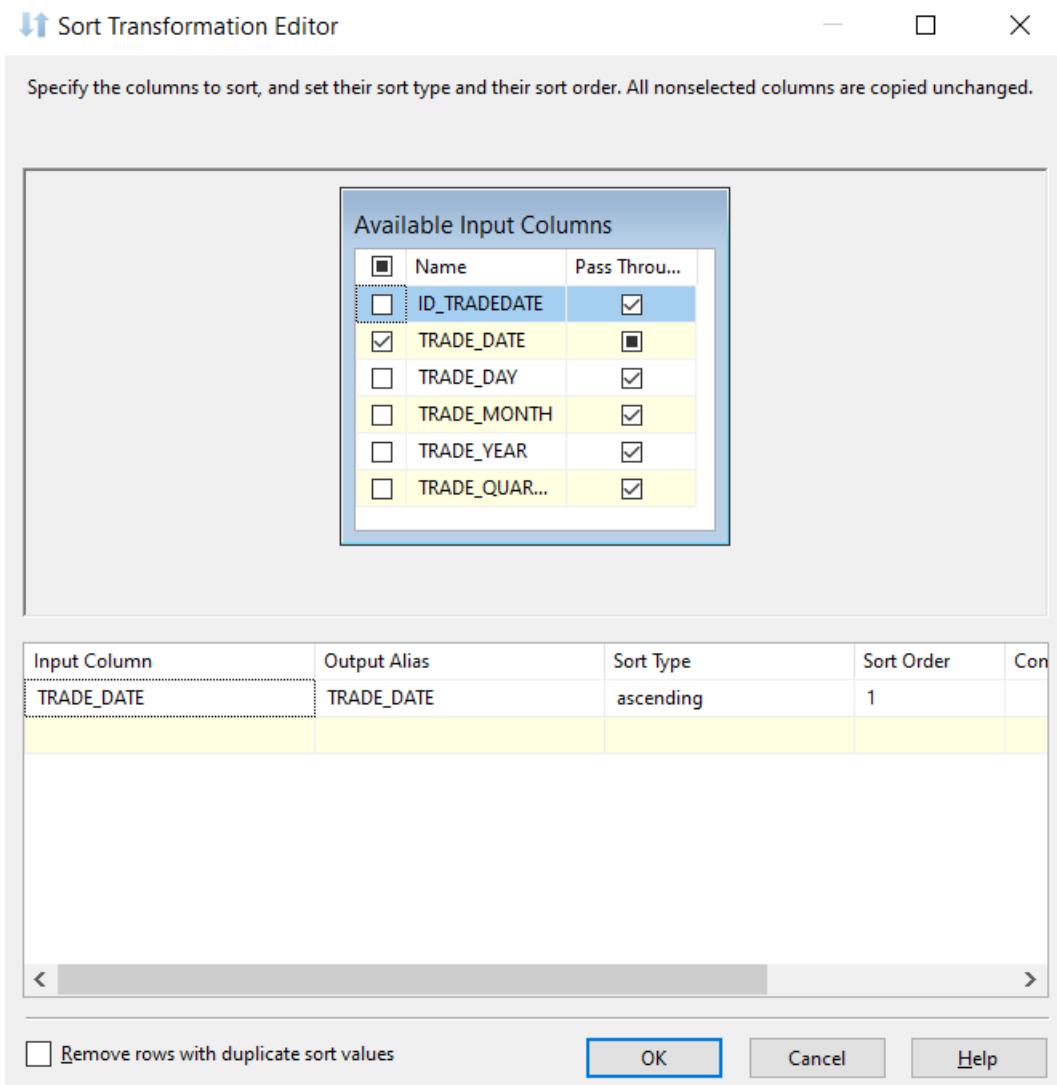
Hình 127: Tạo 2 Sort tương ứng với mỗi Source

Bước 6. Tại **Sort**, click chuột phải chọn **Edit** và chọn cột TRADE_DATE giống với bảng DIM_DATE để chuẩn bị cho quá trình merge



Hình 128: Chọn cột để chuẩn bị cho quá trình merge

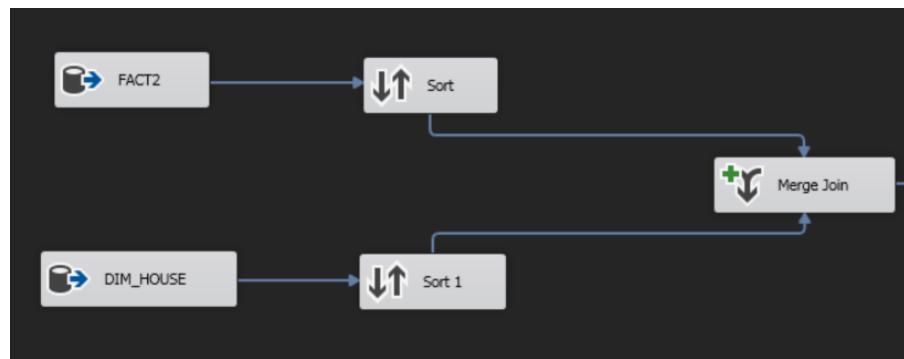
Bước 7. Tương tự ta chọn cột TRADE_DATE cho Sort1



Hình 129: Chọn cột TRADE_DATE cho Sort1

Bước 8. Tạo một **Merge Join** và nối với **Sort**, tiếp theo chọn **Merge Join Left Input** để giữ lại toàn bộ các dòng trong bảng FACT3 bất kể có kết quả khi thực hiện phép kết trái với cột ID của bảng **TRADE_DATE** hay không.

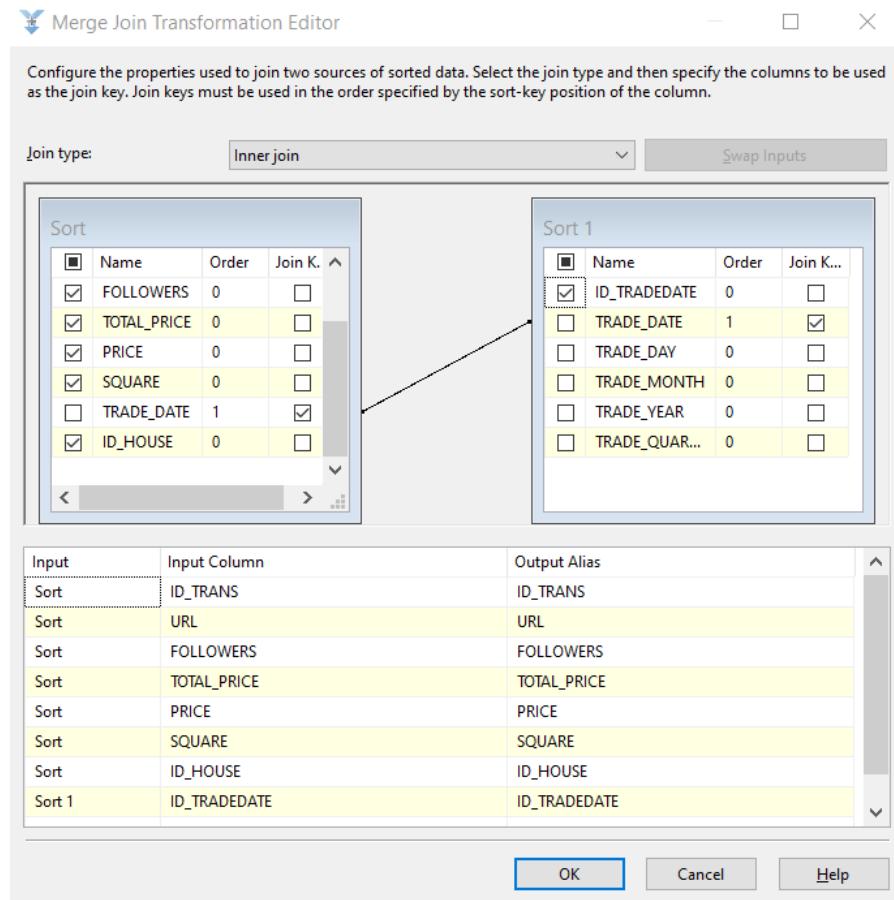
- Nối Sort1 với Merge Join



Hình 130: Nối Sort1 với Merge Join

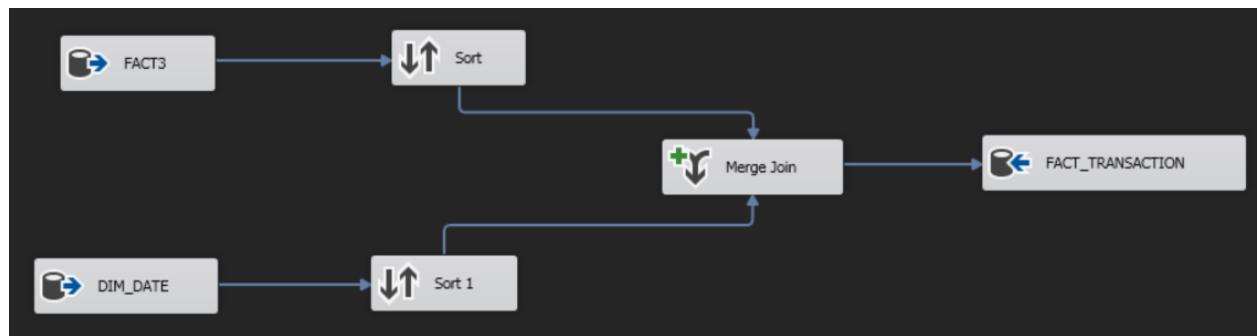
Bước 9. Chuột phải vào **Merge Join** và nhấn **Edit**, một hộp thoại **merge editor** xuất hiện: ở đây ta tick chọn tất cả các cột của **Sort** nhưng không lấy thuộc tính **TRADE_DATE**.

- Tiếp theo ta chọn **ID_TRADEDATE** ở **Sort1** để merge vào **FACT3**
- Kết quả sau khi merge là bảng **FACT3** không còn thuộc tính **TRADE_DATE** và có thêm 1 thuộc tính mới là **ID_TRADEDATE**

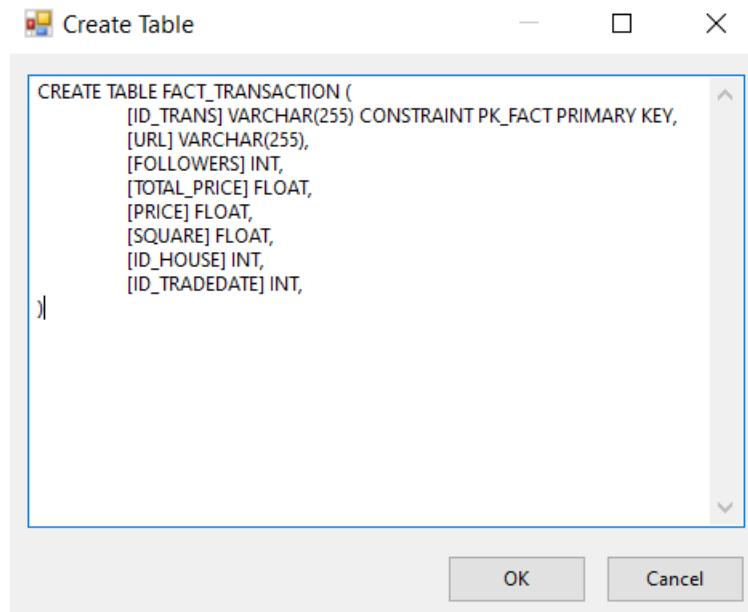


Hình 131: Kết quả sau khi merge

Bước 10. Tạo bảng FACT_TRANSACTION từ một OLE DB Destination để chứa tất cả những gì đã merge



Hình 132: Tạo bảng FACT_TRANSACTION để chứa tất cả những gì đã merge



Hình 133: Câu lệnh SQL tạo bảng FACT_TRANSACTION

*Nội dung câu lệnh SQL tạo bảng FACT_TRANSACTION như sau:

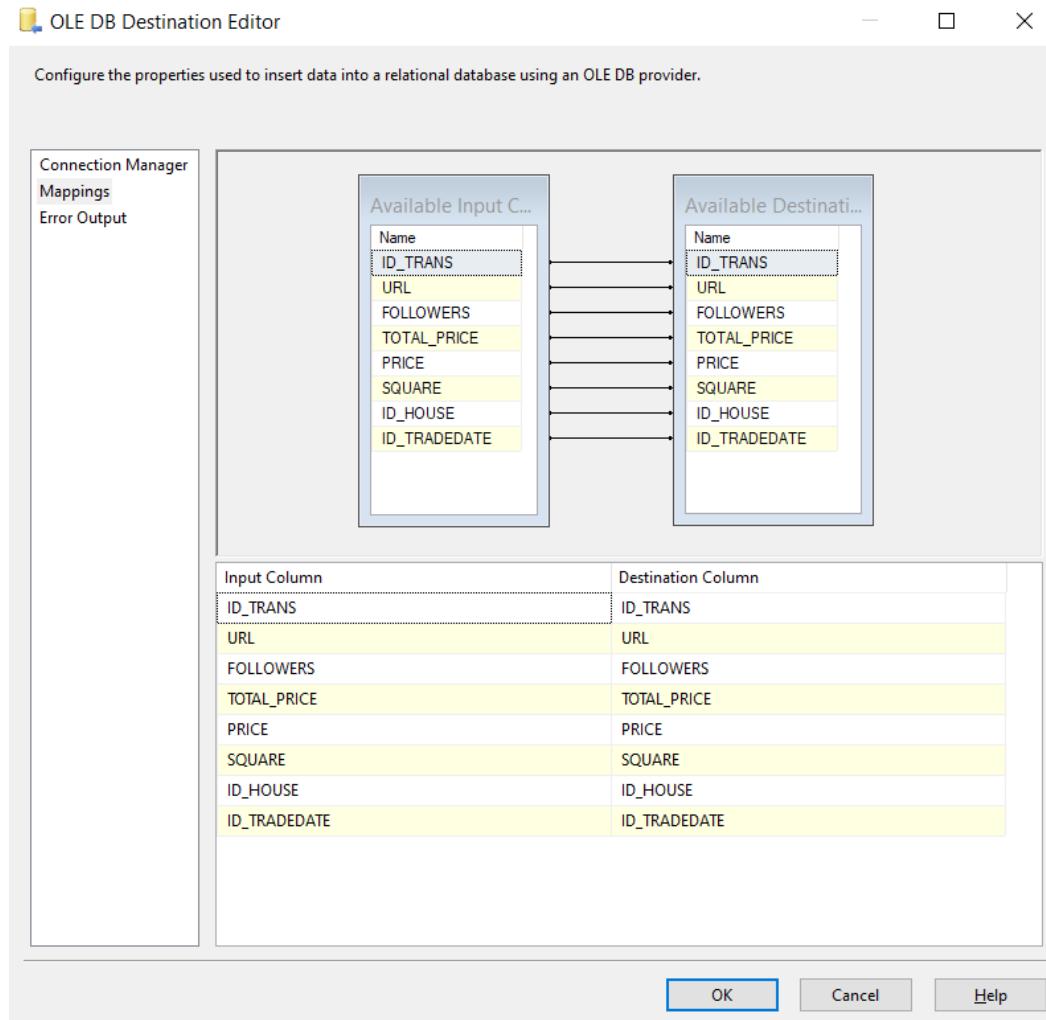
CREATE TABLE FACT_TRANSACTION (

```

[ID_TRANS] VARCHAR(255) CONSTRAINT PK_FACT PRIMARY KEY,
[URL] VARCHAR(255),
[FOLLOWERS] INT,
[TOTAL_PRICE] FLOAT,
[PRICE] FLOAT,
  
```

[SQUARE] FLOAT,
 [ID_HOUSE] INT,
 [ID_TRADEDATE] INT,
)

- Chọn mục **Mappings** để xem xét việc ánh xạ các cột dữ liệu



Hình 134: Xem xét việc ánh xạ các cột dữ liệu

- Nhấn **OK**

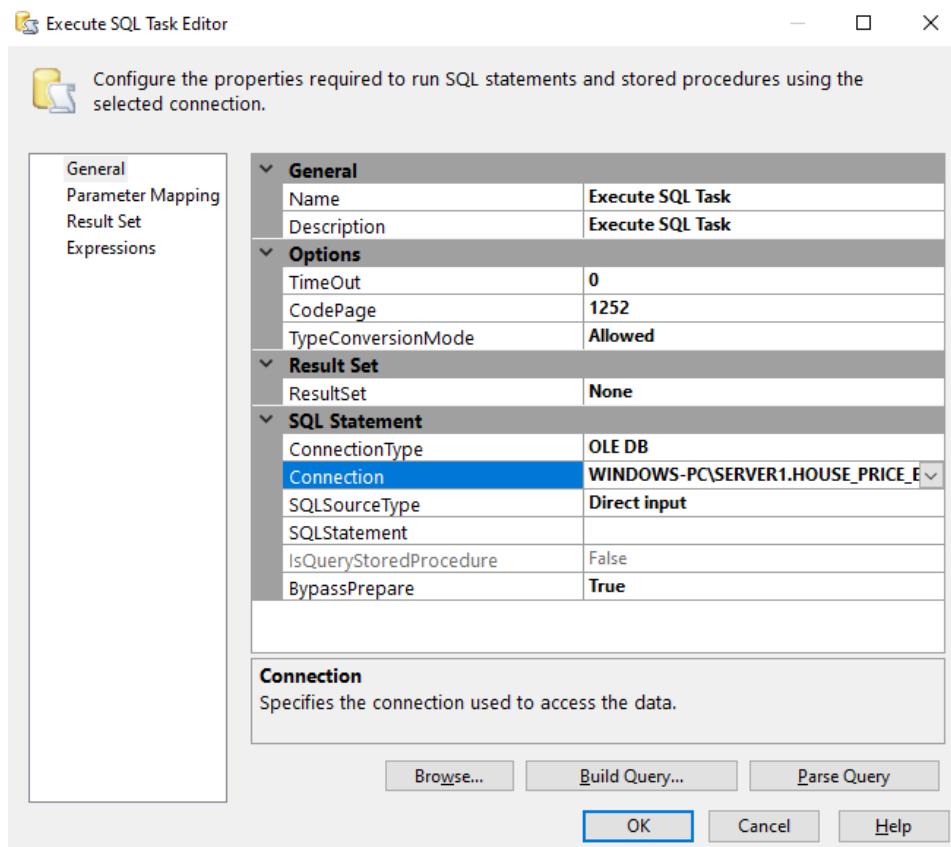
4.12. Tạo khóa ngoại từ bảng Fact đến các Dimension

Bước 1. Tạo một Execute SQL Task để thực thi các câu lệnh SQL tạo các khóa ngoại từ các Dimension đến bảng Fact.



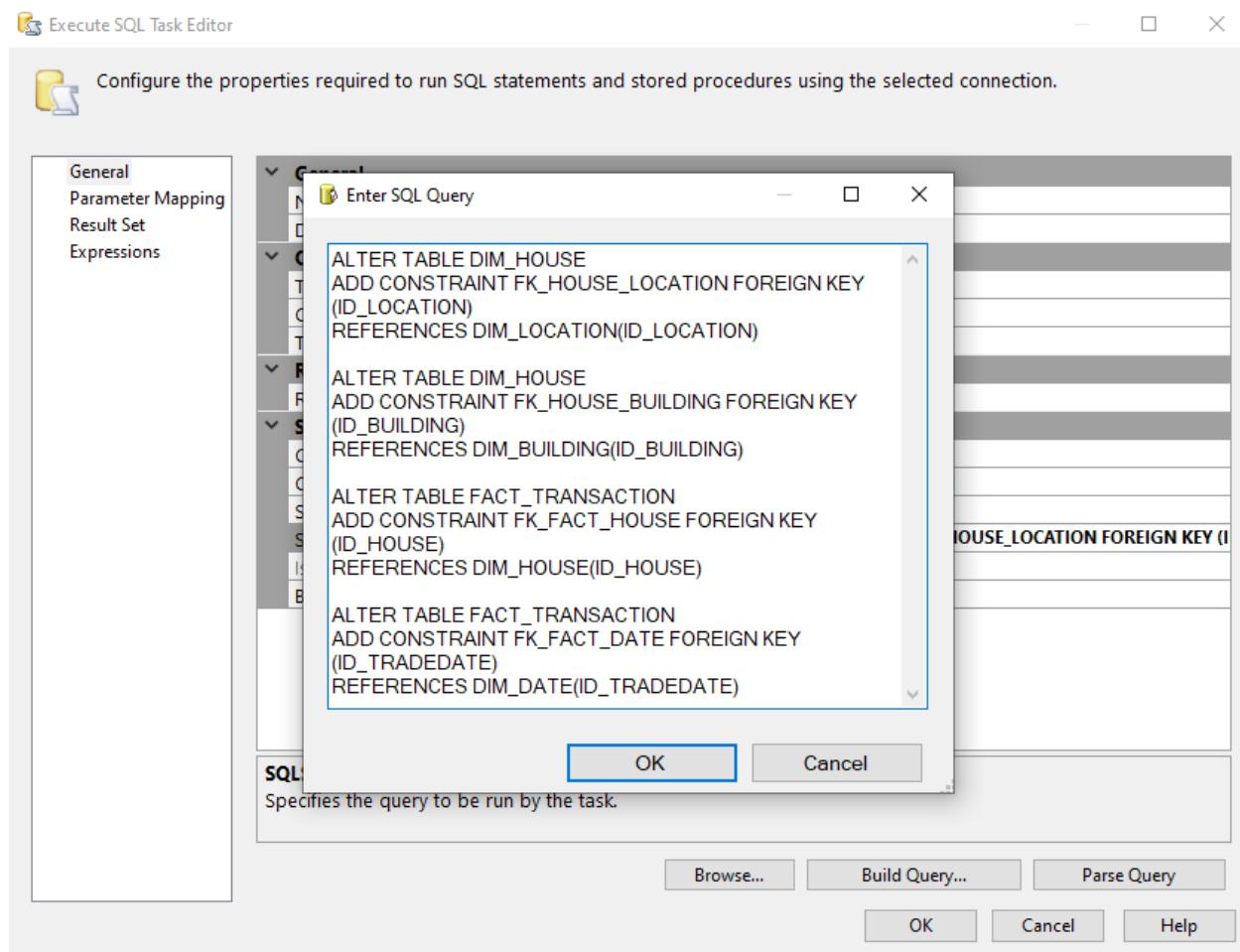
Hình 135: Tạo một Execute SQL Task

Bước 2. Nhấn chuột phải vào **Execute SQL Task** này và chọn **Edit**. Ở ô **Connection**, chọn **connection** đã thiết lập đến **data warehouse** trong **SQL Server**



Hình 136: Chọn connection đã thiết lập đến data warehouse

Bước 3. Ở ô **SQLStatement**, thêm các câu truy vấn SQL thực hiện tạo các khóa ngoại từ các **Dimension** đến bảng **Fact**.

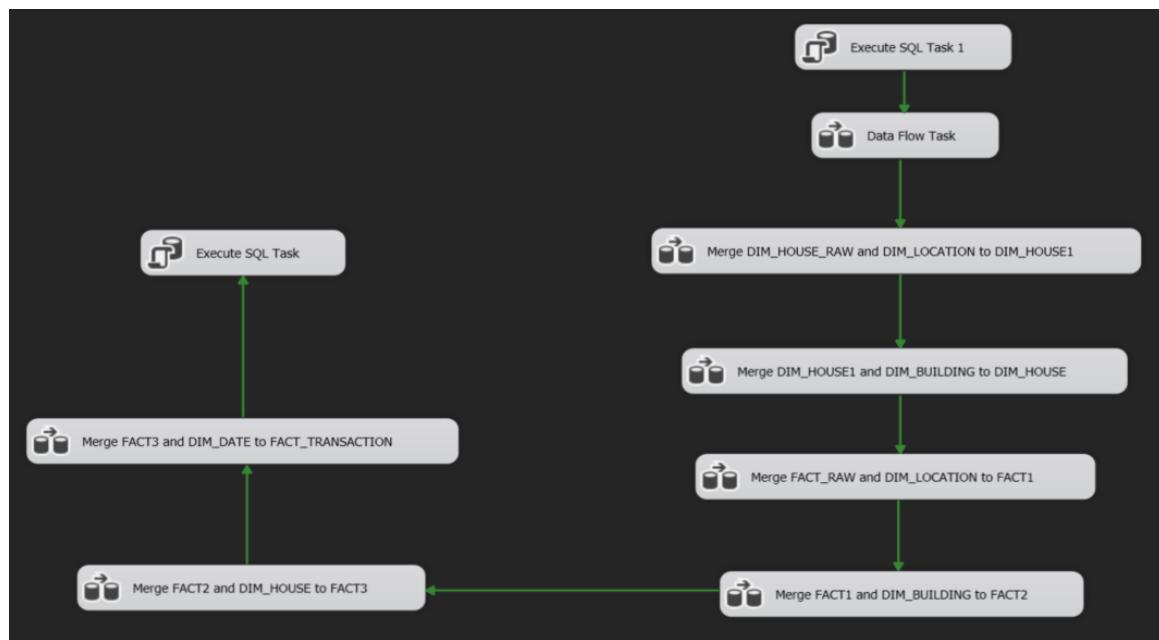


Hình 137: Thêm các câu truy vấn SQL thực hiện tạo các khóa ngoại từ các Dimension đến bảng Fact

- Nhấn **OK** để hoàn tất quá trình.

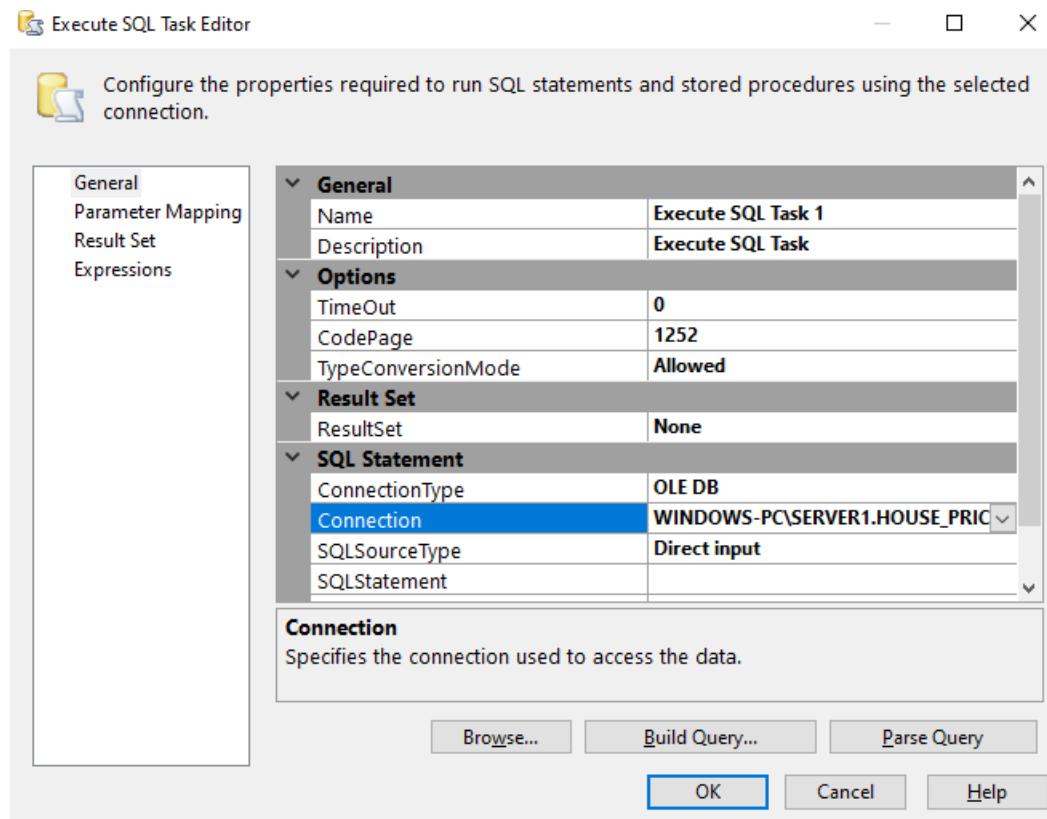
4.13. Chạy dự án SSIS

Bước 1. Thêm vào một **Execute SQL Task** nhằm thực hiện nhiệm vụ đảm bảo dữ liệu mới hoàn toàn (không bị chòng chéo dữ liệu cũ) mỗi khi chạy project, trước quá trình chia bảng **Fact** và các **Dimension**



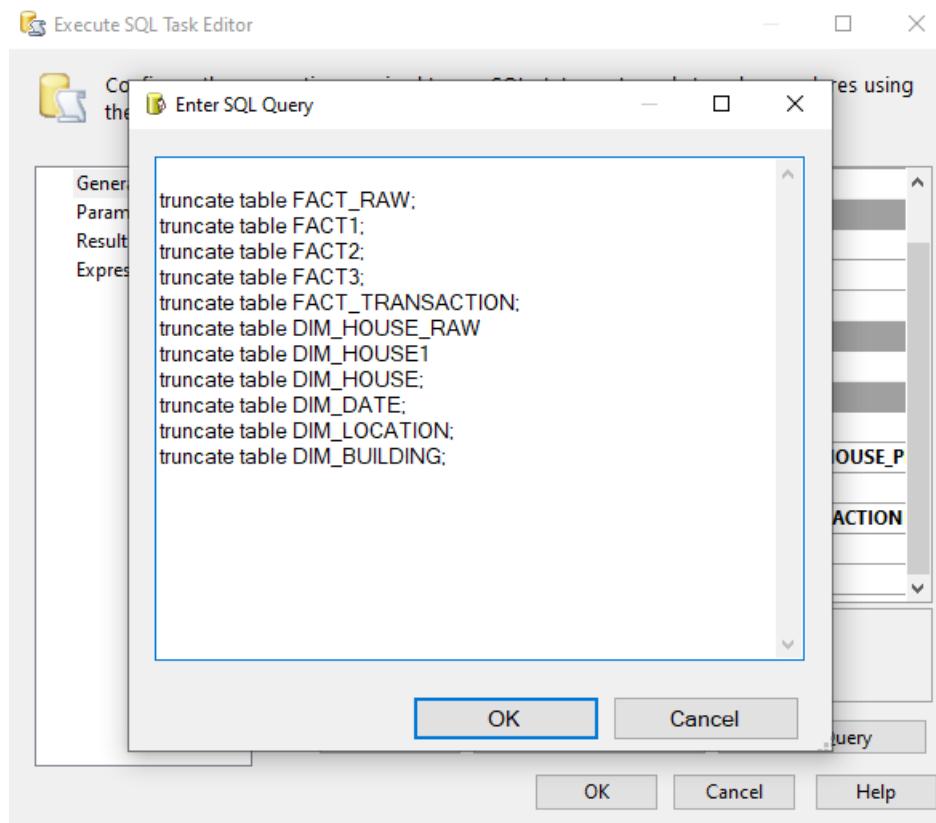
Hình 138: Thêm vào một Execute SQL Task

Bước 2. Nhấn chuột phải vào Execute SQL Task này và chọn Edit. Ở ô Connection, chọn connection đã thiết lập đến data warehouse trong SQL Server



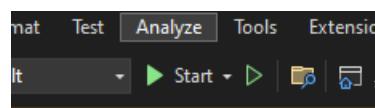
Hình 139: Chọn connection đã thiết lập đến data warehouse

Bước 3. Ở ô **SQLStatement**, thêm các câu truy vấn SQL thực hiện xóa dữ liệu cũ trong các bảng mỗi khi chạy project.



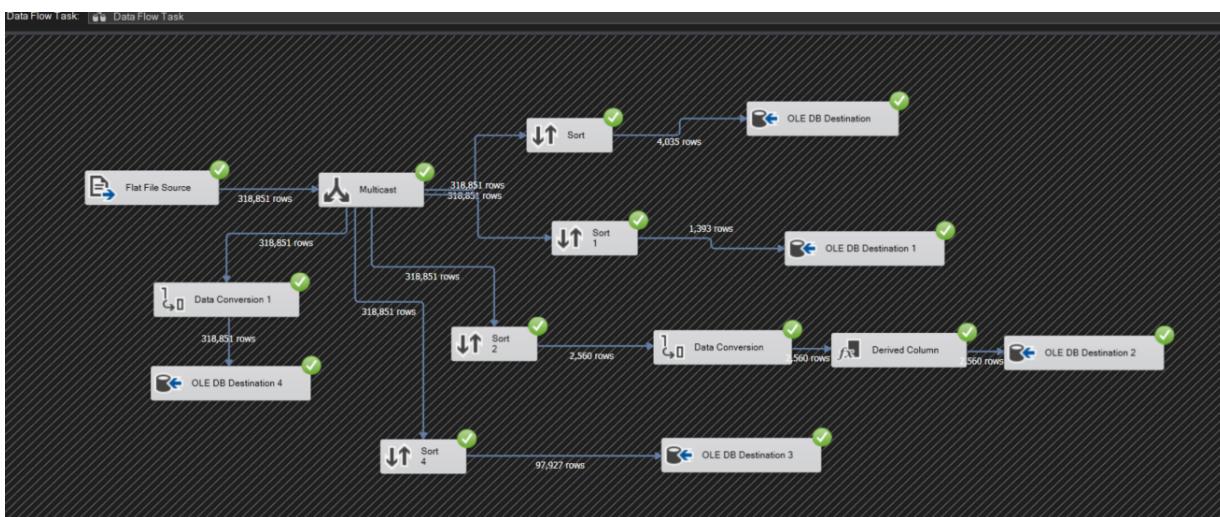
Hình 140: Thêm các câu truy vấn SQL thực hiện xóa dữ liệu cũ trong các bảng

Bước 4. Nhấn nút **Start** trên thanh menu để tiến hành chạy project

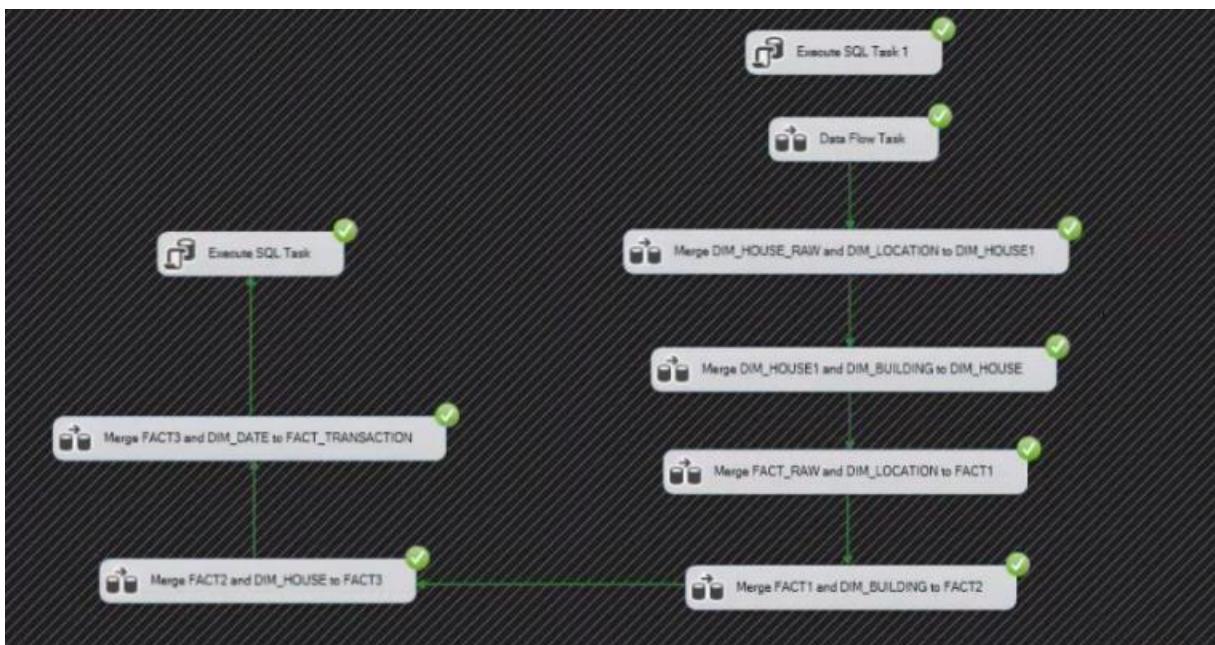


Hình 141: Nút Start

- Kết quả chạy project:

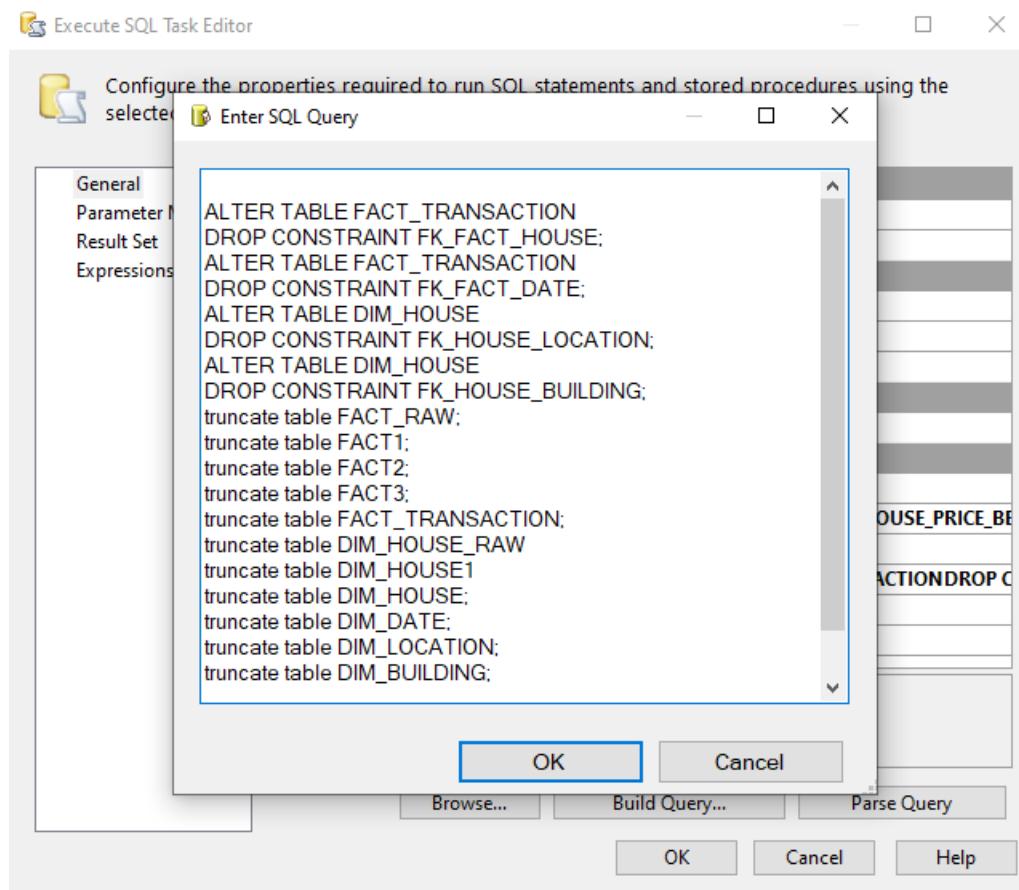


Hình 142: Kết quả chạy project (1)



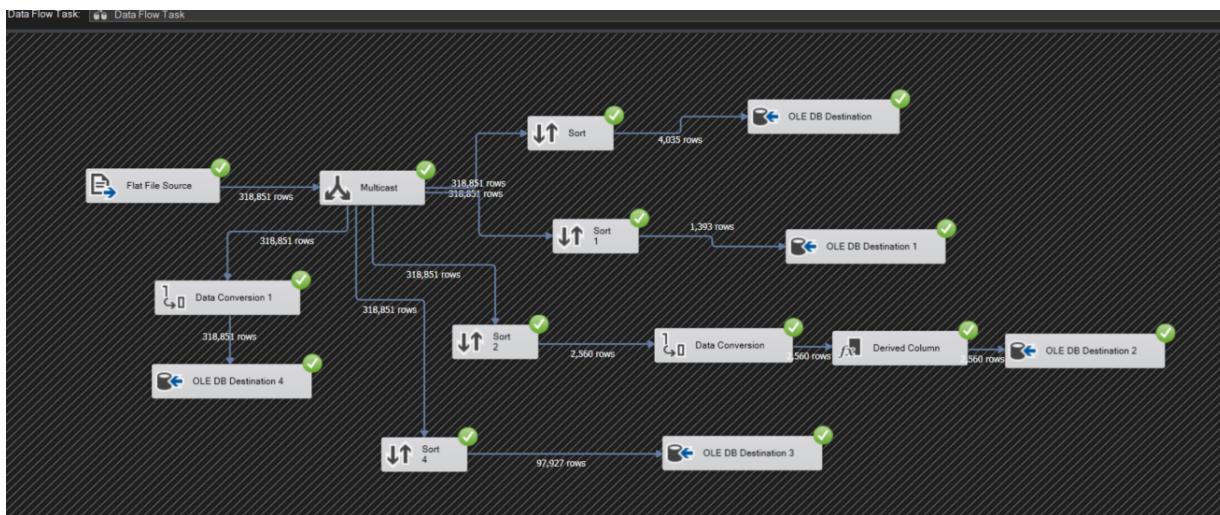
Hình 143: Kết quả chạy project (2)

Bước 5. Sau khi đã chạy project lần đầu thành công, khóa ngoại của bảng **Fact** tham chiếu đến các **Dimension** đã được tạo. Ta tiến hành thêm các lệnh SQL để xóa và thêm lại các khóa ngoại vào mỗi lần khởi chạy lại project tiếp theo.

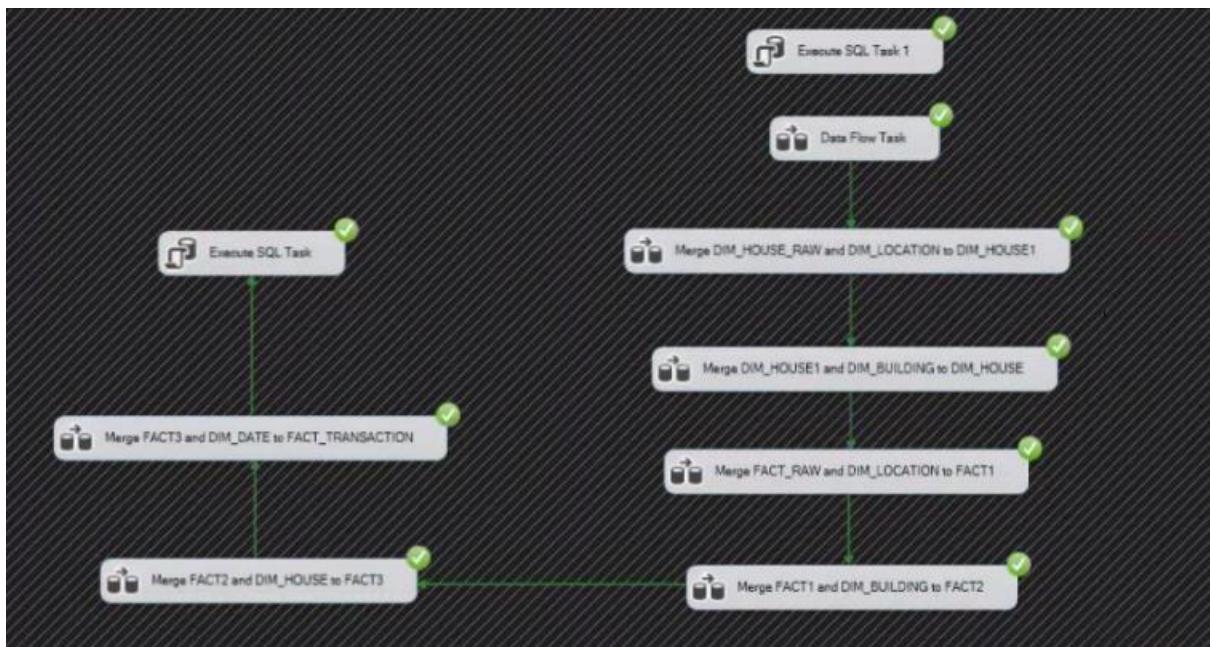


Hình 144: Thêm các lệnh SQL để xóa và thêm lại các khóa ngoại

Bước 6. Tiến hành chạy lại project



Hình 145: Chạy lại project (1)



Hình 146: Chạy lại project (2)

4.14. Kiểm tra dữ liệu

- Kiểm tra dữ liệu bảng DIM_LOCATION

ID_LOCATION	LNG	LAT	CID	DISTRICT
1	116.072514	39.678855	1111055906573	5
2	116.073082	39.660414	1111055947771	5
3	116.077958	39.656478	1111027379544	5
4	116.081646	39.720928	1111055565875	5
5	116.111318	39.949921	1111027377794	12
6	116.115806	40.235586	1111057423750	6
7	116.116904	39.910865	1111047727551	12
8	116.117019	39.93386	1111052073923	12
9	116.117724	39.941152	1111052110644	12
10	116.118409	39.937888	1111043529456	12
11	116.119651	39.934504	1111027382024	12
12	116.121949	40.215899	1111027378569	6
13	116.121964	39.939762	1111027375862	12
14	116.122115	39.932268	1111027378707	12
15	116.122354	39.937384	1111027376723	12
16	116.122491	39.939735	1111027377936	12
17	116.128533	39.723864	1111027377376	5
18	116.128779	40.242216	1111059830724	6
19	116.129677	40.252758	1111100788735	6
20	116.131543	40.25064	1111060303550	6
21	116.132633	40.240934	1111062314453	6
22	116.132908	40.248933	1111063030340	6
23	116.134232	40.251337	1111059823659	6
24	116.134792	39.743391	1111027375971	5
25	116.136098	40.241519	1111099837711	6
26	116.137765	40.243738	1111100893076	6
27	116.140363	40.177269	1111027380914	6
28	116.140365	40.242829	1111100070014	6

Hình 147: Kiểm tra dữ liệu bảng DIM_LOCATION

- Kiểm tra dữ liệu bảng DIM_BUILDING

	ID_BUILDING	BUILDINGTYPE	CONSTRUCTIONTIME	RENOVATIONCONDITION	BUILDINGSTRUCTURE
1	1	1	1970	1	2
2	2	1	1970	3	2
3	3	1	1970	3	6
4	4	1	1970	4	6
5	5	1	1979	1	6
6	6	1	1979	2	6
7	7	1	1979	3	6
8	8	1	1979	4	6
9	9	1	1980	1	2
10	10	1	1980	1	4
11	11	1	1980	1	6
12	12	1	1980	2	6
13	13	1	1980	3	2
14	14	1	1980	3	4
15	15	1	1980	3	6
16	16	1	1980	4	2
17	17	1	1980	4	6
18	18	1	1981	1	2
19	19	1	1981	1	6
20	20	1	1981	2	2
21	21	1	1981	2	6
22	22	1	1981	3	2
23	23	1	1981	3	6
24	24	1	1981	4	2
25	25	1	1981	4	6

Hình 148: Kiểm tra dữ liệu bảng DIM_BUILDING

- Kiểm tra dữ liệu bảng DIM_HOUSE

	ID_HOUSE	NUM_LVROOM	NUM_DWROOM	NUM_KITCHEN	NUM_BATHROOM	ELEVATOR	ID_LOCATION	ID_BUILDING
1	1	1	1	1	1	0	1	1363
2	2	2	1	1	1	0	1	1363
3	3	2	1	1	1	0	1	1366
4	4	2	1	1	1	0	1	1368
5	5	2	1	1	1	0	1	1371
6	6	3	1	1	1	0	1	1366
7	7	3	1	1	1	0	1	1371
8	8	3	1	1	1	0	1	1375
9	9	3	2	1	1	0	1	1363
10	10	3	2	1	1	0	1	1366
11	11	3	2	1	1	0	1	1368
12	12	3	2	1	2	0	1	656
13	13	3	2	1	2	0	1	1366
14	14	3	2	1	2	0	1	1368
15	15	3	2	1	2	0	1	1371
16	16	2	1	1	1	0	2	1238
17	17	2	1	1	1	0	2	1239
18	18	2	1	1	1	0	2	1255
19	19	2	1	1	1	0	2	1256
20	20	2	1	1	1	0	2	1351

Hình 149: Kiểm tra dữ liệu bảng DIM_HOUSE

- Kiểm tra dữ liệu bảng DIM_DATE

	ID_TRADEDATE	TRADE_DATE	TRADE_DAY	TRADE_MONTH	TRADE_QUARTER	TRADE_YEAR
1	1	2010-01-01	1	1	1	2010
2	2	2011-01-01	1	1	1	2011
3	3	2012-01-01	1	1	1	2012
4	4	2013-01-01	1	1	1	2013
5	5	2014-01-01	1	1	1	2014
6	6	2015-01-01	1	1	1	2015
7	7	2016-01-01	1	1	1	2016
8	8	2017-01-01	1	1	1	2017
9	9	2018-01-01	1	1	1	2018
10	10	2011-01-10	10	1	1	2011
11	11	2012-01-10	10	1	1	2012
12	12	2013-01-10	10	1	1	2013
13	13	2014-01-10	10	1	1	2014
14	14	2015-01-10	10	1	1	2015
15	15	2016-01-10	10	1	1	2016
16	16	2017-01-10	10	1	1	2017
17	17	2009-01-11	11	1	1	2009
18	18	2011-01-11	11	1	1	2011
19	19	2012-01-11	11	1	1	2012
20	20	2013-01-11	11	1	1	2013

Hình 150: Kiểm tra dữ liệu bảng DIM_DATE

- Kiểm tra dữ liệu bảng FACT_TRANSACTION

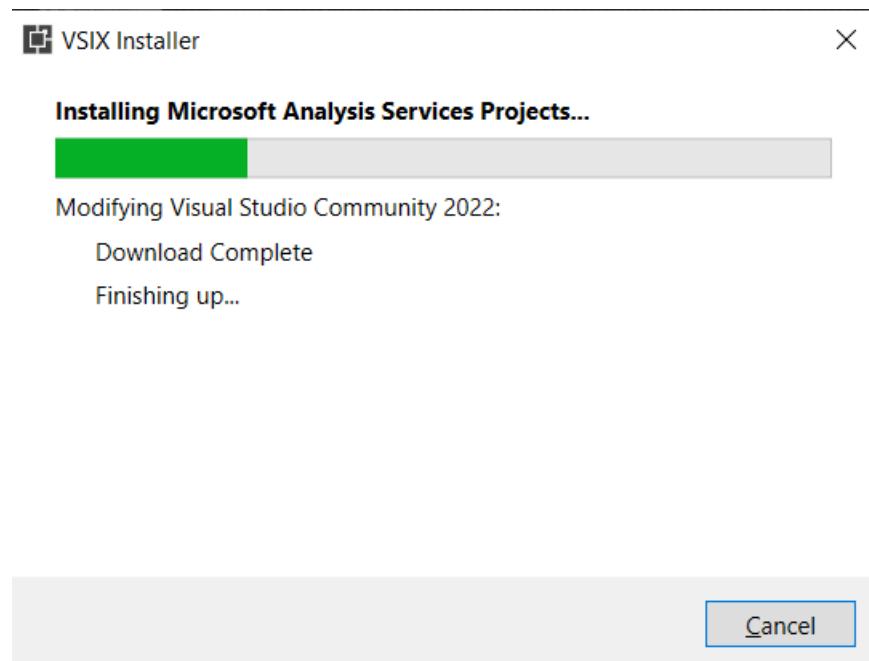
	ID_TRANS	URL	FOLLOWERS	TOTAL_PRICE	PRICE	SQUARE	ID_HOUSE	ID_TRADEDATE
1	101084782030	https://bj.lianjia.com/chengjiao/101084782030.html	106	4150000	31680	131	74561	2336
2	101086012217	https://bj.lianjia.com/chengjiao/101086012217.html	126	5750000	43436	132.38	66920	2032
3	101086041636	https://bj.lianjia.com/chengjiao/101086041636.html	48	10300000	52021	198	88497	678
4	101086406841	https://bj.lianjia.com/chengjiao/101086406841.html	138	2975000	22202	134	59962	2513
5	101086920653	https://bj.lianjia.com/chengjiao/101086920653.html	286	3920000	48396	81	55149	2262
6	101087277815	https://bj.lianjia.com/chengjiao/101087277815.html	57	2756000	52000	53	71399	1988
7	101087292623	https://bj.lianjia.com/chengjiao/101087292623.html	167	2750000	37672	73	77290	1925
8	101087303800	https://bj.lianjia.com/chengjiao/101087303800.html	138	8000000	49521	161.55	66922	2543
9	101087463212	https://bj.lianjia.com/chengjiao/101087463212.html	218	1340000	27917	48	87385	2520
10	101087508625	https://bj.lianjia.com/chengjiao/101087508625.html	134	3800000	55883	68	66501	2528
11	101087641313	https://bj.lianjia.com/chengjiao/101087641313.html	95	8400000	31579	266	86043	2232
12	101087793667	https://bj.lianjia.com/chengjiao/101087793667.html	74	1260000	26250	48	87391	2277
13	101087805216	https://bj.lianjia.com/chengjiao/101087805216.html	50	9480000	39832	238	74670	2239
14	101087812155	https://bj.lianjia.com/chengjiao/101087812155.html	65	8960000	91710	97.7	44343	2174
15	101087919464	https://bj.lianjia.com/chengjiao/101087919464.html	90	2850000	35568	80.13	9894	1735
16	101087921597	https://bj.lianjia.com/chengjiao/101087921597.html	76	3600000	60607	59.4	16384	2433
17	101087926378	https://bj.lianjia.com/chengjiao/101087926378.html	39	7000000	39115	178.96	45719	1988
18	101087957433	https://bj.lianjia.com/chengjiao/101087957433.html	6	2120000	32981	64.28	8419	2403
19	101088008763	https://bj.lianjia.com/chengjiao/101088008763.html	130	3360000	30189	111.3	91570	1764
20	101088050742	https://bj.lianjia.com/chengjiao/101088050742.html	41	1746000	33577	52	8419	2225

Hình 151: Kiểm tra dữ liệu bảng FACT_TRANSACTION

CHƯƠNG III: PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS)

1. Tạo Project SSAS

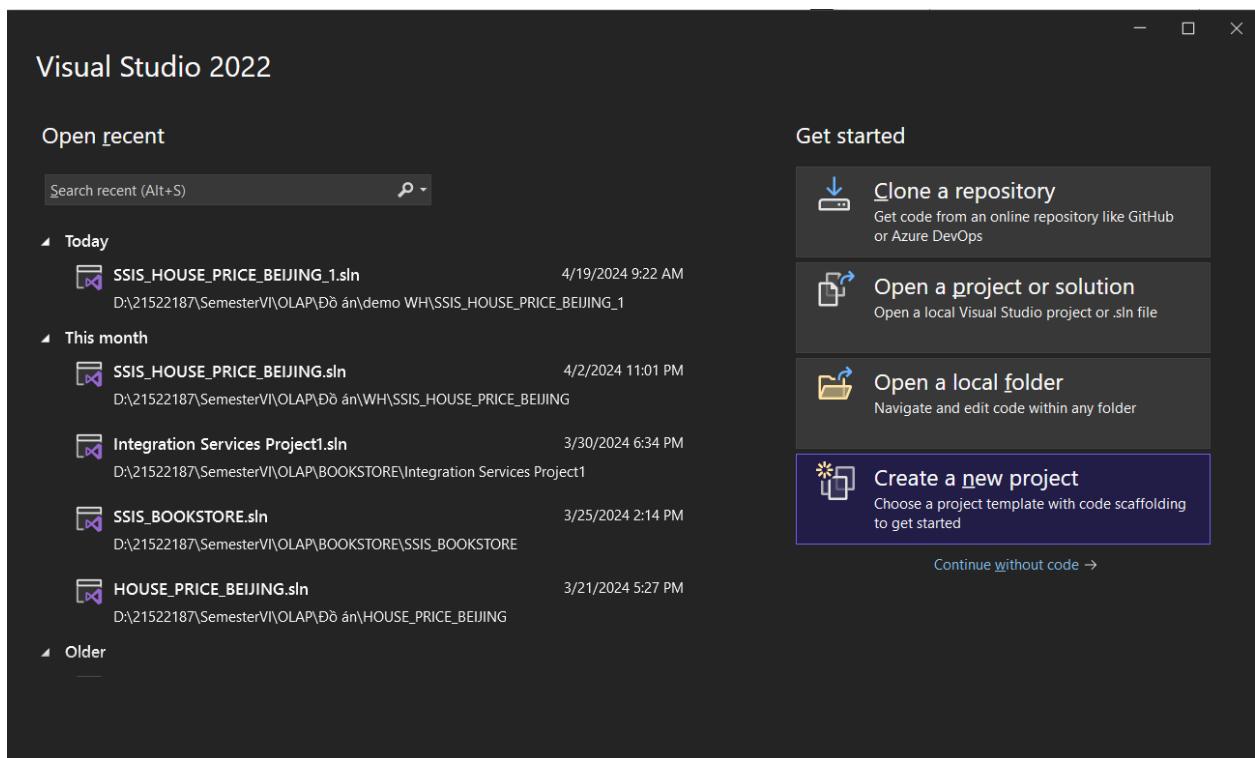
- Tải extention “Microsoft Analysis Services Project” giống như đã làm trong Chương II với SSIS



Hình 152: Tải extention “Microsoft Analysis Services Project”

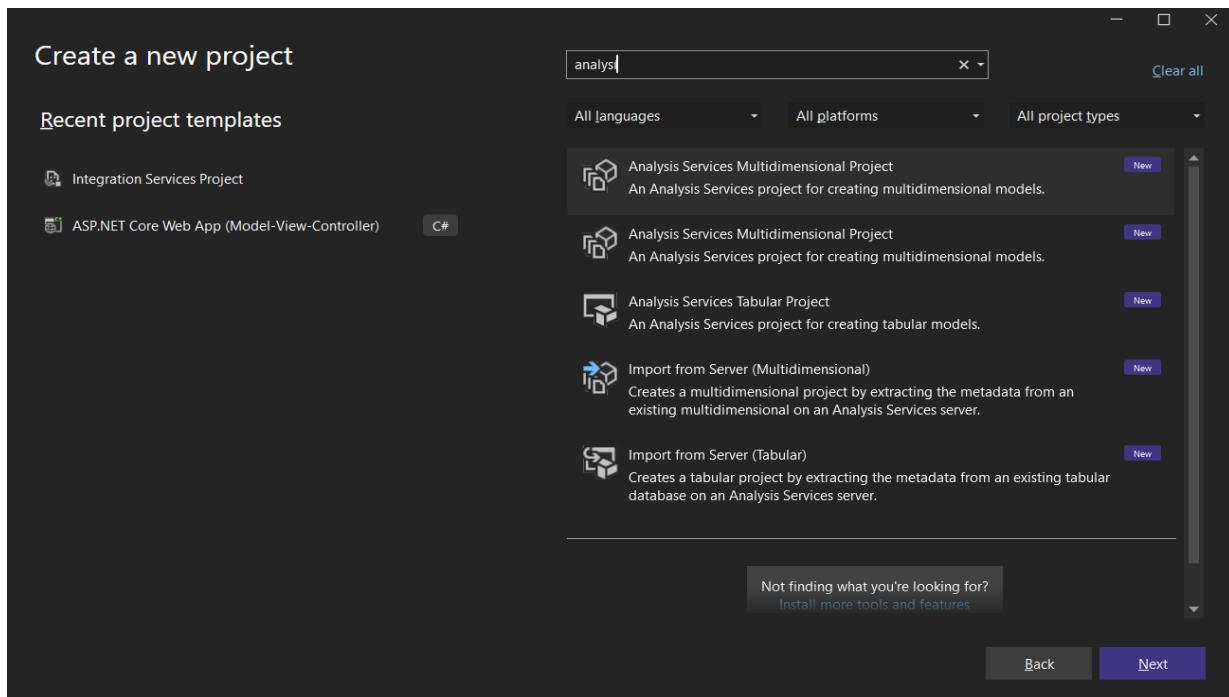
- Sau khi cài thành công, bắt đầu tạo project SSAS:

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



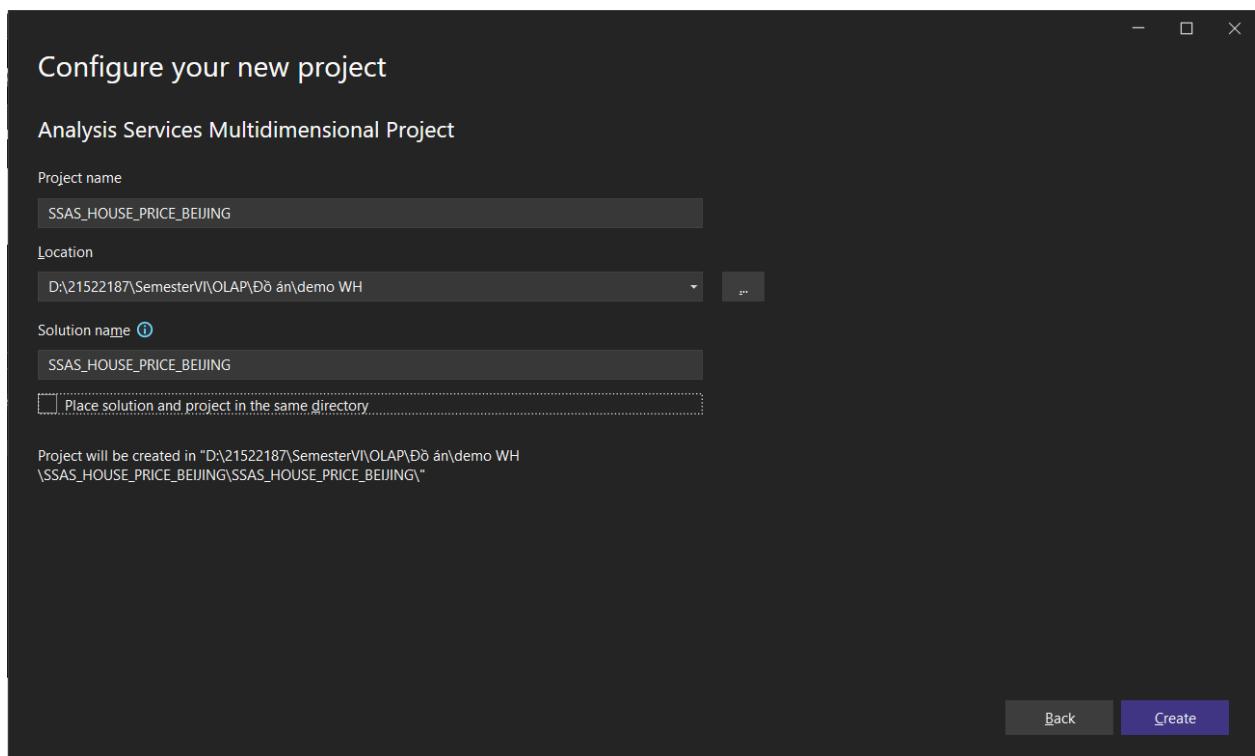
Hình 153: Chọn “Create a new project”

Bước 2: Chọn Analysis Services Multidimensional Project và chọn Next



Hình 154: Chọn Analysis Services Multidimensional Project

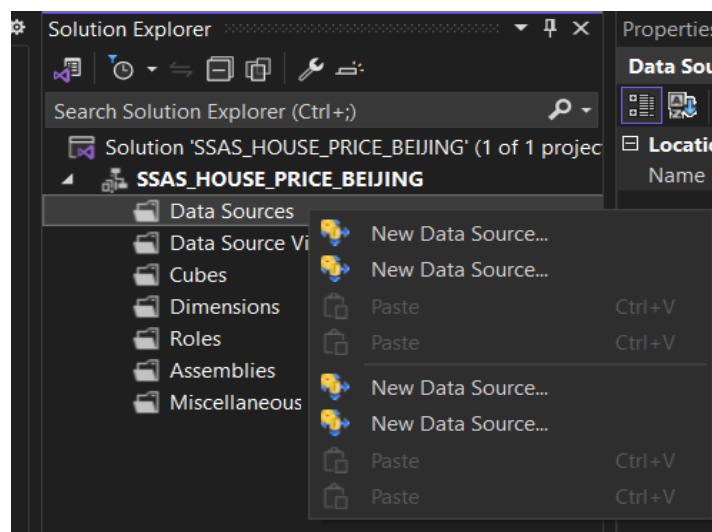
Bước 3: Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn Create.



Hình 155: Đặt tên và thiết lập đường dẫn cho Project

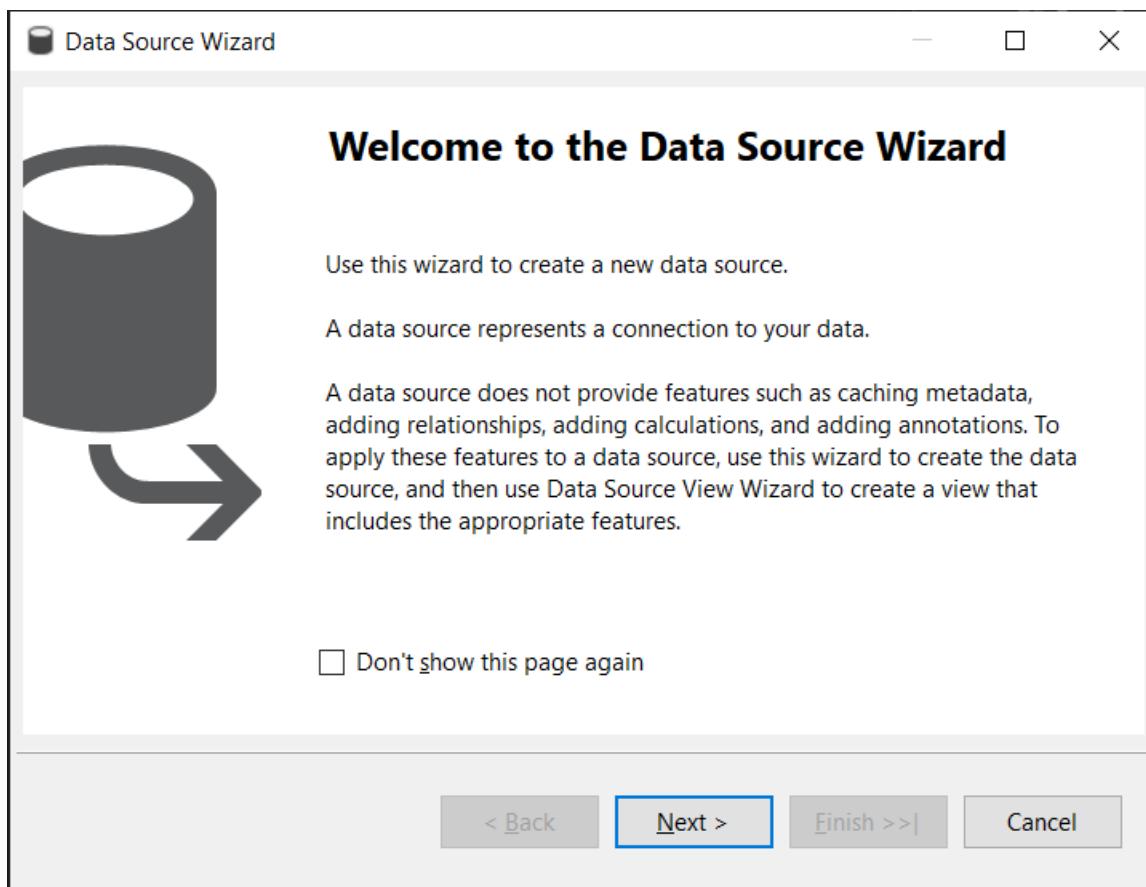
2. Xác định dữ liệu nguồn (Data Sources)

Bước 1: Tại **Solution Explorer**, ta click chuột phải vào thư mục **Data Sources** và chọn **New Data Source**.



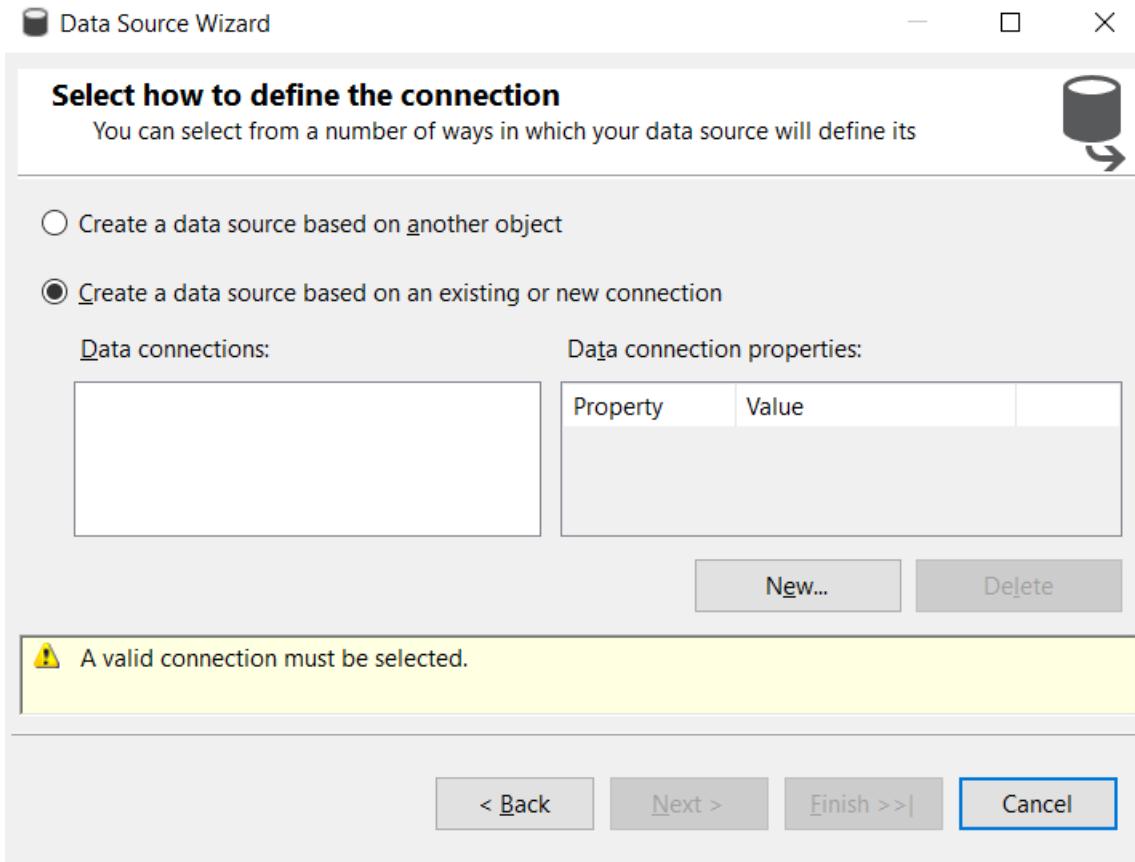
Hình 156: Chọn New Data Source

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



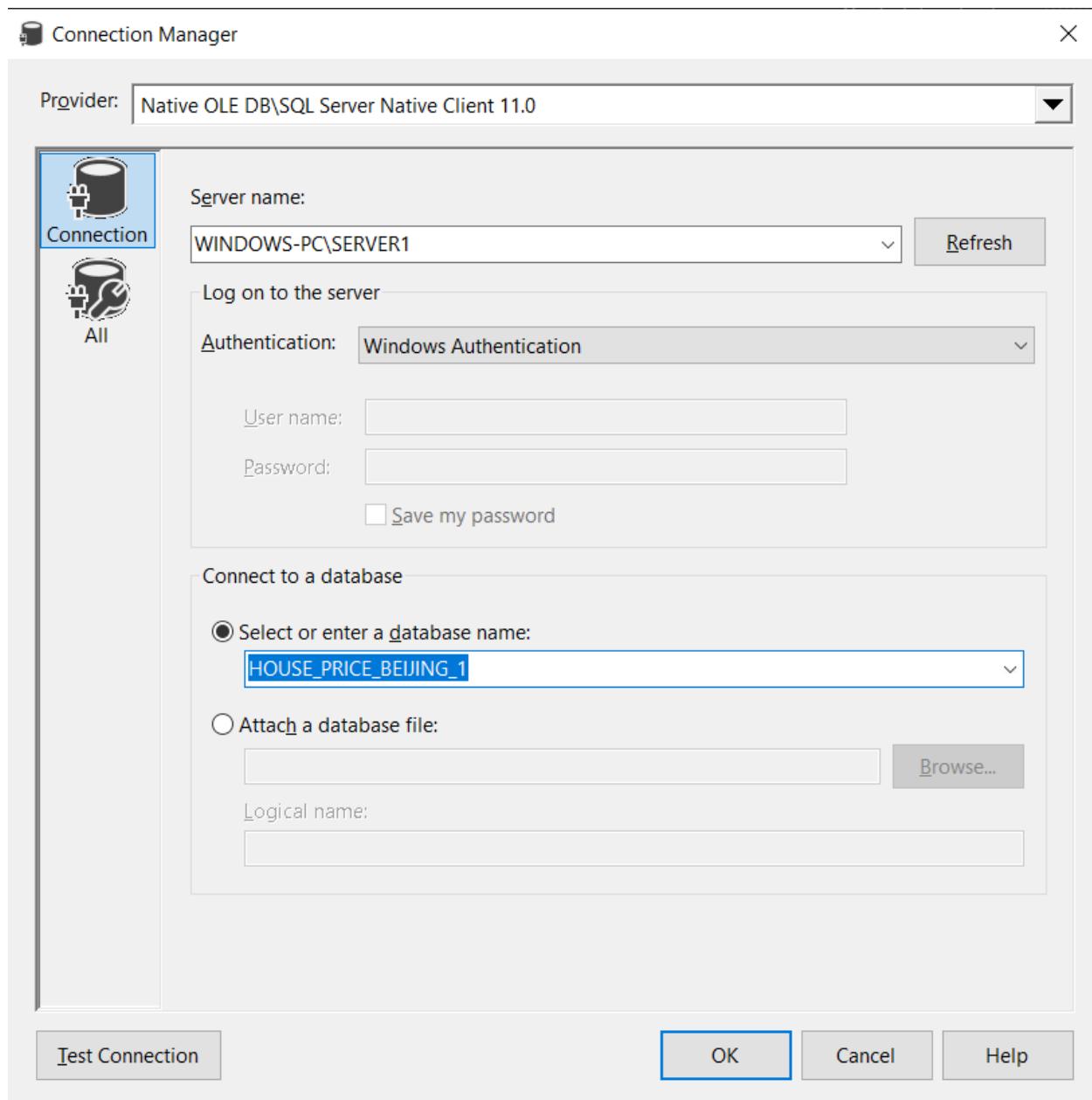
Hình 157: Chọn Next để tiếp tục

Bước 3: 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.



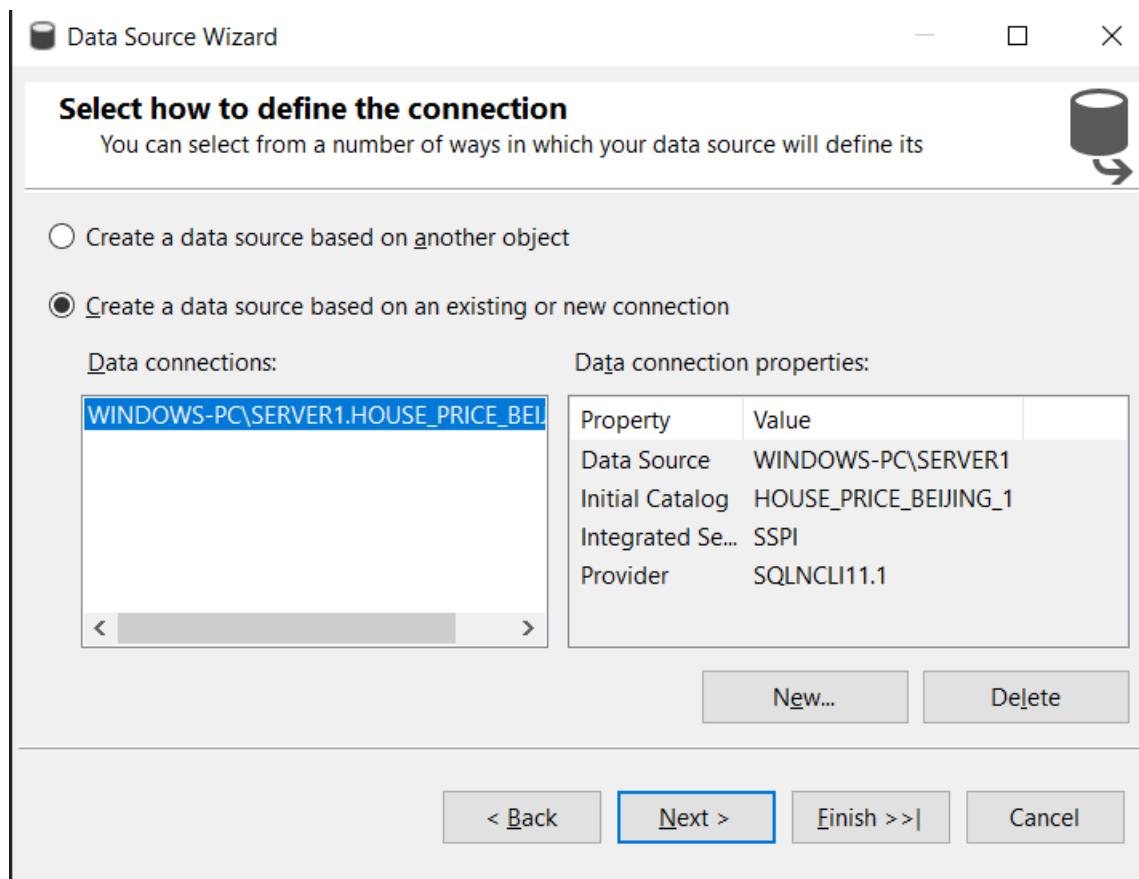
Hình 158: Tạo kết nối với cơ sở dữ liệu đã được tạo từ quá trình SSIS

Bước 4: Hộp thoại **Connection Manager** xuất hiện, ta nhập **Server name** và chọn cơ sở dữ liệu mà ta đã tạo ra từ quá trình SSIS. Sau đó chọn **OK**.



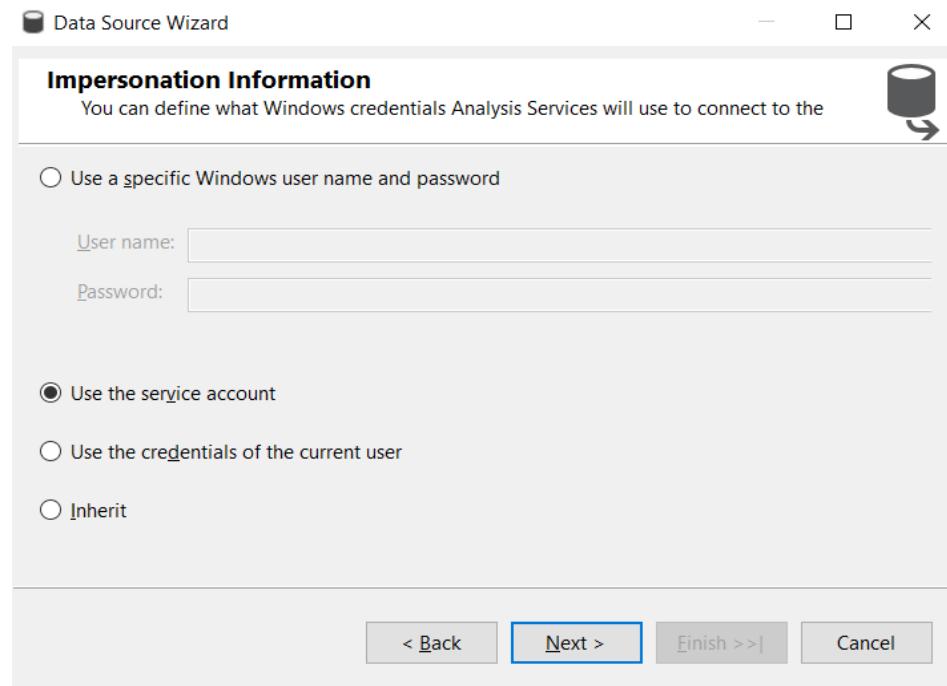
Hình 159: Tạo nhập Server name và chọn cơ sở dữ liệu

- Chọn data source vừa tạo và chọn **Next** để tiếp tục.



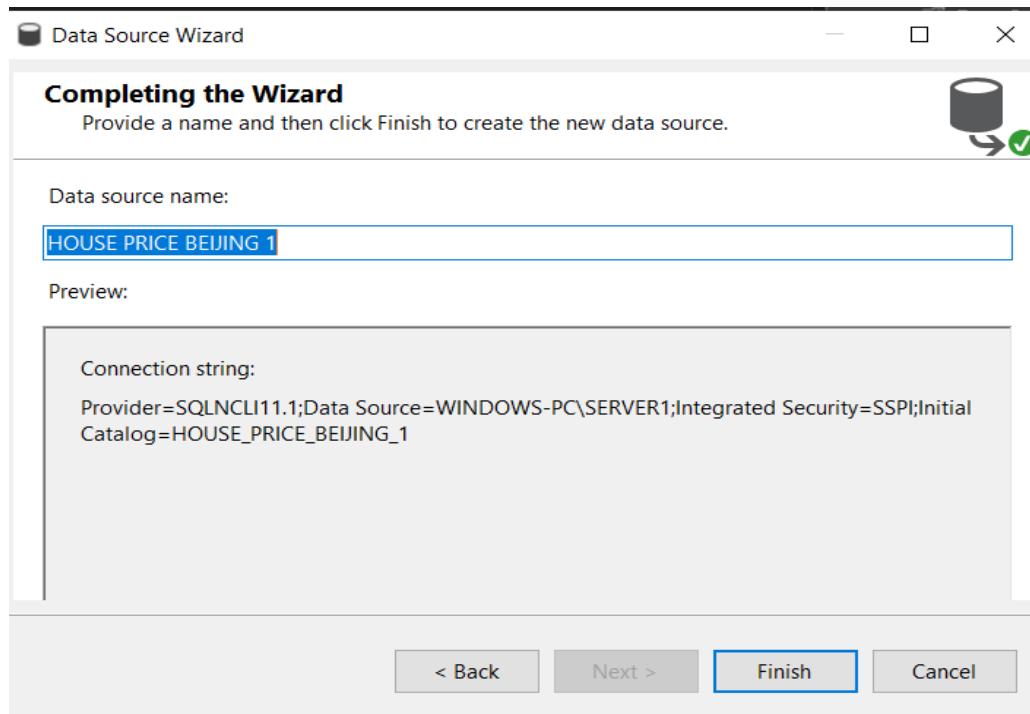
Hình 160: Chọn data source vừa tạo

Bước 5: Chọn “Use the service account”, sau đó chọn Next để tiếp tục.



Hình 161: Chọn “Use the service account”

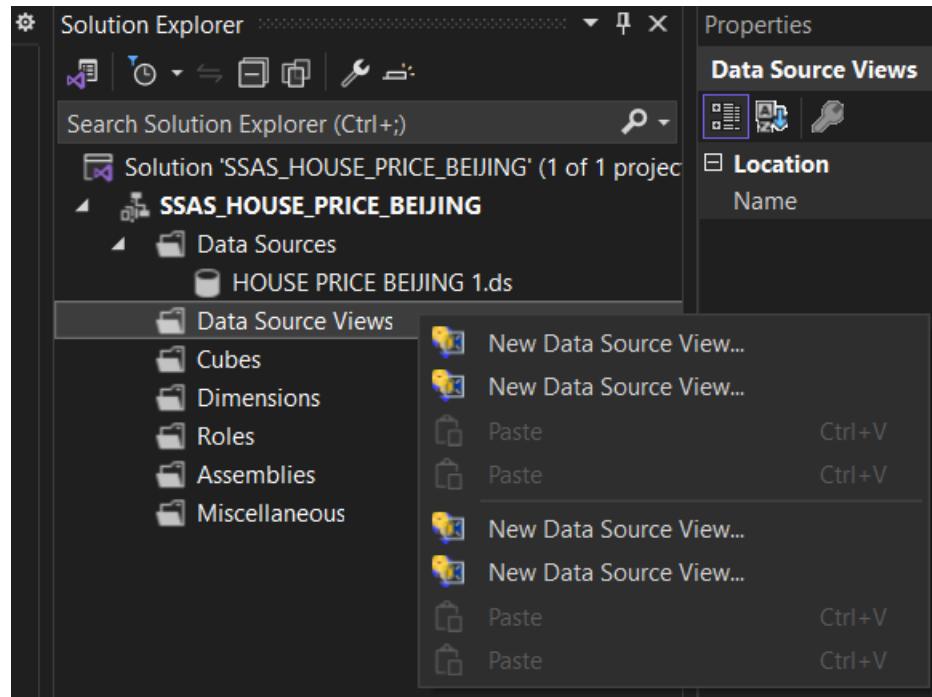
Bước 6: Cuối cùng ta chọn **Finish** để hoàn tất quy trình định nghĩa nguồn dữ liệu.



Hình 162: Chọn Finish để hoàn tất

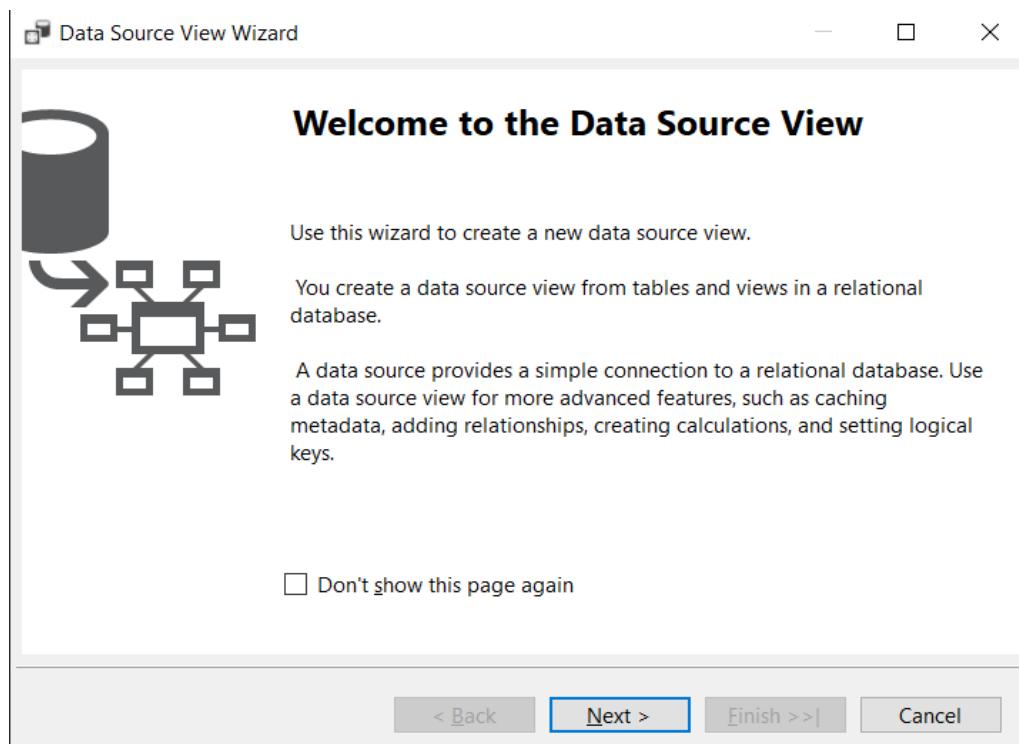
3. Xác định khung nhìn dữ liệu nguồn (Data Source Views)

Bước 1: Tại Solution Explorer, ta click chuột phải vào thư mục **Data Source Views** và chọn **New Data Source View**.



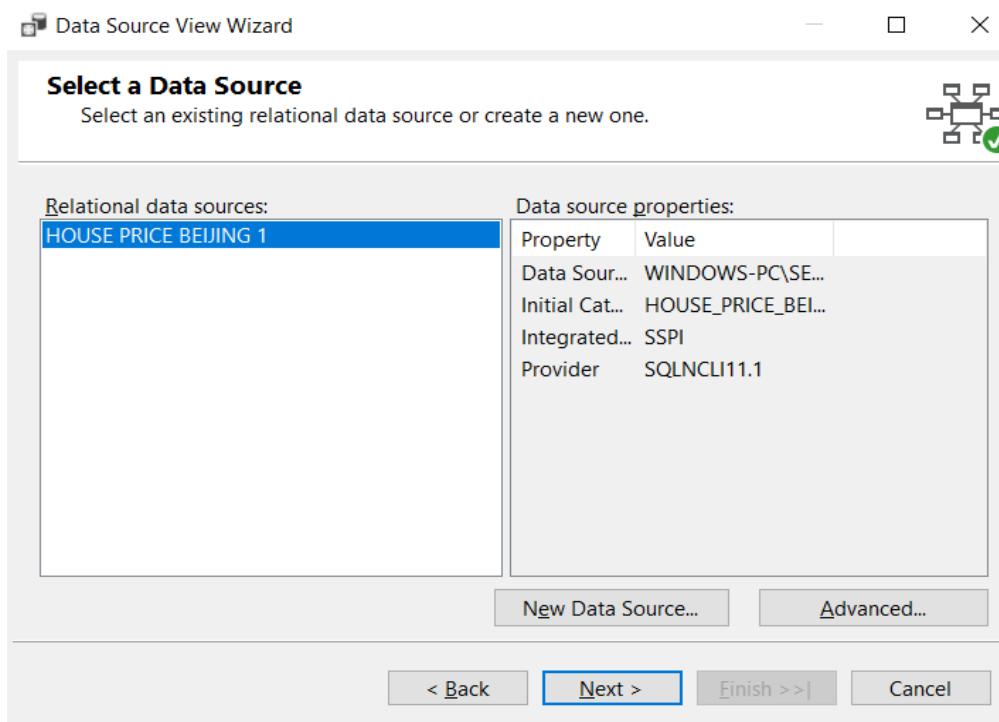
Hình 163: Chọn New Data Source View

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



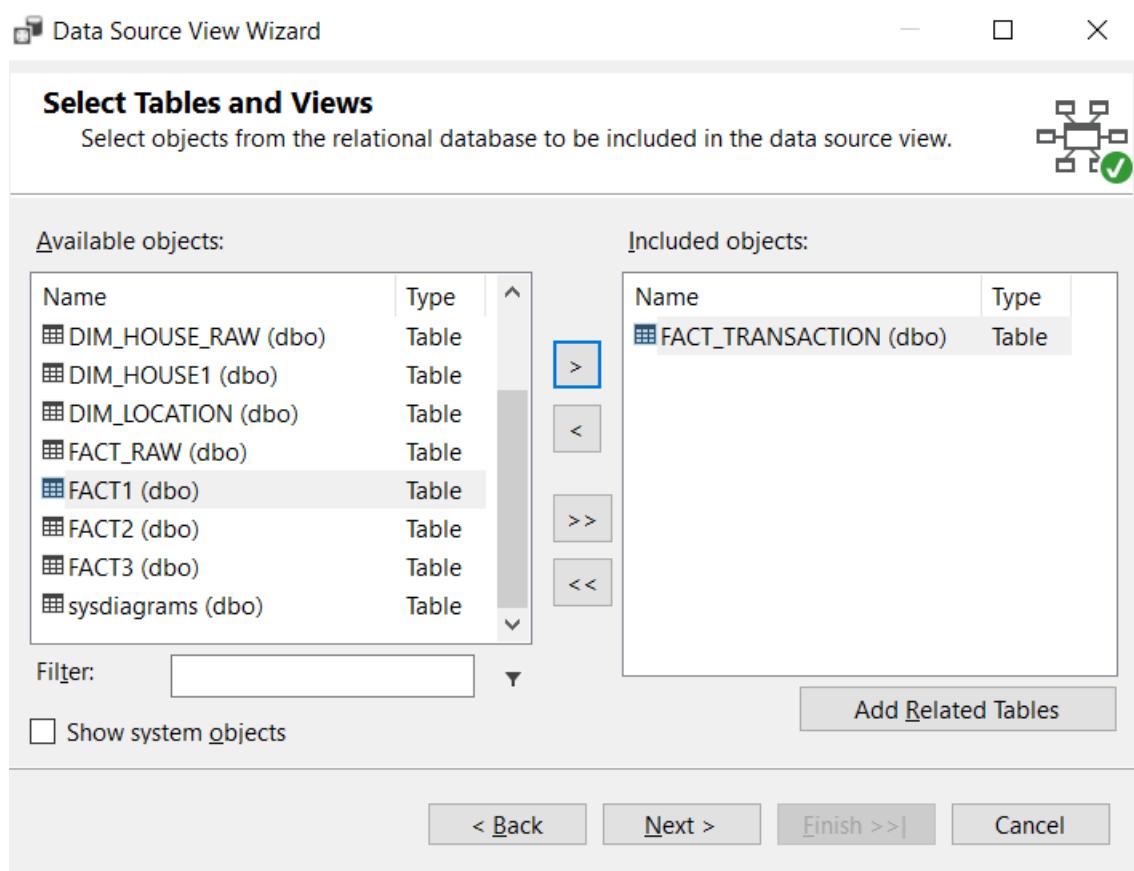
Hình 164: Chọn Next để tiếp tục

Bước 3: Chọn data source vừa tạo, sau đó chọn **Next** để tiếp tục.



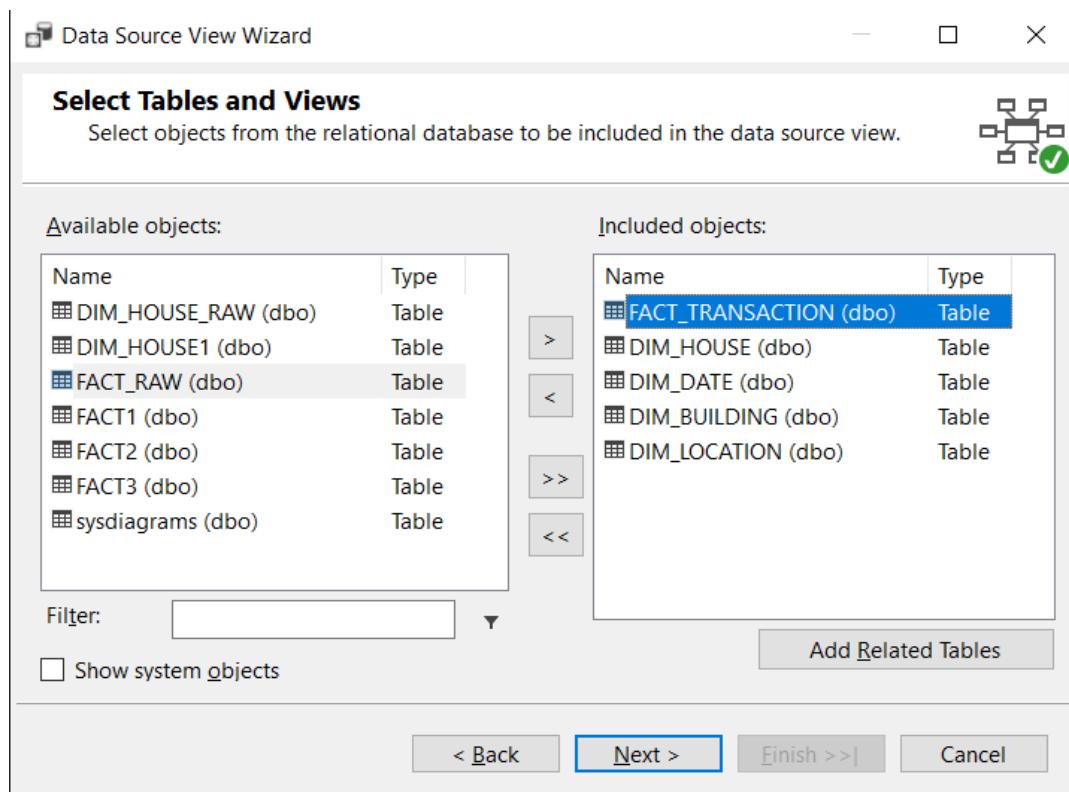
Hình 165: Chọn data source vừa tạo

Bước 4: Chọn bảng **Fact**, sau đó chọn nút > để thêm bảng Fact vào data source view.



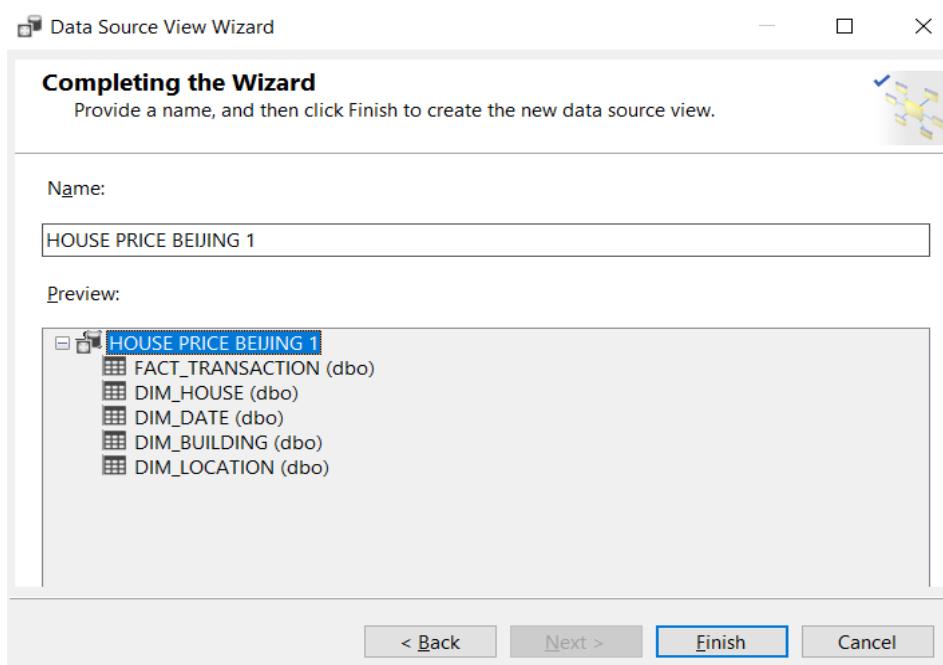
Hình 166: Thêm bảng Fact vào data source

Bước 5: Tiếp theo, chọn nút **Add Related Tables** để thêm tất cả các bảng Dim vào data source view. Sau đó chọn **Next >** để tiếp tục.



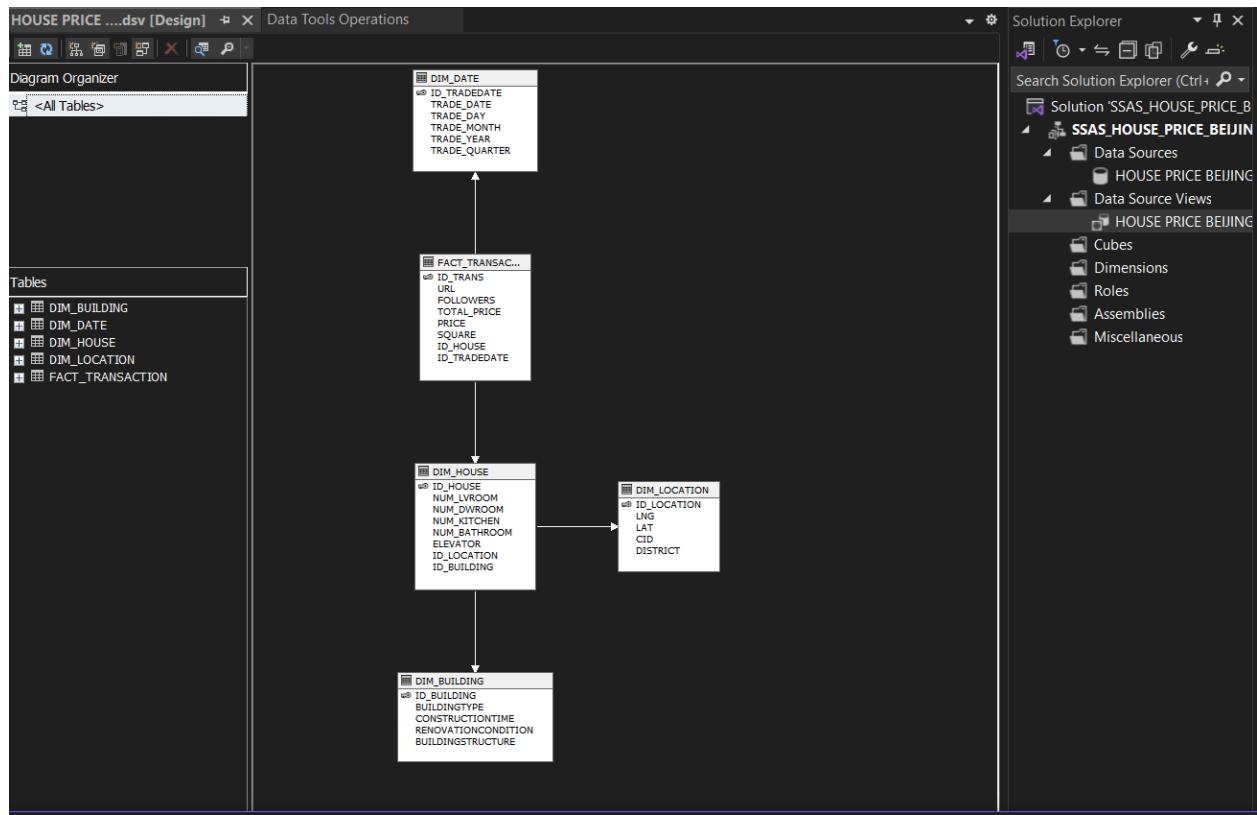
Hình 167: Thêm tất cả các bảng Dim vào data source view.

- Chọn **Finish** để hoàn tất



Hình 168: Chọn Finish

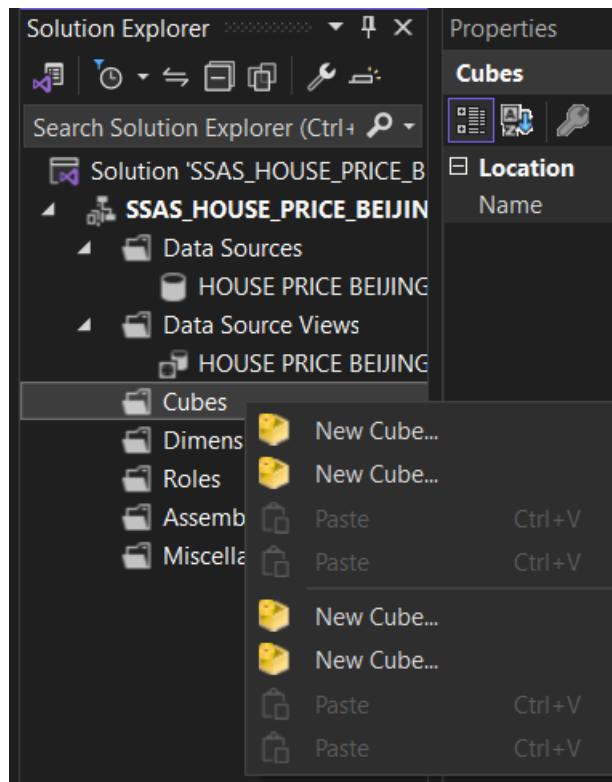
- Phía dưới là tất cả các bảng đã được chọn và thể hiện sự liên kết giữa chúng:



Hình 169: Tất cả các bảng đã được chọn và thể hiện sự liên kết giữa chúng

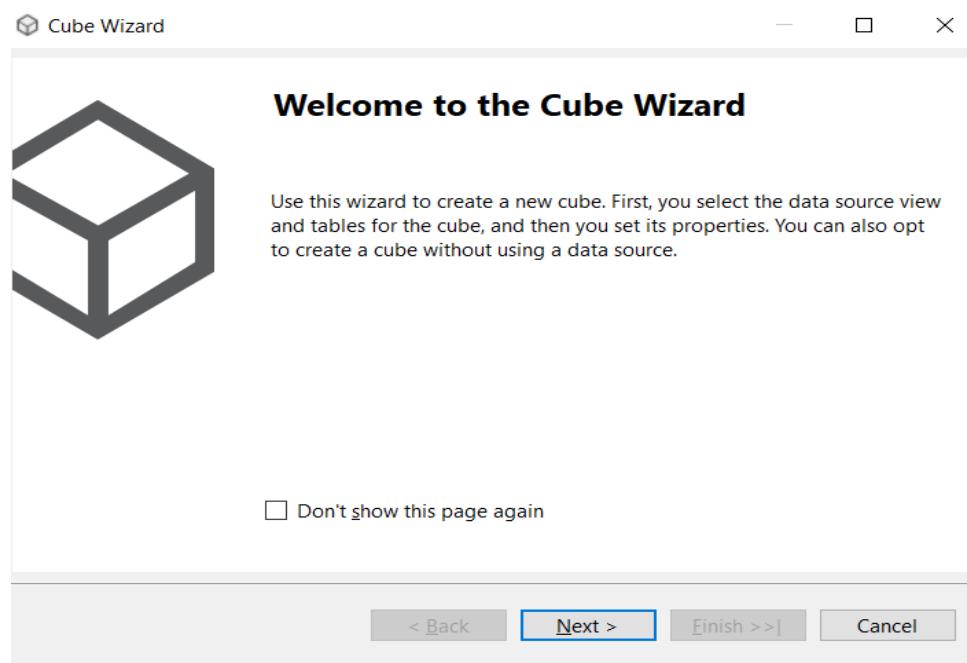
4. Xây dựng các khối (Cubes) và xác định các độ đo (Measures)

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



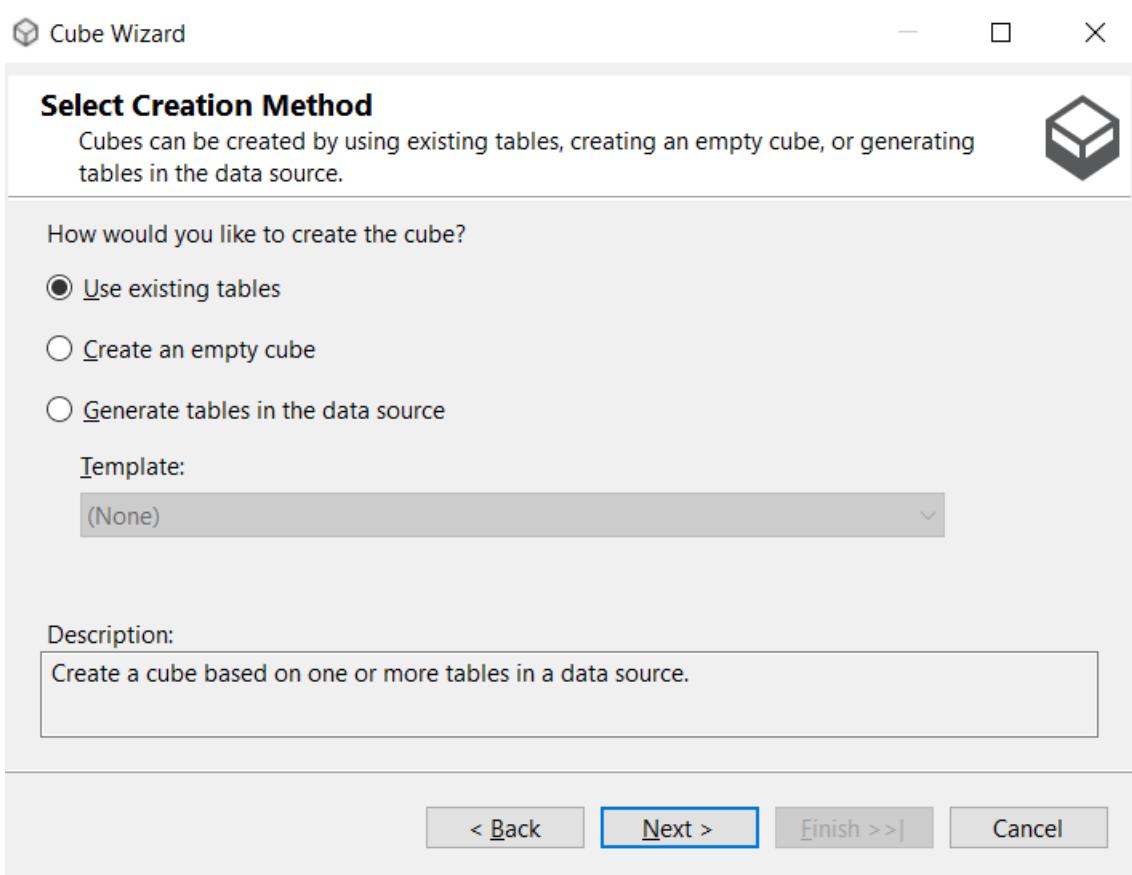
Hình 170: Chọn New Cube

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



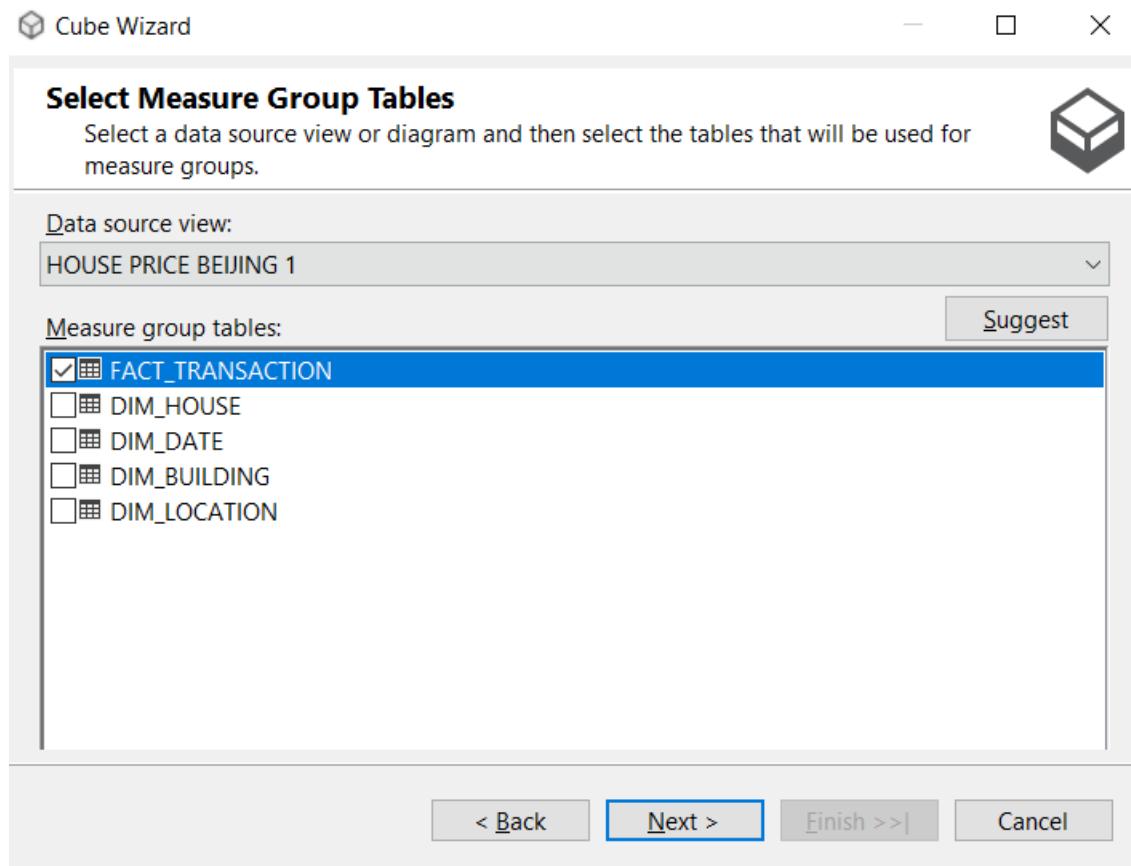
Hình 171: Chọn Next để tiếp tục

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



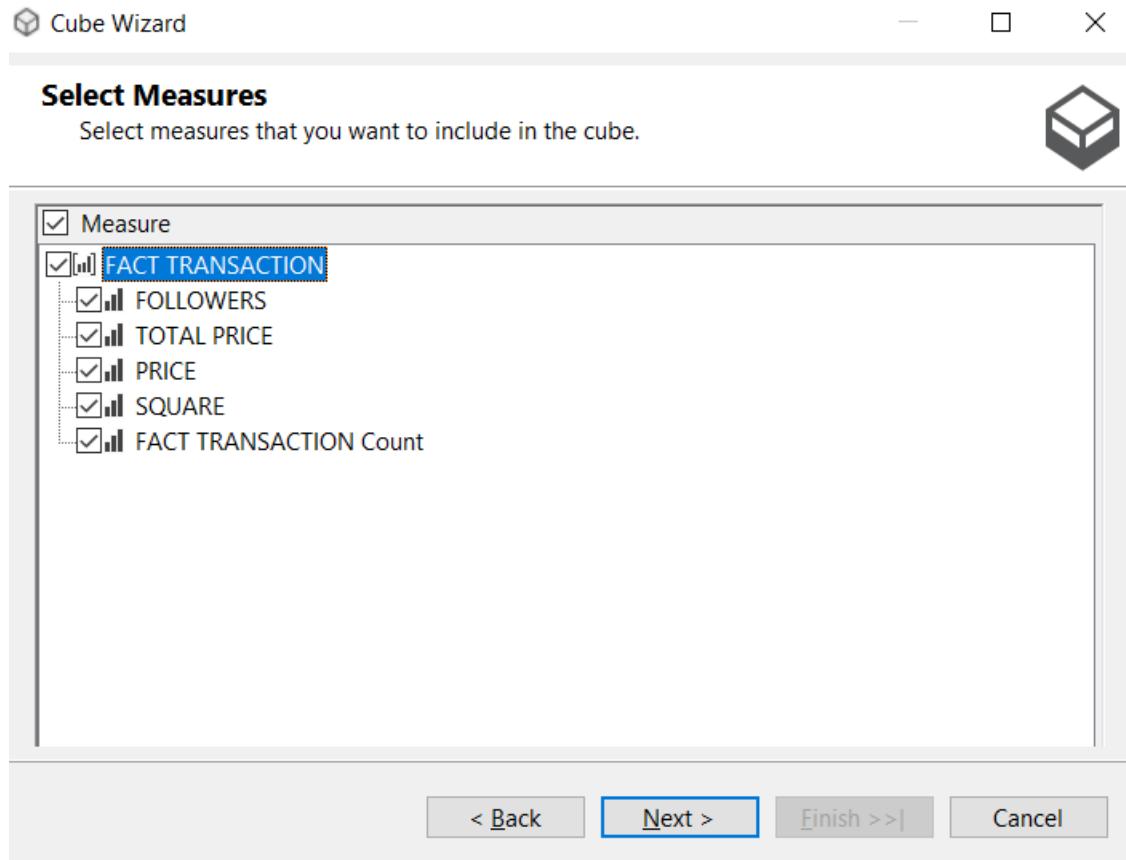
Hình 172: Chọn use existing tables

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



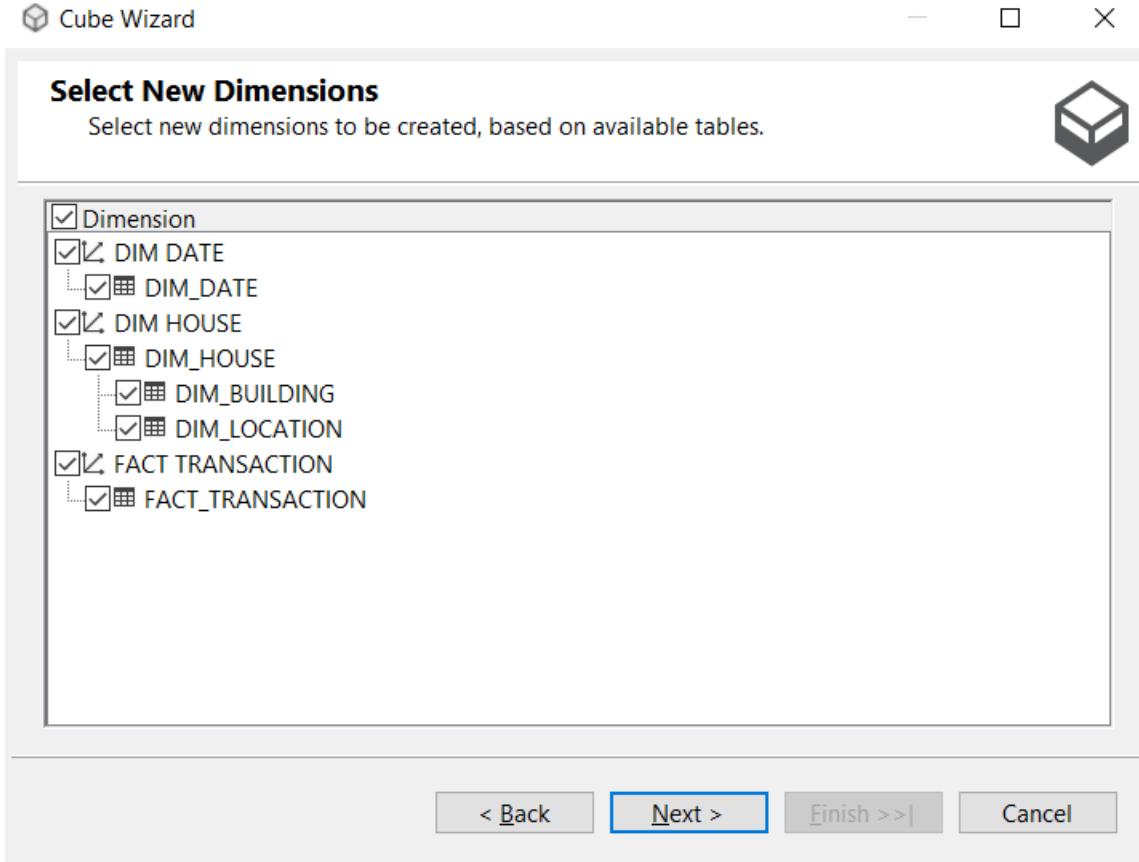
Hình 173: Chọn Fact để phân chia các measure group

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



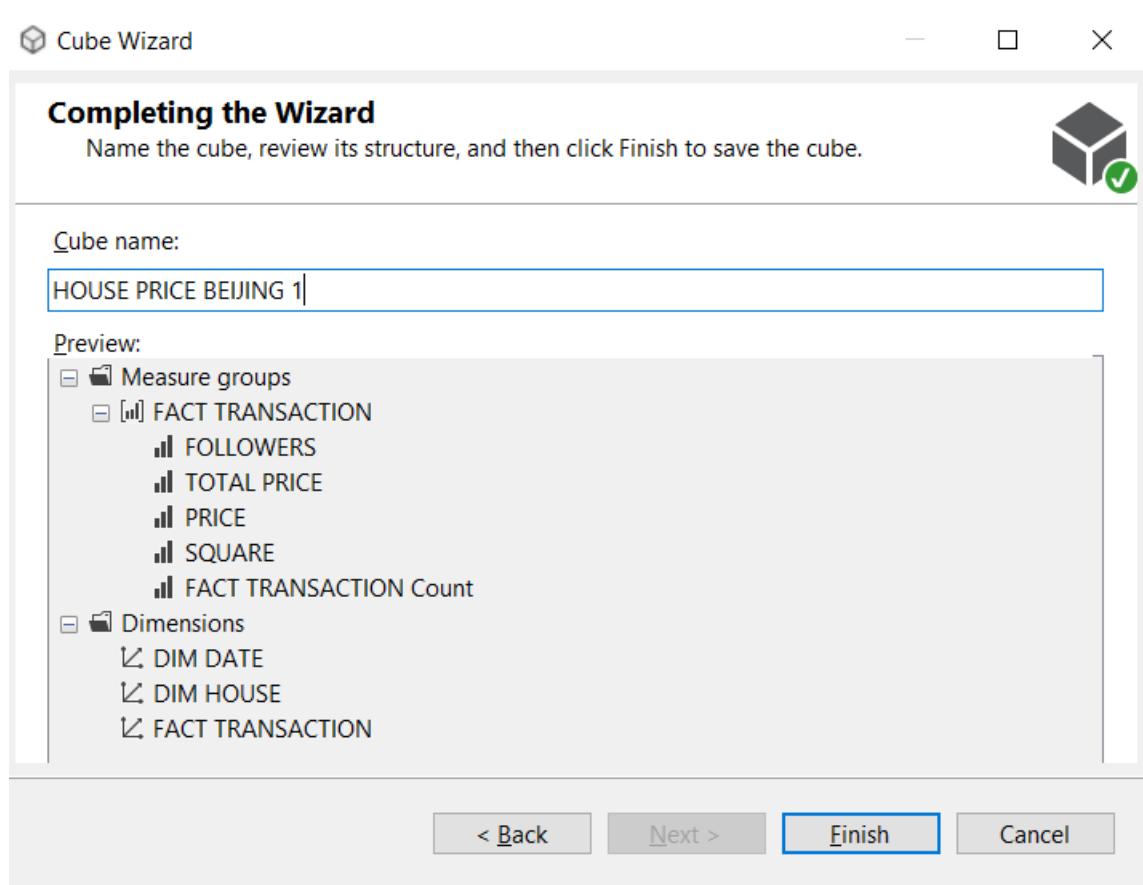
Hình 174: Chọn những độ đo để xuất

Bước 6: Chọn danh sách các bảng Dimension, sau đó chọn **Next** để tiếp tục.



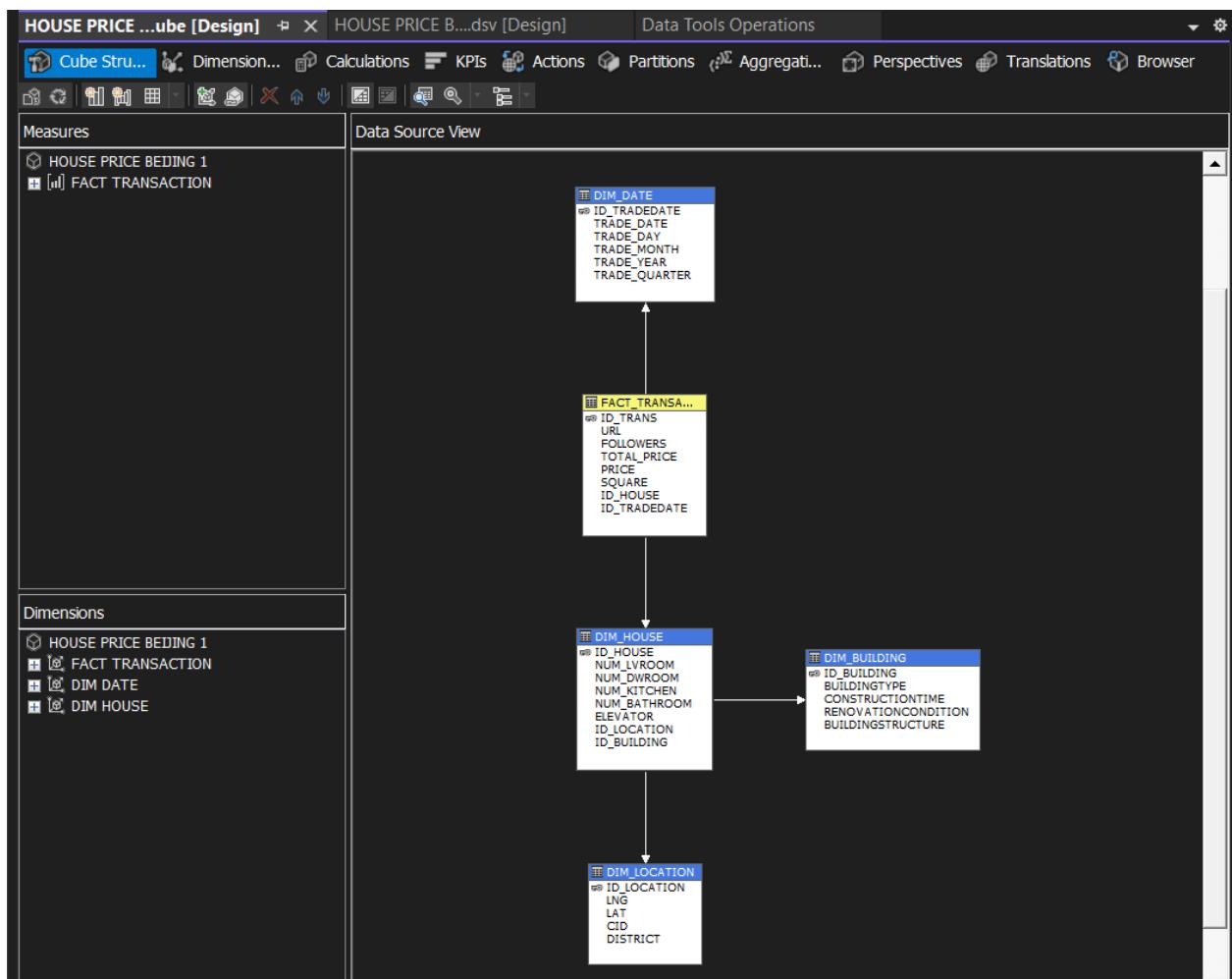
Hình 175: Chọn danh sách các bảng Dimension

Bước 7: Chọn **Finish** để hoàn tất quy trình xây dựng các khối (Cubes) và xác định các độ đo (Measures).



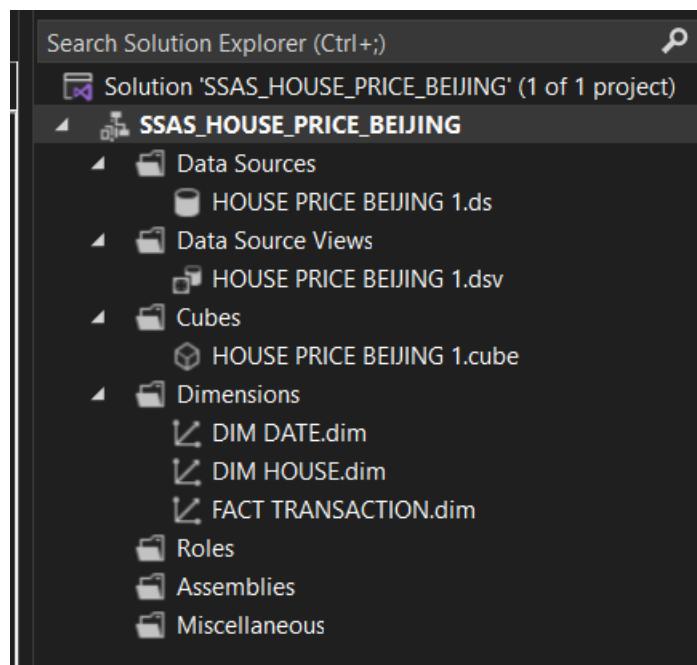
Hình 176: Chọn Finish để hoàn tất quy trình

- Sau khi kết thúc quá trình này, ta sẽ được kết quả như hình sau:



Hình 177: Kết quả như hình

- Ở **Solution Explorer**, lần lượt double-click vào từng bảng dimension để tiến hành thêm các thuộc tính.



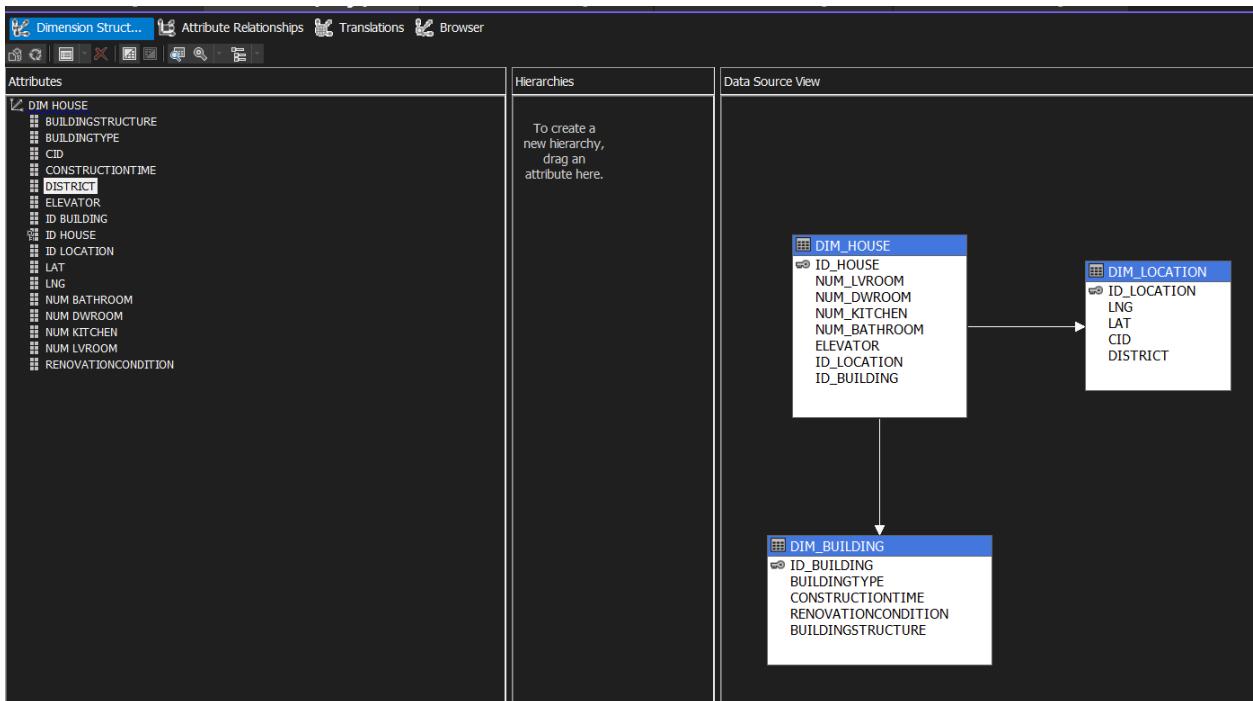
Hình 178: Lần lượt double-click vào từng bảng dimension

Bước 8: Tại folder **Dimensions** trong **Solution Explorer**, ta chọn **DIM DATE.dim**. Sau đó kéo thả các thuộc tính **TRADE_DATE**, **TRADE_DAY**, **TRADE_MONTH**, **TRADE_YEAR**, **TRADE_QUARTER** từ **Data Source View** vào **Attributes**.

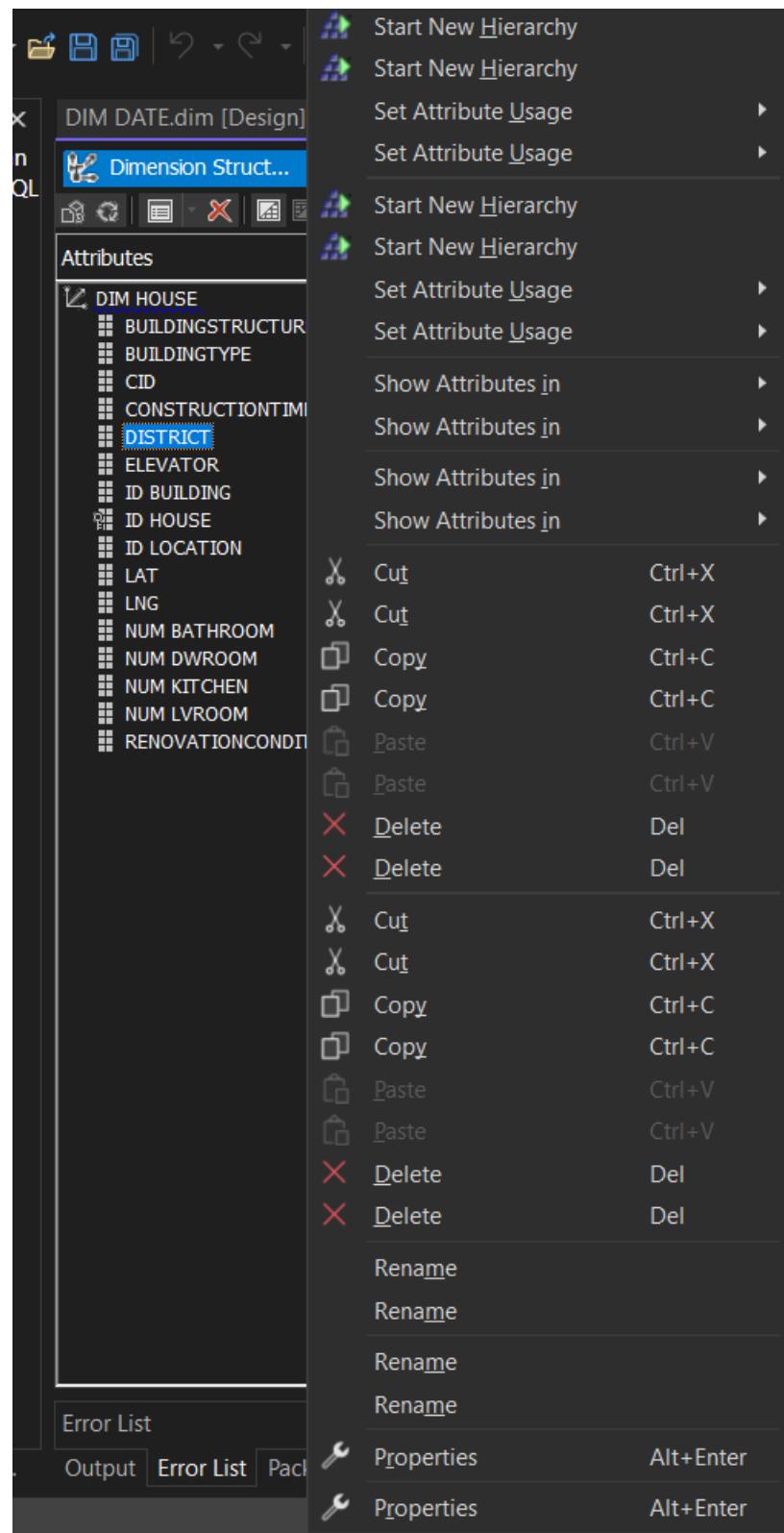
Attributes	Hierarchies	Data Source View							
<ul style="list-style-type: none"> ☒ DIM_DATE <ul style="list-style-type: none"> ☒ ID_TRADEDATE ☒ TRADE_DATE ☒ TRADE_DAY ☒ TRADE_MONTH ☒ TRADE_QUARTER ☒ TRADE_YEAR 	To create a new hierarchy, drag an attribute here.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">☒ DIM_DATE</td></tr> <tr> <td style="padding: 5px;">☒ ID_TRADEDATE</td></tr> <tr> <td style="padding: 5px;">☒ TRADE_DATE</td></tr> <tr> <td style="padding: 5px;">☒ TRADE_DAY</td></tr> <tr> <td style="padding: 5px;">☒ TRADE_MONTH</td></tr> <tr> <td style="padding: 5px;">☒ TRADE_YEAR</td></tr> <tr> <td style="padding: 5px;">☒ TRADE_QUARTER</td></tr> </table>	☒ DIM_DATE	☒ ID_TRADEDATE	☒ TRADE_DATE	☒ TRADE_DAY	☒ TRADE_MONTH	☒ TRADE_YEAR	☒ TRADE_QUARTER
☒ DIM_DATE									
☒ ID_TRADEDATE									
☒ TRADE_DATE									
☒ TRADE_DAY									
☒ TRADE_MONTH									
☒ TRADE_YEAR									
☒ TRADE_QUARTER									

Hình 179: Sau đó kéo thả các thuộc tính từ Data Source View vào Attributes

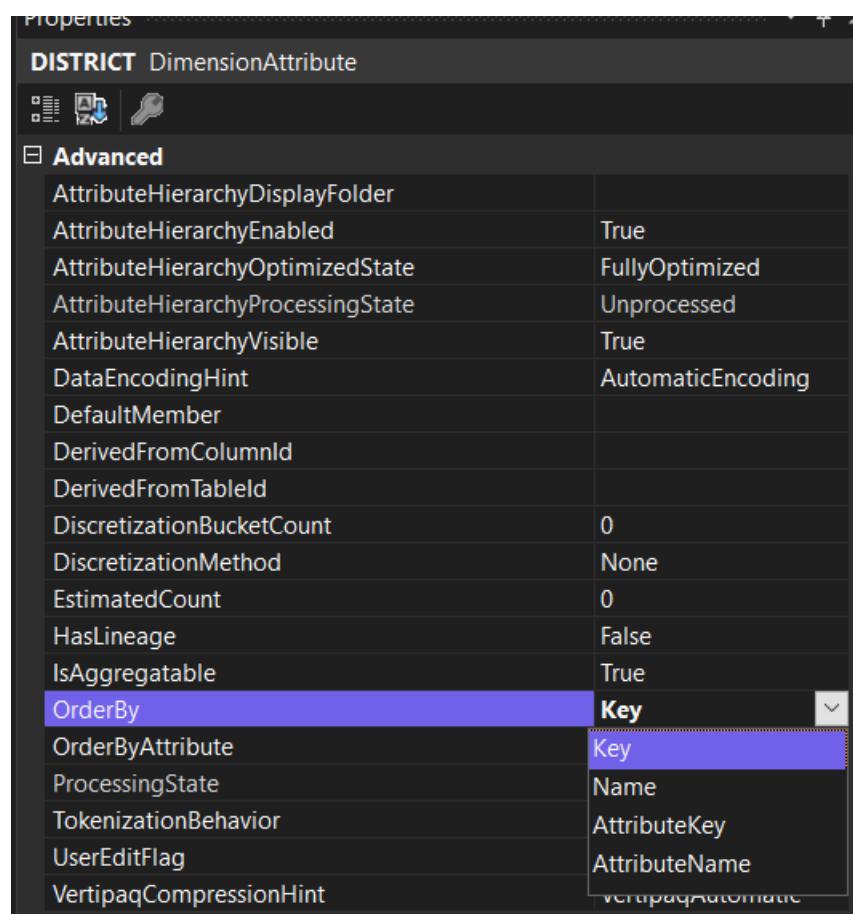
Bước 9: Tại folder Dimensions trong Solution Explorer, ta chọn **DIM HOUSE.dim**. Sau đó kéo thả các thuộc tính **NUM_BATHROOM**, **NUM_DWROOM**, **NUM_KITCHEN**, **MUM_LVROOM** từ Data Source View vào Attributes.



Hình 180: Kéo thả các thuộc tính từ Data Source View vào Attributes

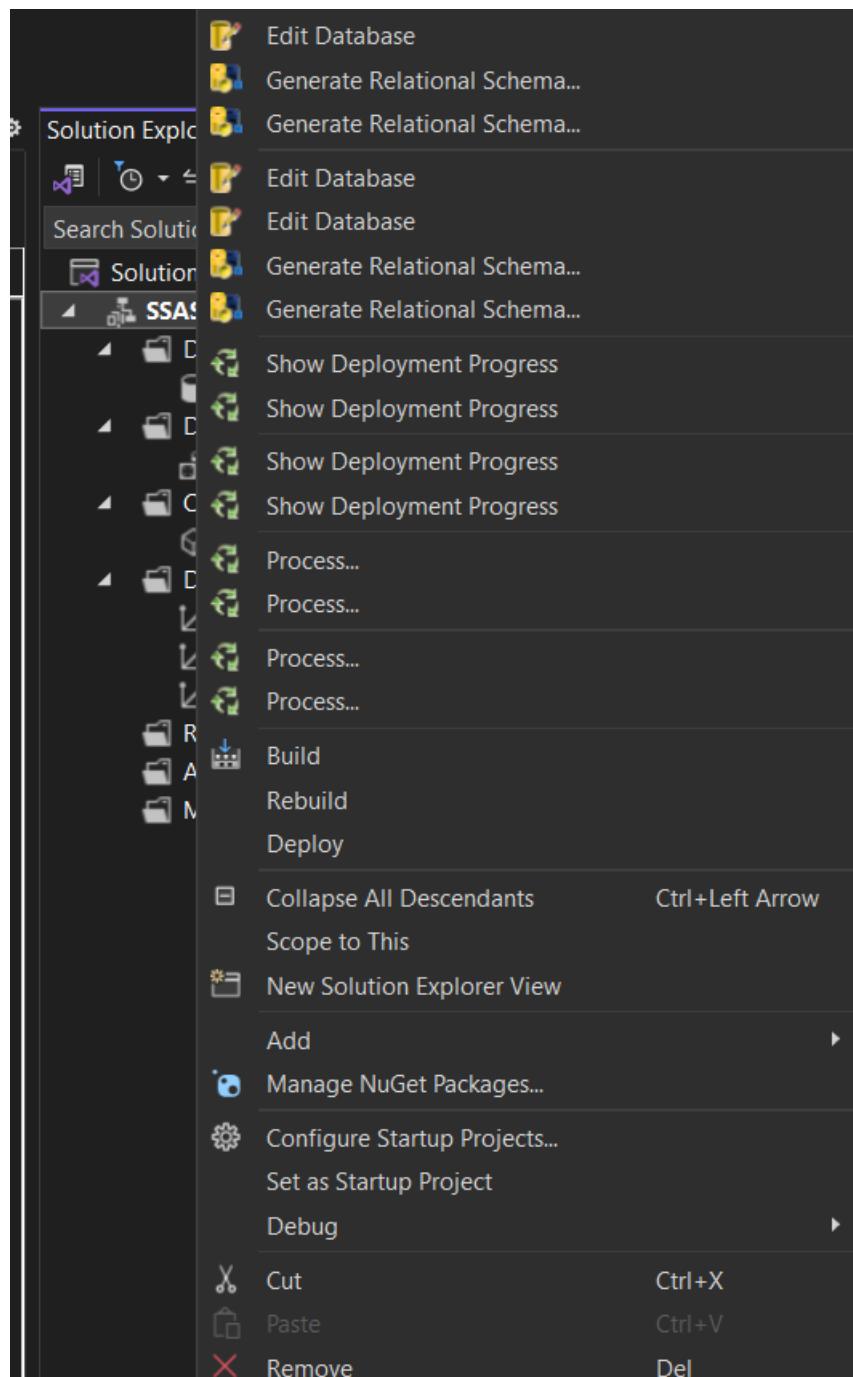


Hình 181: Chọn Properties



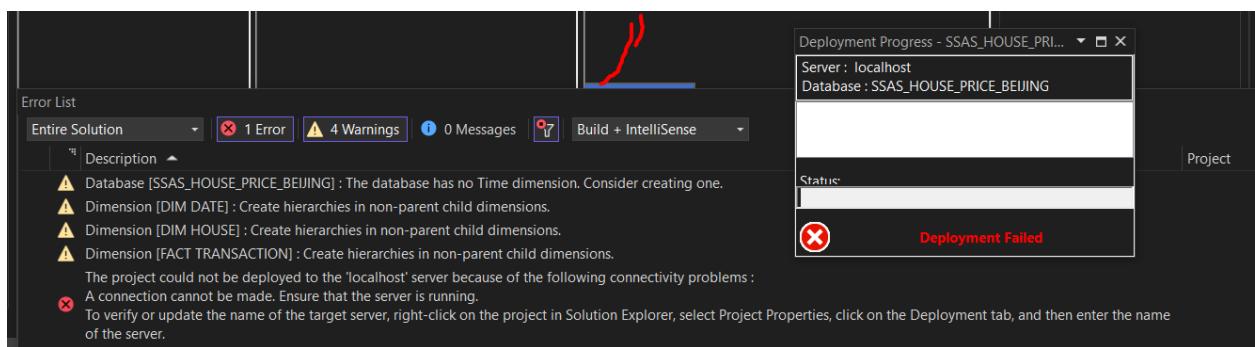
Hình 182: OrderBy "Key"

Bước 10: Ta chọn **Start** để deploy project. Khi deploy thành công ta sẽ nhận được kết quả như sau:



Hình 183: Kết quả khi deploy thành công

- Xuất hiện lỗi:

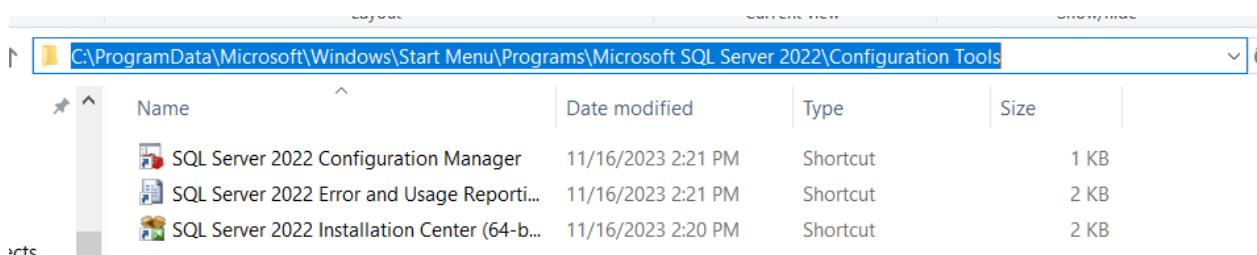


Hình 184: Xuất hiện lỗi

***Nguyên nhân:** Project không deploy được do **chưa kết nối được với server database**:

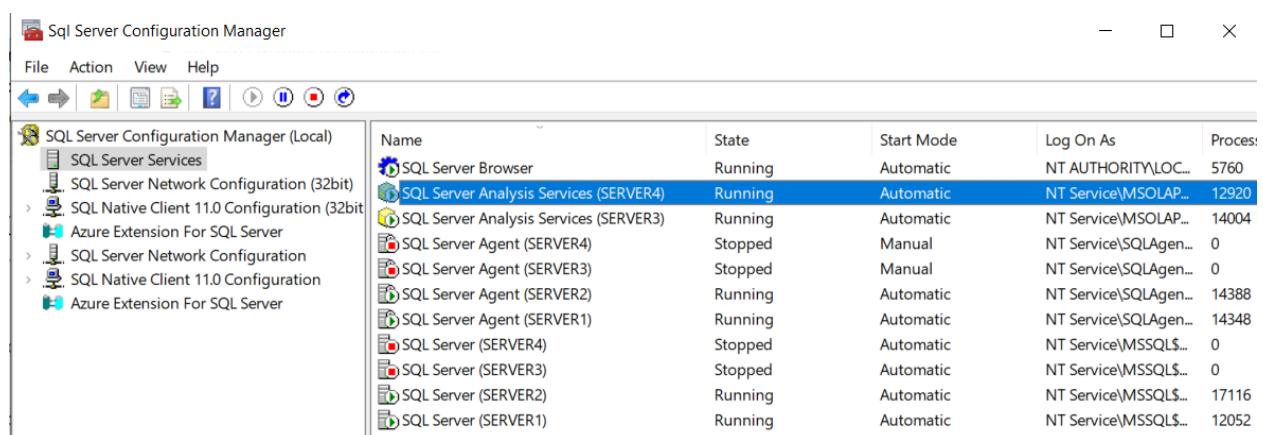
Truy cập vào đường dẫn: “**C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Microsoft SQL Server 2022\Configuration Tools**”

- Mở “**SQL Server 2022 Configuration Manager**”;



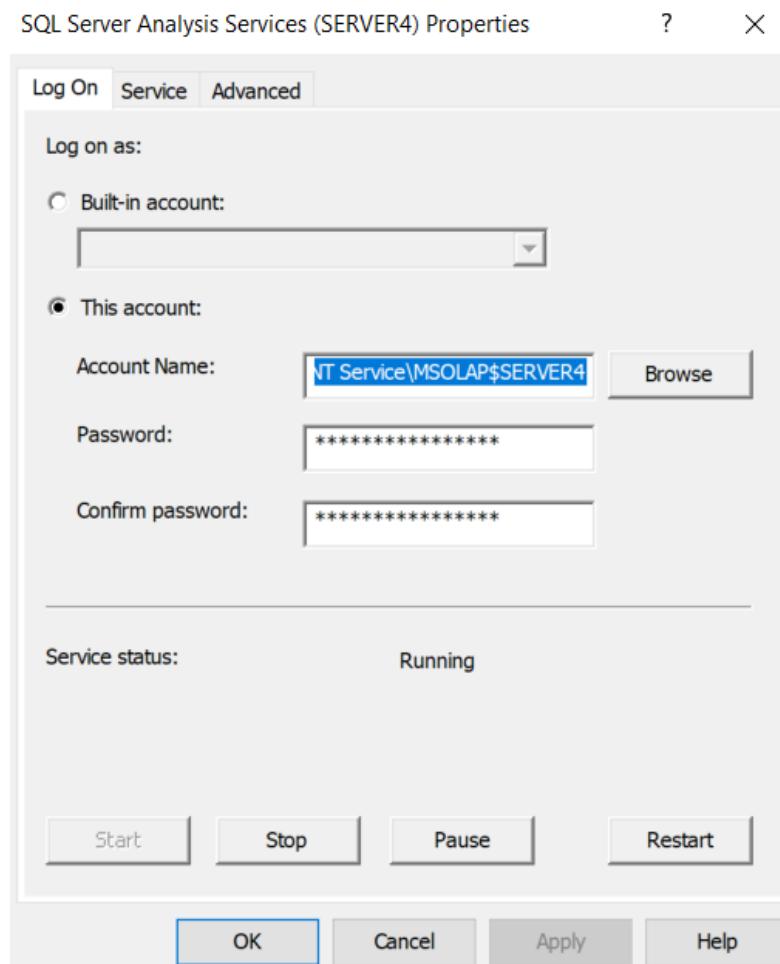
Hình 185: Đường dẫn “SQL Server 2022 Configuration Manager”

- Double-click vào “**SQL Server Analysis Services**”



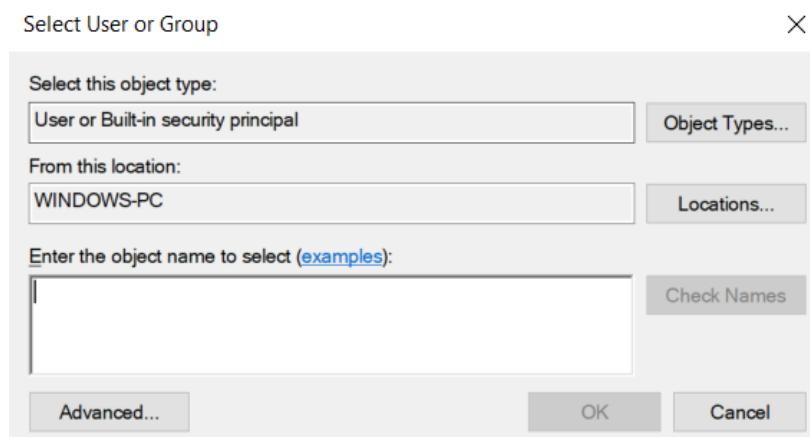
Hình 186: Giao diện SQL Server 2022 Configuration Manager

- R-click vào **SQL Server Analysis Services** chọn **Properties**



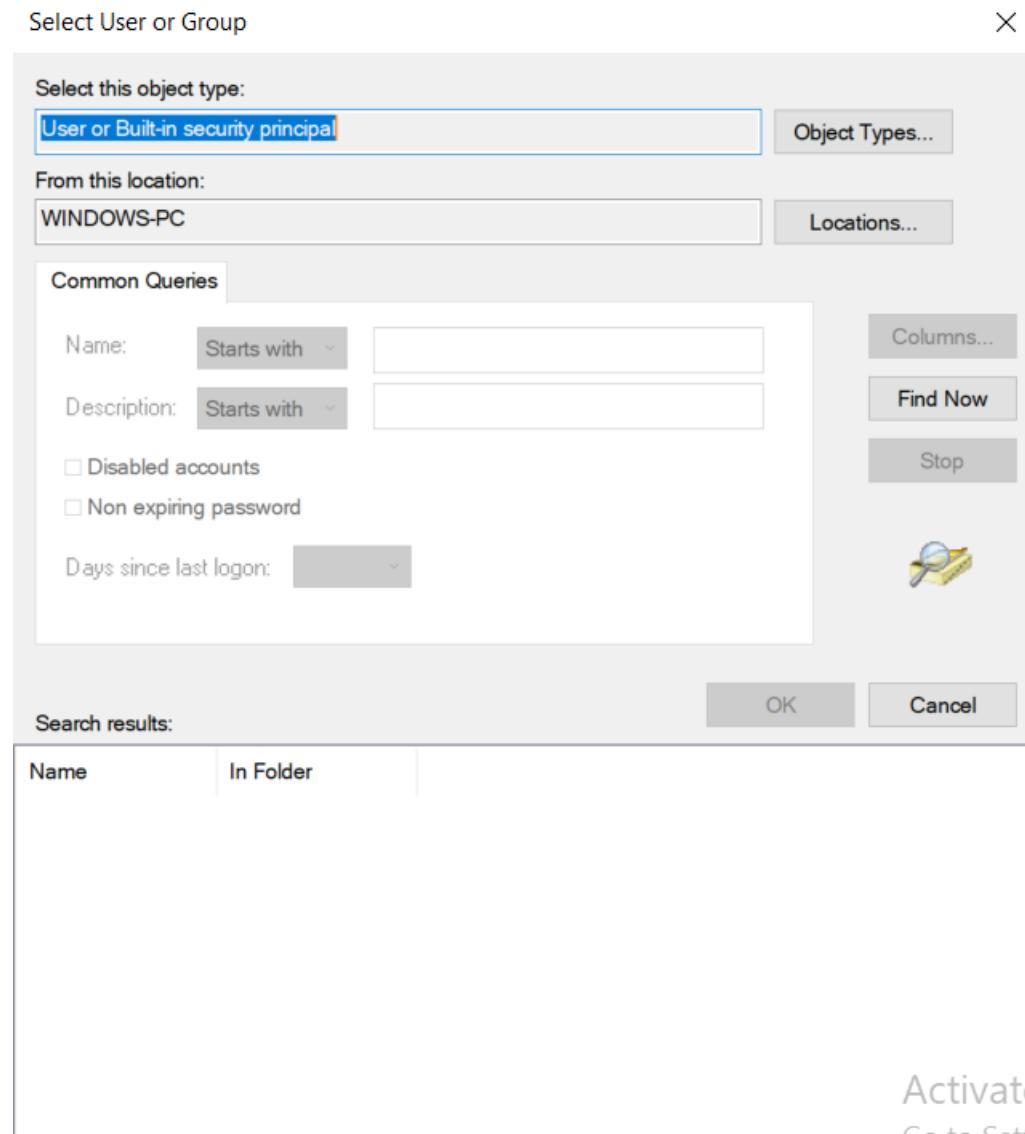
Hình 187: Vào trong Properties

- Tiếp tục chọn **Browse**

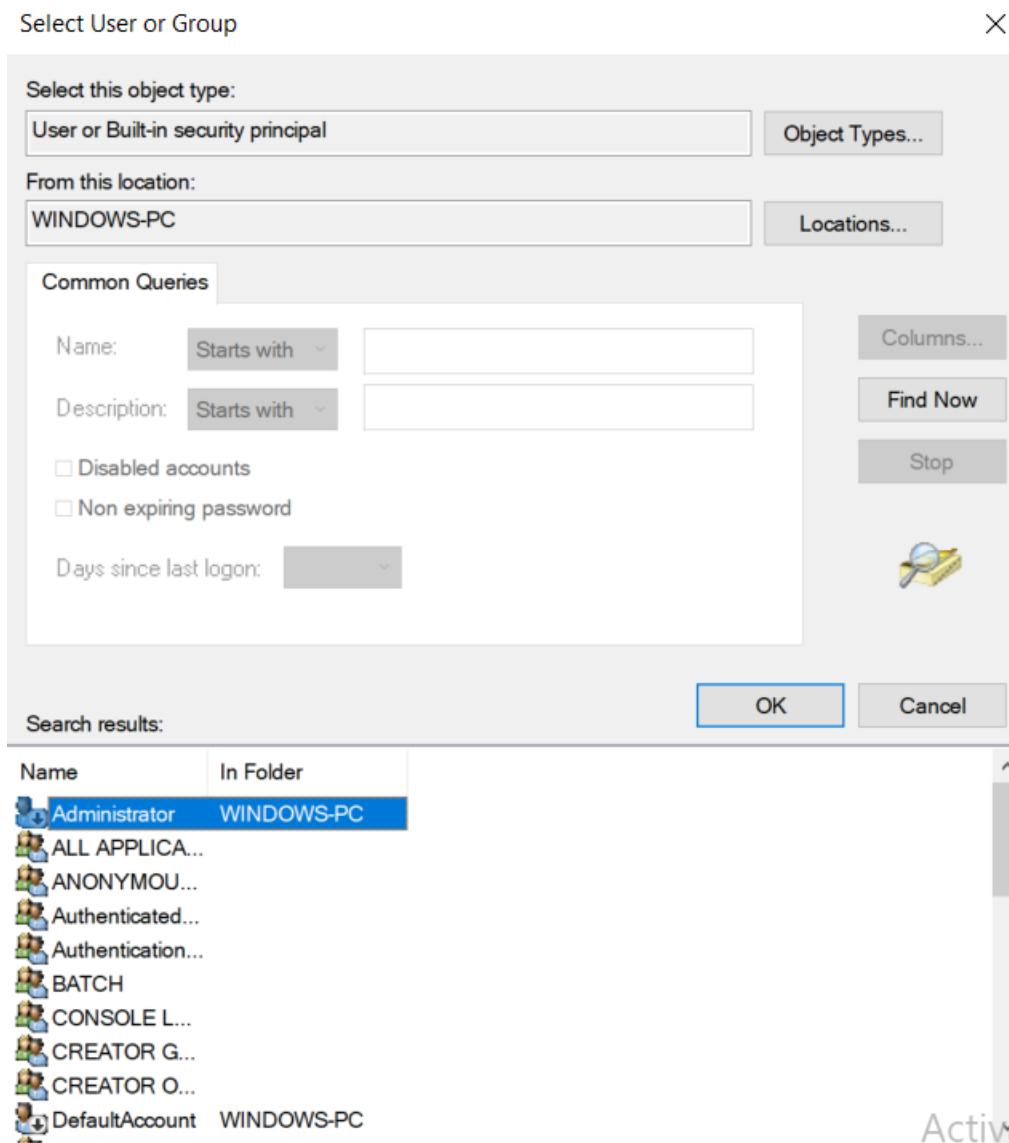


Hình 188: Chọn Browse

- Sau đó chọn **Advanced** -> **Find Now**



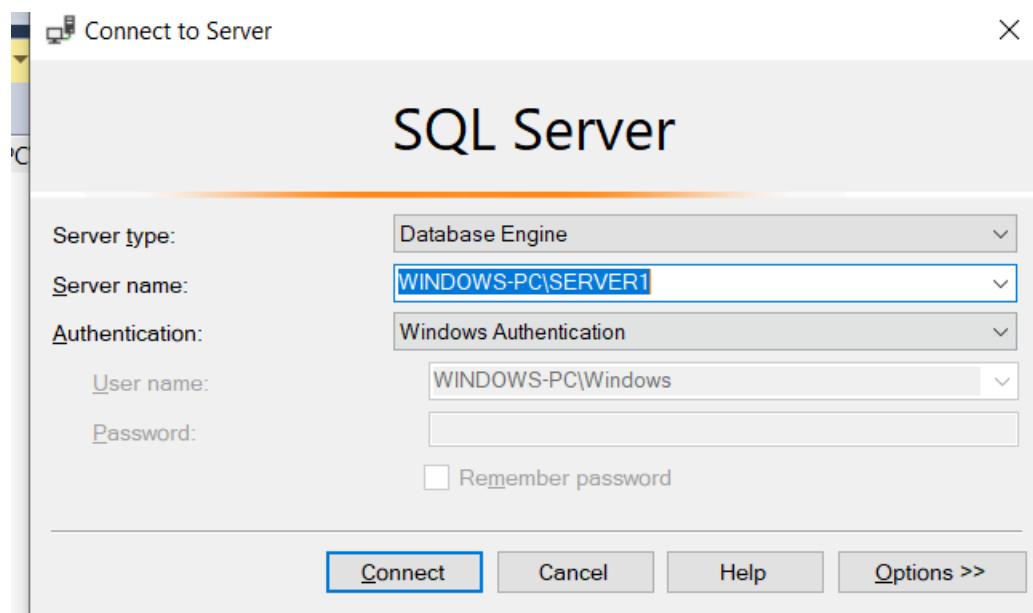
Hình 189: Chọn Advanced -> Find Now



Hình 190: Chọn một tài khoản có đủ quyền

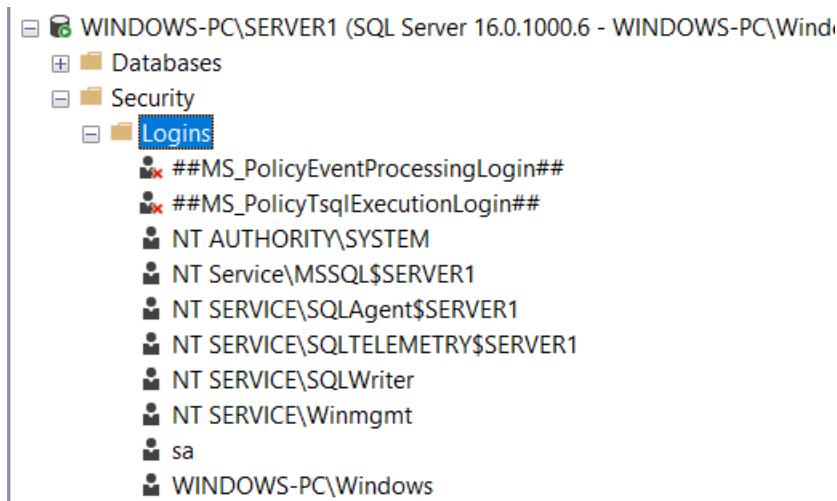
- Sau đó chọn một tài khoản có đủ quyền. Rồi trở về màn hình trước.

Nhập mật khẩu của tài khoản hiện tại rồi **OK**. Services sẽ được restart và thử **Deploy** lại. Lưu ý, nên sử dụng user đăng nhập SSMS hiện tại.



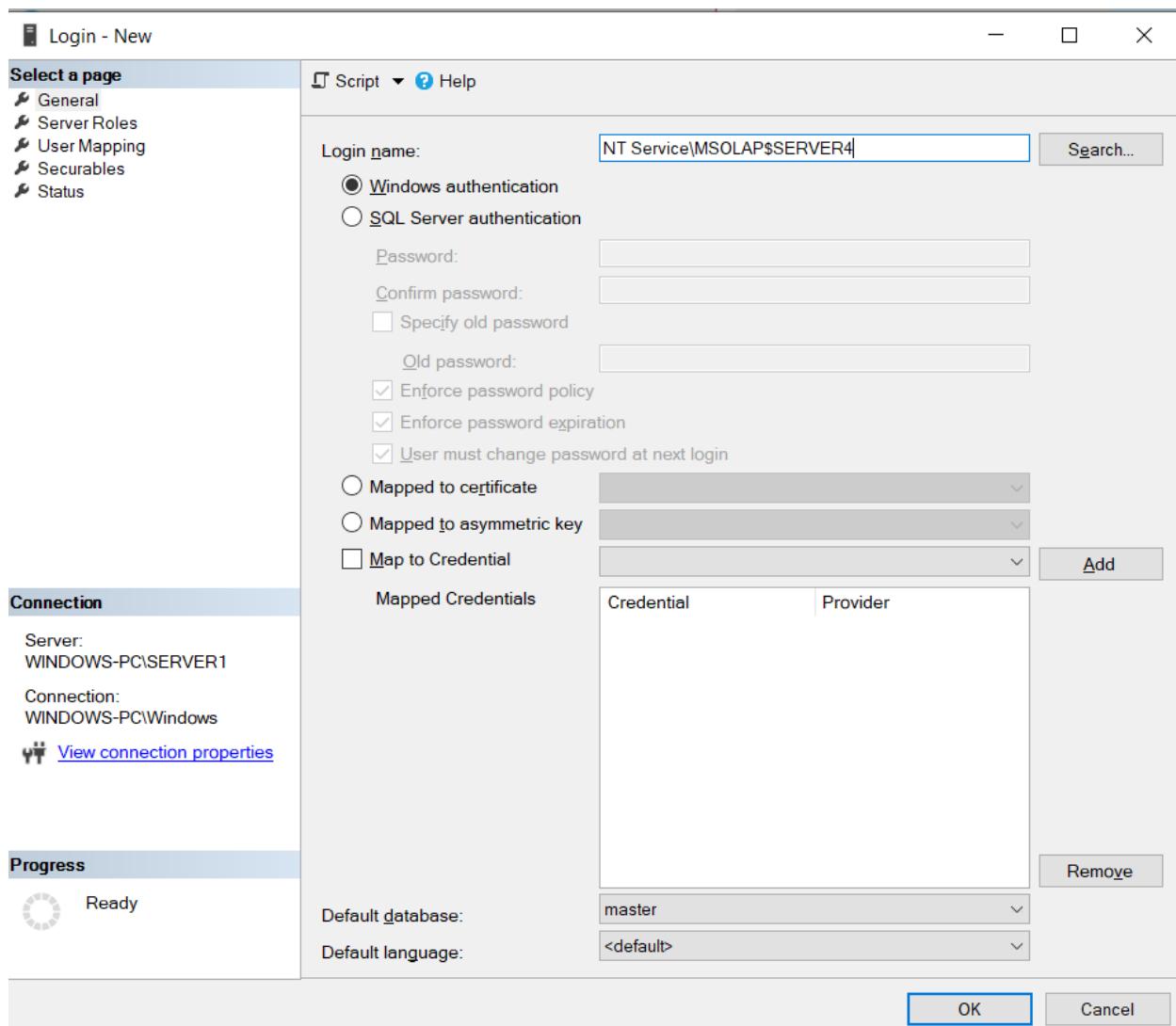
Hình 191: Connect

Vào mục **Logins** của server, chọn **server Analysis** mình sẽ deploy



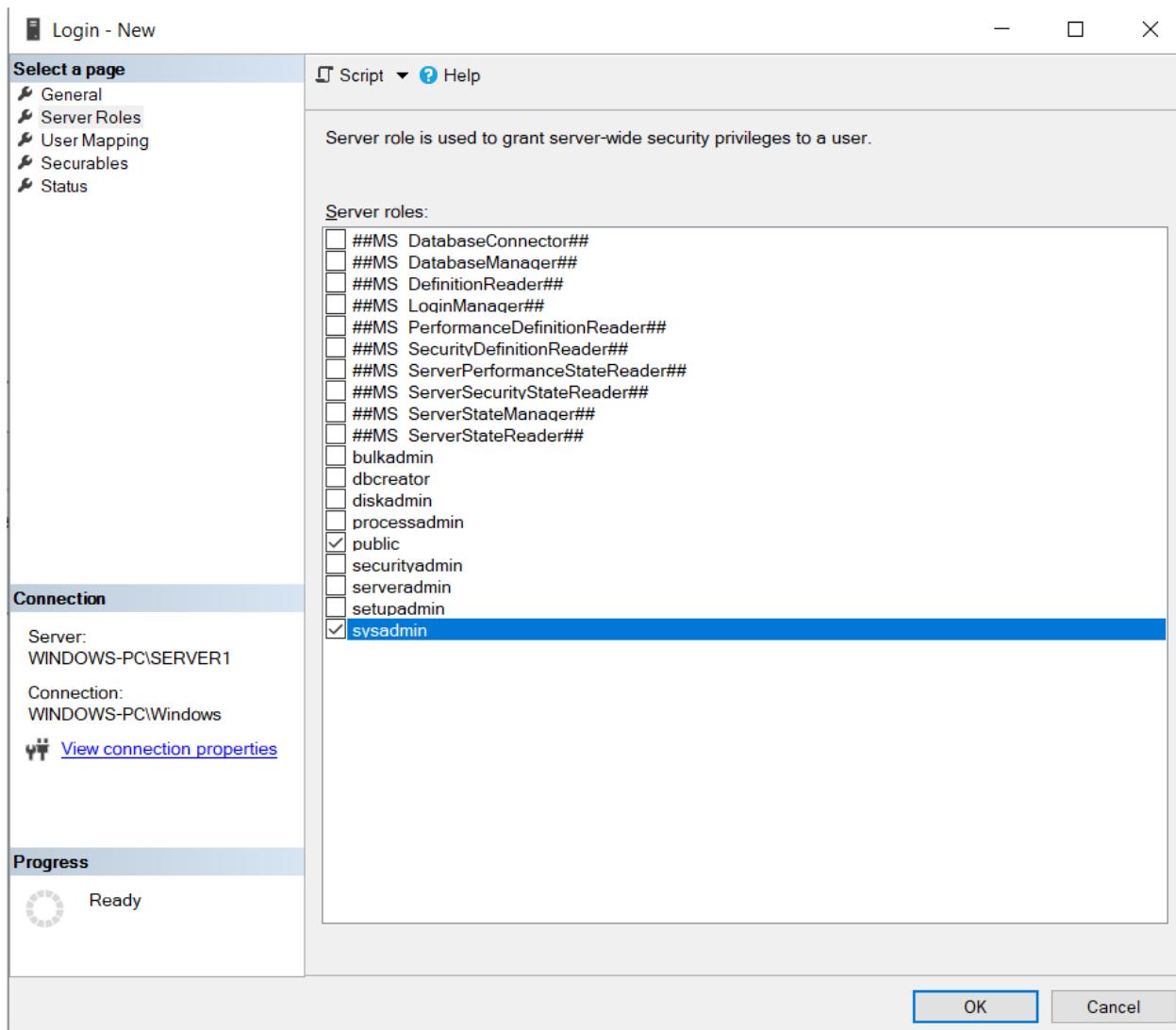
Hình 192: Mục Logins của server

Kiểm tra đã chọn **Windows authentication** hay chưa



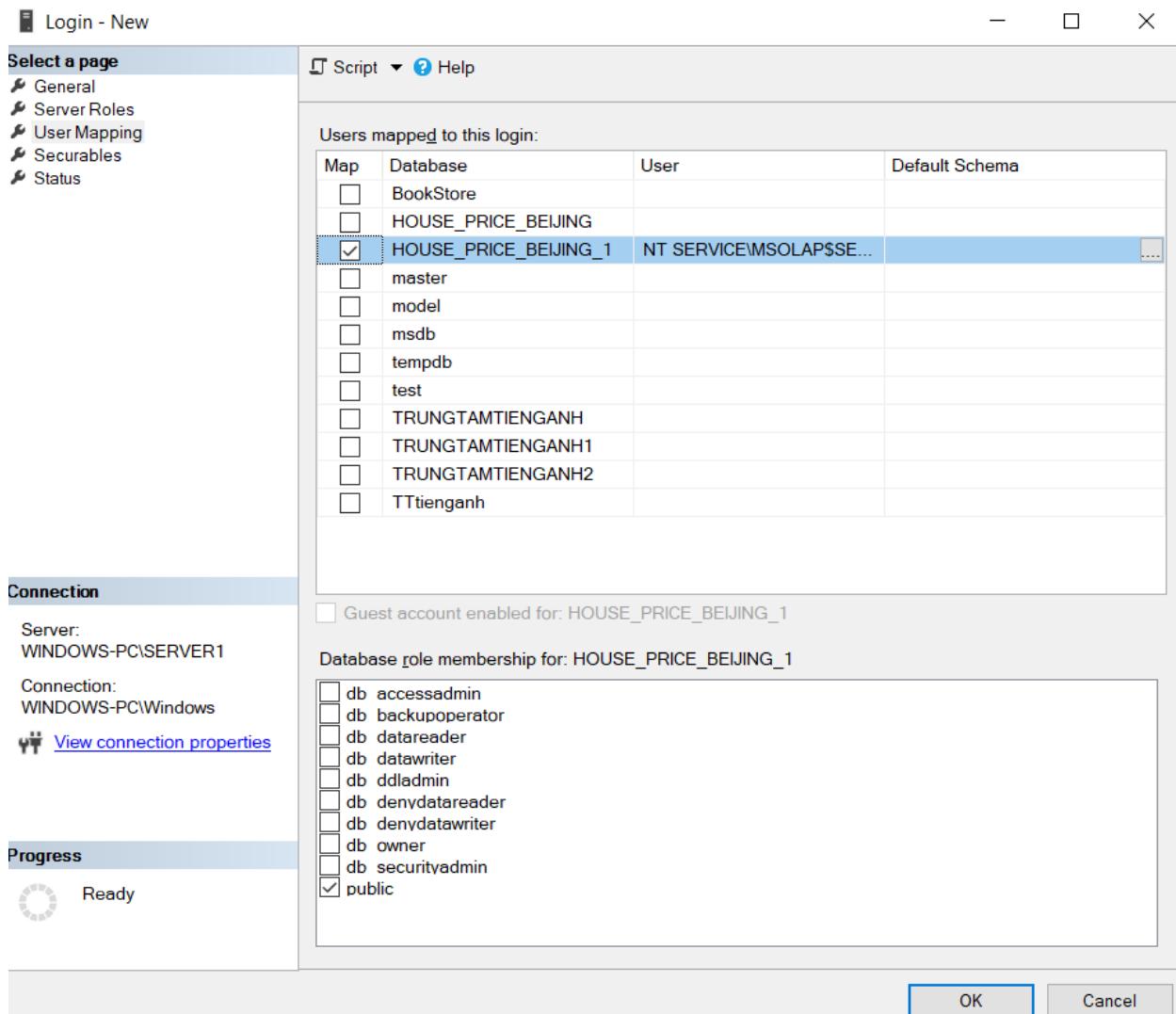
Hình 193: Kiểm tra Windows authentication

Qua mục **Server Roles** tích chọn **public** và **sysadmin**



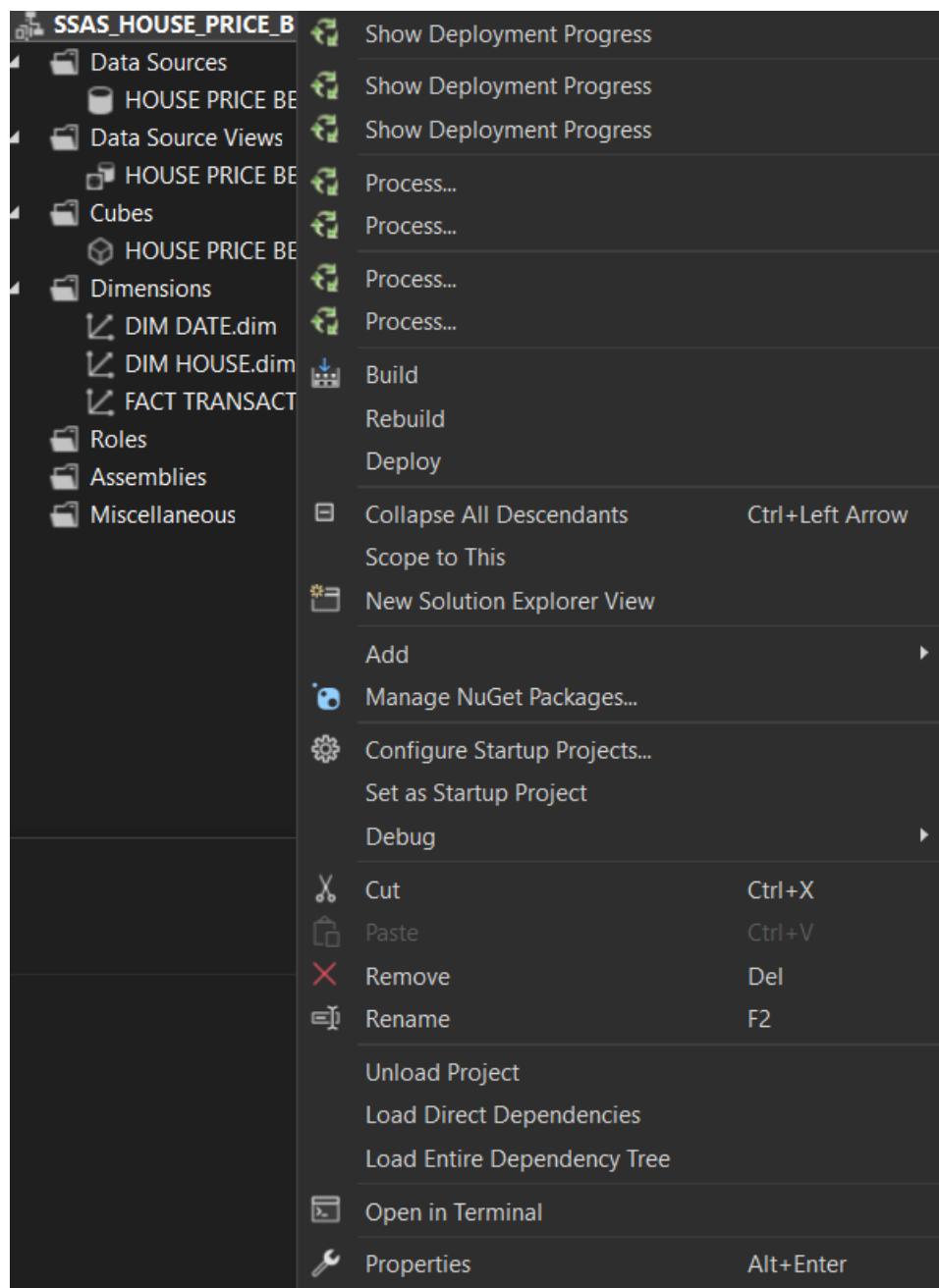
Hình 194: Tích chọn public và sysadmin

Tích chọn **User** như hình tích **public**



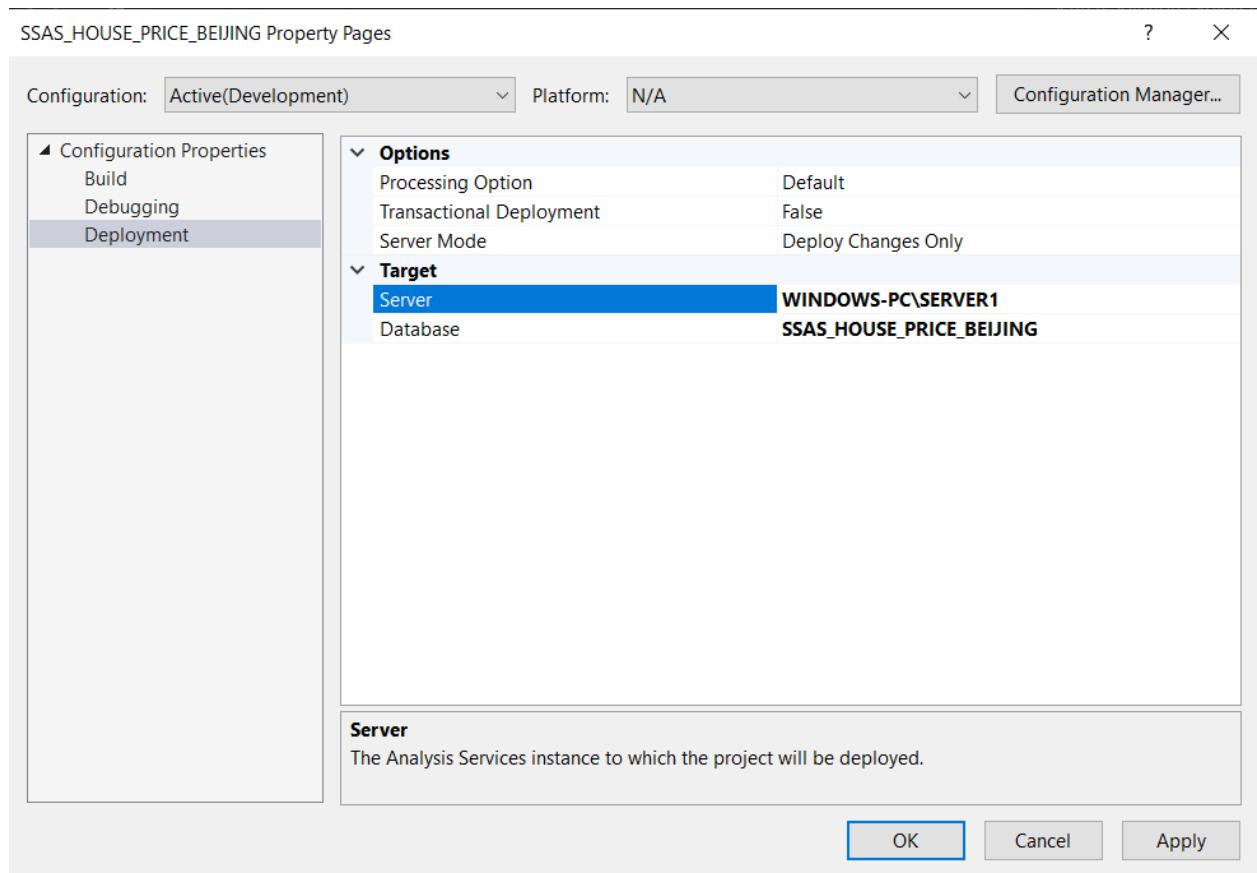
Hình 195: Tích chọn như hình

Trở lại project dự án, chuột phải chọn **Properties**



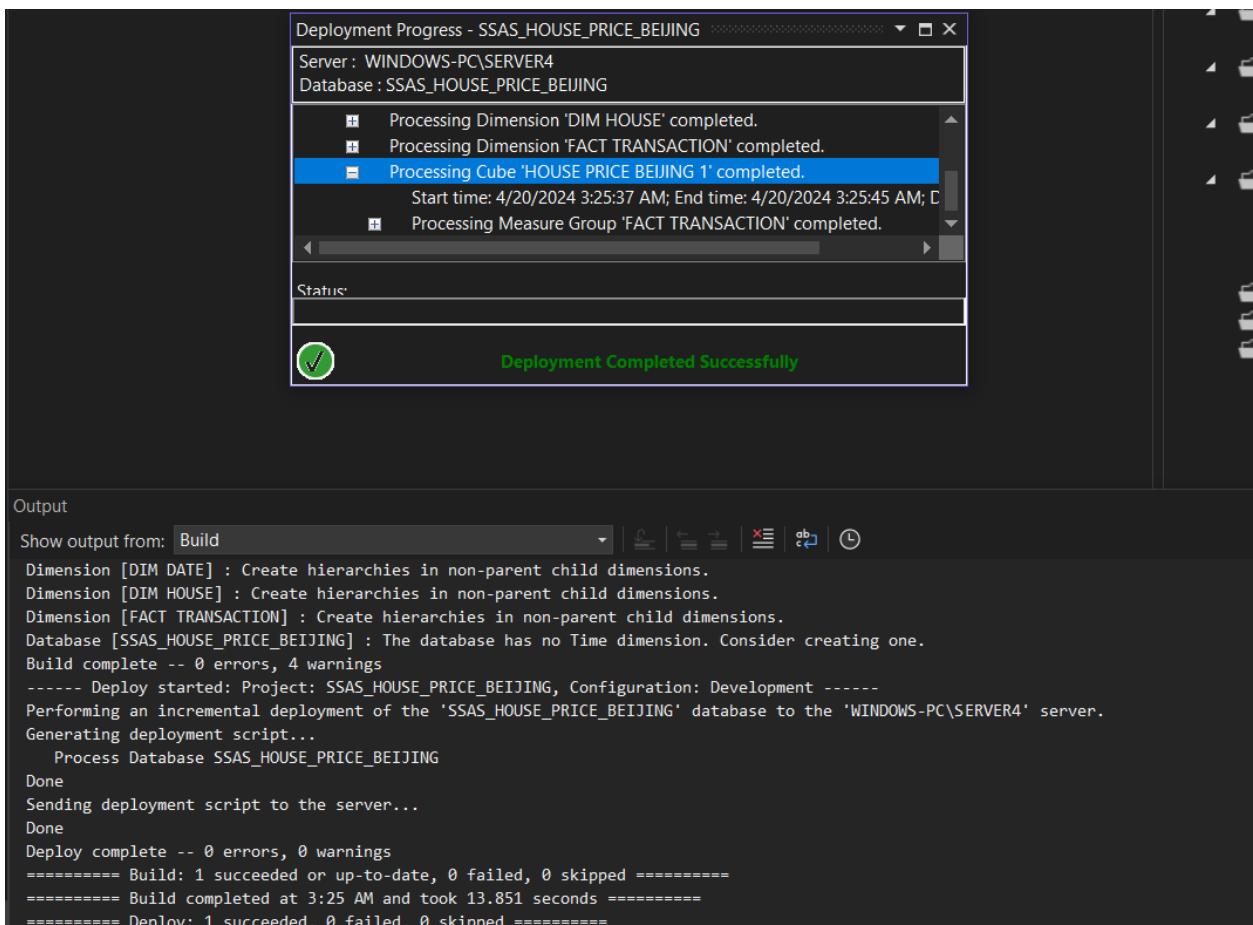
Hình 196: Chọn Properties

Vào mục Deployment điền Server sẽ Deploy và nhấn OK



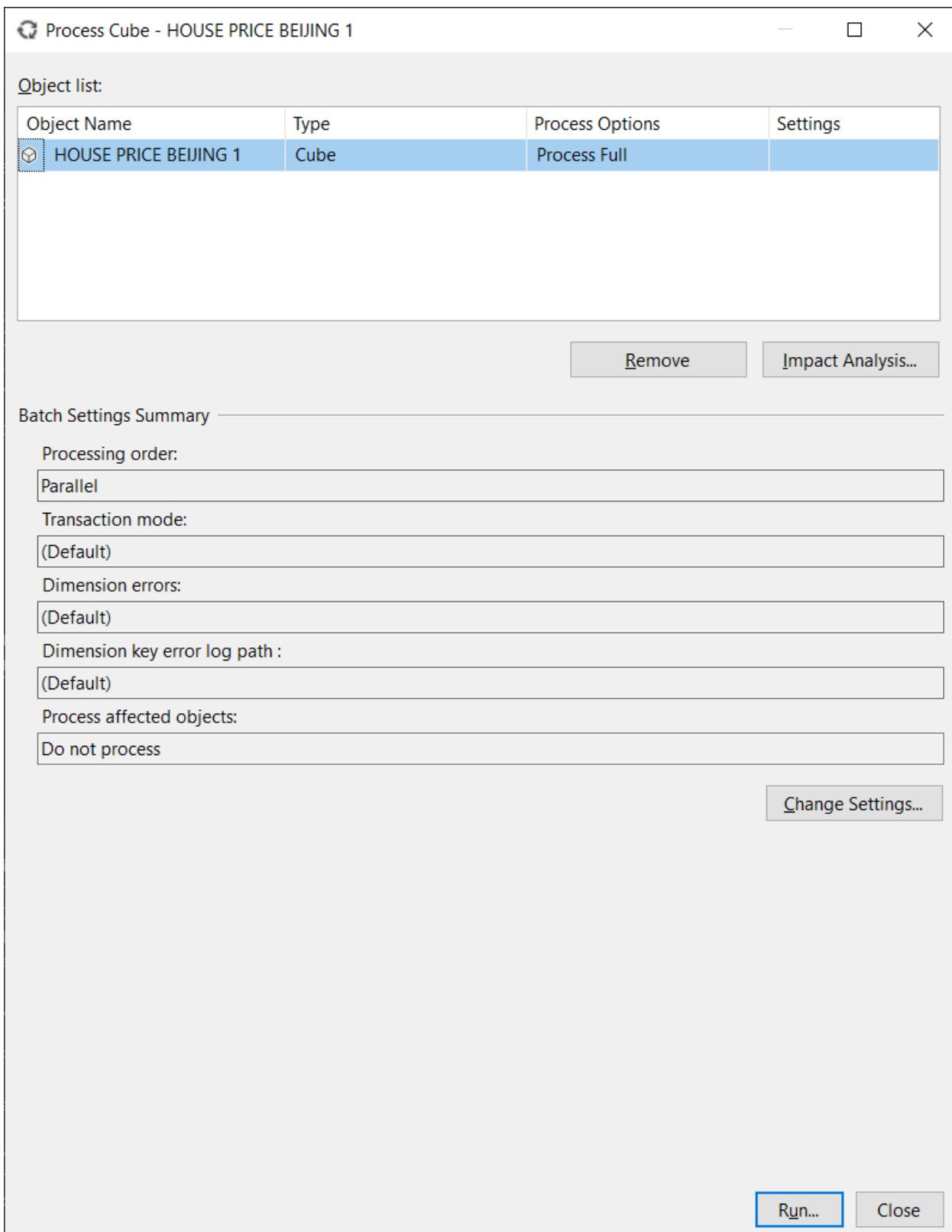
Hình 197: Điện Server sẽ Deploy

Deploy dự án:

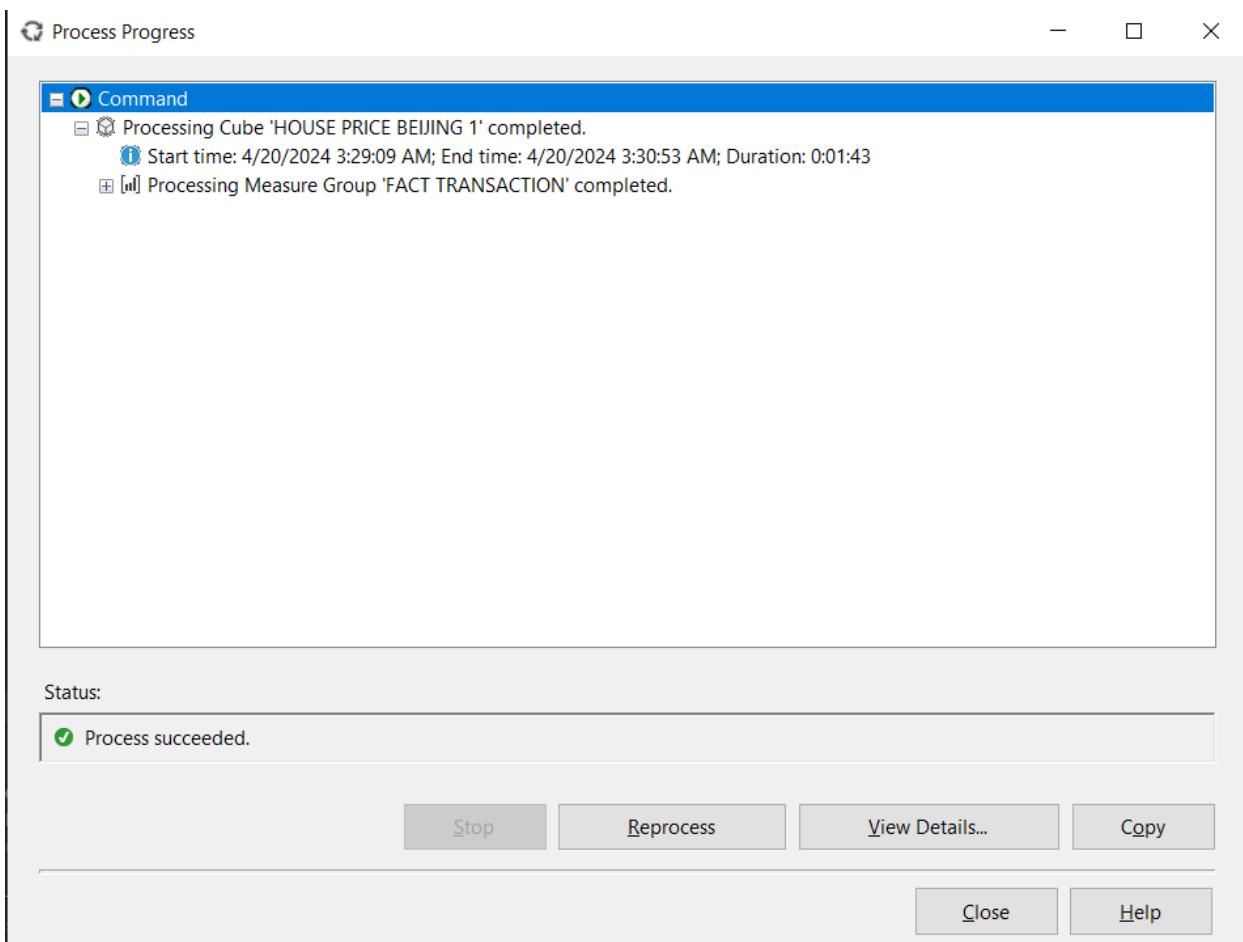


Hình 198: Deploy dự án

Nhấn Run:



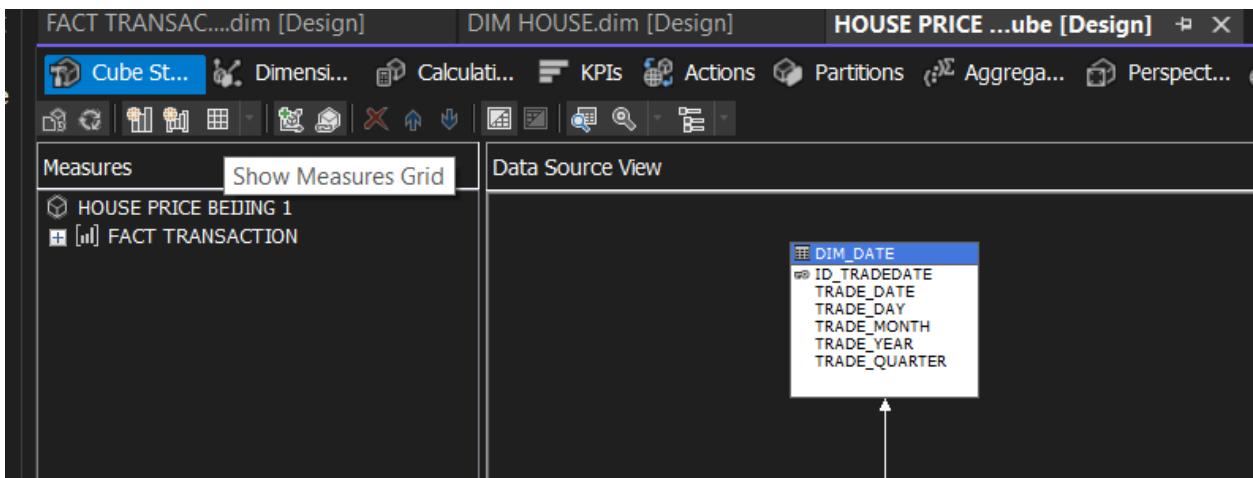
Hình 199: Chọn Run



Hình 200: Thành công

5. Xác định các độ đo (Measures)

Bước 1: Tại khôi vừa tạo, chọn **Show Measures Grid** để hiển thị chi tiết các độ đo.



Hình 201: Chọn Show Measures Grid

- Chi tiết các độ đo sẽ hiển thị dưới dạng bảng, dễ dàng để tương tác.

Measures				
	Name	Measure Group	Data Type	Aggregation
	FOLLOWERS	FACT TRANSACTION	Integer	Sum
	TOTAL PRICE	FACT TRANSACTION	Double	Sum
	PRICE	FACT TRANSACTION	Double	Sum
	SQUARE	FACT TRANSACTION	Double	Sum
	FACT TRANSACTION Count	FACT TRANSACTION	Integer	Count
Add new measure...				

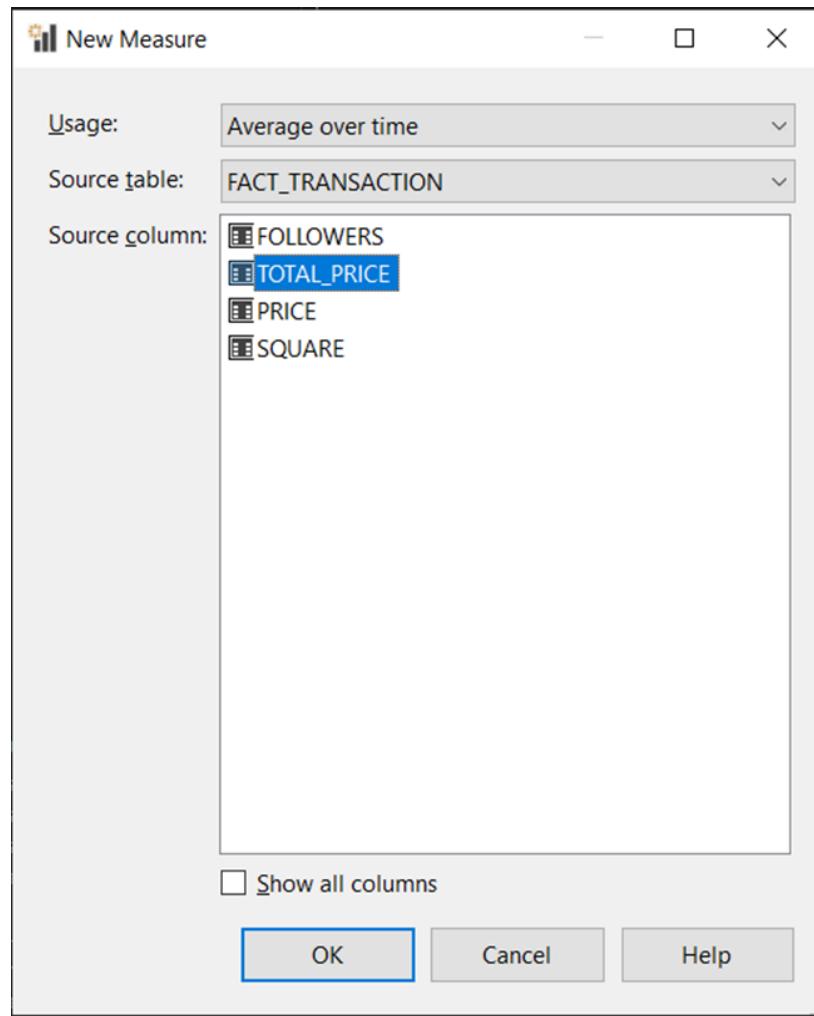
Hình 202: Chi tiết các độ đo

Bước 2: Ta đổi tên và thuộc tính các độ đo hiện tại theo các hàm tổng hợp (aggregation) như sau:

- **FOLLOWER** (Số người theo dõi giao dịch) đổi thành **SUM OF FOLLOWERS** (tổng số người theo dõi giao dịch).
- **TOTAL PRICE** (tổng tiền của giao dịch) đổi thành **SUM OF TOTAL PRICE** (tổng của các tổng tiền giao dịch).
- **PRICE** (Giá trên một m²) đổi thành **MAX PRICE** (Giá trên một m² cao nhất), sau đó đổi Aggregation từ Sum sang Max.
- **Fact TRANSACTION Count** (Số lượng giao dịch) giữ nguyên
- Và thêm các **Measures** tương ứng:

Measures				
	Name	Measure Group	Data Type	Aggregation
■	SUM OF FOLLOWERS	FACT TRANSACTION	Integer	Sum
■	SUM OF TOTAL PRICE	FACT TRANSACTION	Double	Sum
■	MAX OF PRICE	FACT TRANSACTION	Double	Max
■	MAX OF SQUARE	FACT TRANSACTION	Double	Max
■	FACT TRANSACTION Count	FACT TRANSACTION	Integer	Count
■	MIN OF TOTAL PRICE	FACT TRANSACTION	Double	Min
■	AVERAGE OF TOTAL PRICE	FACT TRANSACTION	Double	AverageOfChildren
■	AVERAGE OF FOLLOWERS	FACT TRANSACTION	Integer	AverageOfChildren
■	AVERAGE OF SQUARE	FACT TRANSACTION	Double	Sum
■	URL Count	FACT TRANSACTION	WChar	Count
	Add new measure...			

Hình 203: Thêm các Measures tương ứng



Hình 204: Nhấn Ok

Measures				
	Name	Measure Group	Data Type	Aggregation
■	SUM OF FOLLOWERS	FACT TRANSACTION	Integer	Sum
■	SUM OF TOTAL PRICE	FACT TRANSACTION	Double	Sum
■	MAX OF PRICE	FACT TRANSACTION	Double	Max
■	MAX OF SQUARE	FACT TRANSACTION	Double	Max
■	FACT TRANSACTION Count	FACT TRANSACTION	Integer	Count
■	AVERAGE OF TOTAL PRICE	FACT TRANSACTION	Double	AverageOfChildren
■	MIN OF TOTAL PRICE	FACT TRANSACTION	Double	Min
Add new measure...				

Hình 205: Các Measures

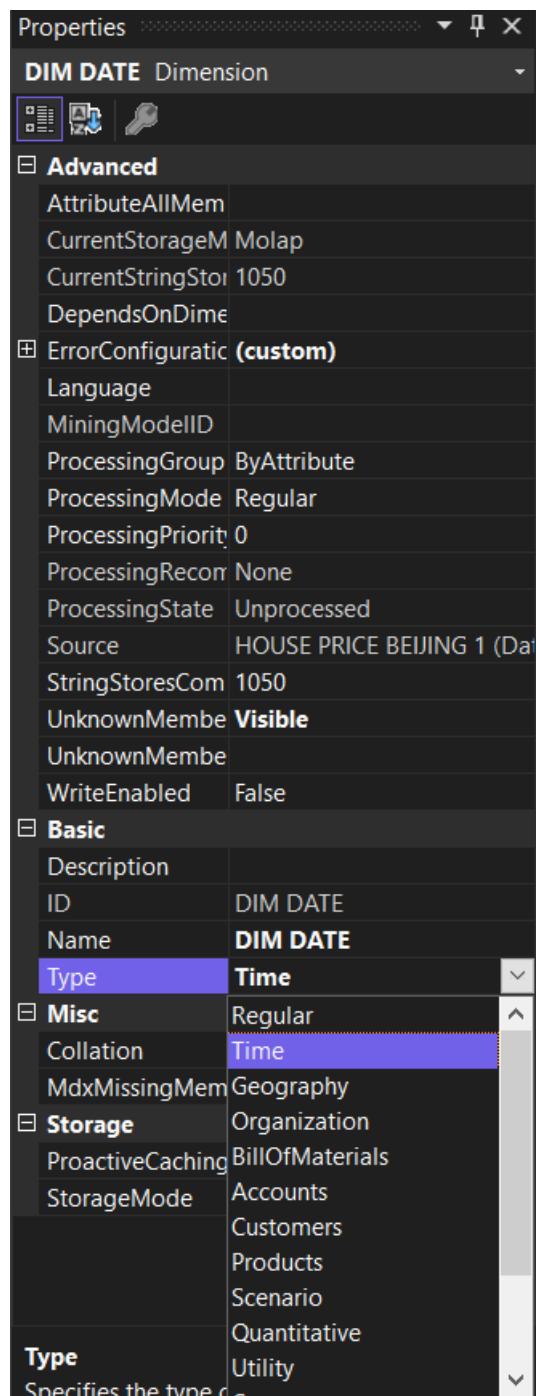
- Sau khi đổi tên và thuộc tính các độ đo ban đầu. Một thông báo xuất hiện yêu cầu ta phải có một time dimension.

Measures				
	Name	Measure Group	Data Type	Aggregation
■	SUM OF FOLLOWERS	FACT TRANSACTION	Integer	Sum
■	SUM OF TOTAL PRICE	FACT TRANSACTION	Double	Sum
■	MAX OF PRICE	FACT TRANSACTION	Double	Max
■	MAX OF SQUARE	FACT TRANSACTION	Double	Max
■	FACT TRANSACTION Count	FACT TRANSACTION	Integer	Count
■	AVERAGE OF TOTAL PRICE	FACT TRANSACTION	Double	AverageOfChildren
■	MIN OF TOTAL PRICE	FACT TRANSACTION	Double	Min
Add new measure...				

Semi-additive measure requires a time dimension.

Hình 206: Thông báo xuất hiện yêu cầu ta phải có một time dimension

Bước 3: Mở Dim Time.dim. Tại cửa sổ **Properties**, ta đổi kiểu bảng từ **Regular** sang **Time**.



Hình 207: Kiểu bảng từ Regular sang Time.

- Quá trình hoàn tất, ta có được các độ đo:

Measures				
	Name	Measure Group	Data Type	Aggregation
■	SUM OF FOLLOWERS	FACT TRANSACTION	Integer	Sum
■	SUM OF TOTAL PRICE	FACT TRANSACTION	Double	Sum
■	MAX OF PRICE	FACT TRANSACTION	Double	Max
■	MAX OF SQUARE	FACT TRANSACTION	Double	Max
■	FACT TRANSACTION Count	FACT TRANSACTION	Integer	Count
■	MIN OF TOTAL PRICE	FACT TRANSACTION	Double	Min
■	AVERAGE OF TOTAL PRICE	FACT TRANSACTION	Double	AverageOfChildren
■	AVERAGE OF FOLLOWERS	FACT TRANSACTION	Integer	AverageOfChildren
■	AVERAGE OF SQUARE	FACT TRANSACTION	Double	Sum
■	URL Count	FACT TRANSACTION	WChar	Count
	Add new measure...			

Hình 208: Các độ đo

5.1. Phân cấp bảng Dim_Time

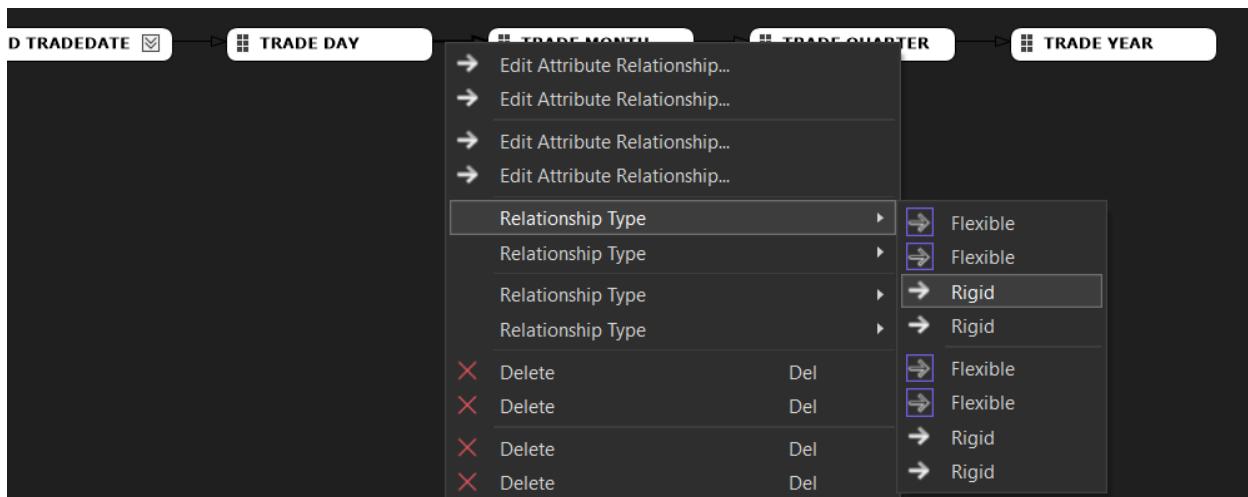
- Kéo những thuộc tính cần phân cấp qua cửa sổ **Hierarchies**

Hình 209: Kéo những thuộc tính cần phân cấp qua cửa sổ Hierarchies

- Tiến hành phân cấp theo thứ tự từ trái sang phải



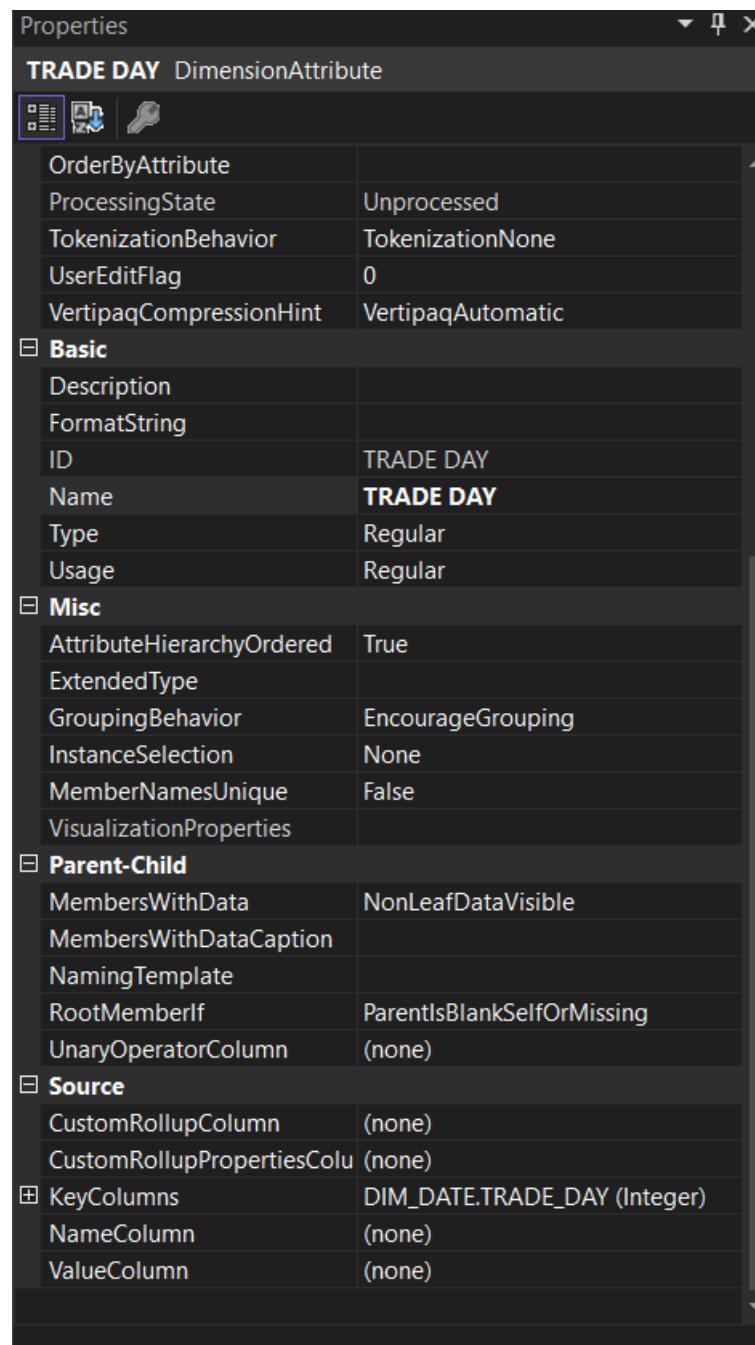
Hình 210: Phân cấp theo thứ tự từ trái sang phải



Hình 211: Chọn Rigid

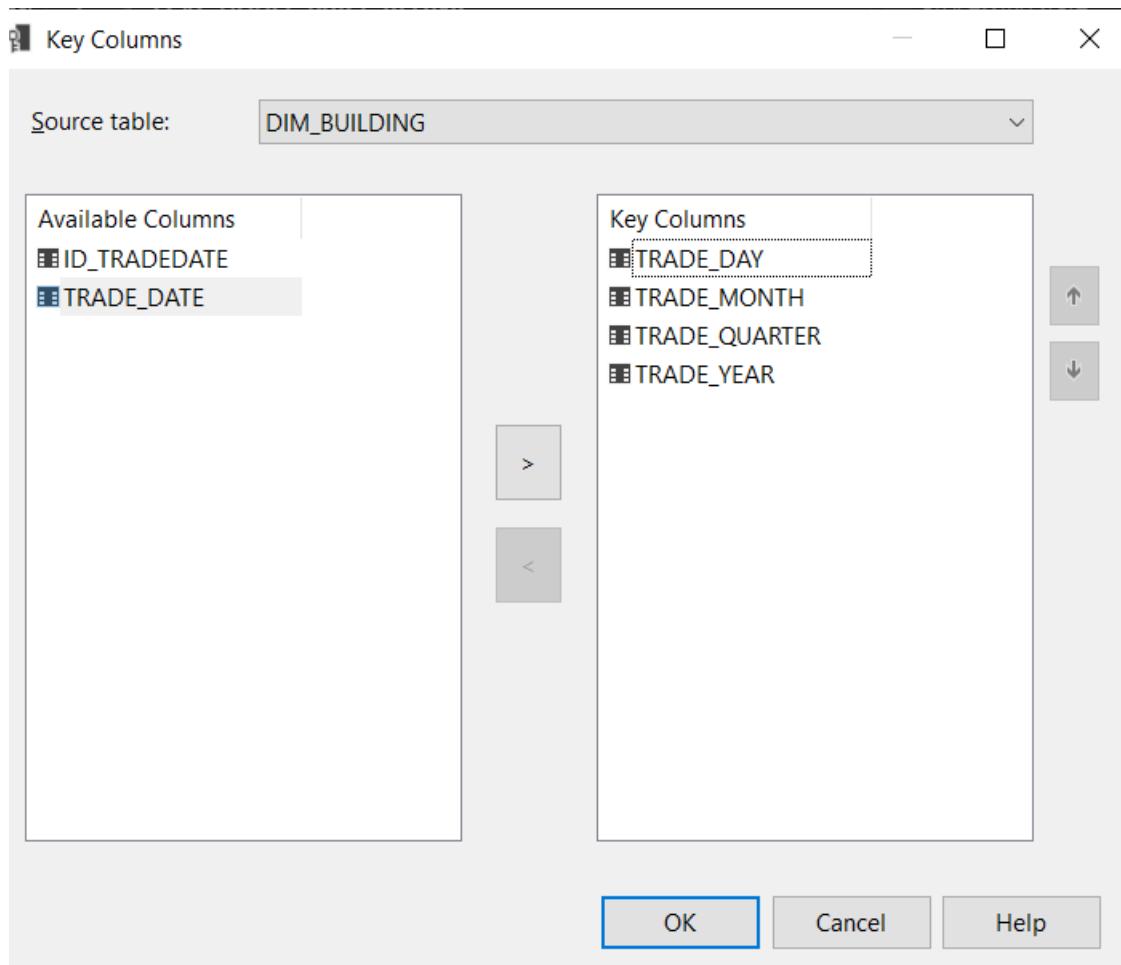
Bước 4: Chính khóa cột (**KeyColumns**) và tên cột (**Name Column**) của thuộc tính **Start Time Minute**. Vì thuộc tính **Start Time Minute** là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

- Tại cửa sổ **Properties** của thuộc tính **Start Time Minute**, chọn **KeyColumns**.



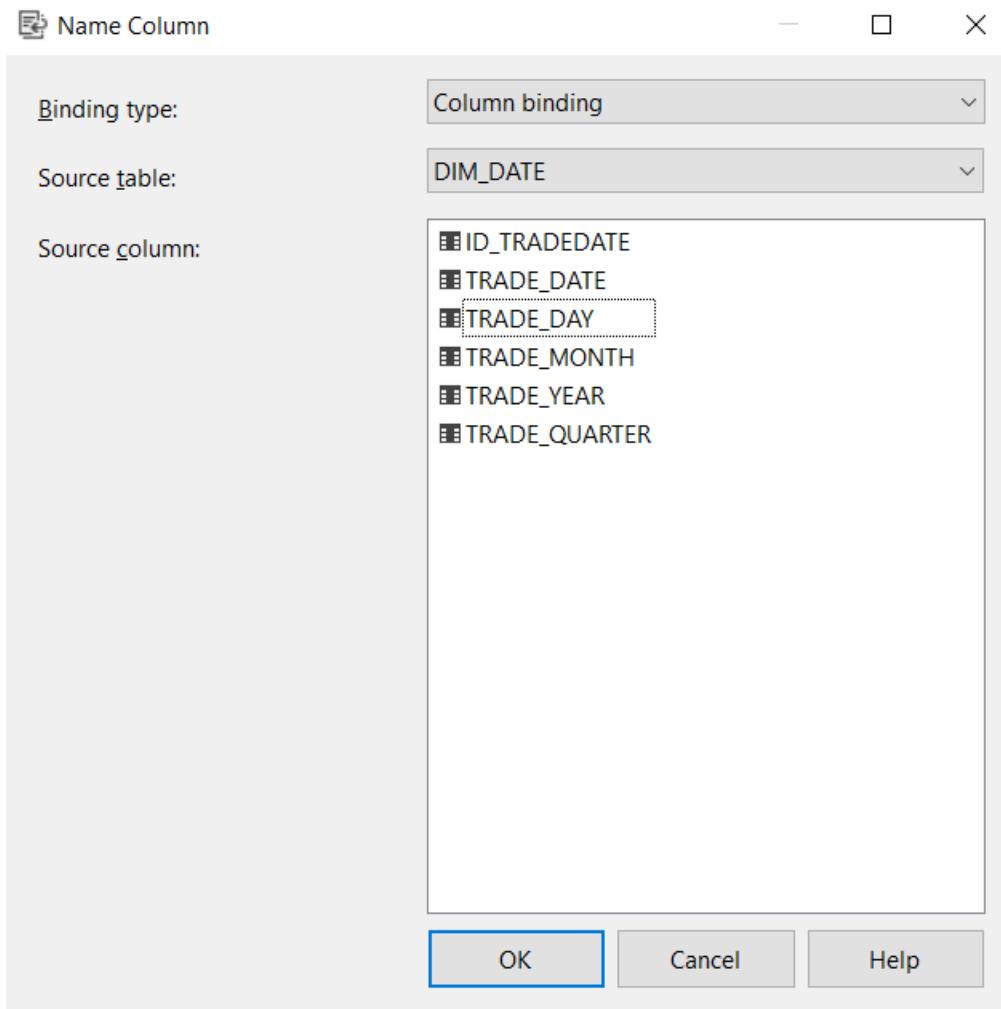
Hình 212: Chọn KeyColumns

- Thêm các thuộc tính cấp cao hơn vào **KeyColumns**, sau đó chọn **OK** để hoàn tất.



Hình 213: Thêm các thuộc tính cấp cao hơn vào KeyColumns

- Tại cửa sổ **Properties** của thuộc tính **Start Time Minute**, ta chọn **Name Column** và chọn tên thuộc tính là **Start Time Minute**.



Hình 214: Chọn Name Column và chọn tên thuộc tính là Start Time Minute

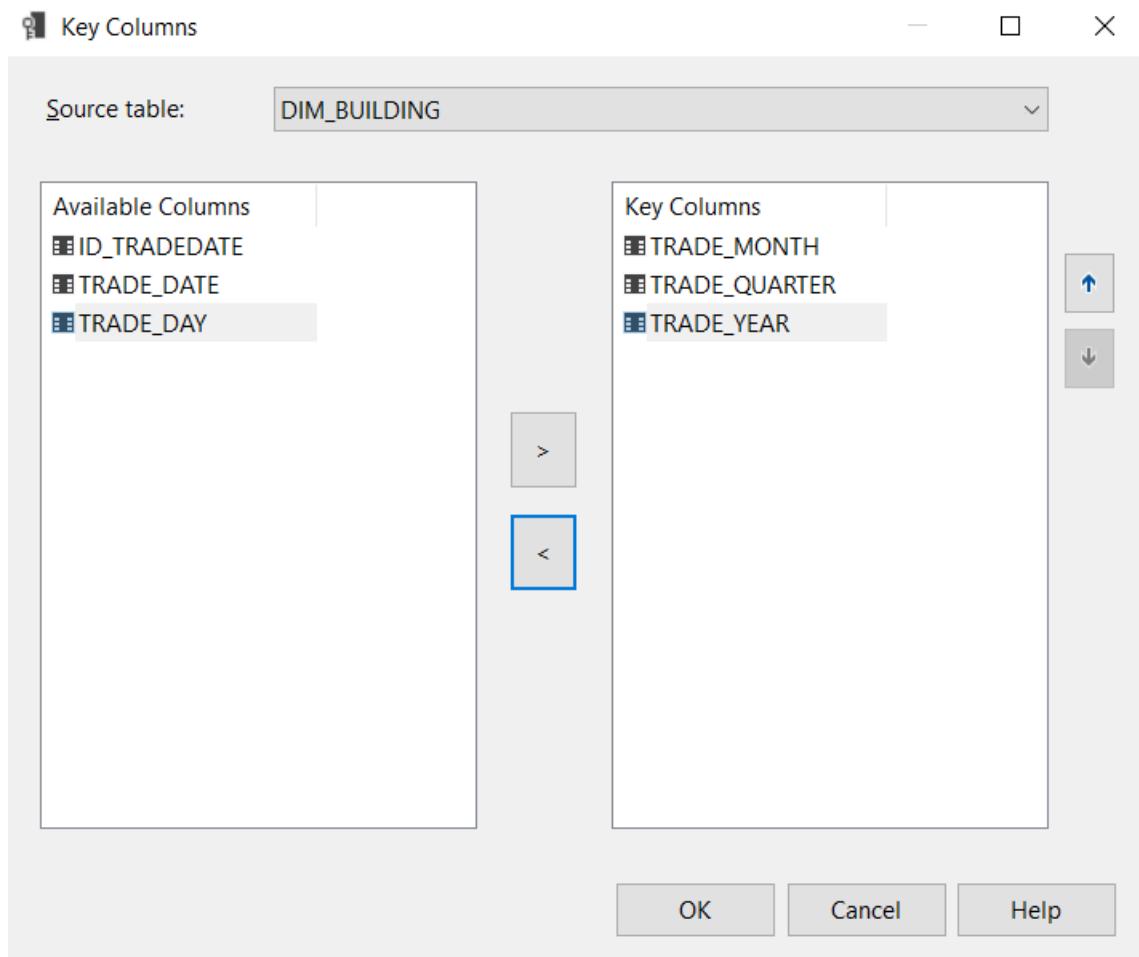
Bước 5: Chính khóa cột (**KeyColumns**) và tên cột (**Name Column**) của thuộc tính **Start Time Hour**. Vì thuộc tính Start Time Hour là thuộc tính cấp nhỏ hơn Start Time Day, Start Time Month, Start Time Quarter, Start Time Year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

- Tại cửa sổ **Properties** của thuộc tính **Start Time Hour**, chọn **KeyColumns**.

TRADE MONTH DimensionAttribute	
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompressionHint	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	TRADE MONTH
Name	TRADE MONTH
Type	Regular
Usage	Regular
Misc	
AttributeHierarchyOrdered	True
ExtendedType	
GroupingBehavior	EncourageGrouping
InstanceSelection	None
MemberNamesUnique	False
VisualizationProperties	
Parent-Child	
MembersWithData	NonLeafDataVisible
MembersWithDataCaption	
NamingTemplate	
RootMemberIf	ParentIsBlankSelfOrMissing
UnaryOperatorColumn	(none)
Source	
CustomRollupColumn	(none)
CustomRollupPropertiesColumn	(none)
KeyColumns	
NameColumn	DIM_DATE.TRADE_MONTH (Integer)
ValueColumn	(none)

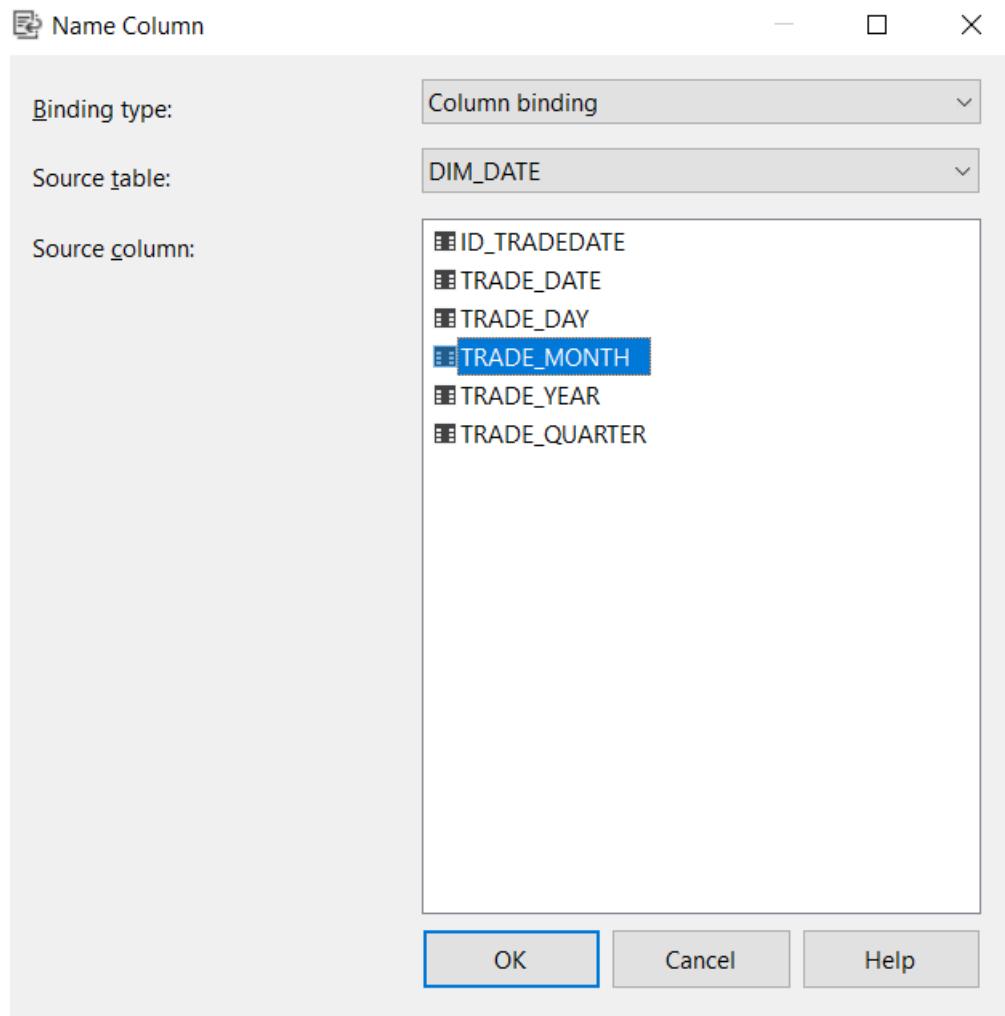
Hình 215: Chọn KeyColumns

- Thêm các những thuộc tính cấp cao hơn vào **KeyColumns**, sau đó chọn **OK** để hoàn tất.



Hình 216: Thêm các những thuộc tính cấp cao hơn vào

- Tại cửa sổ **Properties** của thuộc tính **Start Time Hour**, ta chọn **Name Column** và chọn tên thuộc tính là **Start Time Hour**.



Hình 217: Chọn Name Column và chọn tên thuộc tính là Start Time Hour

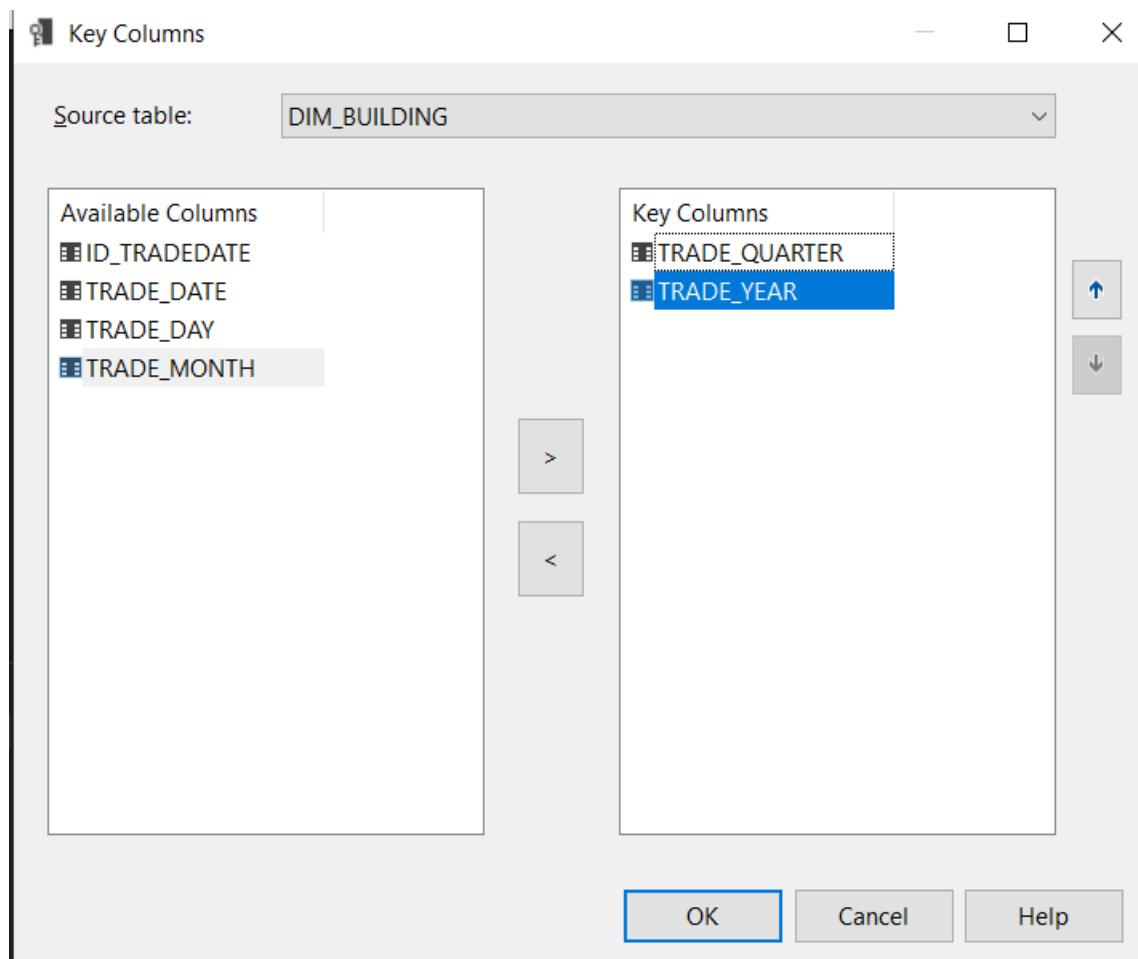
Bước 6: Chính khóa cột (**KeyColumns**) và tên cột (**Name Column**) của thuộc tính **Start Time Day**. Vì thuộc tính Start Time Day là thuộc tính cấp nhỏ hơn Start Time Month, Start Time Quarter, Start Time Year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

- Tại cửa sổ **Properties** của thuộc tính **Start Time Day**, chọn **KeyColumns**.

TRADE QUARTER DimensionAttribute	
OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompressionHint	VertipaqAutomatic
Basic	
Description	
FormatString	
ID	TRADE QUARTER
Name	TRADE QUARTER
Type	Regular
Usage	Regular
Misc	
AttributeHierarchyOrdered	True
ExtendedType	
GroupingBehavior	EncourageGrouping
InstanceSelection	None
MemberNamesUnique	False
VisualizationProperties	
Parent-Child	
MembersWithData	NonLeafDataVisible
MembersWithDataCaption	
NamingTemplate	
RootMemberIf	ParentIsBlankSelfOrMissing
UnaryOperatorColumn	(none)
Source	
CustomRollupColumn	(none)
CustomRollupPropertiesColumn	(none)
KeyColumns	DIM_DATE.TRADE_QUARTER (Integer)
NameColumn	(none)
ValueColumn	(none)

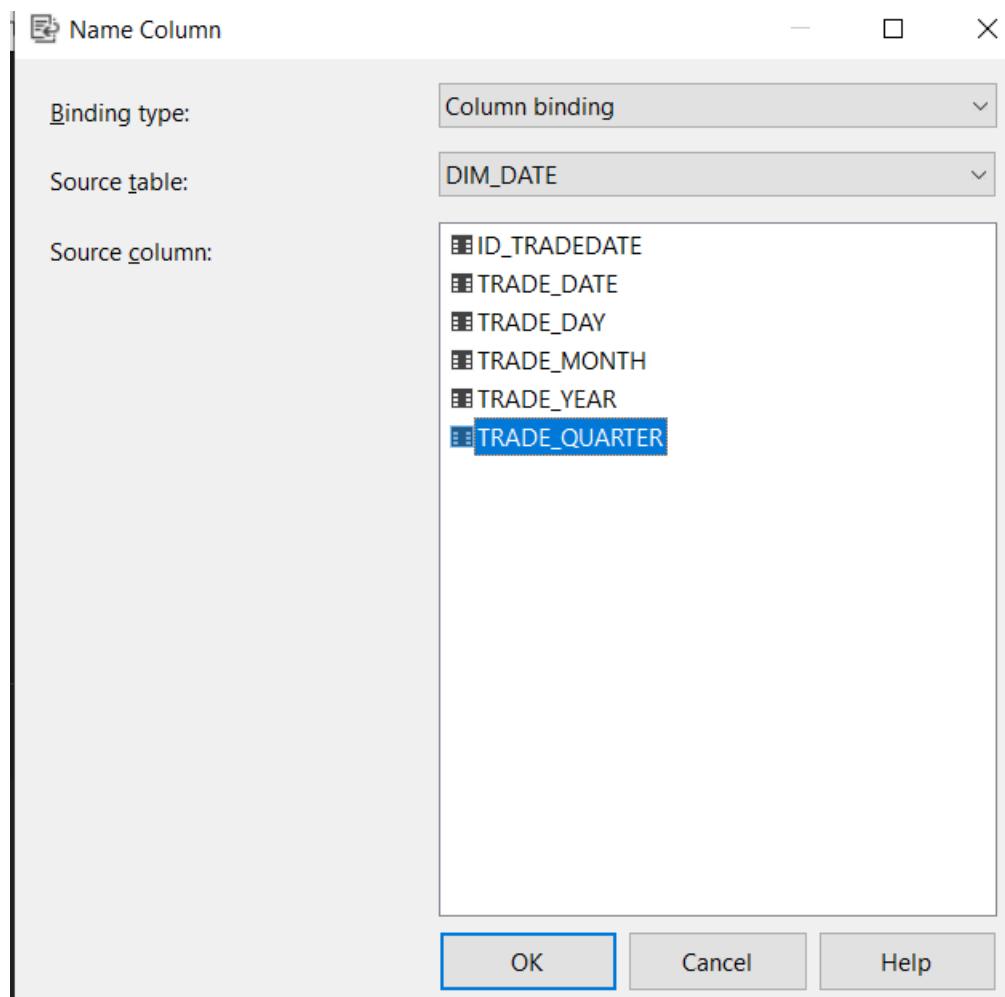
Hình 218: Chọn KeyColumns

- Thêm các những thuộc tính cấp cao hơn vào **KeyColumns**, sau đó chọn **OK** để hoàn tất.



Hình 219: Thêm các những thuộc tính cấp cao hơn vào

- Tại cửa sổ **Properties** của thuộc tính **Start Time Day**, ta chọn **Name Column** và chọn tên thuộc tính là **Start Time Day**.



Hình 220: Chọn Name Column và chọn tên thuộc tính là Start Time Day

- **Deploy** dự án:



Hình 221: Deploy thành công

6. Thực hiện 16 câu truy vấn – Quá trình phân tích dữ liệu bằng thao tác tay trên các khối CUBE, Pivot table trong Excel, ngôn ngữ MDX

6.1. Câu truy vấn 1

Nội dung câu truy vấn: Với mỗi khu vực (**DISTRICT**) đưa ra nhà có diện tích lớn nhất.

Thực hiện truy vấn trên Cube (manual):

Bước 1: Kéo thả thuộc tính [**DISTRICT**] trong bảng [**DIM HOUSE**] và độ đo [**MAX OF SQUARE**]

The screenshot shows the Microsoft SQL Server Analysis Services (SSAS) Management Studio interface. On the left, the 'Cube Browser' pane displays the structure of the 'HOUSE PRICE BEIJING 1' cube. It lists various dimensions and measures, with the 'MAX OF SQUARE' measure highlighted. In the center, the 'Query Editor' pane contains a simple MDX query:

```
DISTRICT | MAX OF SQUARE
```

This query retrieves the maximum square value for each district. The interface includes a toolbar at the top and a status bar at the bottom.

Hình 222: Kéo thả thuộc tính và độ đo

Bước 2: Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

The screenshot shows a software interface for executing MDX queries. On the left, there is a sidebar titled "Metadata" containing a tree view of data structures. The tree includes categories like "Measures" (FACT TRANSACTION), "KPIs", "DIM DATE", "DIM HOUSE", and "FACT TRANSACTION". Under "FACT TRANSACTION", there are several measures such as AVERAGE OF TOTAL PRICE, MAX OF PRICE, and SUM OF FOLLOWERS. The main area displays a table titled "MAX OF SQUARE" with two columns: "DISTRICT" and "MAX OF SQUARE". The data in the table is as follows:

DISTRICT	MAX OF SQUARE
1	1242051.48
2	2294057.86999999
3	234749.71
4	1339260.22
5	264174.16
6	3897511.30000004
7	9050494.76999993
8	3046481.68000001
9	854415.66999999
10	2112975.74
11	1196898.2
12	145393.5
13	862883.19999999

Hình 223: Thực thi câu lệnh và được kết quả

Thực hiện truy vấn trên Pivot table trong Excel:

A	B
DISTRICT	MAX OF SQUARE
1	1242051.48
2	2294057.87
3	234749.71
4	1339260.22
5	264174.16
6	3897511.3
7	9050494.77
8	3046481.68
9	854415.67
10	2112975.74
11	1196898.2
12	145393.5
13	862883.2
Grand Total	26541347.5

Hình 224: Kết quả Pivot table trong Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```
SELECT NON EMPTY { [Measures].[MAX OF SQUARE] } ON COLUMNS,
NON EMPTY { [DIM HOUSE].[DISTRICT].CHILDREN} ON ROWS
FROM [HOUSE PRICE BEIJING 1]
```

- Kết quả:

The screenshot shows the SSAS Script Editor interface. At the top, there is a code editor window containing the following DAX query:

```
SELECT NON EMPTY { [Measures].[MAX OF SQUARE] } ON COLUMNS,
NON EMPTY { [DIM HOUSE].[DISTRICT].CHILDREN } ON ROWS
FROM [HOUSE PRICE BEIJING 1]
```

Below the code editor is a results grid titled "Results". The grid has one column labeled "MAX OF SQUARE" and 13 rows, each containing a numerical value. The first row is highlighted in blue.

	MAX OF SQUARE
1	124205148
2	2294057.86999999
3	234749.71
4	1339260.22
5	264174.16
6	3897511.30000004
7	9050494.76999993
8	3046481.68000001
9	854415.669999999
10	2112975.74
11	1196898.2
12	145393.5
13	862883.199999999

Hình 225: Kết quả

6.2. Câu truy vấn 2

Nội dung câu truy vấn: Liệt kê theo từng năm có giao dịch có giá trên m² lớn nhất và lớn hơn 100000 (Sử dụng bảng phân cấp [TRANS TIME], roll up từ DAY lên YEAR)

Thực hiện truy vấn trên Cube (manual)

Bước 1: Tạo NameSet [FACT_TRANSACTION_MAXOFPRICE_100000] như hình minh họa với biểu thức như sau:

`FILTER({[DIM DATE].[TRANS TIME].[TRADE YEAR].ALLMEMBERS},
[FACT TRANSACTION].[MAX OF PRICE] > 100000)`

The screenshot shows the SSAS Script Organizer. On the left, there is a tree view of scripts. In the center, there is a "Name:" field containing "[FACT_TRANSACTION_MAXOFPRICE_100000]" and an "Expression" field containing the DAX code from the previous step. Below the expression is a message box stating "No issues found". At the bottom, there are "Additional Properties" sections for "Type" (set to "Dynamic") and "Display folder".

Hình 226: Tạo NameSet

Bước 2: Kéo thuộc tính [TRADE YEAR] trong bảng [DIM TIME] và độ đo [FACT TRANSACTION COUNT] vào cửa sổ thực thi.

Hình 227: Kéo thuộc tính và độ đo vào cửa sổ thực thi

Bước 3: Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM DATE	TRANS TIME	In	FACT_TRANSACTION_MAXOFPRICE_100000
<Select dimension>			
			Click to execute the query.
TRADE YEAR	MAX OF PRICE		
2011	125425		
2012	147240		
2013	156250		
2014	147888		
2015	149276		
2016	150000		
2017	150000		
2018	142120		

Hình 228: Được kết quả như hình

Thực hiện truy vấn trên Pivot table trong Excel:

A	B
TRADE YEAR	
2011	125425
2012	147240
2013	156250
2014	147888
2015	149276
2016	150000
2017	150000
2018	142120
Grand Total	156250

Hình 229: Kết quả trong Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT
    NON EMPTY { [Measures].[MAX OF PRICE] } ON COLUMNS,
    NON EMPTY
    FILTER(
        [DIM DATE].[TRANS TIME].[TRADE YEAR].ALLMEMBERS,
        [Measures].[MAX OF PRICE] > 100000
    )
    ON ROWS
FROM
    [HOUSE PRICE BEIJING 1]
  
```

Kết quả:

	MAX OF PRICE
2011	125425
2012	147240
2013	156250
2014	147888
2015	149276
2016	150000
2017	150000
2018	142120

Hình 230: Kết quả MDX

6.3. Câu truy vấn 3

Nội dung câu truy vấn: Liệt kê TOP 3 giao dịch (ID_TRANS, ID_DATE, ID_HOUSE) có tổng giá thấp nhất.

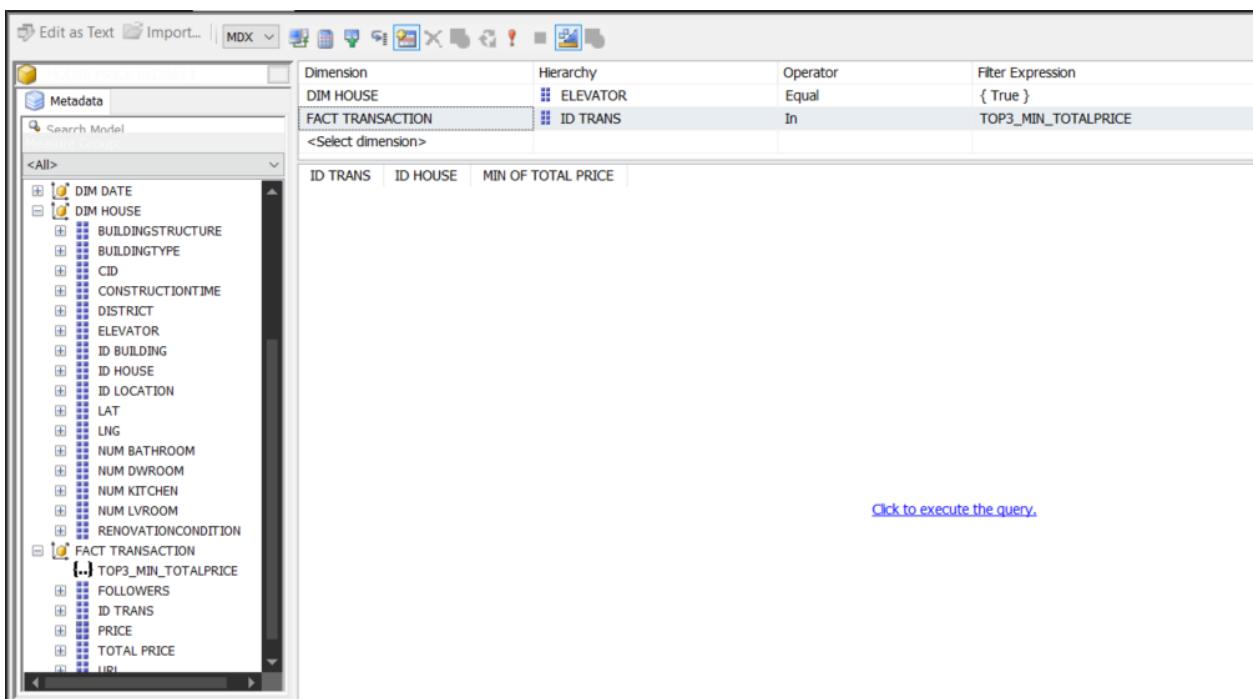
Thực hiện truy vấn trên Cube (manual):

Bước 1: Tạo NameSet [BOTTOM3_MIN_TOTALPRICE] như hình minh họa với biểu thức như sau:

The screenshot shows the 'Script Organizer' window. On the left, there's a list of calculated members with step numbers and descriptions. Step 6 is highlighted and shows the creation of a named set. The right panel displays the properties of this named set. The 'Name' field is set to '[BOTTOM3_MIN_TOTALPRICE]'. The 'Expression' field contains the MDX code: 'BOTTOMCOUNT([FACT TRANSACTION].[ID TRANS].CHILDREN, 3, [Measures].[MIN OF TOTAL PRICE])'. Below the expression, a message says 'No issues found'. Under 'Additional Properties', the 'Type' is set to 'Dynamic'. The bottom right corner shows the line count as 'Ln: 1' and character count as 'Ch: 86'.

Hình 231: Tạo NameSet

Bước 2: Kéo thuộc tính [ID TRANS] trong bảng [FACT], [ID HOUSE] trong bảng [DIM HOUSE] và [MIN OF TOTAL PRICE] vào cửa sổ thực thi.



Hình 232: Kéo vào thuộc tính

Bước 3: Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
FACT TRANSACTION	ID TRANS	In	BOTTOM3_MIN_TOTALPRICE
<Select dimension>			
ID TRANS	ID HOUSE	MIN OF TOTAL PRICE	
BJXC85122015	23367	1000	
BJXC85986083	40267	1000	
BJXW00406500	40868	1000	

Hình 233: Được kết quả như hình

Thực hiện truy vấn trên Pivot table trong Excel:

ID TRANS/ID HOUSE	MIN OF TOTAL PRICE
BJXC85122015	
23367	1000
BJXC85986083	
40267	1000
BJXW00406500	
40868	1000
Grand Total	1000

Hình 234: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT
NON EMPTY {
    [Measures].[MIN OF TOTAL PRICE]
} ON COLUMNS,
NON EMPTY {
    BOTTOMCOUNT([FACT TRANSACTION].[ID
TRANS].CHILDREN,3,[Measures].[MIN OF TOTAL PRICE])
    * [DIM HOUSE].[ID HOUSE].CHILDREN
} ON ROWS
FROM [HOUSE PRICE BEIJING 1]

```

Kết quả:

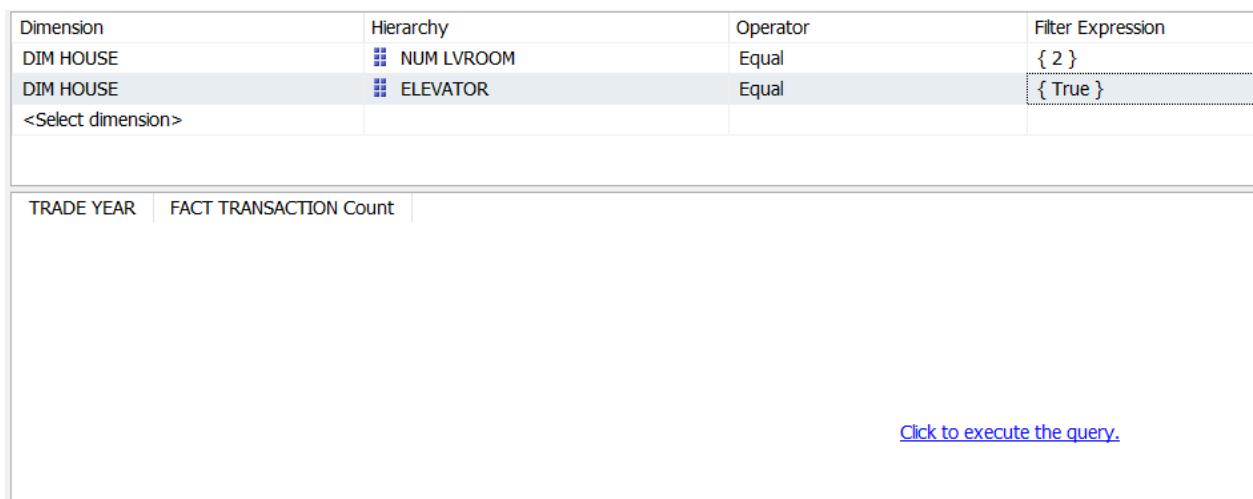
Messages	Results	MIN OF TOTAL PRICE
BJXW00406500	40868	1000
BJXC85986083	40267	1000
BJXC85122015	23367	1000

Hình 235: Kết quả bằng MDX

6.4. Câu truy vấn 4

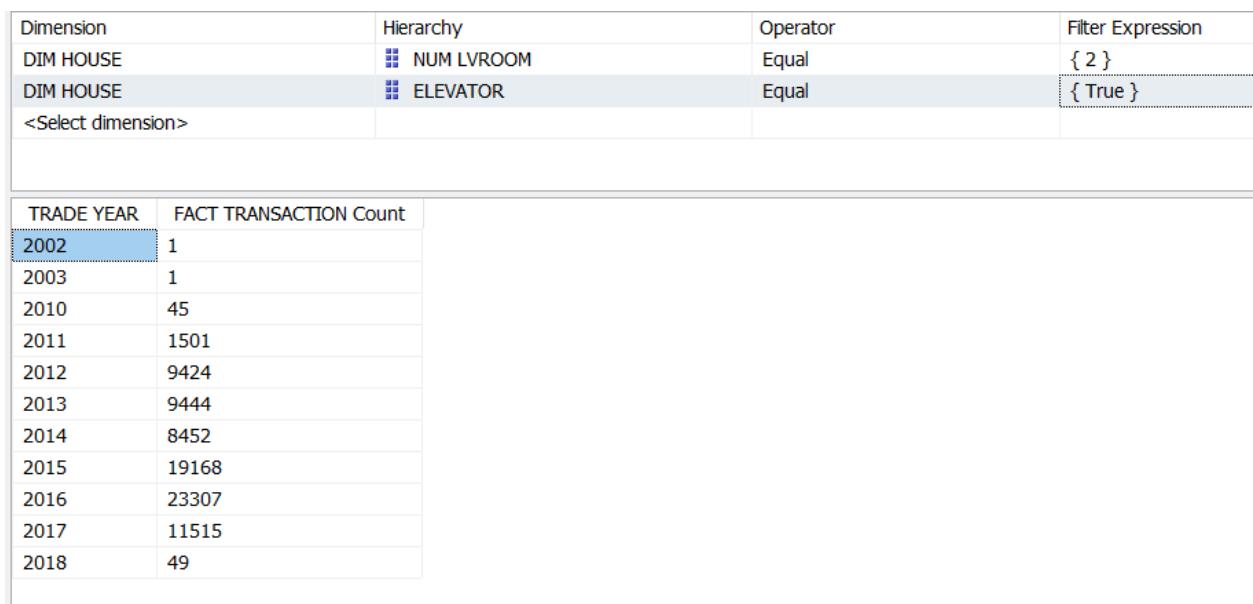
Nội dung câu truy vấn: Thống kê theo các năm đếm có bao nhiêu nhà có số phòng khách (NUM_LVROOM) bằng 2 và có thang máy (ELEVATOR)

Thực hiện truy vấn trên Cube (manual):



Hình 236: Kéo thuộc tính

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:



Hình 237: Kết quả trên Cube

Thực hiện truy vấn trên Pivot table trong Excel:

A	B
NUM LVROOM 2	
ELEVATOR	True
Row Labels	FACT TRANSACTION Count
2002	1
2003	1
2010	45
2011	1501
2012	9424
2013	9444
2014	8452
2015	19168
2016	23307
2017	11515
2018	49
Grand Total	82907

Hình 238: Kết quả trên Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT
[Measures].[FACT TRANSACTION Count] ON COLUMNS,
NON EMPTY [DIM DATE].[TRADE YEAR].MEMBERS ON ROWS
FROM
[HOUSE PRICE BEIJING 1]
WHERE
(
    [DIM HOUSE].[NUM LVROOM].&[2],
    [DIM HOUSE].[ELEVATOR].&[True]
)

```

Kết quả:

FACT TRANSACTION Count	
All	82907
2002	1
2003	1
2010	45
2011	1501
2012	9424
2013	9444
2014	8452
2015	19168
2016	23307
2017	11515
2018	49

Hình 239: Kết quả bằng MDX

6.5. Câu truy vấn 5

Nội dung câu truy vấn: Trong năm 2016, cho biết trung bình số người theo dõi (**FOLLOWERS**) , của nhà là loại công trình là nhà cấp bốn (**BUILDINGTYPE=2**) (**roll up TRANS TIME** từ **TRADE DAY** sang **TRADE YEAR**)

Thực hiện truy vấn trên Cube (manual)

Dimension	Hierarchy	Operator	Filter Expression
DIM DATE	TRANS TIME	Equal	{ 2016 }
DIM HOUSE	BUILDINGTYPE	Equal	{ 2 }
<Select dimension>			

AVERAGE OF FOLLOWERS

Click to execute the query.

Hình 240: Kết quả thuộc tính

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM DATE	TRANS TIME	Equal	{ 2016 }
DIM HOUSE	BUILDINGTYPE	Equal	{ 2 }
<Select dimension>			
AVERAGE OF FOLLOWERS			
23.5581395348837			

Hình 241: Kết quả trên Cube

Thực hiện truy vấn trên Pivot table trong Excel:

A	B	C
1 TRADE YEAR	2016	
2 BUILDINGTYPE	2	
3		
4 AVERAGE OF FOLLOWERS		
5	23.55813953	
6		

Hình 242: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX: (Roll up từ TRADE DAY lên TRADE YEAR)

```
SELECT [Measures].[AVERAGE OF FOLLOWERS] ON COLUMNS
FROM [HOUSE PRICE BEIJING 1]
WHERE ([DIM DATE].[TRANS TIME].[TRADE YEAR].[2016],
[DIM HOUSE].[BUILDINGTYPE].&[2])
```

Kết quả:

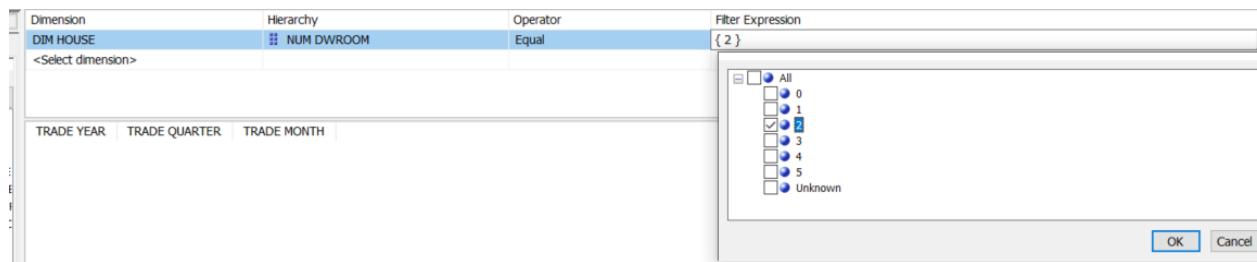
AVERAGE OF FOLLOWERS
23.5581395348837

Hình 243: Kết quả bằng MDX

6.6. Câu truy vấn 6

Nội dung câu truy vấn: Cho biết tổng số tiền các nhà có số phòng vẽ bằng 2 (DRAWINGROOM) theo từng tháng, quý, năm.

Thực hiện truy vấn trên Cube (manual):



Hình 244: Chọn thuộc tính

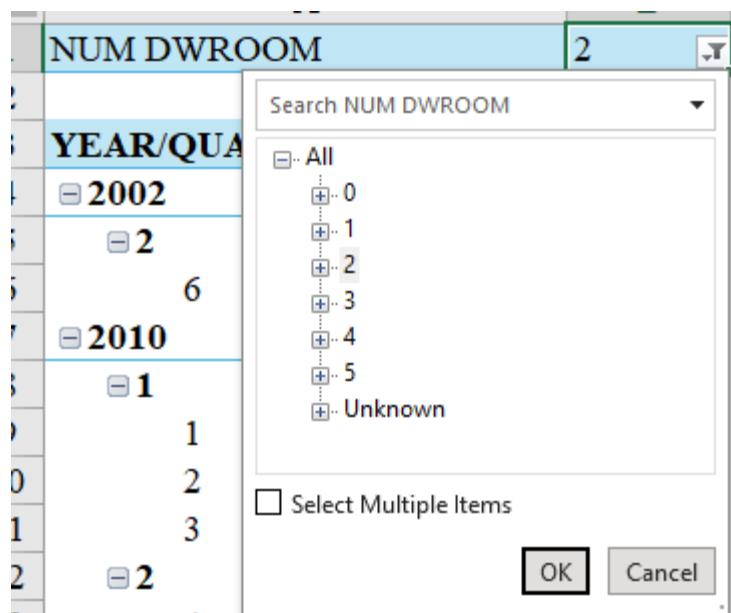
Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

TRADE YEAR	TRADE QUARTER	TRADE MONTH	SUM OF TOTAL PRICE
2002	2	6	1525000
2010	1	1	2255000
2010	1	2	3840000
2010	1	3	6035000
2010	2	4	9410000
2010	2	6	1820000
2010	3	7	1260000
2010	3	8	3325000
2010	3	9	4290000
2010	4	10	6360000
2010	4	11	7719000
2010	4	12	23330000
2011	1	1	13170000
2011	1	3	8930000
2011	2	5	10180000
2011	2	6	37676000
2011	3	7	247071000
2011	3	8	103147000
2011	3	9	651563000
2011	4	10	841634000
2011	4	11	887599000
2011	4	12	1055605000
2012	1	1	458005000
2012	1	2	1520315000
2012	1	3	1904879000
2012	2	4	1724722000
2012	2	5	2242475000
2012	2	6	2601394000
2012	3	7	2561628000
2012	3	8	2221892000
2012	3	9	2116583000
2012	4	10	2380136000
2012	4	11	2785071000
2012	4	12	3964852000
2013	1	1	3972826000

Hình 245: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Filter lớn hơn 2 drawing room:



Hình 246: Sử dụng Filter

Kết quả pivot excel:

NUM DWROOM	2
YEAR/QUARTER/MONTH ▾ SUM OF TOTAL PRICE	
2002	
2	
6	1525000
2010	
1	
1	2255000
2	3840000
3	6035000
2	
4	9410000
6	1820000
3	
7	1260000
8	3325000
9	4290000
4	
10	6360000
11	7719000
12	23330000
2011	
1	
1	13170000
3	8930000
2	
5	10180000
6	37676000
3	
7	247071000

Hình 247: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

SELECT [Measures].[SUM OF TOTAL PRICE] ON COLUMNS,

NON EMPTY {

[DIM DATE].[TRADE YEAR].**CHILDREN***[DIM DATE].[TRADE QUARTER].**CHILDREN***[DIM DATE].[TRADE MONTH].**CHILDREN**}

ON ROWS

FROM [HOUSE PRICE BEIJING 1]

WHERE [DIM HOUSE].[NUM DWROOM].[2]

Kết quả:

The screenshot shows a Microsoft Analysis Services (MDX) query results window. The window has two tabs at the top: 'Messages' (selected) and 'Results'. The results table has a header row 'SUM OF TOTAL PRICE'. The data consists of 20 rows, each containing a date (Year, Month, Day), followed by three numbers (representing dimensions), and a total price value. The data spans from 2002 to 2011.

SUM OF TOTAL PRICE			
2002	2	6	1525000
2010	1	1	2255000
2010	1	2	3840000
2010	1	3	6035000
2010	2	4	9410000
2010	2	6	1820000
2010	3	7	1260000
2010	3	8	3325000
2010	3	9	4290000
2010	4	10	6360000
2010	4	11	7719000
2010	4	12	23330000
2011	1	1	13170000
2011	1	3	8930000
2011	2	5	10180000
2011	2	6	37676000
2011	3	7	247071000
2011	3	8	103147000
2011	3	9	651563000
2011	4	10	841634000

Hình 248: Kết quả bằng MDX

6.7. Câu truy vấn 7

Nội dung câu truy vấn: Cho biết khu vực nào có giá tiền trên m² (**PRICE**) cao nhất từ năm 2016 đến năm 2018.

Thực hiện truy vấn trên Cube (manual):

The screenshot shows the SSAS Management Studio interface. On the left, there is a tree view of the metadata, including a folder named 'HOUSE PRICE BEIJING 1' which contains 'Measures' (FACT TRANSACTION, MAX OF PRICE, etc.) and 'KPIs' (DIM DATE). In the center, there is a table with columns 'Dimension', 'Hierarchy', 'Operator', and 'Filter Expression'. The 'Dimension' column has 'DIM DATE' selected. The 'Hierarchy' column has 'TRANS TIME' selected. The 'Operator' column has 'Range (Inclusive)' selected. The 'Filter Expression' column has '2016 : 2018' entered. Below this table, there is a grid with columns 'DISTRICT' and 'MAX OF PRICE'. The first row of the grid shows '1' and '575713912'. A blue box highlights the 'MAX OF PRICE' value. At the bottom right of the grid, there is a link 'Click to execute the query.'

Hình 249: Chọn thuộc tính

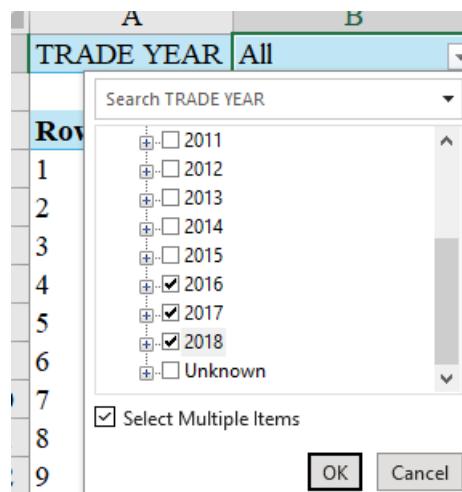
Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM DATE	TRANS TIME	Range (Inclusive)	2016 : 2018
<Select dimension>			
DISTRICT		MAX OF PRICE	
1		575713912	
2		602818544	
3		51294025	
4		263732015	
5		73790418	
6		670505934	
7		2564259257	
8		1064853978	
9		199429500	
10		1098971698	
11		190840346	
12		26334770	
13		115338867	

Hình 250: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Chọn TRADE YEAR từ 2016 – 2018



Hình 251: Chọn các năm tương ứng

Kết quả:

A	B
TRADE YEAR (Multiple Items)	
DISTRICT	MAX OF PRICE
1	575713912
2	602818544
3	51294025
4	263732015
5	73790418
6	670505934
7	2564259257
8	1064853978
9	199429500
10	1098971698
11	190840346
12	26334770
13	115338867
Grand Total	7497883264

Hình 252: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT [Measures].[MAX OF PRICE] ON COLUMNS,
NON EMPTY [DIM HOUSE].[DISTRICT].CHILDREN ON ROWS
FROM [HOUSE PRICE BEIJING 1]

```

WHERE ([DIM DATE].[TRANS TIME].[TRADE YEAR].[2016] : [DIM DATE].[TRANS TIME].[TRADE YEAR].[2018])

Kết quả:

	MAX OF PRICE
1	575713912
2	602818544
3	51294025
4	263732015
5	73790418
6	670505934
7	2564259257
8	1064853978
9	199429500
10	1098971698
11	190840346
12	26334770
13	115338867

Hình 253: Kết quả bằng MDX

6.8. Câu truy vấn 8

Nội dung câu truy vấn: Với kết cấu xây dựng là gạch và bê tông (**BUILDINGSTRUCTURE=4**) hoặc gạch và gỗ (**BUILDINGSTRUCTURE=3**), cho biết trung bình của tổng giá (**TOTALPRICE**) cho mỗi nhà là bao nhiêu, theo từng năm, tháng.

Thực hiện truy vấn trên Cube (manual):

The screenshot shows the SSAS Data Source View (DSV) interface. On the left, there is a tree view of the metadata, including 'Metadata', 'HOUSE PRICE BEIJING 1' (which contains 'Measures' like 'FACT TRANSACTION' with various metrics such as 'AVERAGE OF FOLLOWERS', 'MAX OF PRICE', etc., and 'KPIs'), 'DIM DATE' (with 'ID TRADEDATE', 'TRADE DATE', 'TRADE DAY', 'TRADE MONTH', 'TRADE QUARTER', 'TRADE YEAR', 'TRANS TIME', and 'Members'), and 'DIM HOUSE'. At the top right, there is a configuration panel for a dimension selection:

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	BUILDINGSTRUCTURE	Equal	{ 3, 4 }

Below this, a table is displayed with columns: TRADE YEAR, TRADE MONTH, and AVERAGE OF TOTAL PRICE. The table contains data for years 2010 through 2013. At the bottom right of the table area, there is a link: "Click to execute the query."

Hình 254: Chọn thuộc tính và độ đo

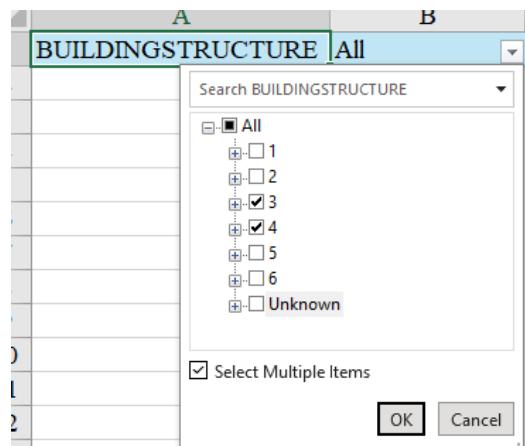
Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	BUILDINGSTRUCTURE	Equal	{ 3, 4 }
<Select dimension>			
TRADE YEAR	TRADE MONTH	AVERAGE OF TOTAL PRICE	
2010	3	780000	
2010	7	1050000	
2010	11	1000000	
2010	12	1285000	
2011	2	1530000	
2011	6	1624400	
2011	7	2281866.66666667	
2011	8	1310333.33333333	
2011	9	3538176.47058824	
2011	10	4685461.53846154	
2011	11	3227560	
2011	12	5345880	
2012	1	4644166.66666667	
2012	2	7214074.07407407	
2012	3	8710833.33333333	
2012	4	8268576.92307692	
2012	5	9070100	
2012	6	11260533.3333333	
2012	7	9645700	
2012	8	8190862.06896552	
2012	9	9947566.66666667	
2012	10	9271033.33333333	
2012	11	11707857.1428571	
2012	12	17239900	
2013	1	15214064.516129	

Hình 255: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Chọn BUILDINGSTRUCTURE bằng 4 hoặc bằng 3 trong filter.



Hình 256: Chọn 3 và 4

Kết quả:

BUILDINGSTRUCTURE	(Multiple Items)
TRADE YEAR/MONTH	AVERAGE OF TOTAL PRICE
2010	
3	780000
7	1050000
11	1000000
12	1285000
2011	
2	1530000
6	1624400
7	2281866.667
8	1310333.333
9	3538176.471
10	4685461.538
11	3227560
12	5345880
2012	
1	4644166.667
2	7214074.074
3	8710833.333
4	8268576.923
5	9070100
6	11260533.33
7	9645700
8	8190862.069
9	9947566.667
10	9271033.333
11	11707857.14
12	17239900
2013	

Hình 257: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```
SELECT [Measures].[AVERAGE OF TOTAL PRICE] ON COLUMNS,  
NON EMPTY {DrillDownLevel(DrillDownLevel([DIM DATE].[TRANS YEAR  
MONTH]))} ON ROWS  
FROM [HOUSE PRICE BEIJING 1]  
WHERE Union(  
 {[DIM HOUSE].[BUILDINGSTRUCTURE].&[4]},  
 {[DIM HOUSE].[BUILDINGSTRUCTURE].&[3]}  
)
```

Kết quả:

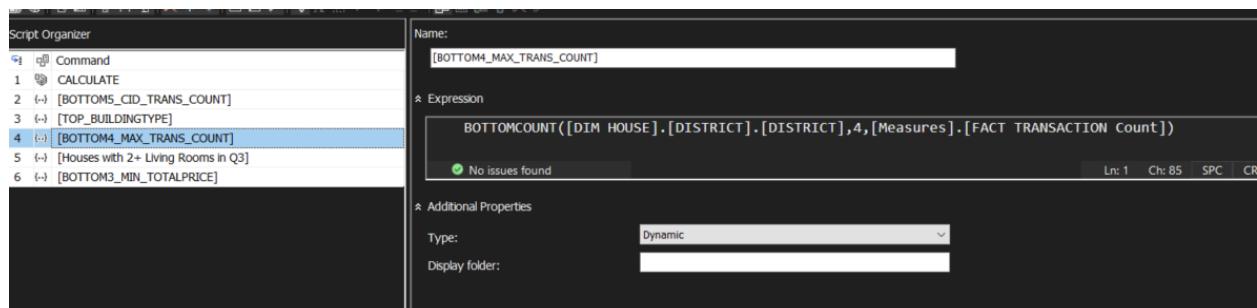
	AVERAGE OF TOTAL PRICE
All	20491429.6839212
2010	1028750
3	780000
7	1050000
11	1000000
12	1285000
2011	3733741.66666667
2	1530000
6	1624400
7	2281866.66666667
8	1310333.33333333
9	3538176.47058824
10	4685461.53846154
11	3227560
12	5345880
2012	9802020.71005917
1	4644166.66666667
2	7214074.07407407
3	8710833.33333333
4	8268576.92307692
5	9070100
6	11260533.3333333
7	9645700
8	8190862.06896552
9	9947566.66666667
10	9271033.33333333
11	11707857.1428571
12	17239900
2013	14409179.7101449
1	15214064.516129
2	13091947.3684211
3	25375161.2903226
4	7919480
5	13902500
6	12652633.3333333
7	13125483.8709677
8	17164733.3333333
9	16030733.3333333
10	12234866.66666667
11	12777862.0689655
12	11350172.4137931
2014	12851140
1	10015720

Hình 258: Kết quả bảng MDX

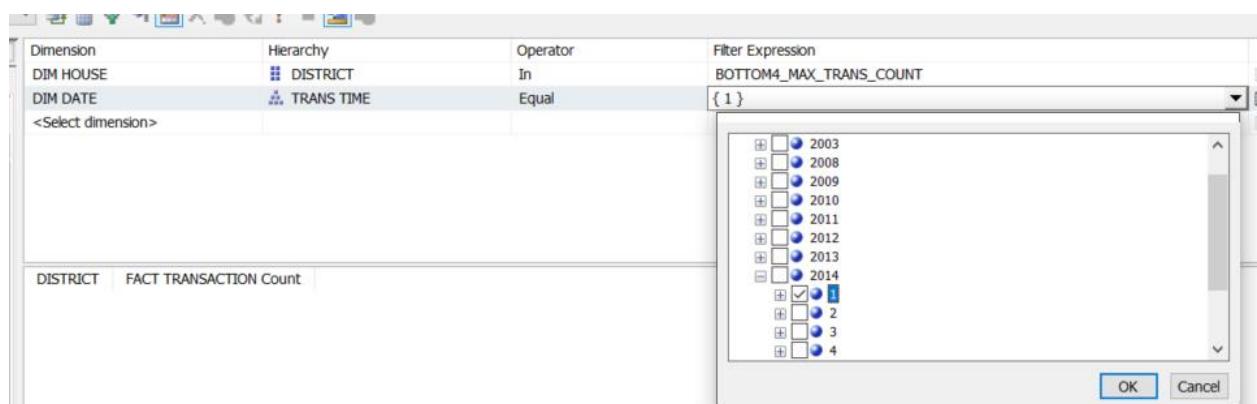
6.9. Câu truy vấn 9

Nội dung câu truy vấn: Cho biết trong quý một năm 2014 (**QUARTER=1**), TOP 2 khu vực nào bán được ít nhà nhất và bán được bao nhiêu.

Thực hiện truy vấn trên Cube (manual):



Hình 259: Chọn NameSet với biểu thức như hình



Hình 260: Chọn thuộc tính và độ đo

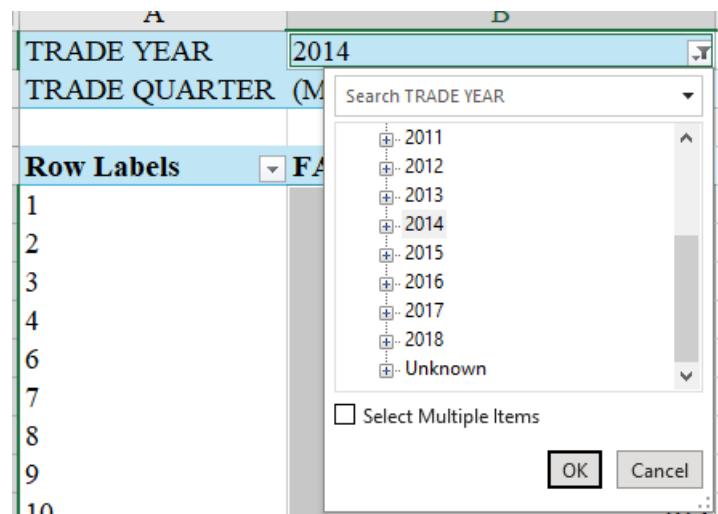
Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	DISTRICT	In	BOTTOM4_MAX_TRANS_COUNT
DIM DATE	TRANS TIME	Equal	{1}
<Select dimension>			
DISTRICT		FACT TRANSACTION Count	
3		43	
12		17	

Hình 261: Kết quả bằng Cube

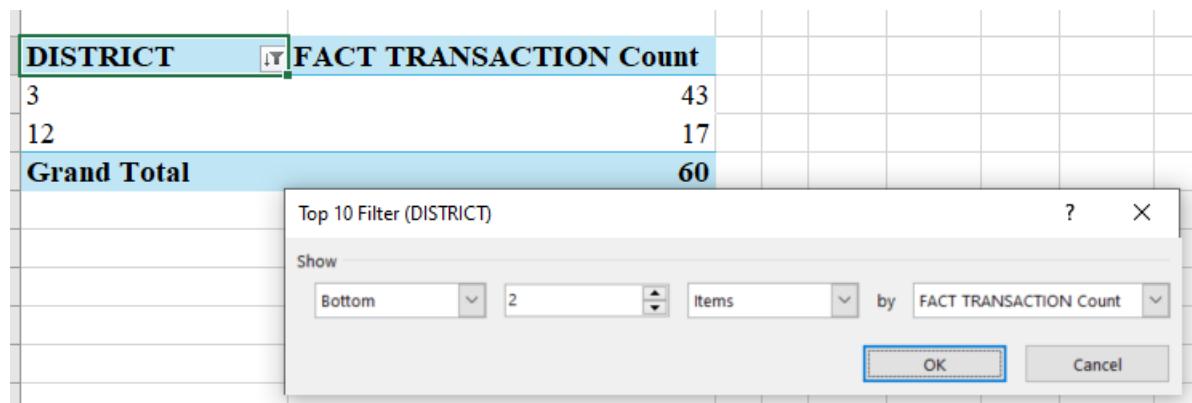
Thực hiện truy vấn trên Pivot table trong Excel:

Filter TRADE YEAR = 2014, TRADE QUARTER = 1



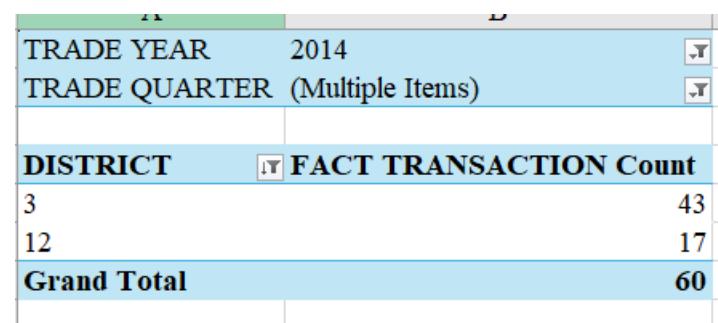
Hình 262: Chọn Filter

DISTRICT TOP 2 ít giao dịch nhất filter:



Hình 263: Chọn Top 10 Filter

Kết quả:



Hình 264: Kết quả bằng Cube

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT NON EMPTY { [Measures].[FACT TRANSACTION Count] } ON COLUMNS,
NON EMPTY
ORDER(
    BOTTOMCOUNT(
        [DIM HOUSE].[DISTRICT].[DISTRICT].MEMBERS,
        4,
        [Measures].[FACT TRANSACTION Count]
    ),
    [Measures].[FACT TRANSACTION Count],
    DESC
) ON ROWS
FROM [HOUSE PRICE BEIJING 1]
WHERE [DIM DATE].[TRANS TIME].[TRADE YEAR].&[2014].&[1]&[2014]

```

Kết quả:

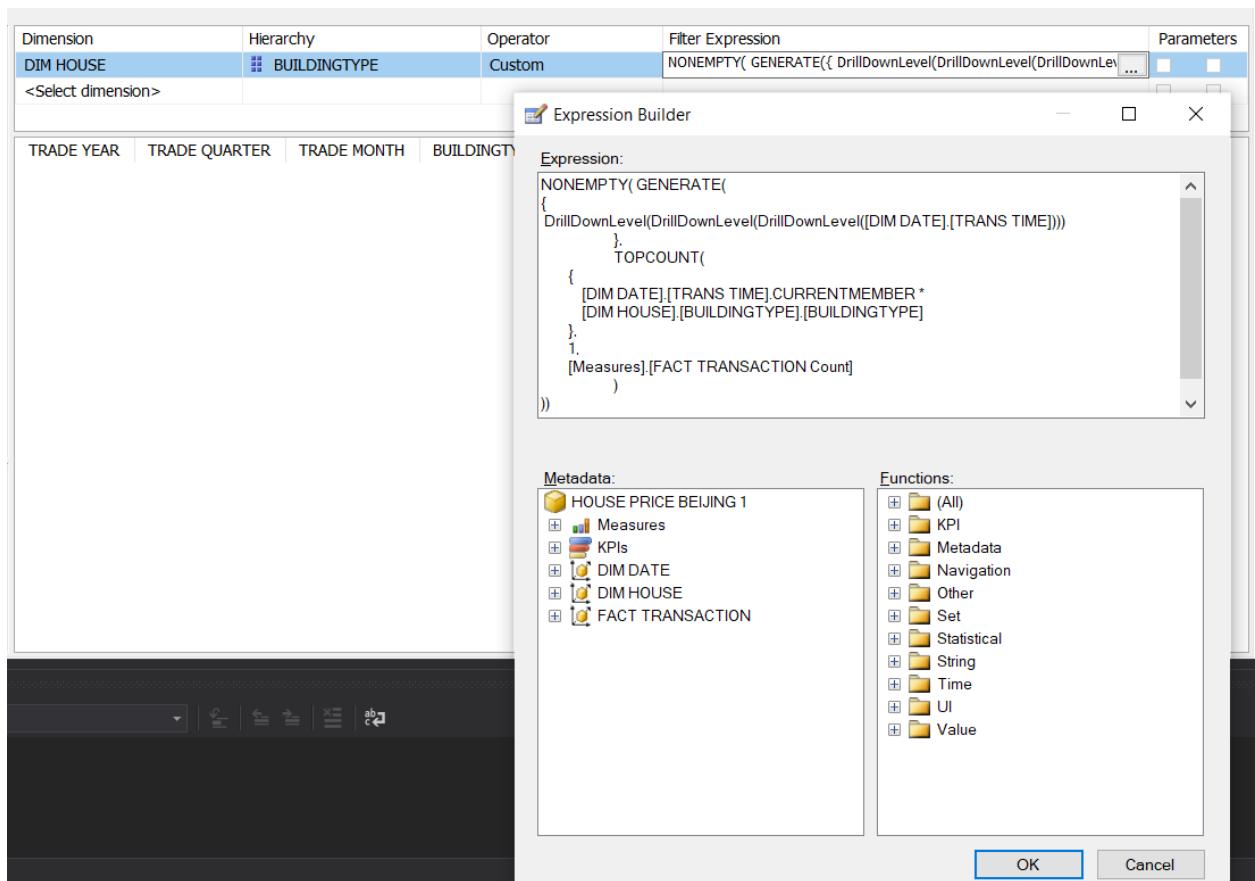
FACT TRANSACTION Count	
3	43
12	17

Hình 265: Kết quả bằng MDX

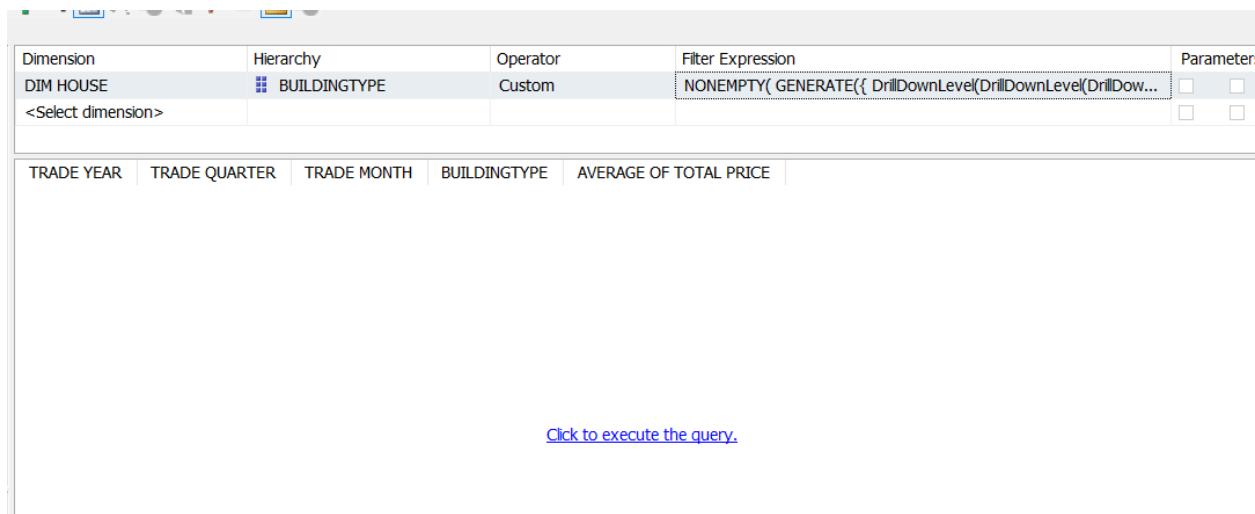
6.10. Câu truy vấn 10

Nội dung câu truy vấn: Thông kê theo năm, quý, tháng, trung bình tổng tiền giao dịch của nhà có loại (**BUILDINGTYPE**) nhiều nhất tại từng thời điểm.

Thực hiện truy vấn trên Cube (manual):



Hình 266: Filter Custom cho thuộc tính BUILDINGTYPE



Hình 267: Chọn thuộc tính

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression	P
DIM HOUSE	BUILDINGTYPE	Custom	NONEMPTY(GENERATE({ DrillDownLevel(DrillDownLevel(DrillDow...)) }))	
<Select dimension>				
TRADE YEAR	TRADE QUARTER	TRADE MONTH	BUILDINGTYPE	AVERAGE OF TOTAL PRICE
2002	2	6	4	1525000
2002	3	7	1	2555000
2003	3	7	1	1607000
2008	4	10	3	1260000
2009	1	1	4	1500000
2010	1	1	4	1097857.14285714
2010	1	2	4	1438400
2010	1	3	4	1780909.09090909
2010	2	4	4	2571357.14285714
2010	2	5	4	1350000
2010	2	6	4	1820000
2010	3	7	4	1692666.66666667
2010	3	8	4	1318500
2010	3	9	4	1680230.76923077
2010	4	10	4	2225200
2010	4	11	4	2030444.44444444
2010	4	12	4	3035000
2011	1	1	4	2074687.5
2011	1	2	4	1876250
2011	1	3	4	1986222.22222222
2011	2	4	4	2168000
2011	2	5	4	2496444.44444444
2011	2	6	4	4073760

Hình 268: Kết quả Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Filter TOP 1 BUILDINGTYPE

The screenshot shows an Excel PivotTable with 'Row Labels' and 'AVERAGE OF TOTAL PRICE' as the value field. A context menu is open over the 'AVERAGE OF TOTAL PRICE' cell, displaying a 'Top 10 Filter (BUILDINGTYPE)' option. A sub-dialog box titled 'Top 10 Filter (BUILDINGTYPE)' is displayed, showing the 'Show' section with 'Top' selected, '1' items, and 'by' 'AVERAGE OF TOTAL PRICE'. The 'OK' button is highlighted.

Hình 269: Sử dụng Filter

Kết quả:

A	B
YEAR/QUARTER/MONTH/BUILDINGTYPE	AVERAGE OF TOTAL PRICE
2002	
2	1525000
6	
4	
3	2555000
7	
1	
2003	
3	1607000
7	
1	
2008	
4	1260000
10	
3	
2009	
1	1500000
1	
4	
2010	
1	1097857.143
1	
4	
2	1438400
4	
3	
4	1780909.091
2	
4	
4	2571357.143

Hình 270: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT [Measures].[AVERAGE OF TOTAL PRICE] ON COLUMNS,
NON EMPTY GENERATE(
{
    DrillDownLevel(DrillDownLevel(DrillDownLevel([DIM DATE].[TRANS
TIME]))))
},

```

```

TOPCOUNT(
{
    [DIM DATE].[TRANS TIME].CURRENTMEMBER *
    [DIM HOUSE].[BUILDINGTYPE].[BUILDINGTYPE]
},
1,
[Measures].[FACT TRANSACTION Count]
)
) ON ROWS
FROM [HOUSE PRICE BEIJING 1]

```

Kết quả:

AVERAGE OF TOTAL PRICE		
All	4	228907065.14015
2002	1	2555000
2	4	1525000
6	4	1525000
3	1	2555000
7	1	2555000
2003	1	1607000
3	1	1607000
7	1	1607000
2008	3	1260000
4	3	1260000
10	3	1260000
2009	4	1500000
1	4	1500000
1	4	1500000
2010	4	1997474.57627119
1	4	1570487.17948718
1	4	1097857.14285714
2	4	1438400
3	4	1780909.09090909
2	4	2448062.5
4	4	2571357.14285714
5	4	1350000
6	4	1820000
3	4	1558068.96551724
7	4	1692666.66666667
8	4	1318500
9	4	1680230.76923077

Hình 271: Kết quả bằng MDX

6.11. Câu truy vấn 11

Nội dung câu truy vấn: Top 5 cộng đồng (CID) có tổng số lượng giao dịch thấp nhất, và cho biết tổng tiền giao dịch của các cộng đồng đó.

Thực hiện truy vấn trên Cube (manual):

```

Name: BOTTOM5_CID_TRANS_COUNT
Script:
  BOTTOMCOUNT([DIM HOUSE].[CID].[CID], 5, [Measures].[FACT TRANSACTION Count])
  
```

Hình 272: Tạo NameSet

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	CID	In	BOTTOM5_CID_TRANS_COUNT
<Select dimension>			
CID	FACT TRANSACTION Count	SUM OF TOTAL PRICE	
1111063830085	1	14000000	
1111099418663	1	13500000	
1111100692723	1	1400000	
1114619720585020	1	1880000	
117258115420294	1	3660000	

Hình 273: Kết quả trên Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Filter TOP 5 CID giao dịch ít nhất

CID	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1111027373683	48	255235000
1111027373684		
1111027373685		
1111027373686		
1111027373689		
1111027373690		

Hình 274: Tao Filter Top 5

Kết quả:

CID	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1111063830085	1	14000000
1111099418663	1	13500000
1111100692723	1	1400000
1114619720585020	1	1880000
117258115420294	1	3660000
Grand Total	5	34440000

Hình 275: Kết quả trên Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```
SELECT {[Measures].[FACT TRANSACTION Count],[Measures].[SUM OF
TOTAL PRICE]} ON COLUMNS,
NON EMPTY BOTTOMCOUNT([DIM HOUSE].[CID].[CID], 5, [Measures].[FACT
TRANSACTION Count]) ON ROWS
FROM [HOUSE PRICE BEIJING 1]
```

Kết quả:

	FACT TRANSACTION Count	SUM OF TOTAL PRICE
117258115420294	1	3660000
1114619720585020	1	1880000
1111100692723	1	1400000
1111099418663	1	13500000
1111063830085	1	14000000

Hình 276: Kết quả trên MDX

6.12. Câu truy vấn 12

Nội dung câu truy vấn: Kiến trúc hiện đại là phong cách thống trị giới kiến trúc phương Tây từ những năm 1933 đến những năm 1960, cho biết trung bình giá, trung bình diện tích của những ngôi nhà được xây dựng trong khoảng thời gian này (**CONSTRUCTIONTIME**) theo từng năm.

Thực hiện truy vấn trên Cube (manual):

The screenshot shows a data analysis interface. On the left, there is a sidebar titled 'Metadata' containing sections for 'Search Model', 'All', 'Measures' (with sub-options like 'FACT TRANSACTION' and various aggregation functions), and 'KPIs'. On the right, there is a main panel with a table for defining a query. The table has four columns: 'Dimension' (set to 'DIM HOUSE'), 'Hierarchy' (set to 'CONSTRUCTIONTIME'), 'Operator' (set to 'Range (Inclusive)'), and 'Filter Expression' (set to '1933 : 1960'). Below the table is a blue link labeled 'Click to execute the query'.

Hình 277: Chọn thuộc tính và độ đo

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	CONSTRUCTIONTIME	Range (Inclusive)	1933 : 1960
<Select dimension>			
CONSTRUCTIONTIME	AVERAGE OF TOTAL PRICE	AVERAGE OF SQUARE	
1933	1643000	56.7	
1934	2092500	15.305	
1944	14066500	132.9	
1950	3468181.81818182	66.7345454545455	
1952	4357600	72.48	
1953	4356846.15384615	66.8853846153846	
1954	5044444.44444444	68.58111111111111	
1955	4628870.96774194	76.3874193548387	
1956	4393714.28571429	75.2950649350649	
1957	3235712.12121212	63.6536363636364	
1958	4082136.75213675	64.7309401709402	
1959	3349160	67.9386	
1960	4275461.07784431	70.9643113772456	

Hình 278: Kết quả trên Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Filter CONSTRUCTION TIME chọn between: 1933 đến 1960

A	B	C
CONSTRUCTIONTIME	AVERAGE OF TOTAL PRICE	AVERAGE OF SQUARE
1933	1643000	56.7
1934		105
1944		9
1950		15
1952		18
1953		52
1954		52

Hình 279: Tạo Filter between

Kết quả:

A	B	C
CONSTRUCTIONTIME	AVERAGE OF TOTAL PRICE	AVERAGE OF SQUARE
1933	1643000	56.7
1934	2092500	15.305
1944	14066500	132.9
1950	3468181.818	66.73454545
1952	4357600	72.48
1953	4356846.154	66.88538462
1954	5044444.444	68.58111111
1955	4628870.968	76.38741935
1956	4393714.286	75.29506494
1957	3235712.121	63.65363636
1958	4082136.752	64.73094017
1959	3349160	67.9386
1960	4275461.078	70.96431138
Grand Total	4928412.2	81.69330869

Hình 280: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```
SELECT {[Measures].[AVERAGE OF TOTAL PRICE],[Measures].[AVERAGE OF
SQUARE]} ON COLUMNS,
```

```
NON EMPTY DESCENDANTS(
```

```
 {[DIM HOUSE].[CONSTRUCTIONTIME].[1933] : [DIM
HOUSE].[CONSTRUCTIONTIME].[1960]},
```

```
[DIM HOUSE].[CONSTRUCTIONTIME].[CONSTRUCTIONTIME]
```

```
) ON ROWS
```

```
FROM [HOUSE PRICE BEIJING 1]
```

Kết quả:

	AVERAGE OF TOTAL PRICE	AVERAGE OF SQUARE
1933	1643000	56.7
1934	2092500	15.305
1944	14066500	132.9
1950	3468181.81818182	66.7345454545455
1952	4357600	72.48
1953	4356846.15384615	66.8853846153846
1954	5044444.44444444	68.58111111111111
1955	4628870.96774194	76.3874193548387
1956	4393714.28571429	75.2950649350649
1957	3235712.12121212	63.6536363636364
1958	4082136.75213675	64.7309401709402
1959	3349160	67.9386
1960	4275461.07784431	70.9643113772455

Hình 281: Kết quả bằng MDX

6.13. Câu truy vấn 13

Nội dung câu truy vấn: Liệt kê số lượng và tổng số tiền giao dịch đạt chuẩn chất lượng cao qua từng quý của các năm Được biết giao dịch đạt chuẩn chất lượng cao chỉ bàn giao những ngôi nhà có tình trạng cải tạo đạt mức “Hardcover” (**RENOVATIONCONDITION**), được làm bởi các vật liệu có chất lượng cao kết hợp thép và bê tông (**BUILDINGSTRUCTURE**)

Thực hiện truy vấn trên Cube (manual):

Hình 282: Chọn thuộc tính và độ đo

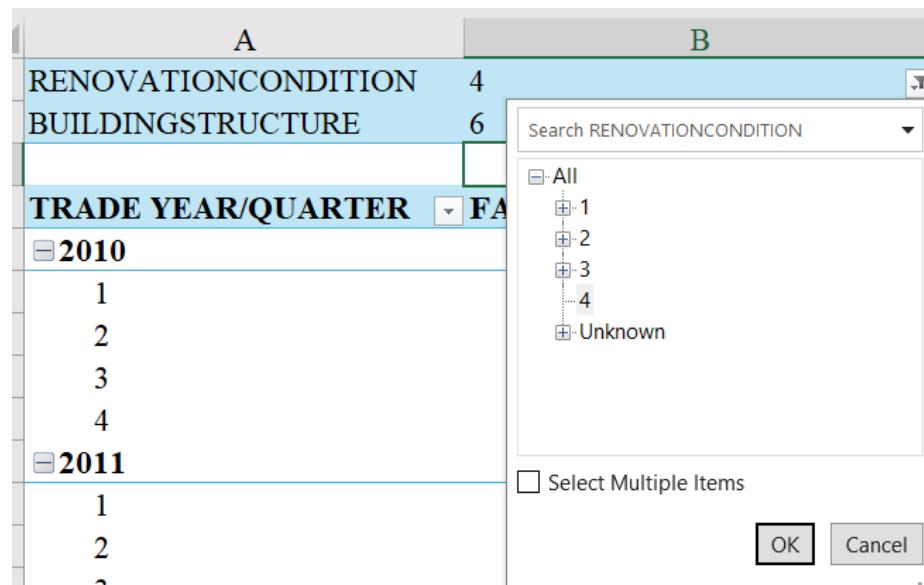
Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

Dimension		Hierarchy	Operator	Filter Expression
DIM HOUSE		RENOVATIONCONDITION	Equal	{ 4 }
DIM HOUSE		BUILDINGSTRUCTURE	Equal	{ 6 }
<Select dimension>				
TRADE YEAR	TRADE QUARTER	FACT TRANSACTION COUNT	SUM OF TOTAL PRICE	
2010	1	10	11935000	
2010	2	8	11210000	
2010	3	12	14759000	
2010	4	17	31302000	
2011	1	6	16325000	
2011	2	4	12460000	
2011	3	3	3826000	
2011	4	6	24225000	
2012	1	7	11835000	
2012	2	5	6920000	
2012	3	8	16680000	
2012	4	12	18590000	
2013	1	12	22318000	
2013	2	6	13318000	
2013	3	6	15805000	
2013	4	7	20017000	
2014	1	8	16358000	
2014	2	1168	3974467000	
2014	3	2485	7978404000	
2014	4	4319	14809281000	

Hình 283: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Filter RENOVATIONCONDITION, BUILDINGSTRUCTURE chọn tương ứng 4 và 6



Hình 284: Chọn Filter

Kết quả:

A	B	C
RENOVATIONCONDITION	4	
BUILDINGSTRUCTURE	6	
TRADE YEAR/QUARTER	FACT TRANSACTION Count	SUM OF TOTAL PRICE
2010		
1	10	11935000
2	8	11210000
3	12	14759000
4	17	31302000
2011		
1	6	16325000
2	4	12460000
3	3	3826000
4	6	24225000
2012		
1	7	11835000
2	5	6920000
3	8	16680000
4	12	18590000
2013		
1	12	22318000
2	6	13318000
3	6	15805000
4	7	20017000
2014		

Hình 285: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT {[Measures].[FACT TRANSACTION Count], [Measures].[SUM OF
TOTAL PRICE]} ON COLUMNS,
NON EMPTY {FILTER(DrillDownLevel(DrillDownLevel([DIM DATE].[TRANS
TIME])),  

NOT [DIM DATE].[TRANS TIME].CURRENTMEMBER IS [DIM
DATE].[TRANS TIME].[All])} ON ROWS
FROM [HOUSE PRICE BEIJING 1]
WHERE (
    {[DIM HOUSE].[BUILDINGSTRUCTURE].&[6]},  

    {[DIM HOUSE].[RENOVATIONCONDITION].&[4]})  

)

```

Kết quả:

	FACT TRANSACTION Count	SUM OF TOTAL PRICE
2010	47	69206000
1	10	11935000
2	8	11210000
3	12	14759000
4	17	31302000
2011	19	56836000
1	6	16325000
2	4	12460000
3	3	3826000
4	6	24225000
2012	32	54025000
1	7	11835000
2	5	6920000
3	8	16680000
4	12	18590000
2013	31	71458000
1	12	22318000
2	6	13318000
3	6	15805000
4	7	20017000

Hình 286: Kết quả bằng MDX

6.14. Câu truy vấn 14

Nội dung câu truy vấn: Cho biết số giao dịch, trung bình m², của những nhà có thang máy (**ELEVATOR**) giao dịch trong năm 2018, xếp theo thứ tự trung bình m² giảm dần.

Thực hiện truy vấn trên Cube (manual):

The screenshot shows a dimension selection grid at the top with columns for Dimension, Hierarchy, Operator, and Filter Expression. It includes filters for DIM DATE (TRANS TIME, Equal, { 2018 }) and DIM HOUSE (ELEVATOR, Equal, { True }). Below the grid is a table with columns TRADE YEAR, DISTRICT, FACT TRANSACTION Count, and AVERAGE OF SQUARE. A button labeled "Click to execute the query." is visible at the bottom right.

Hình 287: Chọn thuộc tính và độ đo

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

TRADE YEAR	DISTRICT	FACT TRANSACTION Count	AVERAGE OF SQUARE
2018	7	42	3696.21
2018	2	16	1253
2018	11	12	1148.49
2018	8	12	1036.43
2018	6	11	943.11
2018	1	4	614.29
2018	10	6	376.42
2018	4	4	372.58
2018	5	3	221.34
2018	9	2	130.3
2018	13	1	75.26
2018	3	1	53.06

Hình 288: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Kéo các trường dữ liệu vào tương ứng, filter ELEVATOR bằng 1 và kích chuột phải vào giá trị bất kỳ của cột AVERAGE OF SQUARE và chọn Sort -> Sort Largest to Smallest

A	B	C	D	E	F
ELEVATOR	True				
DISTRICT	FACT TRANSACTION Count	AVERAGE OF SQUARE			
2018					
1	4	614.29			
2	16	1253			
3	1	53.06			
4	4	372.58			
5	3	221.34			
6	11	943.11			
7	42	3696.21			
8	12	1036.43			
9	2	130.3			
10	6	376.42			
11	12	1148.49			
13	1	75.26			
Grand Total	114	9920.49			

Hình 289: Chọn trường dữ liệu và filter

Kết quả:

A	B	C
ELEVATOR	True	
DISTRICT	FACT TRANSACTION Count	AVERAGE OF SQUARE
2018		
7	42	3696.21
2	16	1253
11	12	1148.49
8	12	1036.43
6	11	943.11
1	4	614.29
10	6	376.42
4	4	372.58
5	3	221.34
9	2	130.3
13	1	75.26
3	1	53.06
Grand Total	114	9920.49

Hình 290: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT NON EMPTY {
    [Measures].[FACT TRANSACTION Count],
    [Measures].[AVERAGE OF SQUARE]
} ON COLUMNS,
NON EMPTY {
    ORDER(
        [DIM DATE].[TRADE YEAR].[TRADE YEAR].ALLMEMBERS * [DIM
HOUSE].[DISTRICT].[DISTRICT].ALLMEMBERS,
        [Measures].[AVERAGE OF SQUARE],
        BDESC
    )
} ON ROWS
FROM [HOUSE PRICE BEIJING 1]
WHERE (
    [DIM DATE].[TRANS TIME].[TRADE YEAR].&[2018],
    [DIM HOUSE].[ELEVATOR].&[True]
)

```

Kết quả:

		FACT TRANSACTION Count	AVERAGE OF SQUARE
2018	7	42	3696.21
2018	2	16	1253
2018	11	12	1148.49
2018	8	12	1036.43
2018	6	11	943.11
2018	1	4	614.29
2018	10	6	376.42
2018	4	4	372.58
2018	5	3	221.34
2018	9	2	130.3
2018	13	1	75.26
2018	3	1	53.06

Hình 291: Kết quả bằng MDX

6.15. Câu truy vấn 15

Nội dung câu truy vấn: Thống kê cho biết giá trên m² (**PRICE**) lớn nhất qua các năm của những nhà có thể loại công trình (**BUILDINGTYPE**) là nhà cấp 4 và có

điều kiện cài tạo (**RENOVATIONCONDITION**) không phải là đơn giản, xếp theo thứ tự giá trên một m² (**PRICE**) **tăng dần**.

Thực hiện truy vấn trên Cube (manual):

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE	BUILDINGTYPE	Equal	{ 4 }
DIM HOUSE	RENOVATIONCON...	Not Equal	{ 3 }
<Select dimension>			

TRADE YEAR	MAX OF PRICE
------------	--------------

[Click to execute the query.](#)

Hình 292: Chọn các thuộc tính và độ đo

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

TRADE YEAR	MAX OF PRICE
2002	17053
2010	23729
2009	26964
2011	79768
2018	123047
2014	131927
2012	141485
2013	149276
2015	149276
2016	149742
2017	149967

Hình 293: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Kéo các trường dữ liệu vào tương ứng, filter **BUILDINGTYPE** bằng 4, filter **RENOVATIONCONDITION** bằng giá trị khác 3 và kích chuột phải vào giá trị bất kỳ của cột **MAX OF PRICE** và chọn Sort -> Sort Smallest to Largest

YEAR	MAX OF PRICE
2002	17053
2009	26964
2010	23729
2011	79768
2012	141485
2013	149276
2014	131927
2015	149276
2016	149742
2017	149967
2018	123047
Grand Total	149967

Hình 294: Chọn các trường dữ liệu và Filter

Kết quả:

A	B
BUILDINGTYPE	4
RENOVATIONCONDITION	(Multiple Items)
YEAR	MAX OF PRICE
2002	17053
2010	23729
2009	26964
2011	79768
2018	123047
2014	131927
2012	141485
2013	149276
2015	149276
2016	149742
2017	149967
Grand Total	149967

Hình 295: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT
    NON EMPTY { [Measures].[MAX OF PRICE] } ON COLUMNS,
    NON EMPTY
    ORDER(
        {
            [DIM DATE].[TRANS TIME].[TRADE YEAR].ALLMEMBERS
        },
        [Measures].[MAX OF PRICE],
        ASC
    )
    ON ROWS
FROM
    (
        SELECT
            -{ [DIM HOUSE].[RENOVATIONCONDITION].&[3] }
        ON COLUMNS
        FROM
            (
                SELECT

```

```

{ [DIM HOUSE].[BUILDINGTYPE].&[4] }
ON COLUMNS
FROM
[HOUSE PRICE BEIJING 1]
)
)
WHERE
[DIM HOUSE].[BUILDINGTYPE].&[4]

```

Kết quả:

	MAX OF PRICE
2002	17053
2010	23729
2009	26964
2011	79768
2018	123047
2014	131927
2012	141485
2013	149276
2015	149276
2016	149742
2017	149967

Hình 296: Kết quả bằng MDX

6.16. Câu truy vấn 16

Nội dung câu truy vấn: Cho biết thống kê theo từng năm ứng với mỗi huyện (**DISTRICT**) có bao nhiêu đường link giao dịch (**URL**) có kết cấu xây dựng (**BUILDINGSTRUCTURE**) là thép

Thực hiện truy vấn trên Cube (manual):

Dimension	Hierarchy	Operator	Filter Expression
DIM HOUSE <Select dimension>	BUILDINGSTRUCTURE	Equal	{ 5 }
TRADE YEAR	DISTRICT	URL Count	

[Click to execute the query.](#)

Hình 297: Chọn các thuộc tính và độ đo

Nhấn “Click to execute the query” để thực thi câu lệnh và được kết quả như sau:

TRADE YEAR	DISTRICT	URL Count
2011	6	1
2011	9	1
2012	1	4
2012	4	1
2012	6	2
2012	7	4
2012	9	1
2012	10	3
2012	11	1
2012	13	1
2013	4	5
2013	6	3
2013	7	6
2013	9	1
2013	11	3
2013	13	2

Hình 298: Kết quả bằng Cube

Thực hiện truy vấn trên Pivot table trong Excel:

Kéo các trường dữ liệu vào tương ứng, filter **BUILDINGSTRUCTURE** bằng 5

A	B
BUILDINGSTRUCTURE 5	
YEAR/DISTRICT	URL Count
2011	

Hình 299: Chọn các trường dữ liệu và Filter

Kết quả:

A	B
BUILDINGSTRUCTURE 5	
YEAR/DISTRICT	URL Count
2011	
6	1
9	1
2012	
1	4
4	1
6	2
7	4
9	1
10	3
11	1
13	1
2013	
4	5
6	3
7	6
9	1
11	3

Hình 300: Kết quả bằng Excel

Thực hiện truy vấn bằng ngôn ngữ MDX:

```

SELECT
NON EMPTY { [Measures].[URL Count] } ON COLUMNS,
NON EMPTY {
    [DIM DATE].[TRANS TIME].[TRADE YEAR].ALLMEMBERS *
    [DIM HOUSE].[DISTRICT].[DISTRICT].ALLMEMBERS
} ON ROWS
FROM
(SELECT
    { [DIM HOUSE].[BUILDINGSTRUCTURE].&[5] } ON COLUMNS
FROM
    [HOUSE PRICE BEIJING 1]
)
WHERE
( [DIM HOUSE].[BUILDINGSTRUCTURE].&[5] )

```

Kết quả:

The screenshot shows a software interface for running MDX queries. At the top, there are tabs for 'Messages' and 'Results'. The 'Results' tab is active, displaying a table with two columns: 'URL Count' and 'Year'. The data is as follows:

	URL Count
2011	1
2011	1
2012	4
2012	1
2012	2
2012	4
2012	1
2012	3
2012	1
2012	1
2013	5
2013	3
2013	6

Hình 301: Kết quả bằng MDX

CHƯƠNG IV: QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)

1. Cài đặt các công cụ cần thiết

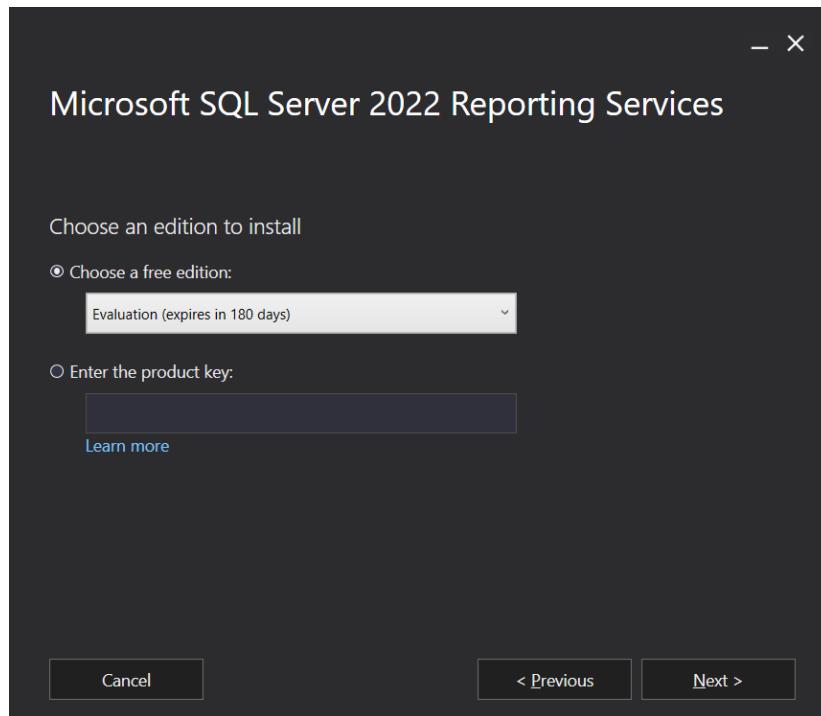
Những công cụ sau đây cần được cài đặt để thực hiện quá trình SSRS.

- **SQLReportingServices.exe:** Cài đặt Service SSRS để tạo server Deploy các báo cáo.
- **ReportBuilder.msi:** Report Builder phần mềm thiết kế Report tương tự với Report Designer trong Microsoft Visual Studio.
- Cài đặt công cụ **Microsoft SQL Server 2022 Reporting Services.**



Hình 302: Cài đặt công cụ Microsoft SQL Server 2022 Reporting Services

- Tại trang **Choose an edition to install**, chọn **choose a free edition** và tiếp tục nhấn **Developer** và nhấp **Next**.



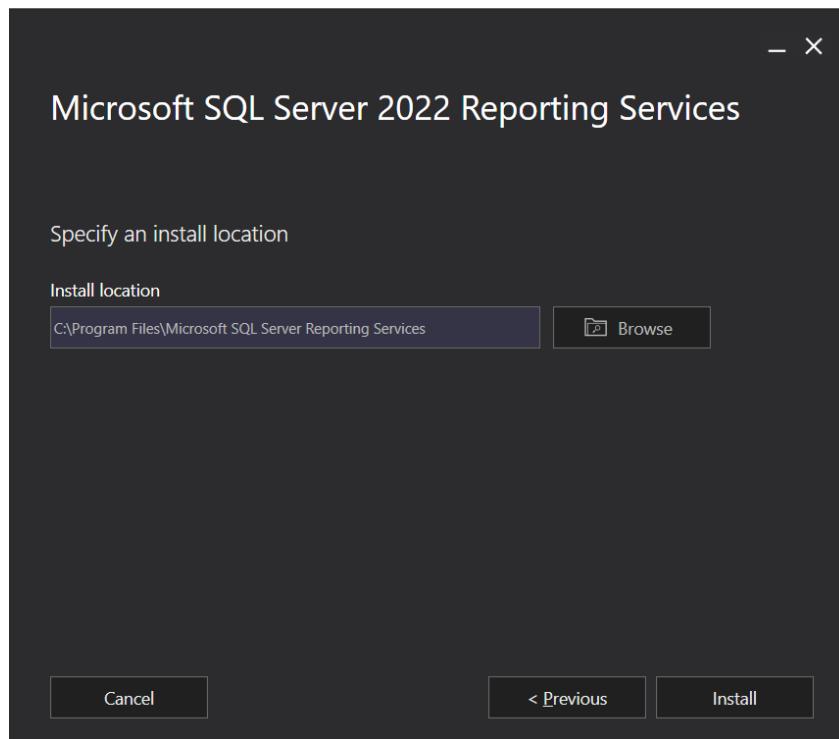
Hình 303: Chọn choose a free edition

- Tại trang giao diện **Review the license terms**, chọn **I accept the license terms** và nhấn **Next**:



Hình 304: Chọn I accept the license terms

- Nhấn Install:



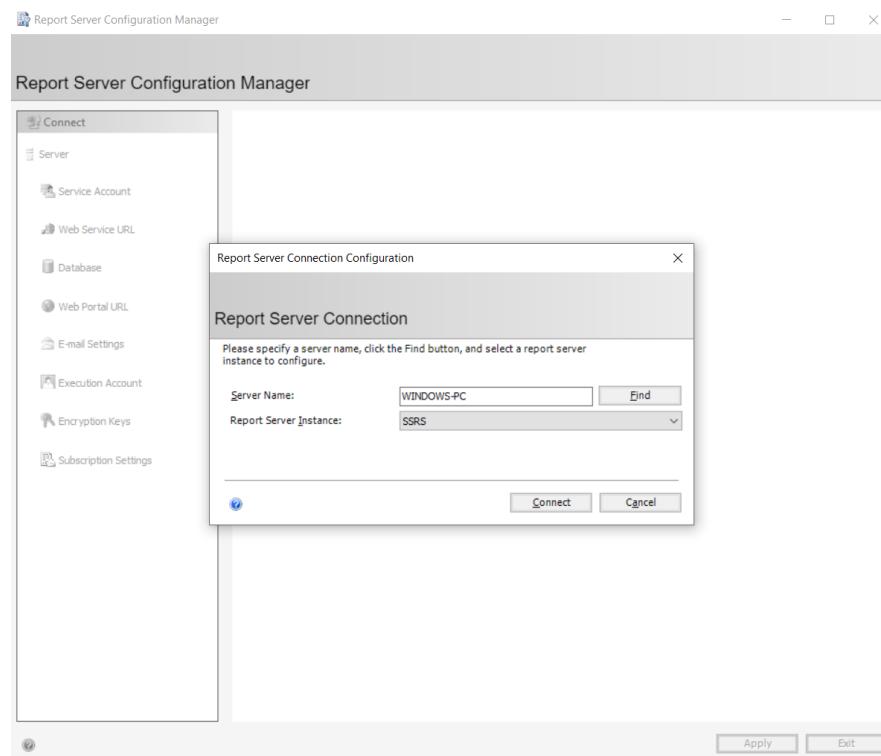
Hình 305: Nhấn Install

- Sau khi cài đặt thành công, click chọn **Configure Report Server** để mở giao diện Setup Completed.



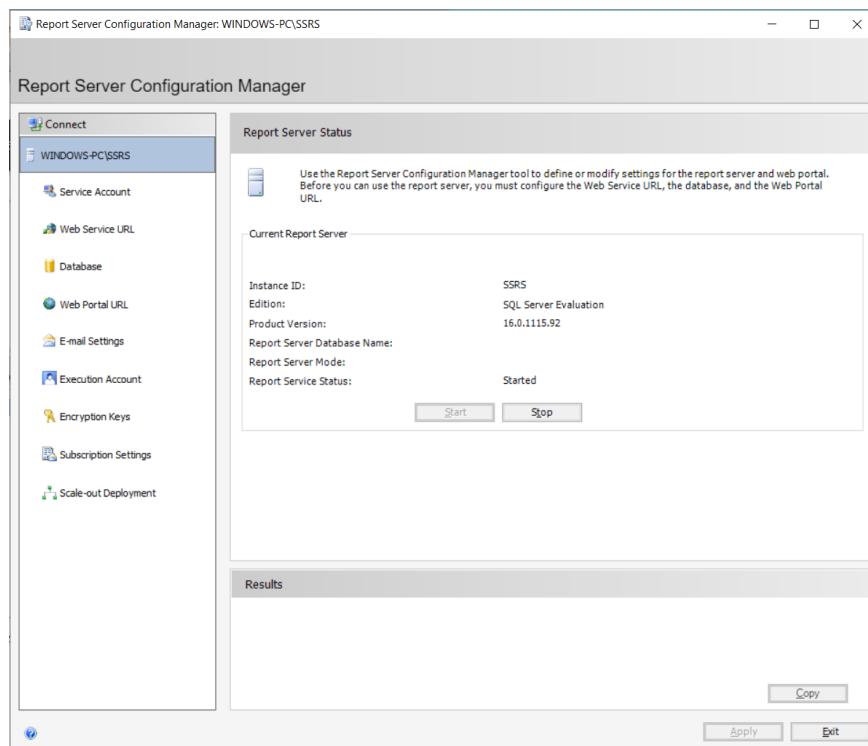
Hình 306: Chọn Configure Report Server

- Các bước cấu hình **Report Server Configuration Manager**.



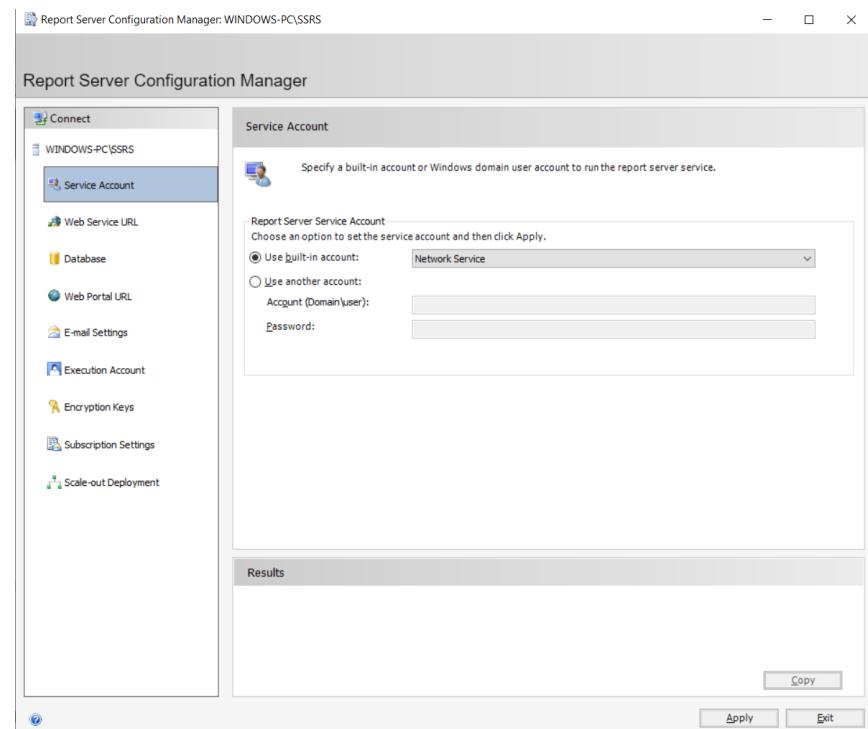
Hình 307: Cấu hình Report Server Configuration Manager

- Giao diện tổng thể của công cụ **Report Server Configuration Manager**



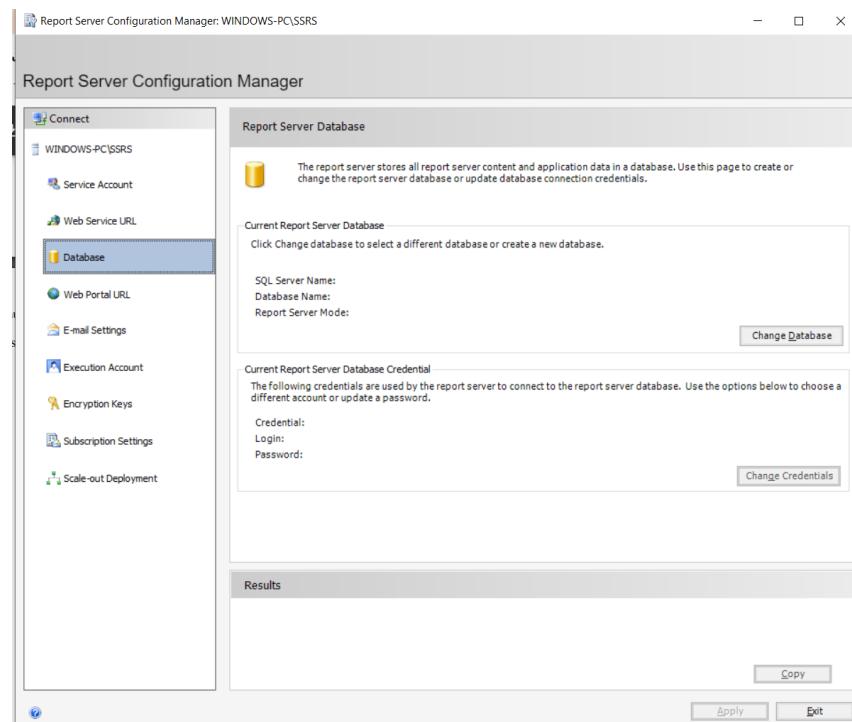
Hình 308: Giao diện tổng thể

- Tại tab **Service Account**, chuyển đổi giá trị trong **Use a built – in account**, chọn **Apply**, ta thu được kết quả như sau:



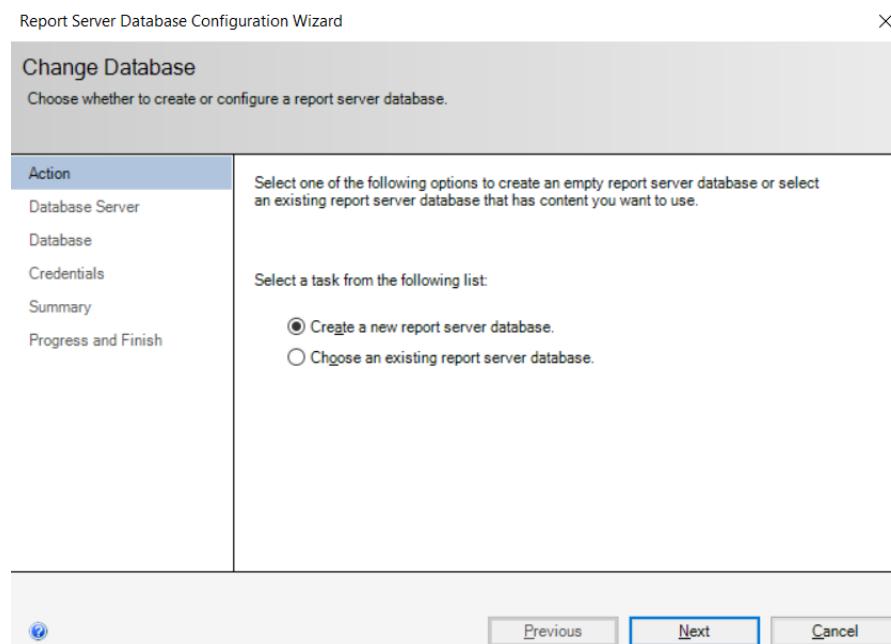
Hình 309: Thu được kết quả như sau

- Tại tab **Database**, tạo mới một dataset bằng cách chọn **Change Database**



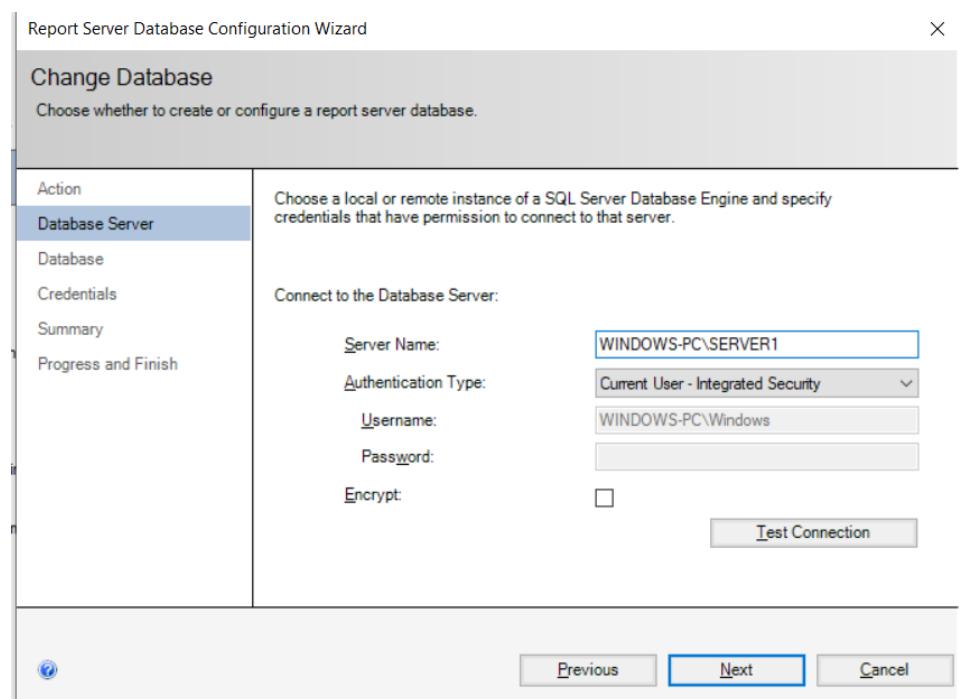
Hình 310: Chọn Change Database

- Chọn **Create a new report server database** và nhấp **Next**.



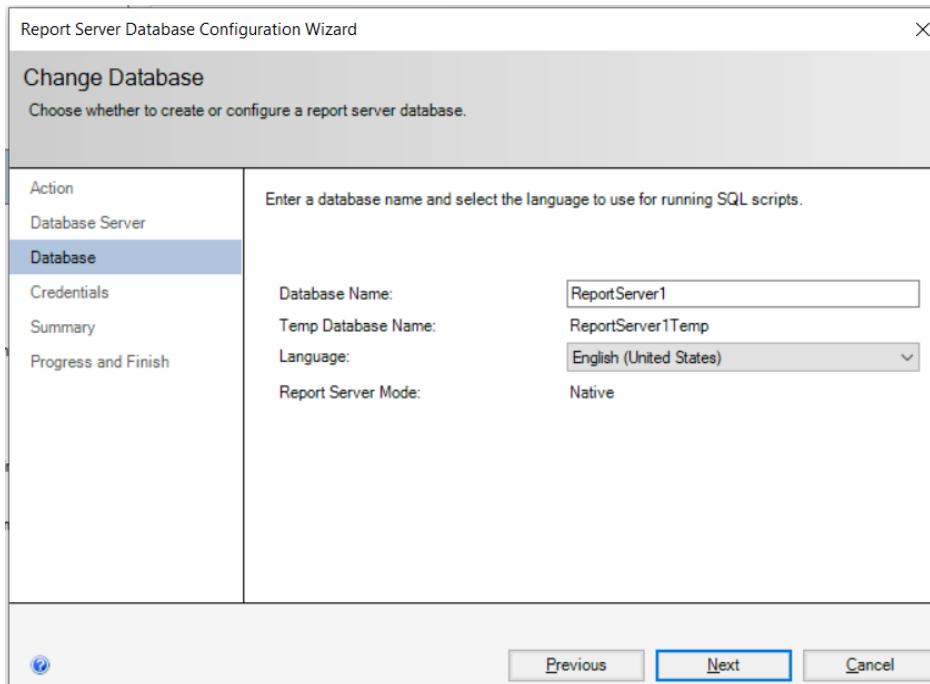
Hình 311: Chọn Create a new report server database

- Điền vào **Server Name** phù hợp và nhấp **Next**.



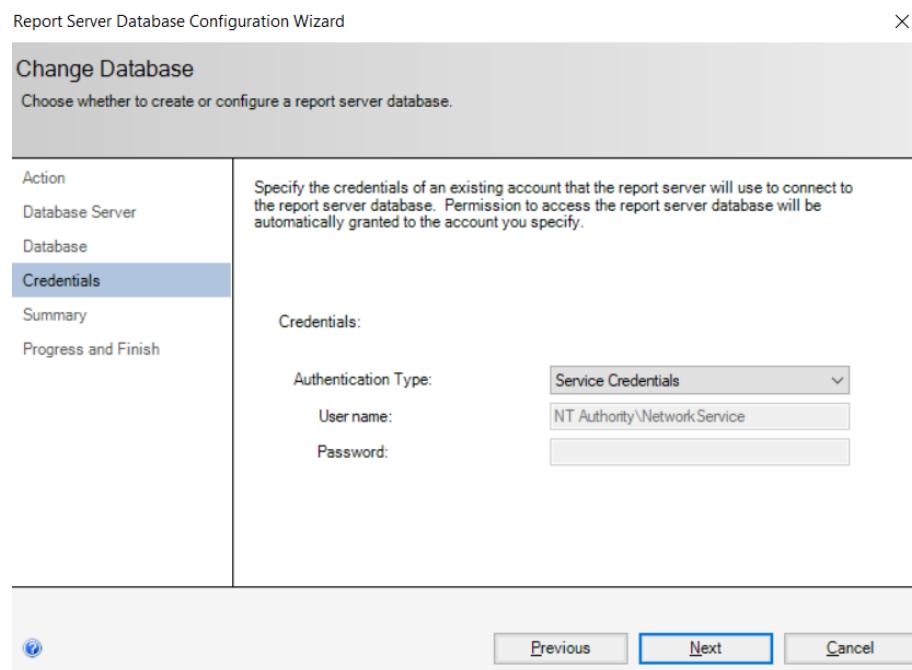
Hình 312: Điện Server Name phù hợp

- Đặt tên cho **Database** và nhấn **Next**



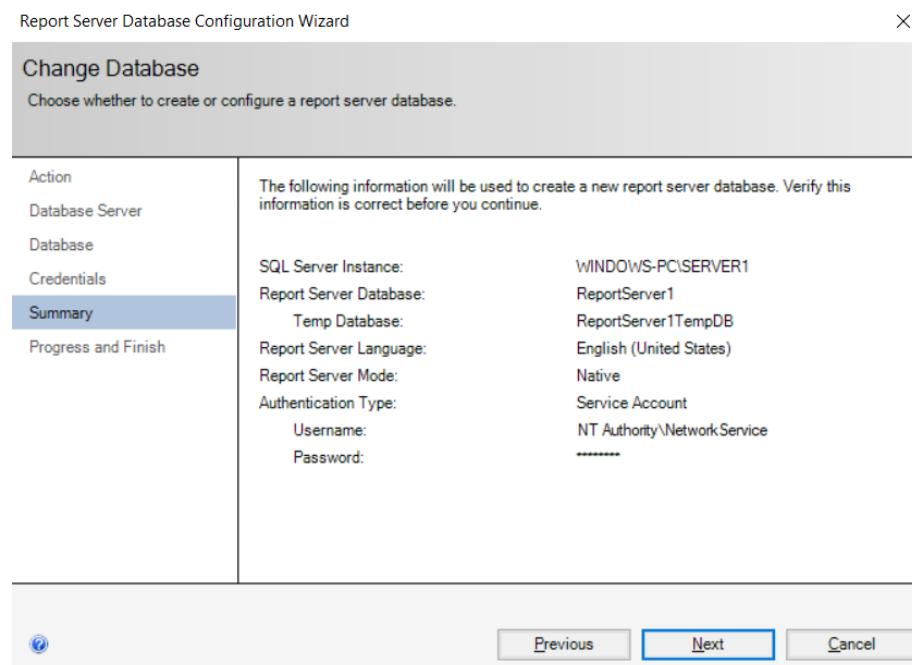
Hình 313: Đặt tên cho Database

- Trong tab **Credentials** tiếp tục nhấn **Next**.



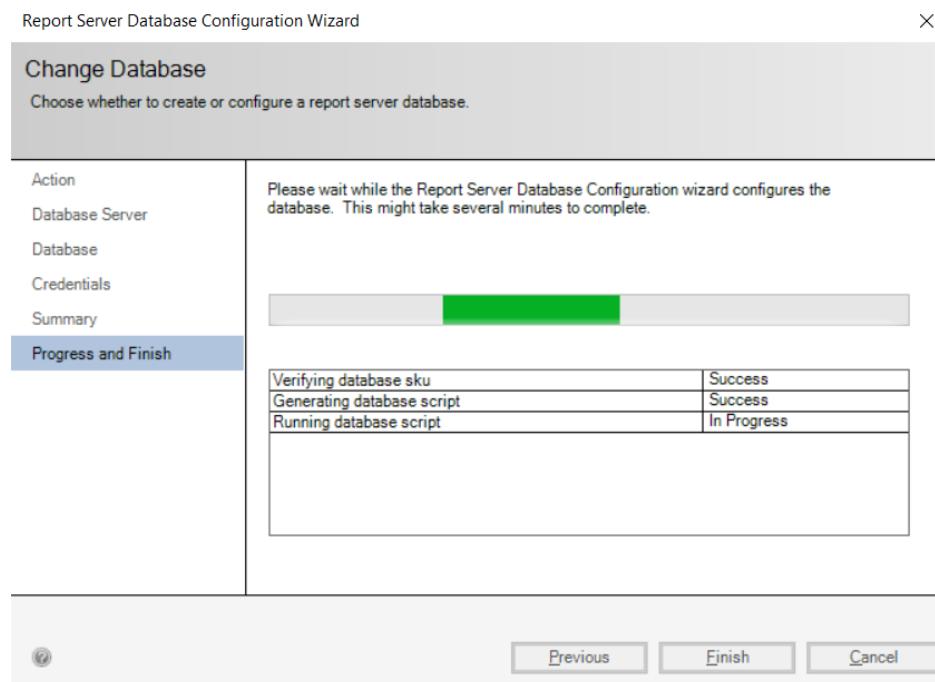
Hình 314: Tiếp tục nhấn Next

- Trong tab **Summary**, tiếp tục nhấn **Next**.



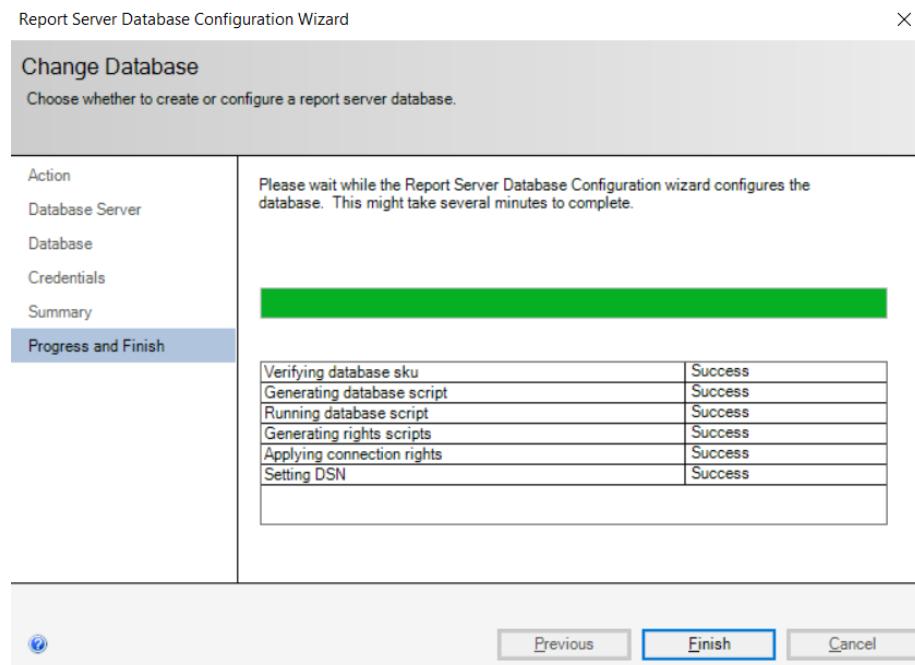
Hình 315: Tiếp tục

- Hệ thống tiến hành tạo **Database**



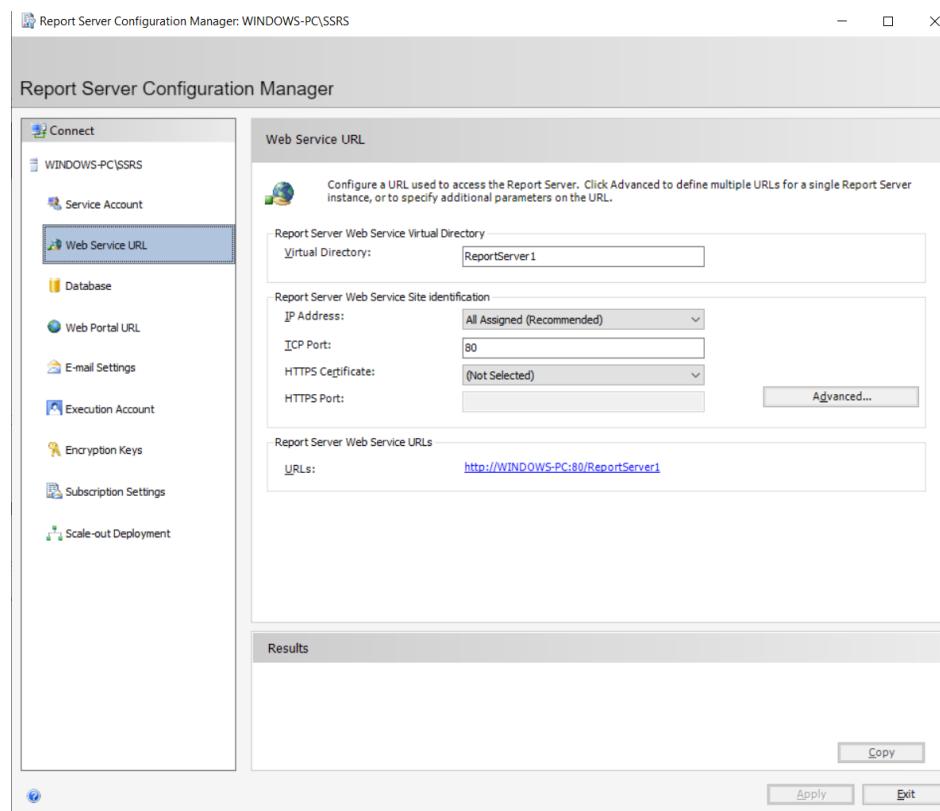
Hình 316: Đang tiến hành tạo Database

- Tạo database thành công.



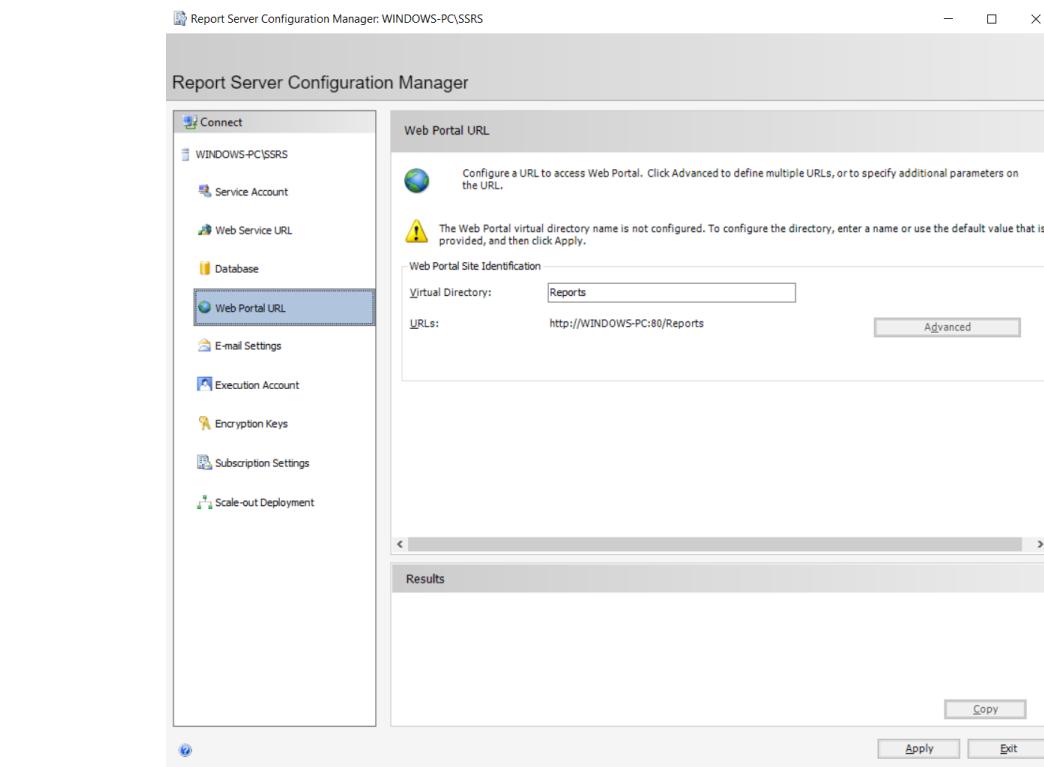
Hình 317: Tạo Database thành công

- Tại giao diện tab **Web Service URL**, sửa tên **Vitural Directory** thành tên **ReportServer1**

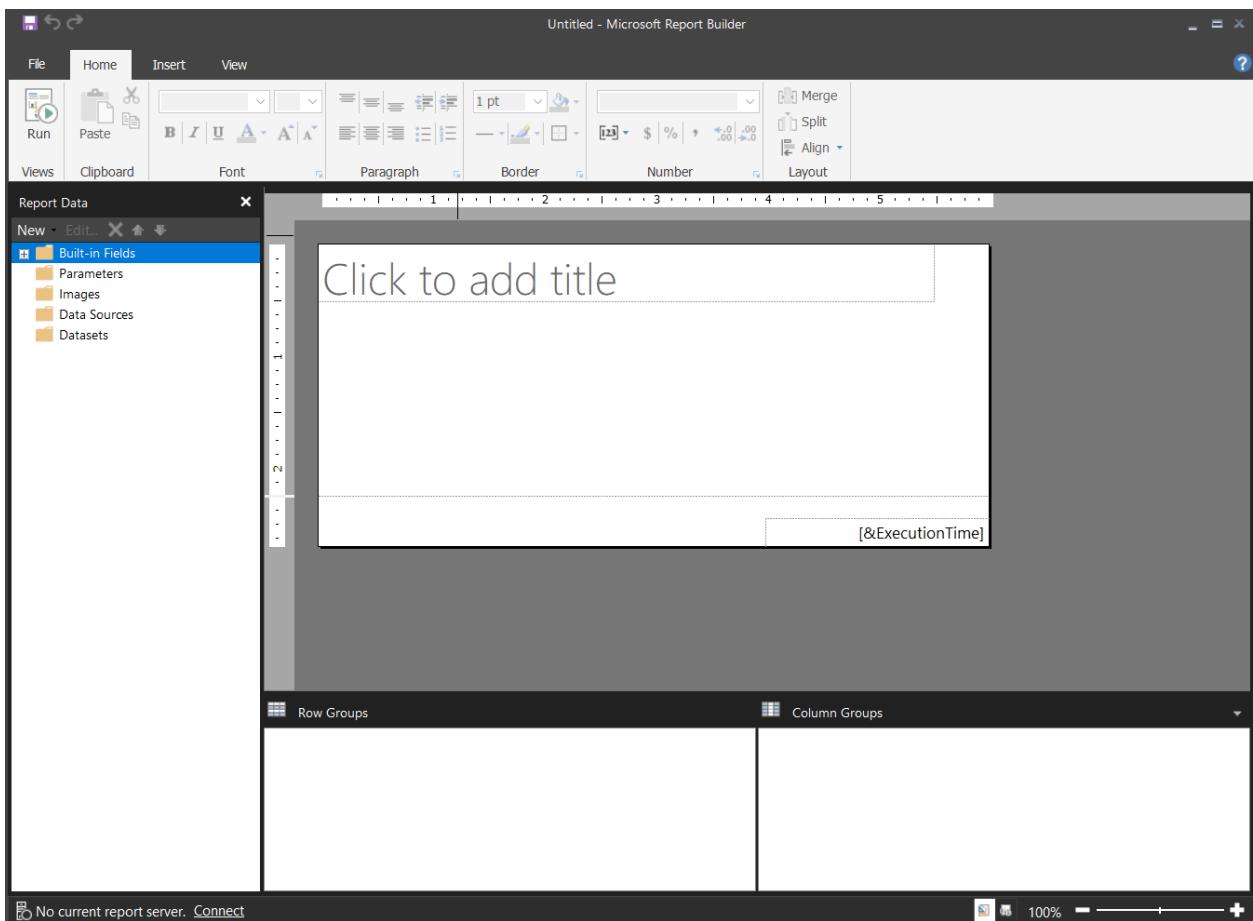


Hình 318: Sửa tên Vitural Directory thành tên ReportServer1

- Tại giao diện tab **Web Portal URL**, chọn **Apply** để xây dựng trang web



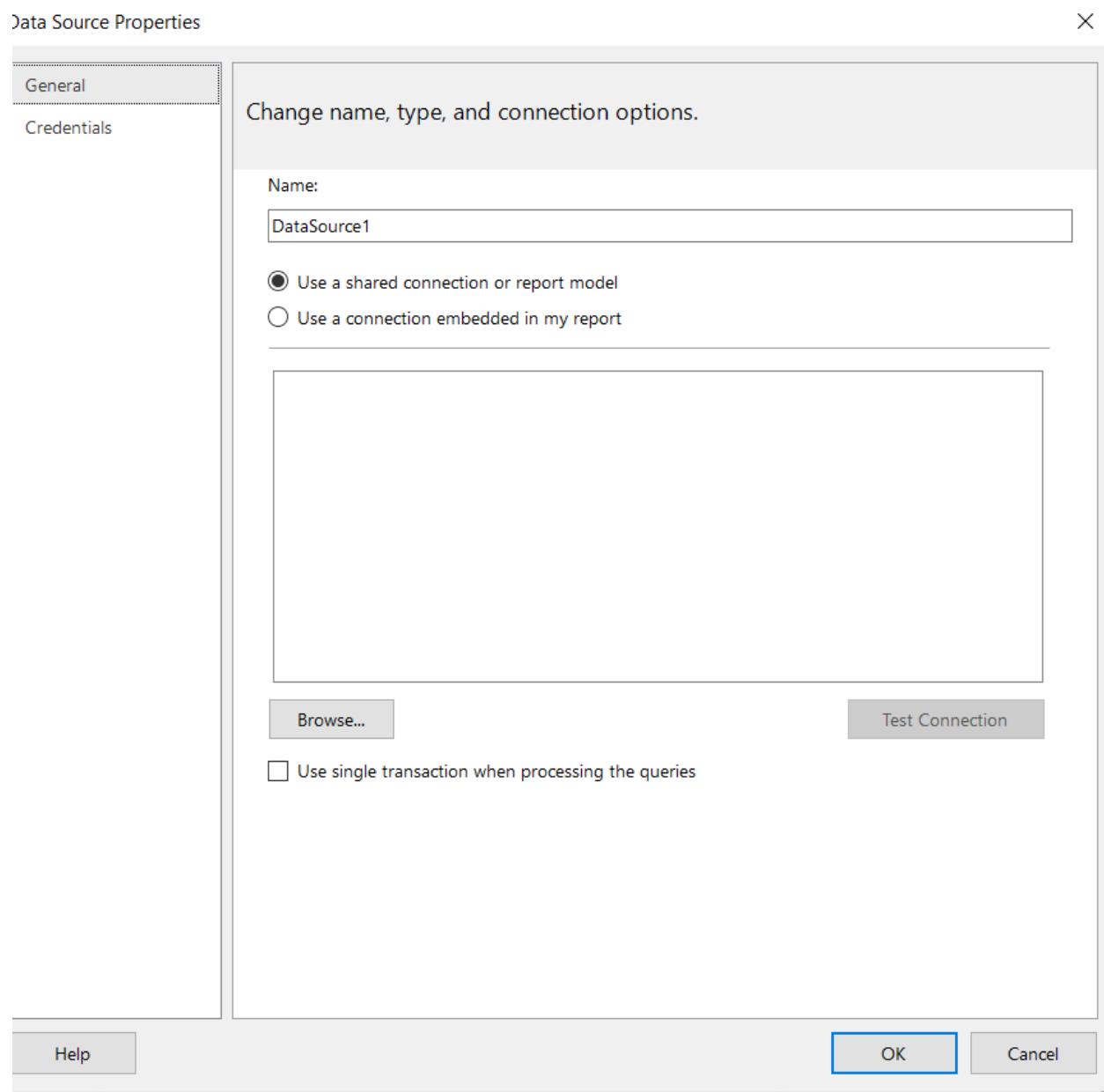
Hình 319: Chọn Apply



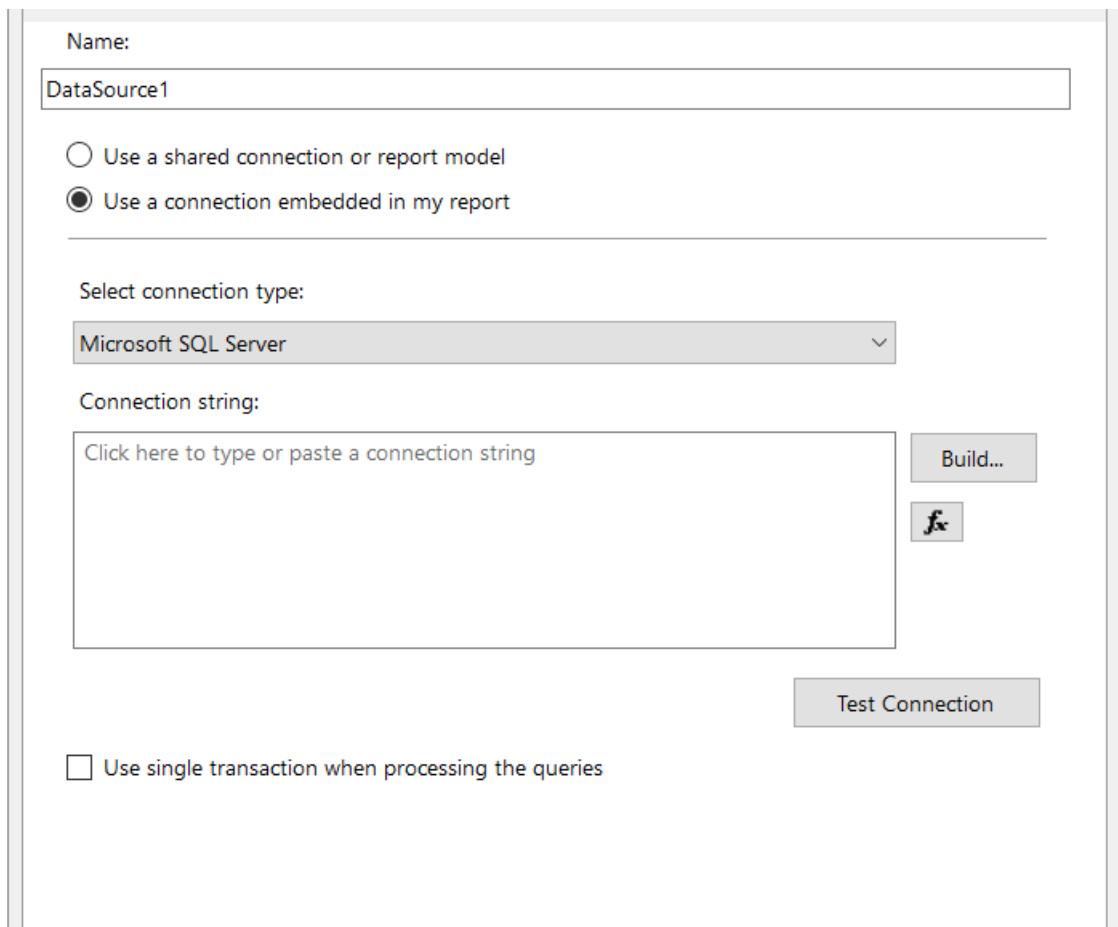
Hình 320: Giao diện



Hình 321: Add Data Source

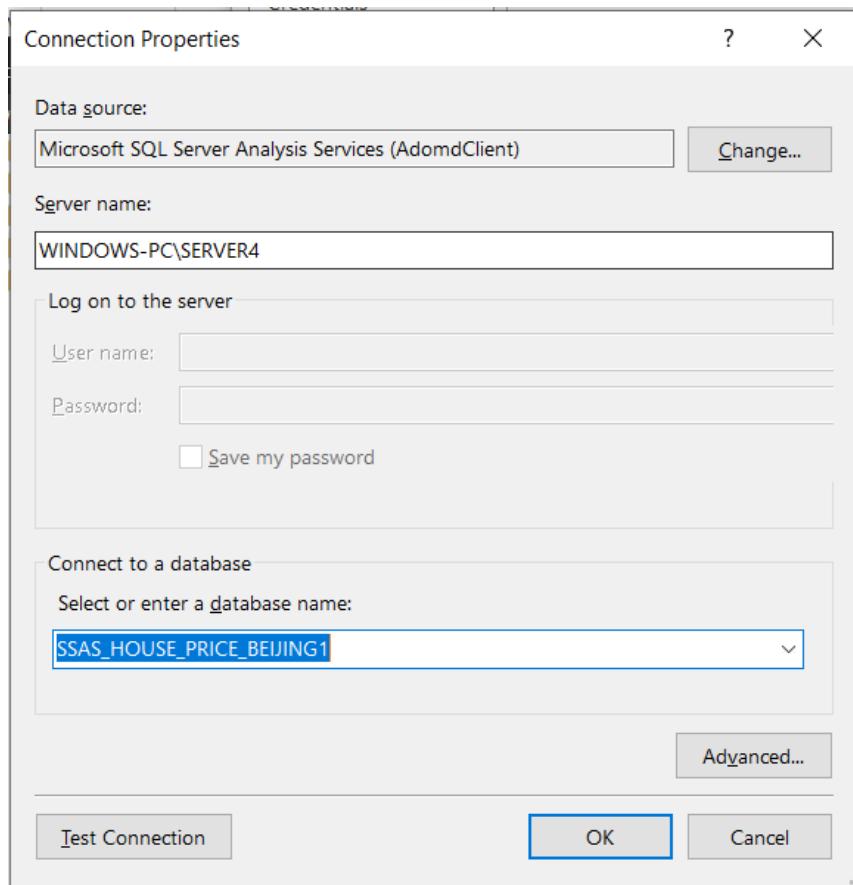


Hình 322: Giao diện Data Source Properties

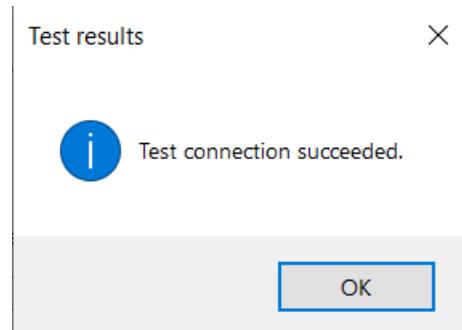


Hình 323: Chọn Use a connection embedded in my report

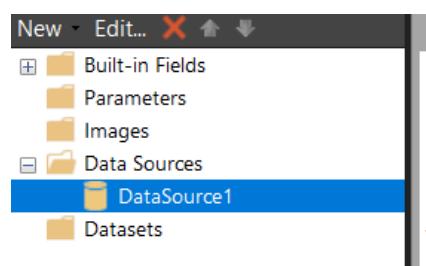
- Chọn **build**



Hình 324: Kết nối với Database



Hình 325: Kết nối thành công



Hình 326: Kiểm tra lại

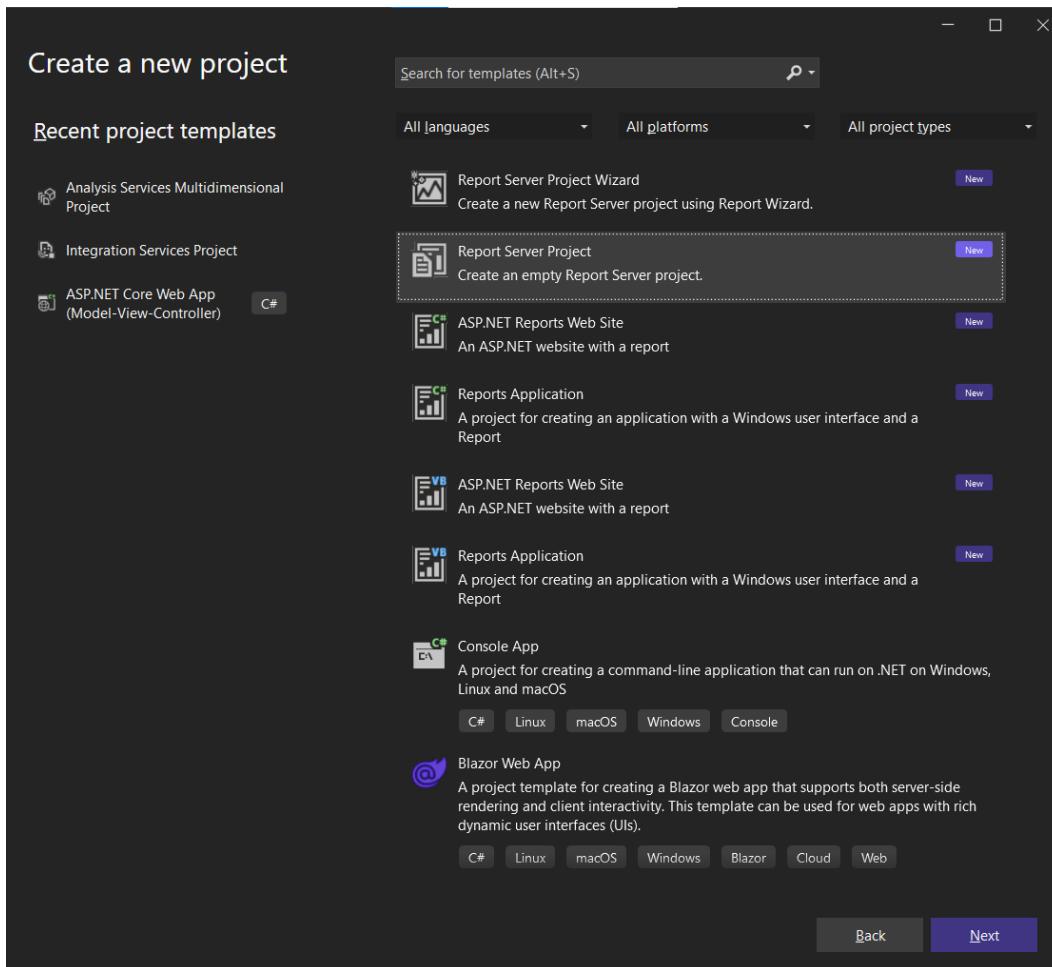
2. Report Designer (Reporting Services)

2.1 Tạo Report đơn giản

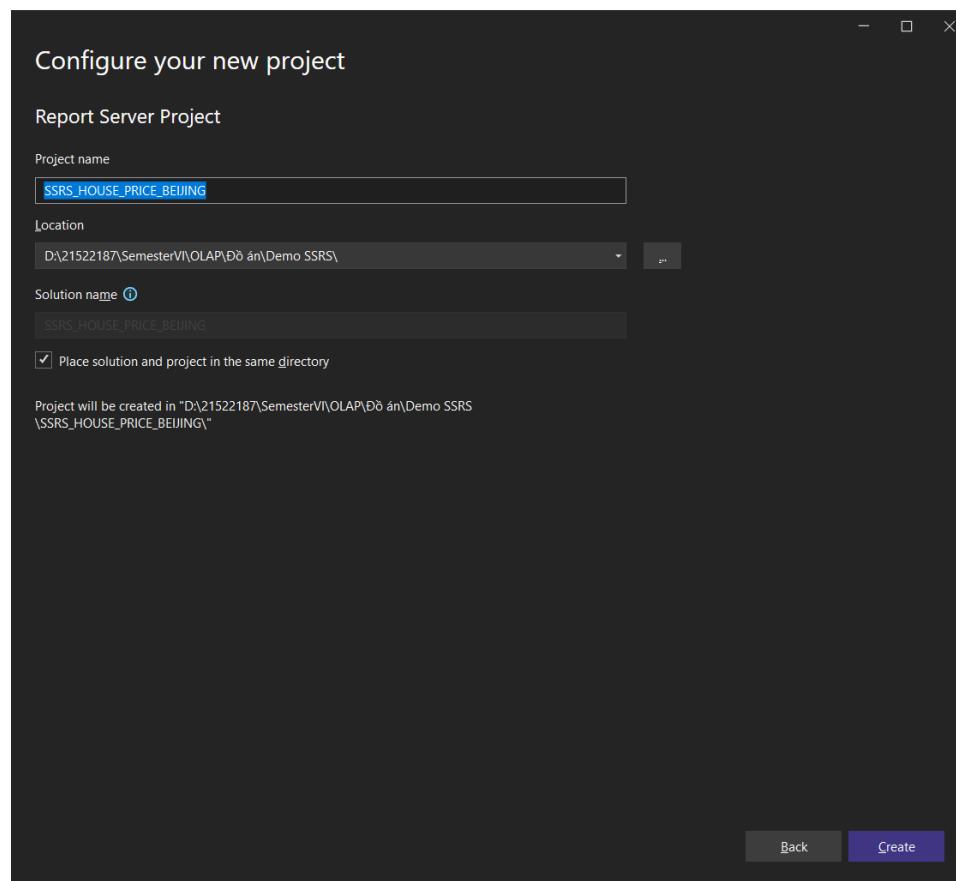
2.1.1 Mục đích, câu truy vấn 1

Tạo một báo cáo để in thông tin số lượng giao dịch của từng huyện có phòng ngũ bằng 2 trong 3 năm 2014, 2015, 2016

2.1.2 Tạo một Report Server Project mới

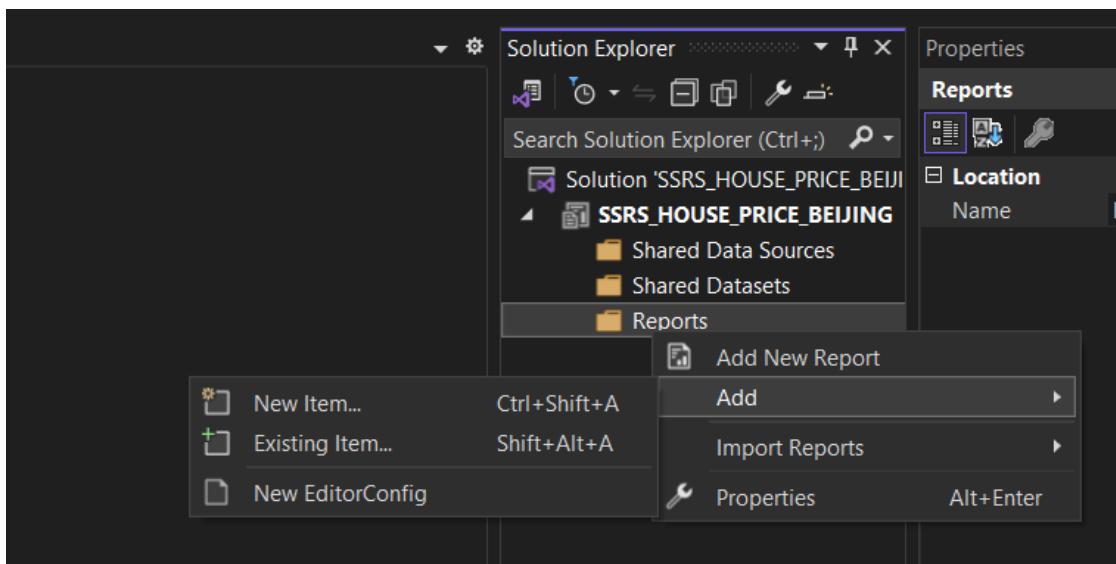


Hình 327: Tạo Report Server Project



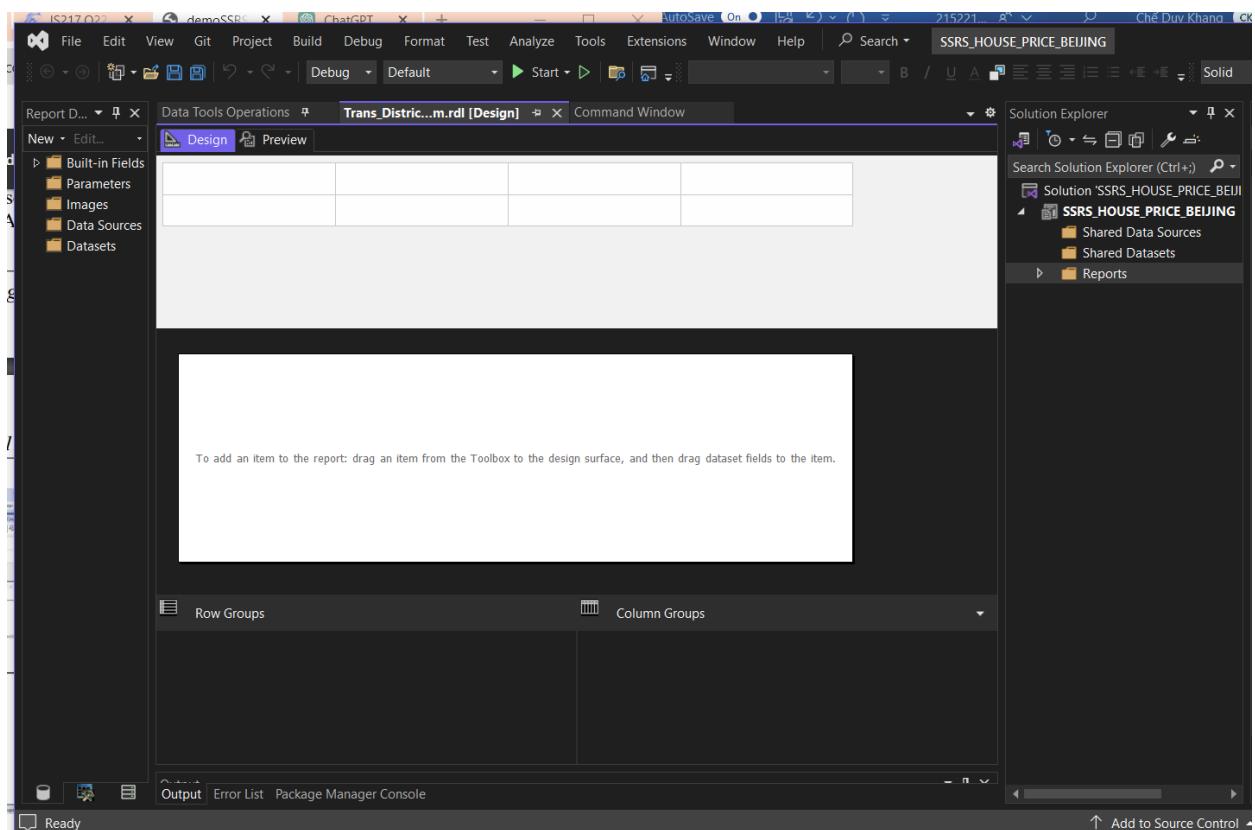
Hình 328: Đặt tên Project

- Để tạo một Report mới, tại SE, ở mục Report r-click chọn Add → New Item



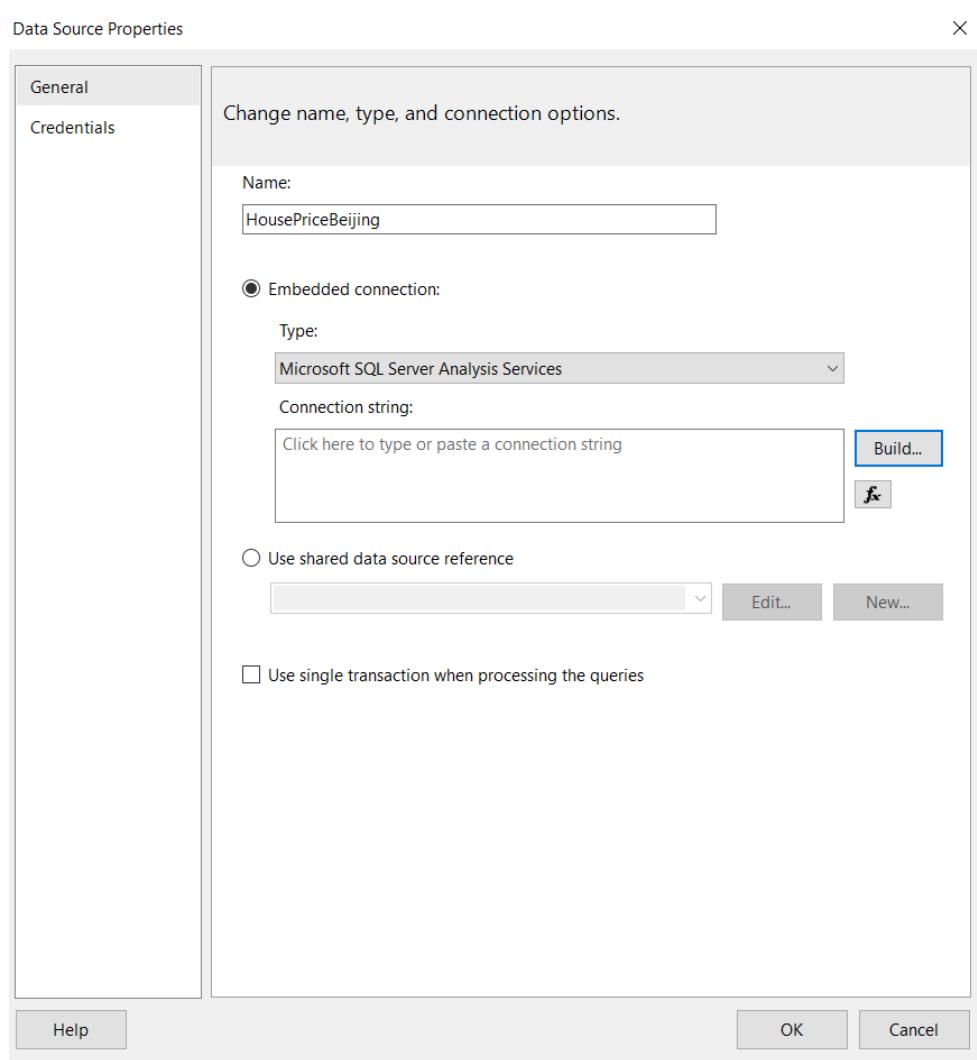
Hình 329: Chọn New Item

- Tại cửa sổ mới xuất hiện, chọn mục **Reports**. Chọn **Add**, một **Report** trống mới được tạo



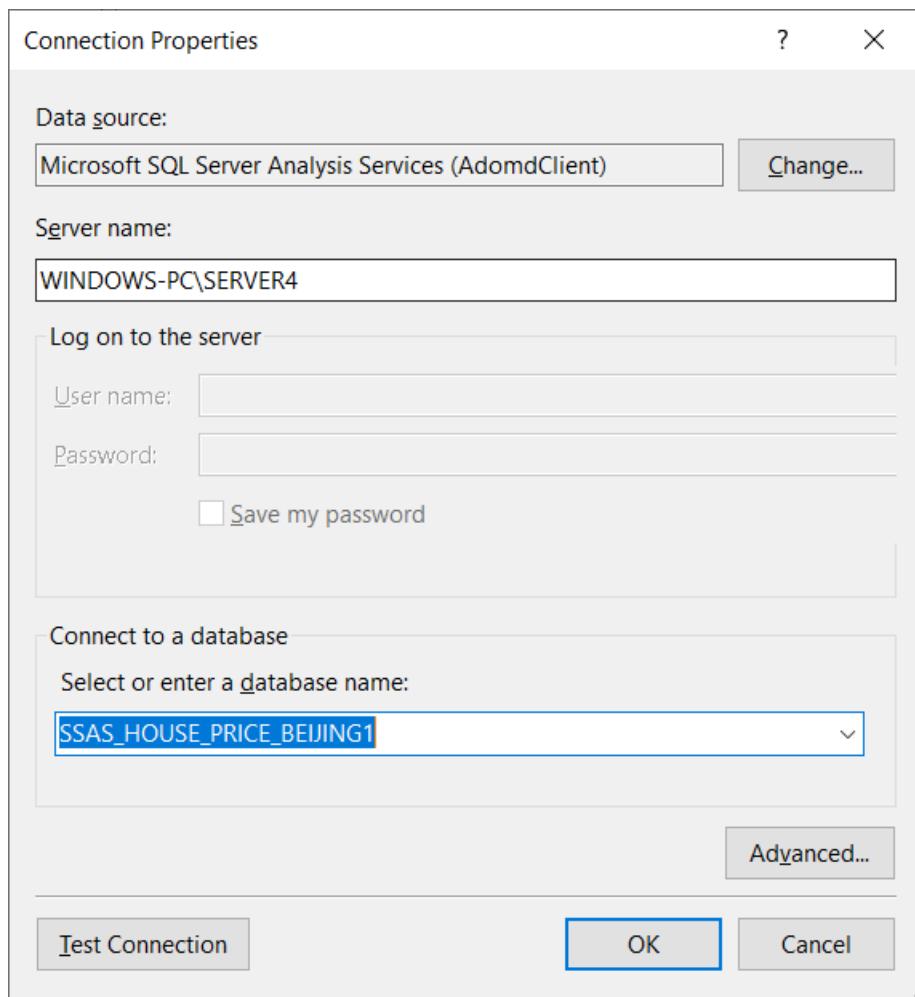
Hình 330: Report trống được tạo

- Kết nối với Cube dữ liệu SSAS
- Ở **Report Data pane**, r-click vào mục **Data Source**, chọn **Add Data Source**.
- Ở cửa sổ **Data Source Properties**, nhập **Hospital Data** vào mục **Name**. Chọn **Embedded connection** sau đó ở **Type** tìm **Microsoft SQL Server Analysis Services**. Tiếp tục chọn **Build** tại phần **Connection String**, cửa sổ mới xuất hiện.



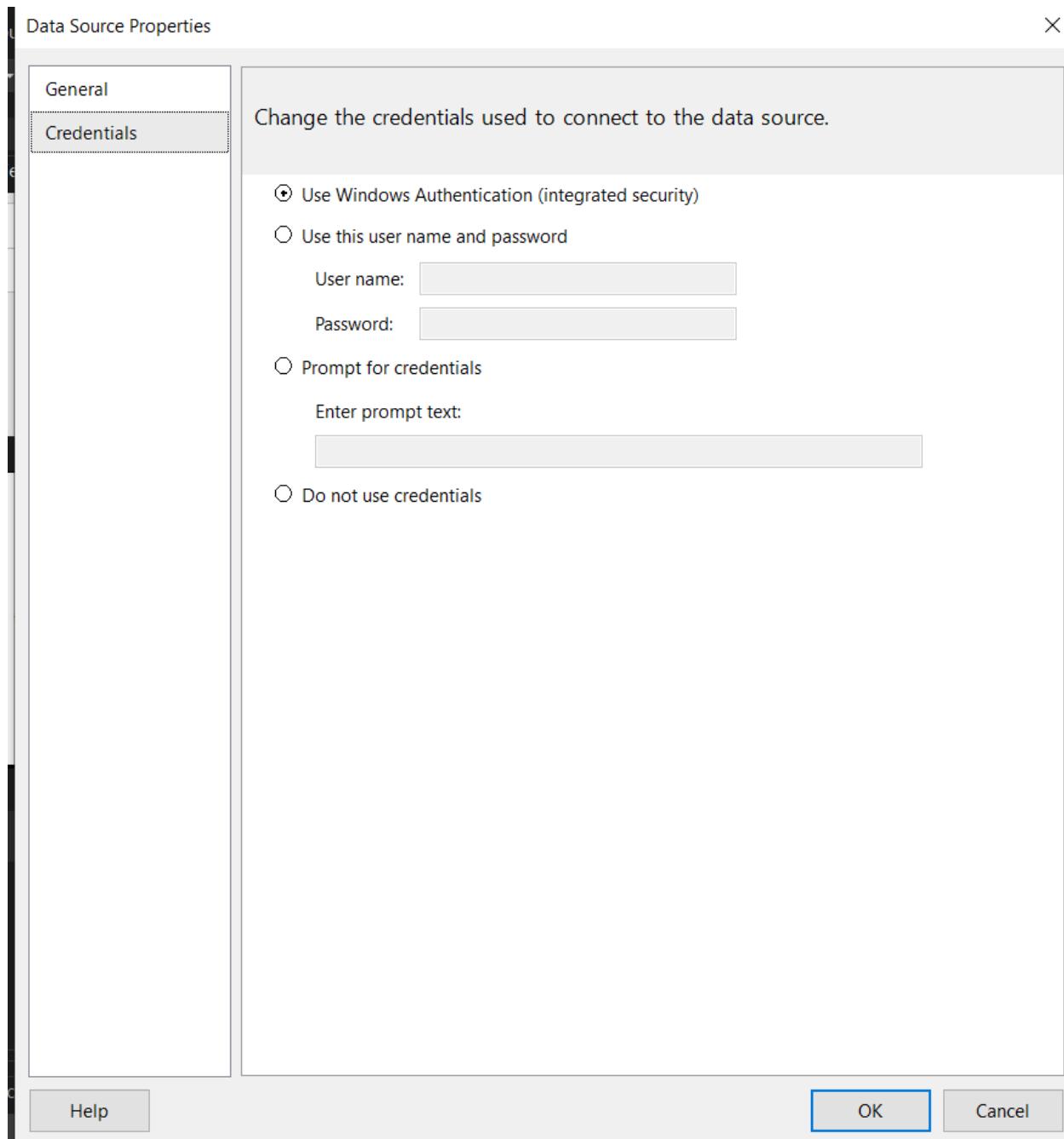
Hình 331: Kết nối với Cube dữ liệu SSAS

- Tại cửa sổ mới, nhập **localhost** vào **Server name**, sau đó tại mục **Connect to a database** nhập tên project SSAS đã hoàn thiện tại phần trước để sử dụng cho SSRS.



Hình 332: Nhập localhost vào Server name và nhập tên project SSAS

- Chọn **OK** để lưu Connection. Chuyển sang phần **Credentials**. Kiểm tra radio button **Use Window Authentication** đã được chọn hay chưa, sau đó chọn **OK**. **Data Source** mới được thêm vào mục **Data Source** tại **Report Data pane**.

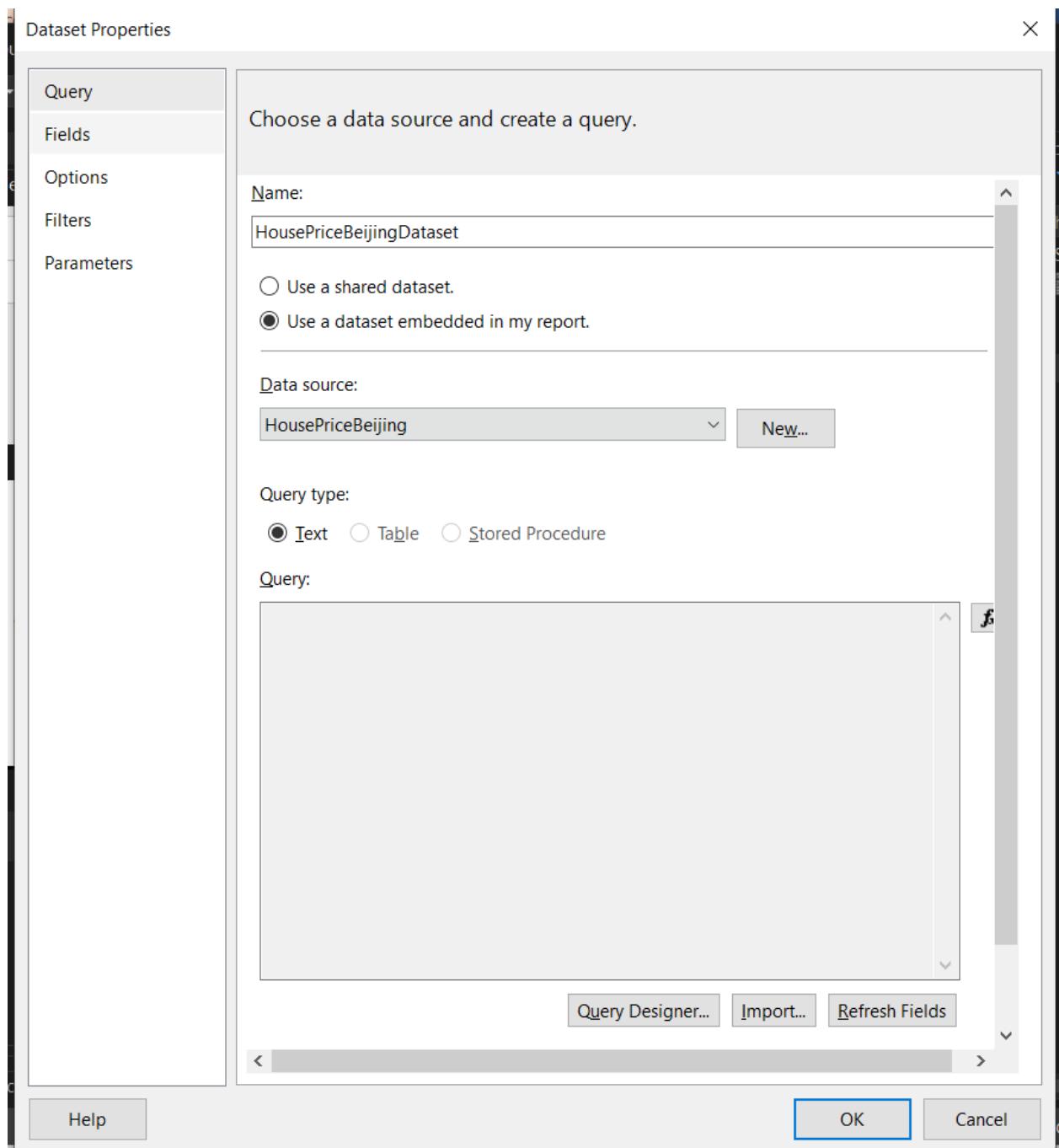


Hình 333: Chọn Use Window Authentication

2.1.3. Tạo Dataset cho Report

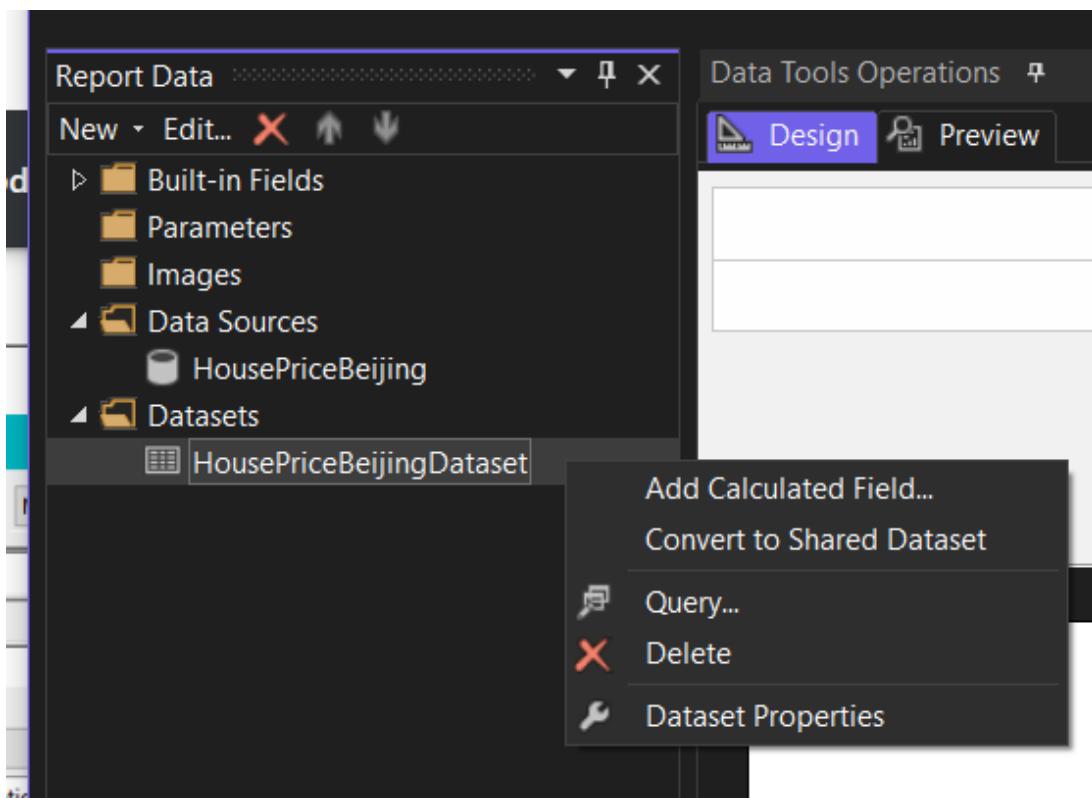
Ở mục **Report Data pane**, r-click vào **Datasets** và chọn **Add Dataset**.

Ở cửa sổ **Dataset Properties** nhập **HousePriceBeijingDataset** vào mục **Name**. Chọn **Use a dataset embedded in my report**.

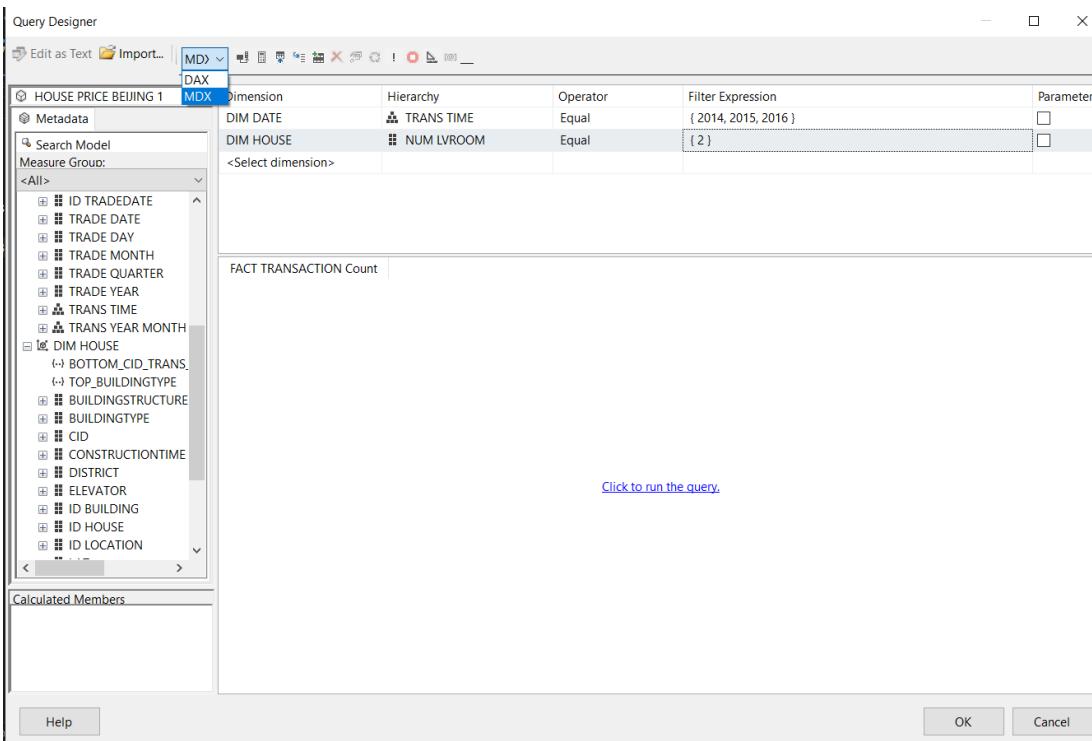


Hình 334: Chọn Use a dataset embedded in my report

- Phần **Data Source** chọn **HousePriceBeijing**, chọn **Query Designer** để tìm những trường thích hợp thêm vào báo cáo phù hợp với câu truy vấn
- Tương tự như ở **Cube Browser**, thực hiện kéo các thuộc tính để thực hiện câu truy vấn.
- Chọn **Query**



Hình 335: Chọn Query



Hình 336: Chọn MDX

The screenshot shows the Microsoft Analysis Services Query Designer interface. On the left, there's a tree view of the dataset structure under 'HOUSE PRICE BEIJING 1'. The 'Measures' node is expanded, showing various calculated measures like 'FACT TRANSACTION' and its sub-measures such as 'AVERAGE OF FOLLOWERS', 'SUM OF TOTAL PRICE', etc. The 'Calculated Members' section is also visible. In the center, there's a table titled 'FACT TRANSACTION Count' with columns 'DISTRICT', 'FACT TRANSACTION Count', and 'SUM OF TOTAL PRICE'. The data is as follows:

DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1	5323	23335373000
2	8949	26676065000
3	504	1434669000
4	5041	11900434000
5	1163	2477573000
6	11787	31124008000
7	32783	113484736000
8	11578	47345486000
9	4341	11131459000
10	9332	43272782000
11	3547	8468808000
12	566	1184297000
13	2351	4831061000

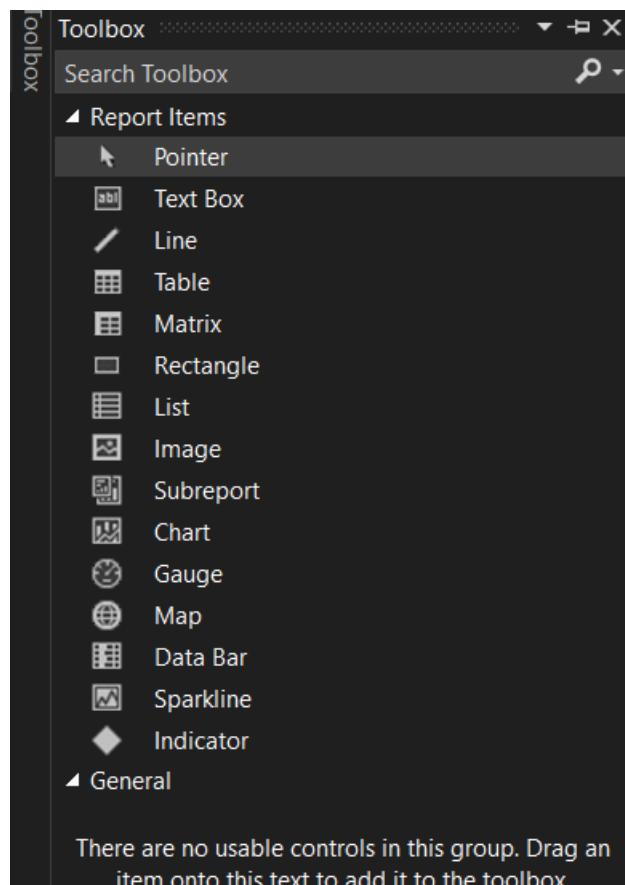
At the top right, there are filter settings for 'DIM DATE' and 'DIM HOUSE'. The 'DIM DATE' filter has 'TRANS TIME' as the hierarchy, 'Equal' as the operator, and a filter expression of '{ 2014, 2015, 2016 }'. The 'DIM HOUSE' filter has 'NUM LVROOM' as the hierarchy, 'Equal' as the operator, and a filter expression of '(2)'. Below the table, there are 'OK' and 'Cancel' buttons.

Hình 337: Kết quả Query

Chọn **OK** để sử dụng các trường. Trở về **Dataset Properties** tiếp tục chọn **OK**. Dataset mới đã được thêm vào mục Dataset với các trường đã sử dụng trong câu truy vấn tại **Query Designer**.

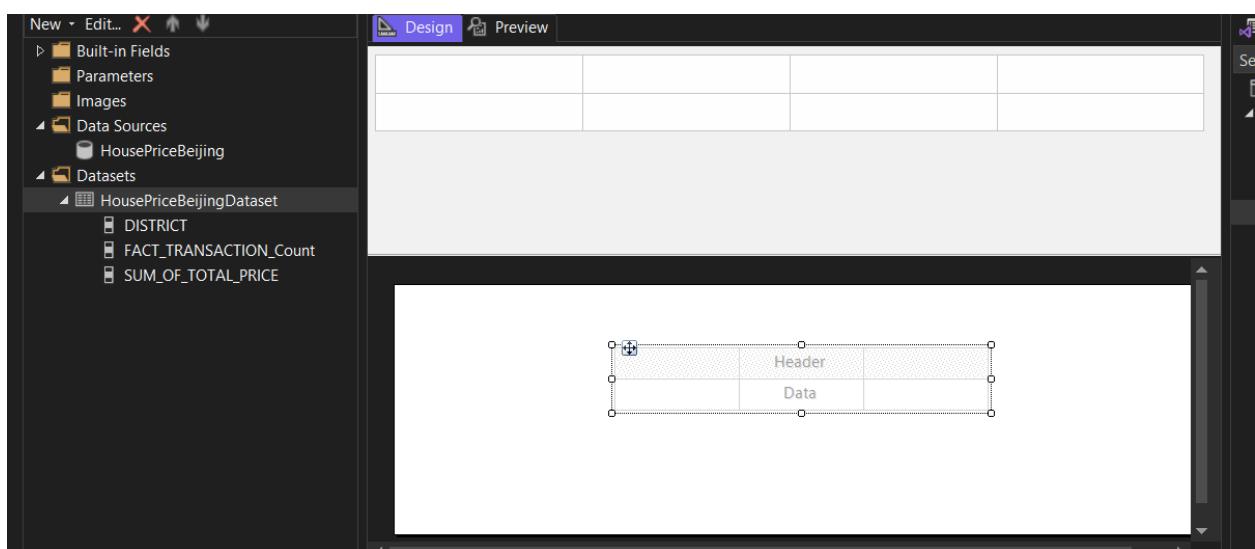
2.1.5 Thêm dữ liệu vào Report

Nếu chưa có **SSRS Toolbox**, tại **Menu → View → Toolbox** hoặc sử dụng tổ hợp phím **Ctrl + Alt + X**



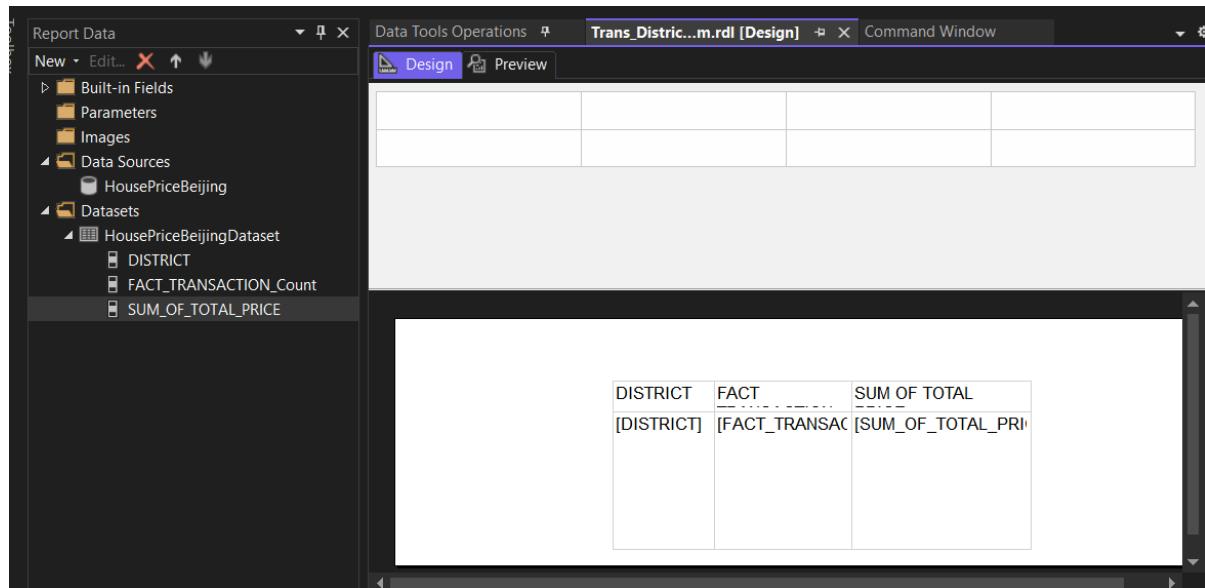
Hình 338: SSRS Toolbox

- Kéo một Table mới từ Toolbox sang giao diện Design.



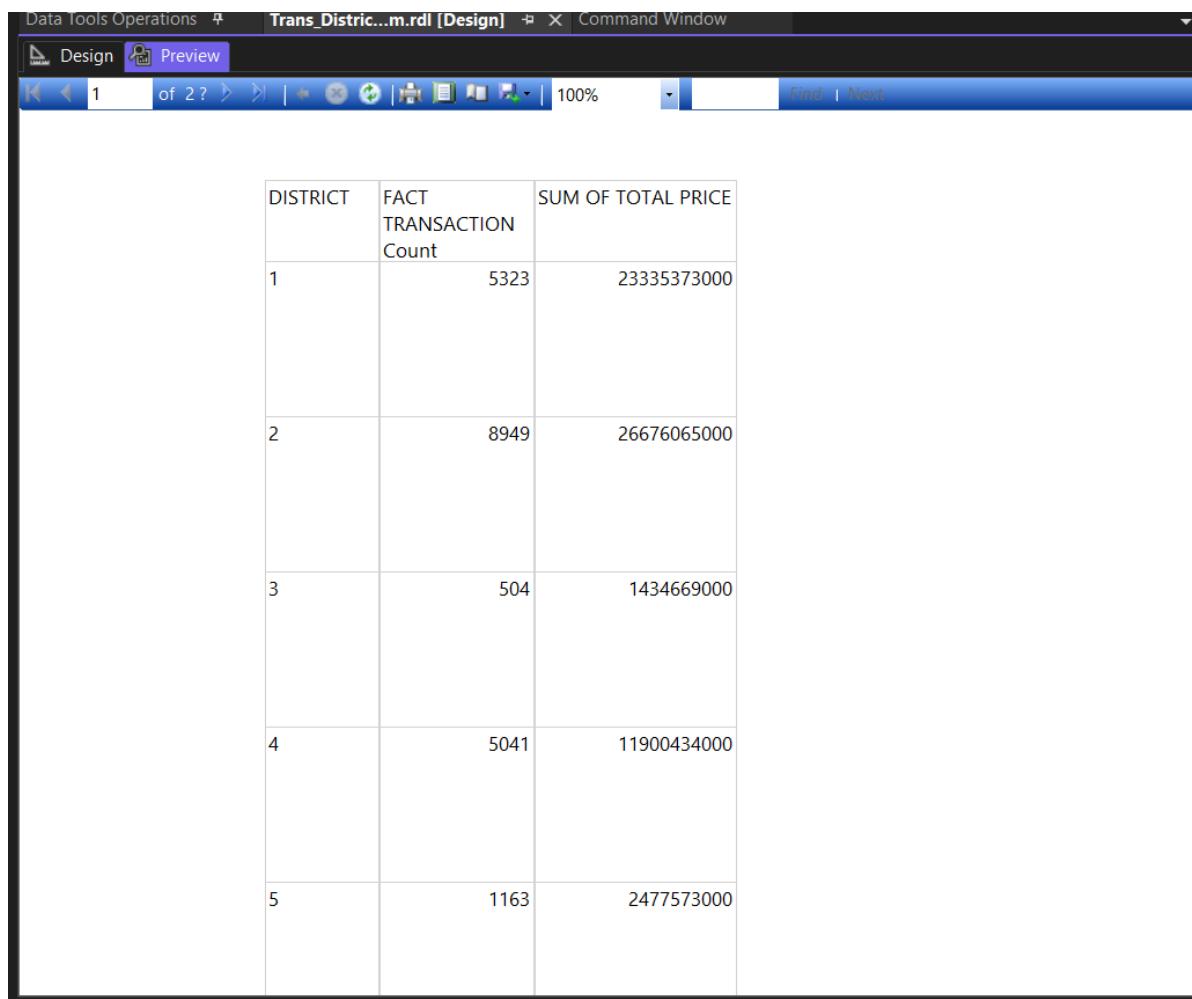
Hình 339: Kéo vào một Table mới

- Mở rộng **HousePriceBeijingDataset** trong mục **Dataset** để hiển thị đầy đủ các trường có thể sử dụng. Lần lượt kéo các trường vào bảng



Hình 340: Kéo lần lượt các trường vào bảng

- Chọn **Tab Preview** để xem báo cáo

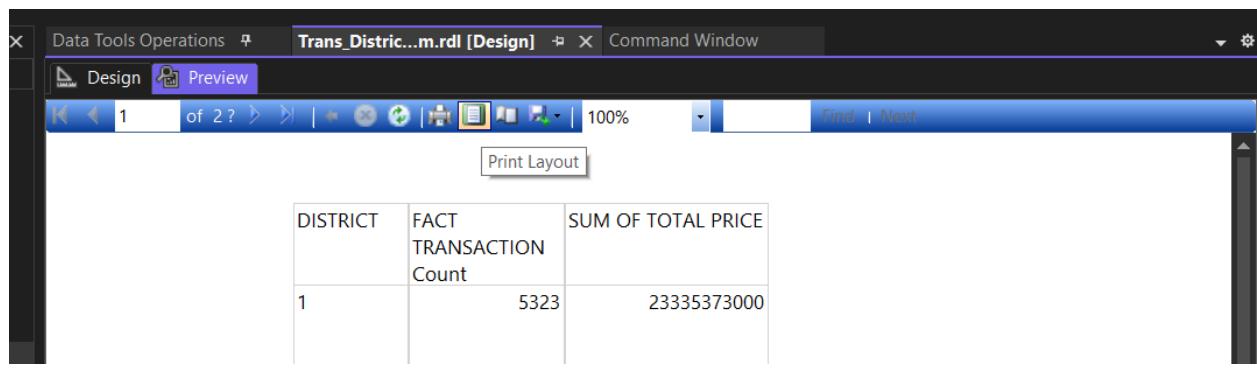


The screenshot shows the SSRS Report Designer interface with the 'Preview' tab selected. The report displays a table with five rows of data. The columns are labeled 'DISTRICT', 'FACT TRANSACTION Count', and 'SUM OF TOTAL PRICE'. The data rows are as follows:

DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1	5323	23335373000
2	8949	26676065000
3	504	1434669000
4	5041	11900434000
5	1163	2477573000

Hình 341: Xem báo cáo

- Chọn **Print Layout** để hiển thị báo cáo giống với trang in.



The screenshot shows the SSRS Report Designer interface with the 'Print Layout' button highlighted in the toolbar. The report displays a table with five rows of data, identical to the one in Figure 341. The columns are labeled 'DISTRICT', 'FACT TRANSACTION Count', and 'SUM OF TOTAL PRICE'. The data rows are as follows:

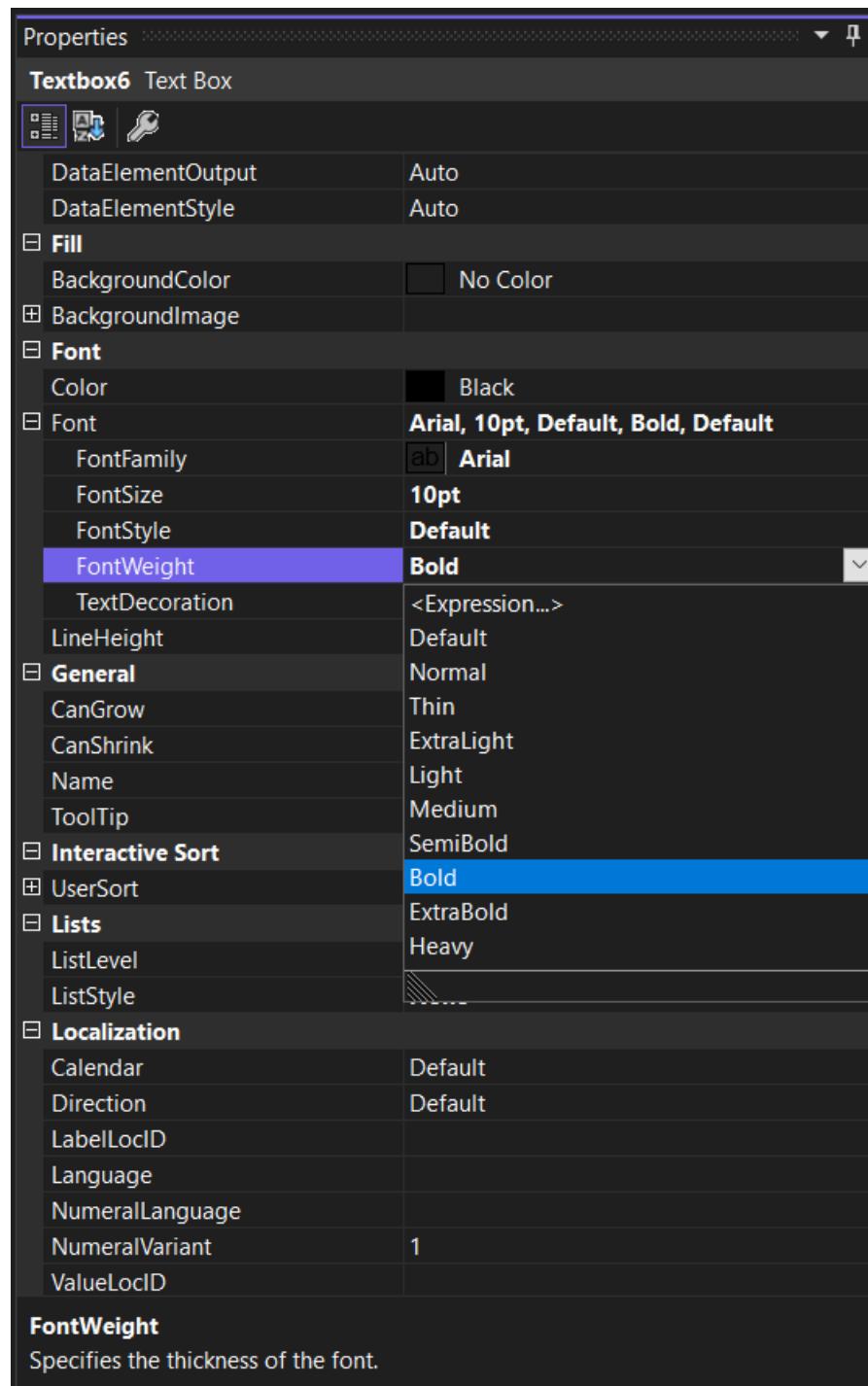
DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1	5323	23335373000

Hình 342: Hiển thị báo cáo giống với trang in

DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1	5323	23335373000
2	8949	26676065000
3	504	1434669000
4	5041	11900434000
5	1163	2477573000
6	11787	31124008000
7	32783	113484736000

2.1.6 Định dạng Report

Trở về tab **Design**, chọn hàng của **Header**. Tại **Properties pane**, mở rộng mục **Font**, tại phần **FontWeight** chọn **Bold**.



Hình 343: Chọn FontWeight Bold

- **R-click** tại giao diện **Design** và chọn **Insert → Page Header**, sử dụng một **Textbox** mới với **FontSize 16pt, FontWeight Bold, TextAlign Center** có nội dung là **APR MDC Patient Count Report**.
- Thay đổi **.TextAlign** cho **Header** của **Table** thành **Center**.

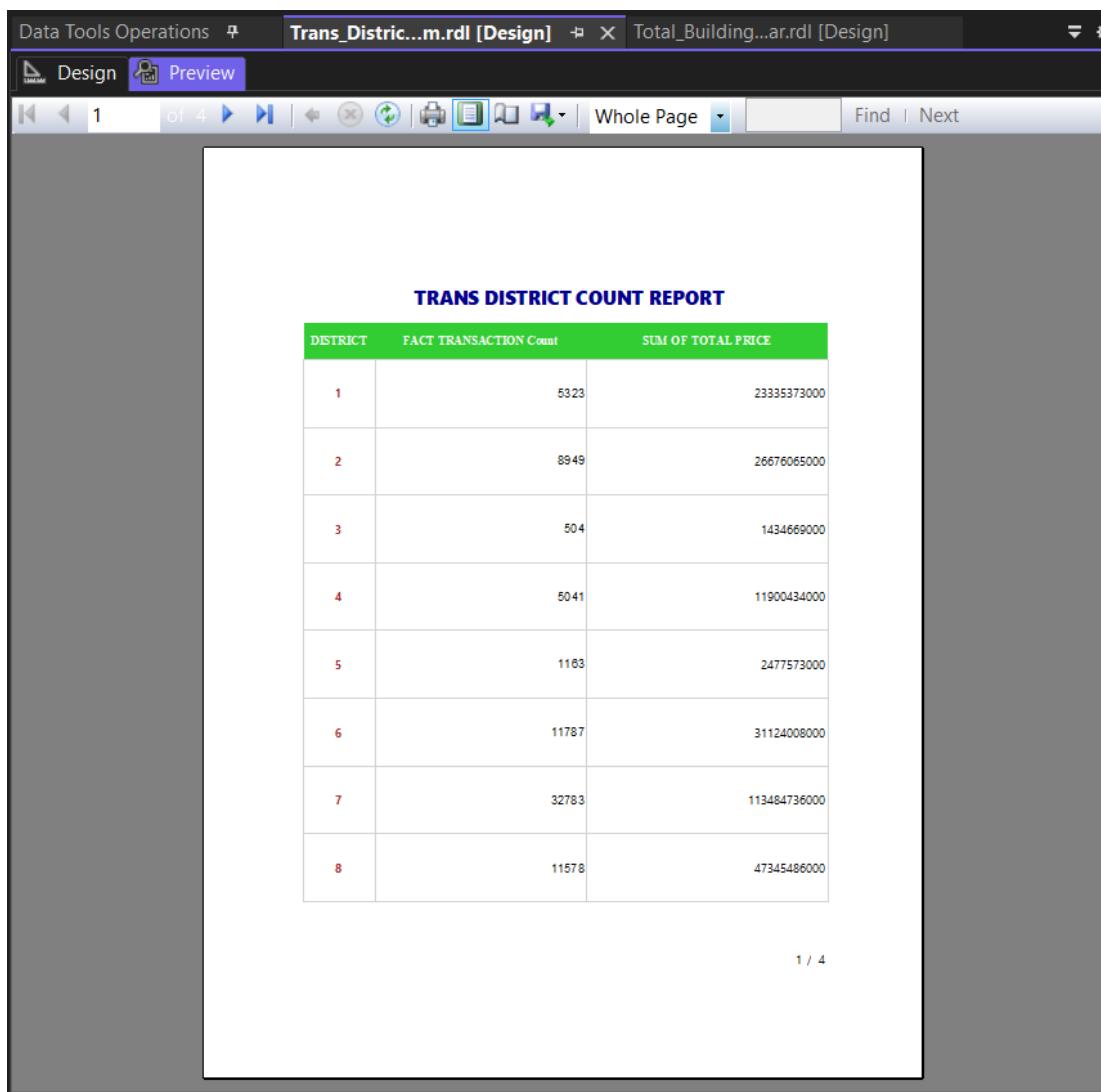
- Định dạng lại report:

The screenshot shows the Microsoft Report Designer interface. At the top center, the title 'TRANS DISTRICT COUNT REPORT' is displayed in bold blue capital letters. Below the title is a table with three columns. The first column is labeled 'DISTRICT' and contains the expression '[DISTRICT]'. The second column is labeled 'FACT TRANSACTION Count' and contains the expression '[FACT_TRANSACTION_Count]'. The third column is labeled 'SUM OF TOTAL PRICE' and contains the expression '[SUM_OF_TOTAL_PRICE]'. The table has a green header row. At the bottom of the designer window, there are two buttons: 'Row Groups' and 'Column Groups'.

DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
[DISTRICT]	[FACT_TRANSACTION_Count]	[SUM_OF_TOTAL_PRICE]

Hình 344: Sau khi định dạng lại

- Report trong Print Layout:



Hình 345: Print Layout của Report

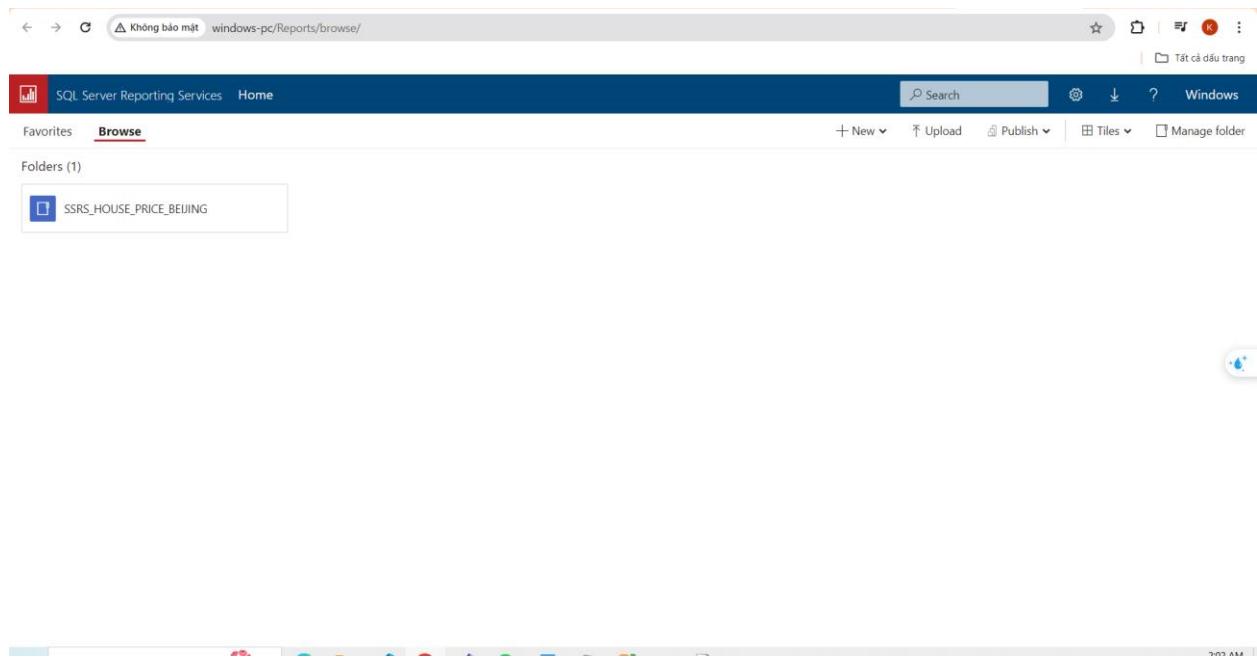
2.1.7 Xuất báo cáo

2.1.7.1 Cấu hình Report Server

Mở **Report Server Configuration Manager** và kết nối đến instance mặc định. Tại **Web Service URL** xem các thông tin cấu hình, kiểm tra nội dung tại mục **Virtual Directory** là **ReportServer** và chọn **Apply**

2.1.7.2 Xuất báo cáo

- Trở lại project SSRS, ở Menu → Build → Deploy. Sau khi Deploy thành công, truy cập vào địa chỉ **localhost/reports** bằng trình duyệt web.



Hình 346: Giao diện trình duyệt

- Mở thư mục **SSRS_HOUSE_PRICE_BEIJING**, sau đó mở **Trans_DistrictRoom**.

Hình 347: Thư mục SSRS_HOUSE_PRICE_BEIJING

The screenshot shows a report titled "TRANS DISTRICT COUNT REPORT" generated by SQL Server Reporting Services. The report displays data for 10 districts, showing the count of transactions and the sum of total price for each. The data is presented in a table with three columns: District, Fact Transaction Count, and Sum of Total Price.

DISTRICT	FACT TRANSACTION Count	SUM OF TOTAL PRICE
1	5323	23335373000
2	8949	26676065000
3	504	1434669000
4	5041	11900434000
5	1163	2477573000
6	11787	31124008000
7	32783	113484736000
8	11578	47345486000
9	4341	11131459000
10	9332	43272782000

Hình 348: Trans_DistrictRoom

2.2 Tạo report có Total và Grouping

2.2.1 Mục đích, câu truy vấn 2

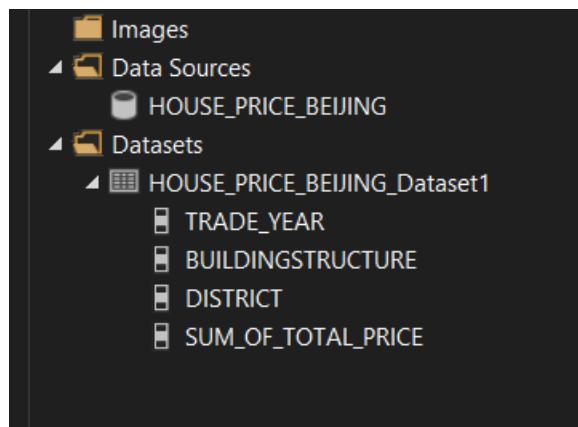
Tạo một báo cáo thể hiện tổng tiền giao dịch, dựa trên từng **BUILDINGSTRUCTURE** theo mỗi huyện, theo từng năm

2.2.2 Tạo các dữ liệu cần thiết cho Report

- Tạo một Report mới việc thêm **Data Source** và **Dataset** tương tự như việc tạo một **Report** đơn giản.

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
2002	6	7	5110000
2003	6	2	1607000
2008	6	6	1260000
2009	2	10	1500000
2010	2	10	1565000
2010	2	12	129120000
2010	4	12	4115000
2010	6	9	966000
2010	6	12	120702000
2011	1	6	1150000
2011	1	7	1240000
2011	2	1	108621000
2011	2	2	230281000
2011	2	3	10756000
2011	2	4	238622000
2011	2	6	636000000
2011	2	7	525751000
2011	2	8	607935000
2011	2	9	136625000
2011	2	10	489320000
2011	2	11	134717000
2011	2	12	64716000

Hình 349: Report mới



Hình 350: Các Datasets

A screenshot of a report template. The title is 'TOTAL PRICE OF BUILDINGSTRUCTURE AND DISTRICT'. Below the title is a table with four columns: 'TRADE YEAR', 'BUILDINGSTRUCTURE', 'DISTRICT', and 'SUM OF TOTAL PRICE'. Each column contains a placeholder bracketed expression: '[TRADE_YEAR]', '[BUILDINGSTRUCTURE]', '[DISTRICT]', and '[SUM_OF_TOTAL_PRICE]'. At the bottom right of the report area, there is a footer placeholder '[&PageNumber]/[&TotalPageNumber]'.

Hình 351: Report thô

TOTAL PRICE OF BUILDINGSTRUCTURE AND DISTRICT

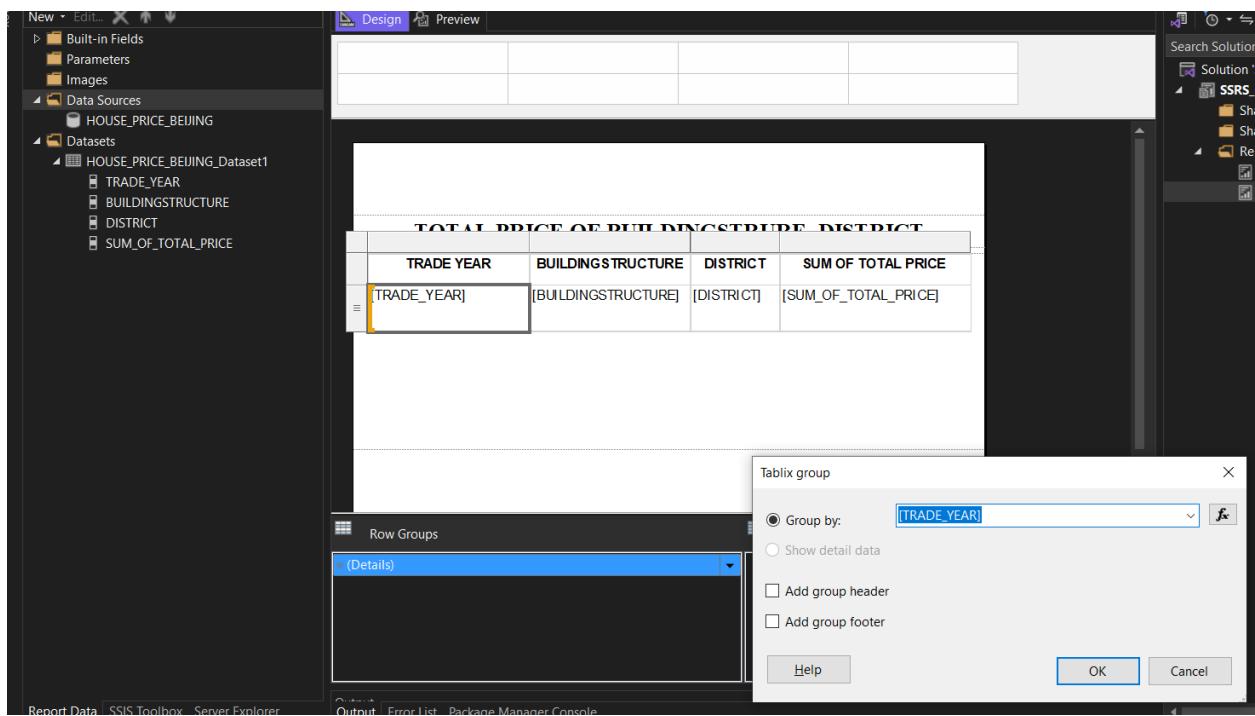
TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
2002	6	7	5110000
2003	6	2	1607000
2008	6	6	1280000
2009	2	10	1500000
2010	2	10	1585000
2010	2	12	129120000
2010	4	12	4115000
2010	6	9	988000
2010	6	12	120702000
2011	1	6	1150000
2011	1	7	1240000
2011	2	1	108821000

1 / 29

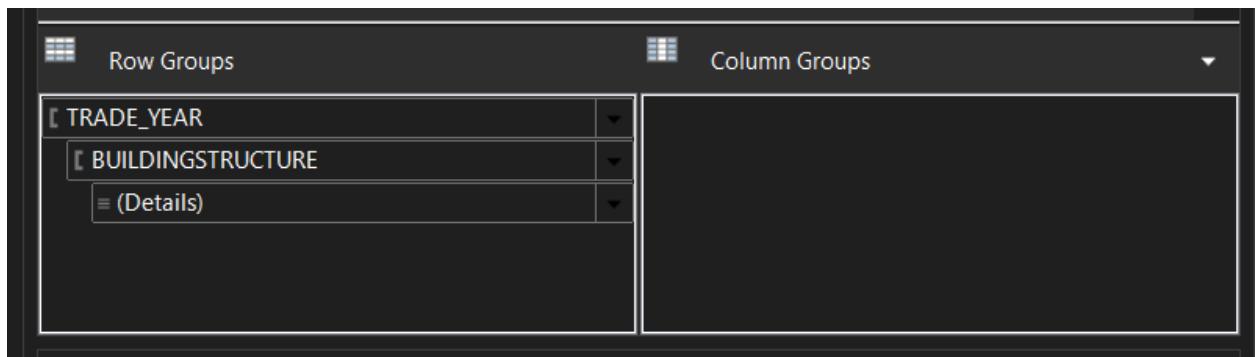
Hình 352: Sau khi được định dạng lại

2.2.3 Group dữ liệu

Từ **Report Data pane**, ta kéo **TRADE_YEAR, BUILDINGSTRUCTURE** đến **Row Groups pane**. Đặt trường trên bên trên vị trí của trường mặc định = (**Details**)



Hình 353: Đặt trường trên bên trên vị trí của trường mặc định (1)



Hình 354: Đặt trường trên bên trên vị trí của trường mặc định (2)

- Sau khi Group:

The screenshot shows a report preview window in Report Designer. The title of the report is "TOTAL PRICE OF BUILDINGSTRUCTURE_DISTRICT". The table has four columns: TRADE YEAR, BUILDING STRUCTURE, DISTRICT, and SUM OF TOTAL PRICE. The data is grouped by year (2002-2011), building structure (1-6), and district (1-12). Grand totals are shown for each group.

TRADE YEAR	BUILDING STRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
		7	5110000
2003	6	2	1607000
2008	6	6	1286000
2009	2	10	1500000
2010	2	10	1565000
		12	129120000
	4	12	4115000
	6	9	966000
		12	120702000
2011	1	6	1150000
		7	1240000
		2	108621000

Hình 355: Kết quả sau khi group

2.2.4 Total và Grand Total

- **R-click** vào ô dữ liệu của **Total Costs**, chọn **Add Total**, **Report Designer** thêm một dòng dữ liệu với một ô dữ liệu mới tính tổng chi phí trong một chi tiết của **Row Group**.

- Thêm các thông tin khác phù hợp để định dạng lại **Report**.
- Bằng cách sử dụng **Add Total → After** ở ô dữ liệu của **BUILDINGSTRUCTURE**

TOTAL PRICE OF BUILDINGSTRUCTURE AND DISTRICT

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
[TRADE_YEAR]	[BUILDINGSTRUCTURE]	[DISTRICT]	[SUM_OF_TOTAL_PRICE]
	Total		[Sum(SUM_OF_TOTAL_PRICE)]

[&PageNumber] / [&To]

Hình 356: Thêm các trường cần thiết

- Cuối cùng thêm **YEAR TOTAL** bằng cách thực hiện tương tự **Add Total → After** ở ô dữ liệu của **TRADE YEAR**.
- Thêm **Background Color** cho từng hàng tính Total để dễ phân biệt.

TOTAL PRICE OF BUILDINGSTRUCTURE AND DISTRICT

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
[TRADE_YEAR]	[BUILDINGSTRUCTURE]	[DISTRICT]	[SUM_OF_TOTAL_PRICE]
	BUILDING TOTAL		[Sum(SUM_OF_TOTAL_PRICE)]
YEAR TOTAL			[Sum(SUM_OF_TOTAL_PRICE)]

[&PageNumber] / [&To]

Hình 357: Sau khi được thêm Background Color

- Kết quả sau khi thêm **total**:

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
		7	5110000
	BUILDING TOTAL		6635000
2003	6	2	1607000
	BUILDING TOTAL		1607000
2008	6	6	1280000
	BUILDING TOTAL		1280000
2009	2	10	1500000
		BUILDING TOTAL	1500000
	2010	10	1585000
		12	129120000
		4	4115000
		6	966000
		12	120702000
	BUILDING TOTAL		256468000

1 / 30

Hình 358: Kết quả sau khi thêm Total (1)

The screenshot shows a Microsoft Access report window titled "TOTAL PRICE OF BUILDINGSTRURE AND DISTRICT". The report displays a table with the following data:

2018	6	10	44760000
		11	51136000
		13	3800000
BUILDING TOTAL			1036621000
YEAR TOTAL			1112886285000

The report interface includes a toolbar at the top with "Design" and "Preview" buttons, and a status bar at the bottom indicating "30 / 30".

Hình 359: Kết quả sau khi thêm Total (2)

- Cuối cùng **Menu → Build → Deploy**, truy cập <http://localhost/reports> để xem báo cáo

The screenshot shows a SQL Server Reporting Services interface. At the top, it displays the path: Home > SSRS_HOUSE_PRICE_BEIJING > TOTAL_BUILDINGSTRUCTURE_YEAR. Below the header are navigation buttons for Favorites, Browse, and search, along with zoom and print options. The main content area contains a table titled "TOTAL PRICE OF BUILDINGSTRUCTURE AND DISTRICT". The table has four columns: TRADE YEAR, BUILDINGSTRUCTURE, DISTRICT, and SUM OF TOTAL PRICE. The data is grouped by year, with specific rows for building structures and totals. The table includes the following data:

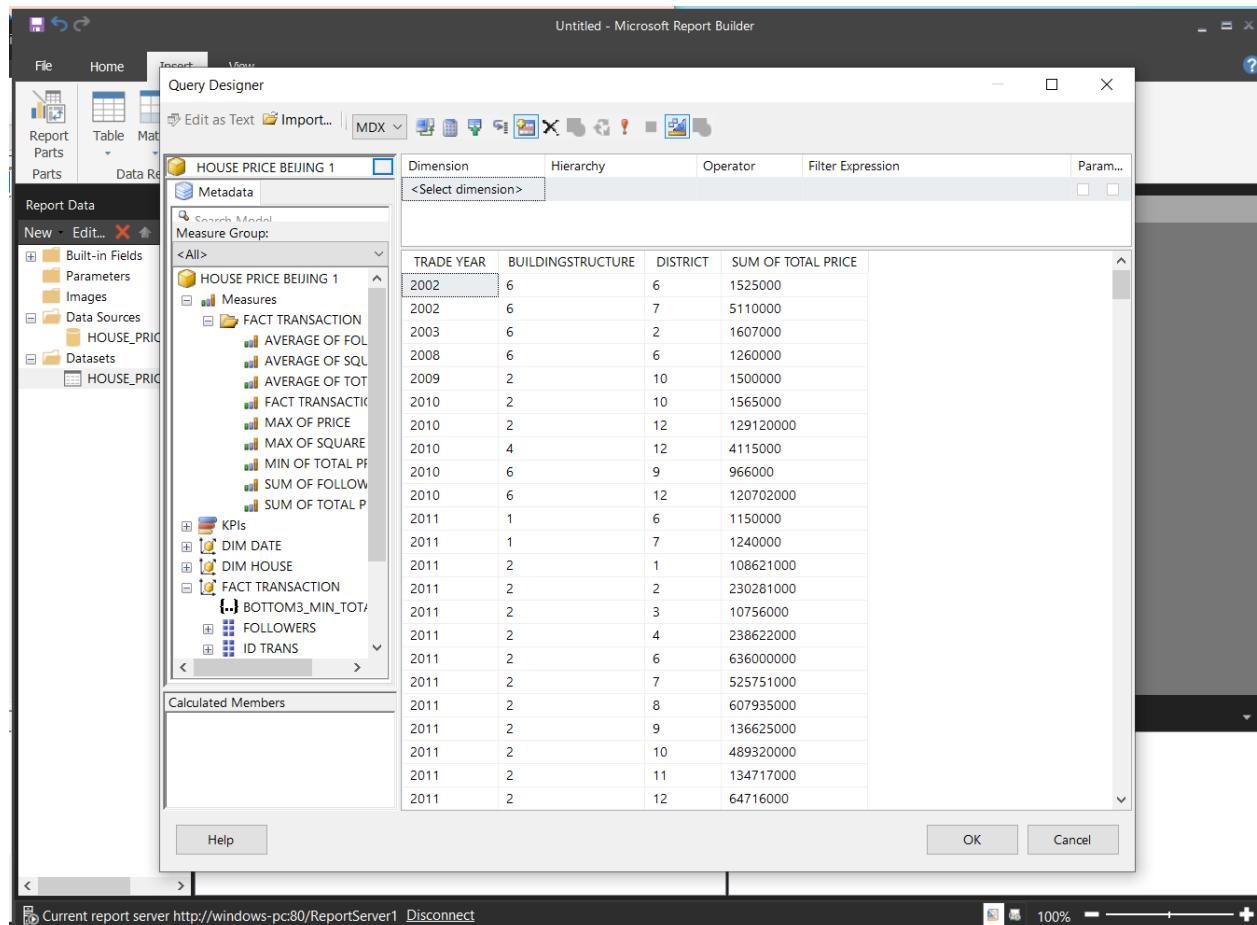
TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
		7	5110000
	BUILDING TOTAL		6635000
2003	6	2	1607000
	BUILDING TOTAL		1607000
2008	6	6	1260000
	BUILDING TOTAL		1260000
2009	2	10	1500000
	BUILDING TOTAL		1500000
2010	2	10	1565000
		12	129120000
	4	12	4115000
	6	9	966000
		12	120702000
	BUILDING TOTAL		256468000
2011	1	6	1150000
		7	1240000
	2	1	108621000

Hình 360: Kết quả trên trình duyệt

3. Report Builder

3.1 Tạo Report có Total và Grouping

Bước 1: Chọn những thuộc tính cần thiết cho báo biểu giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn **OK**

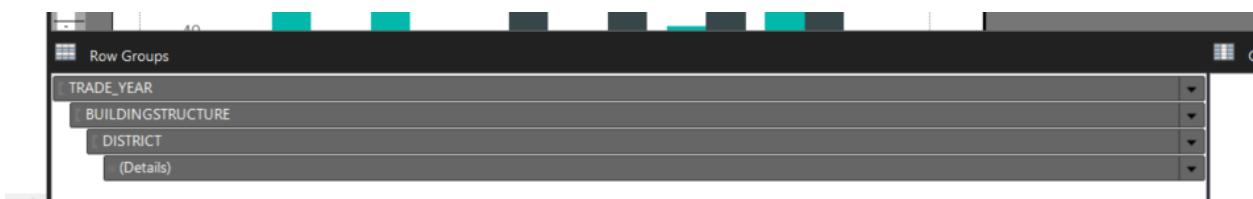


Hình 361: Chọn thuộc tính cần thiết

Bước 2: Insert một Table vào với tùy chọn là **Insert Table** và kéo **table** qua vùng thiết kế (**Design**), kéo những thuộc tính vào bảng table như hình minh họa:

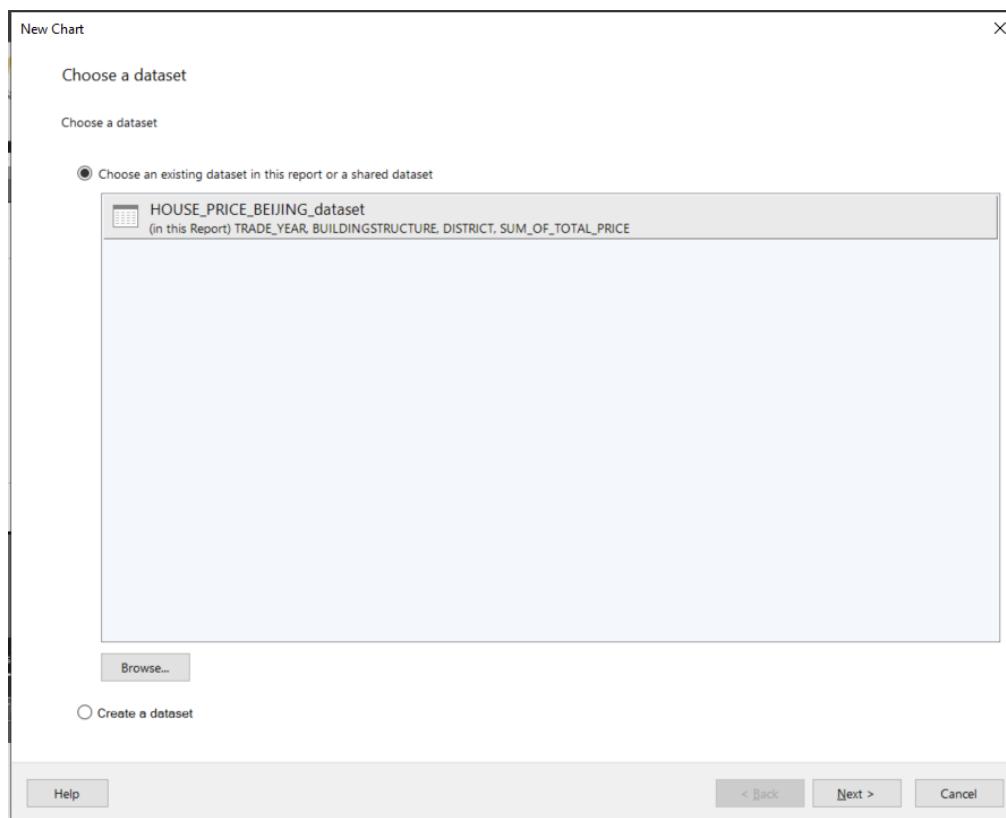
TRADE_YEAR	[BUILDINGSTRUCTURE]	[DISTRICT]	[SUM_OF_TOTAL_PRICE]
	TOTAL BUILDING		[Sum(SUM_OF_TOTAL_PRICE)]
	TOTAL YEAR		[Sum(SUM_OF_TOTAL_PRICE)]
TOTAL			[Sum(SUM_OF_TOTAL_PRICE)]

Hình 362: Insert Table

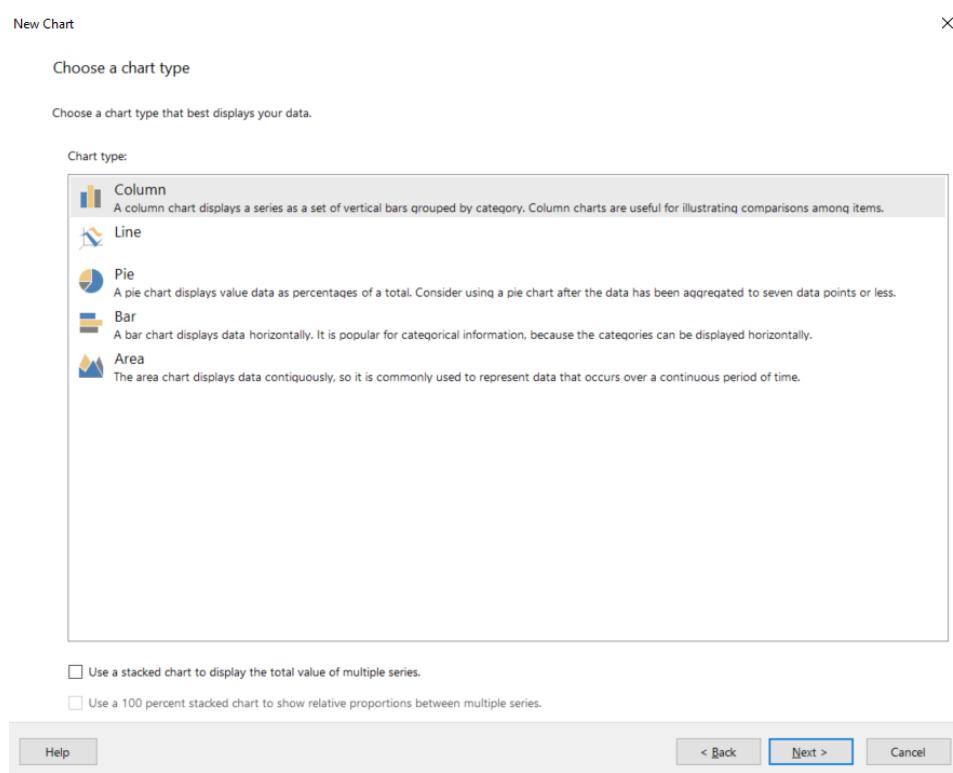


Hình 363: Các thuộc tính được kéo vào

Bước 3: Ta chọn **Insert** một **Chart** → Chọn **Chart Wizard** → Chọn lược đồ dạng **Column**.

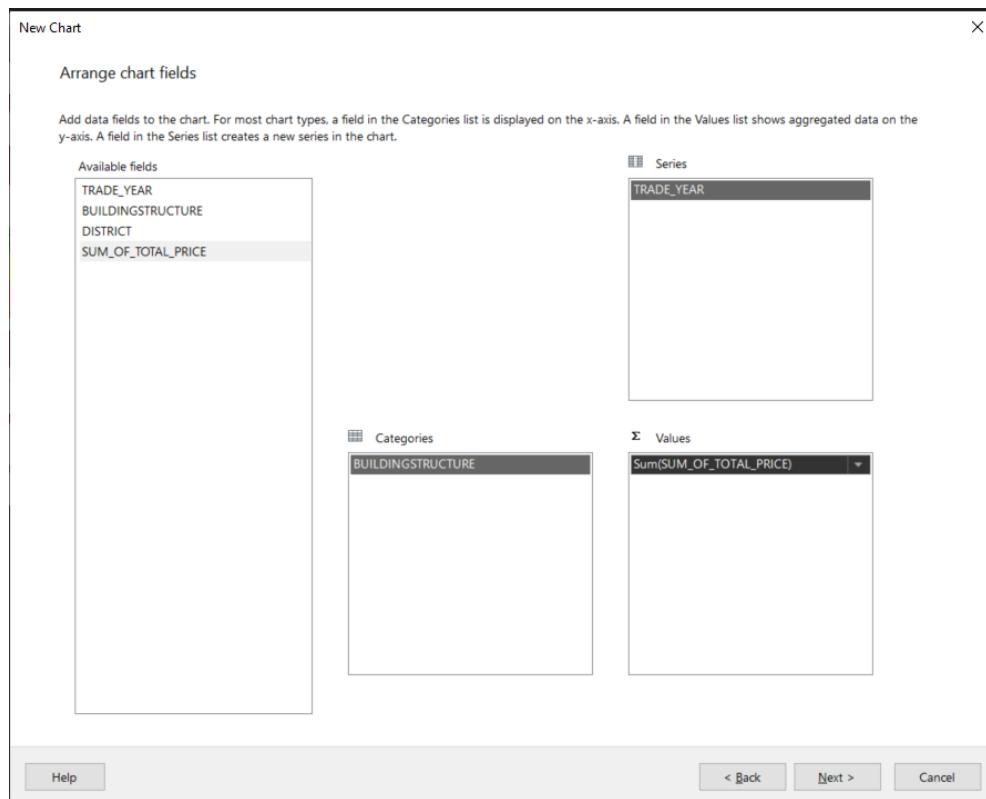


Hình 364: Chart Wizard



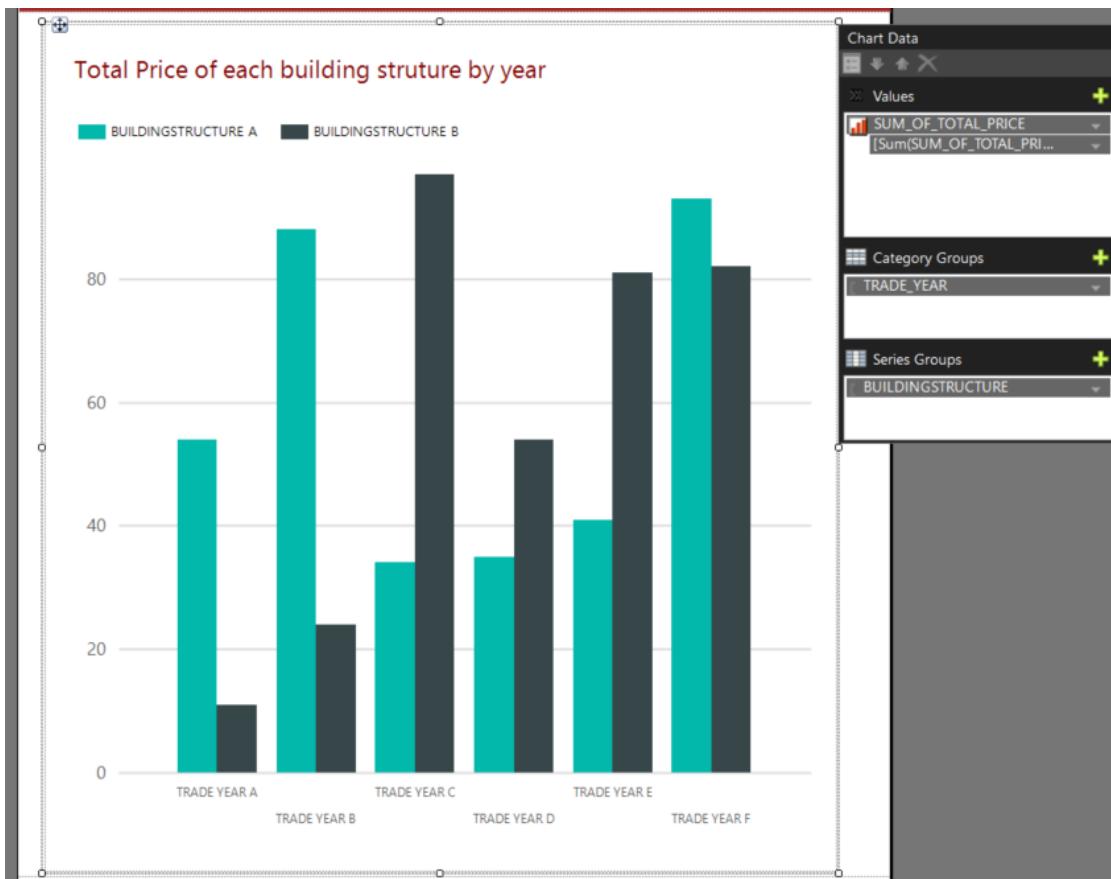
Hình 365: Chọn lược đồ dạng Column

- Tính tổng giá tiền của mỗi **building struture** theo năm



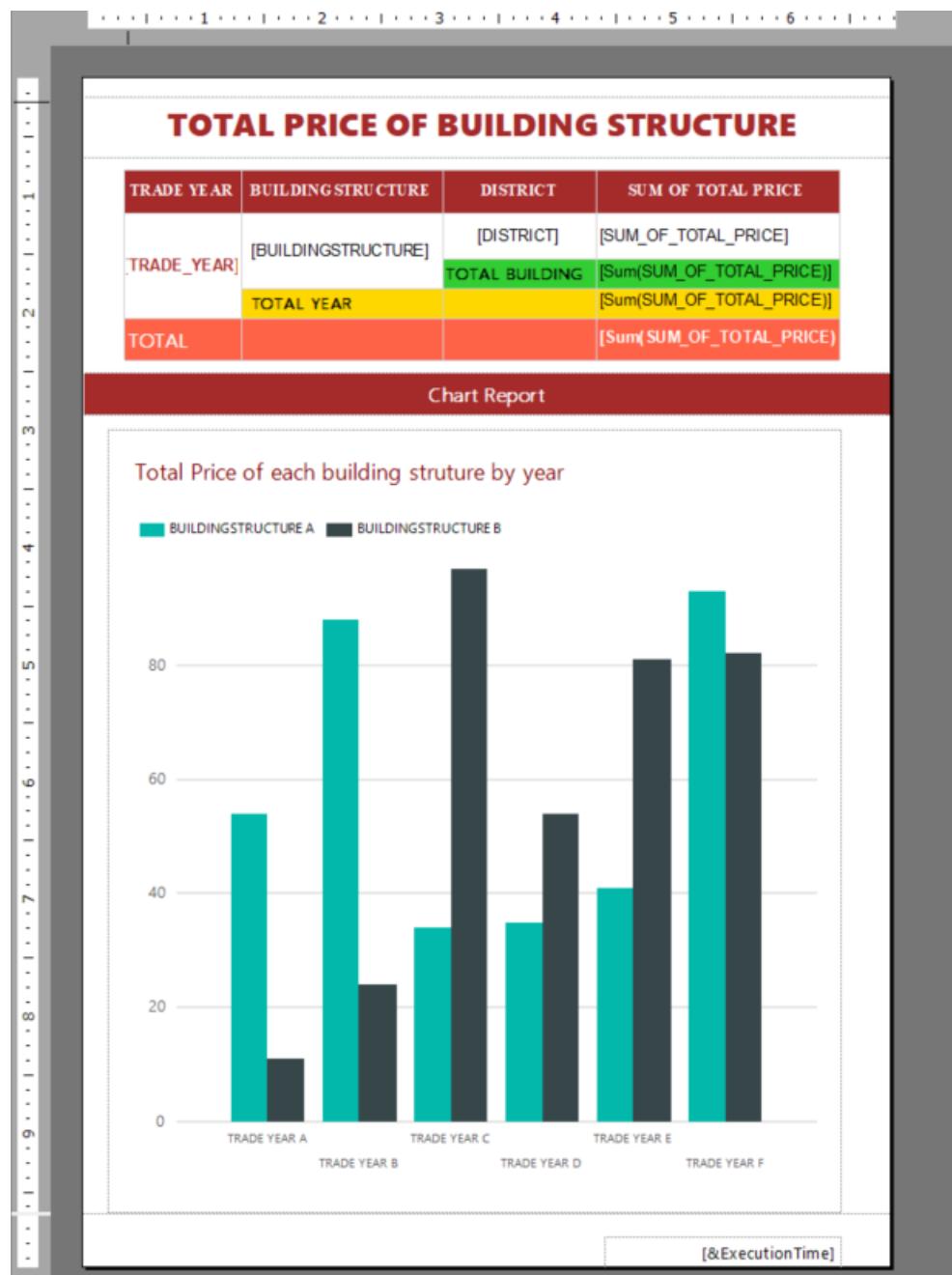
Hình 366: Tạo Chart mới

- Biểu đồ được tạo:



Hình 367: Biểu đồ

- Định dạng report:



Hình 368: Báo cáo sau khi định dạng

- Report trong PrintLayout

The screenshot shows a Microsoft Word document in Print Layout view. The ribbon at the top has tabs for Home, Insert, Page Layout, and View. The Print Layout tab is selected. The ribbon also includes sections for Print, Page Setup, Print Layout, Export, Document Map, Parameters, Options, and Find.

**TOTAL PRICE OF BUILDING STRUCTURE
AND DISTRICT**

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE	
2002	6	6	1525000	
		7	5110000	
		TOTAL BUILDING	6635000	
		TOTAL YEAR	6635000	
2003	6	2	1607000	
			TOTAL BUILDING	1607000
		TOTAL YEAR	1607000	
2008	6	6	1260000	
			TOTAL BUILDING	1260000
		TOTAL YEAR	1260000	
2009	2	10	1500000	
			TOTAL BUILDING	1500000
		TOTAL YEAR	1500000	
2010	2	10	1565000	
		12	129120000	
			TOTAL BUILDING	130685000
	4	12	4115000	
			TOTAL BUILDING	4115000
6	12	120702000		
	9	966000		
		TOTAL BUILDING	121668000	
		TOTAL YEAR	256468000	

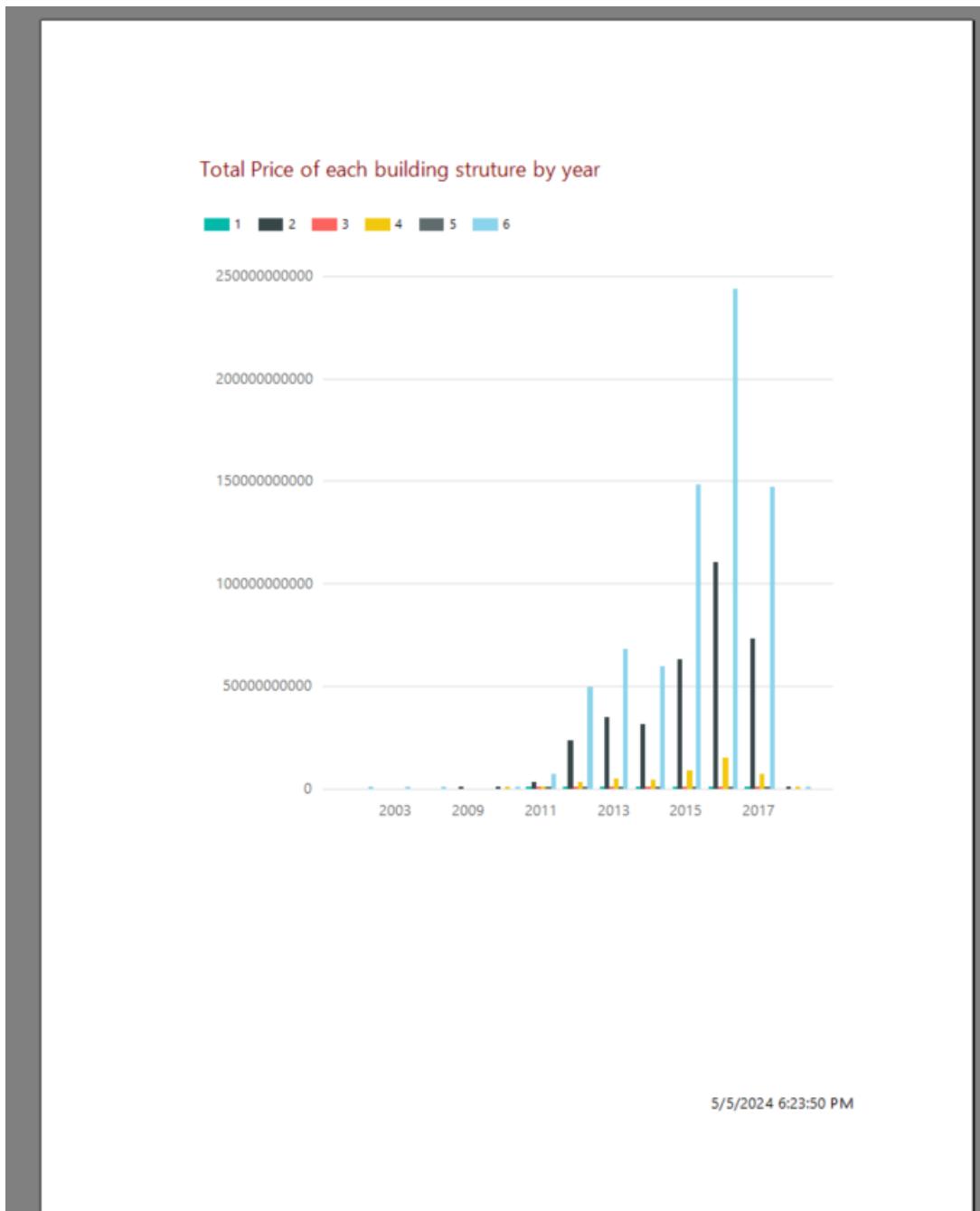
5/5/2024 6:23:50 PM

Hình 369: Report trong PrintLayout (1)

	4	7	3300000
		TOTAL BUILDING	3300000
		1	64630000
		10	44760000
		11	51136000
		13	3800000
		2	77670000
		3	1680000
		4	22380000
		5	8260000
		6	42225000
		7	272585000
		8	77450000
		9	10030000
		TOTAL BUILDING	676606000
		TOTAL YEAR	1036621000
		TOTAL	1112886285000
		Chart Report	

5/5/2024 6:23:50 PM

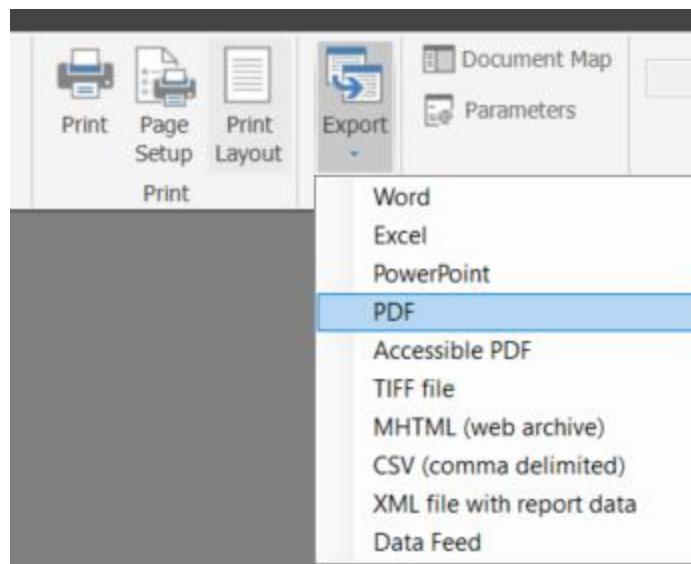
Hình 370: Report trong PrintLayout (2)



Hình 371: Report trong PrintLayout (3)

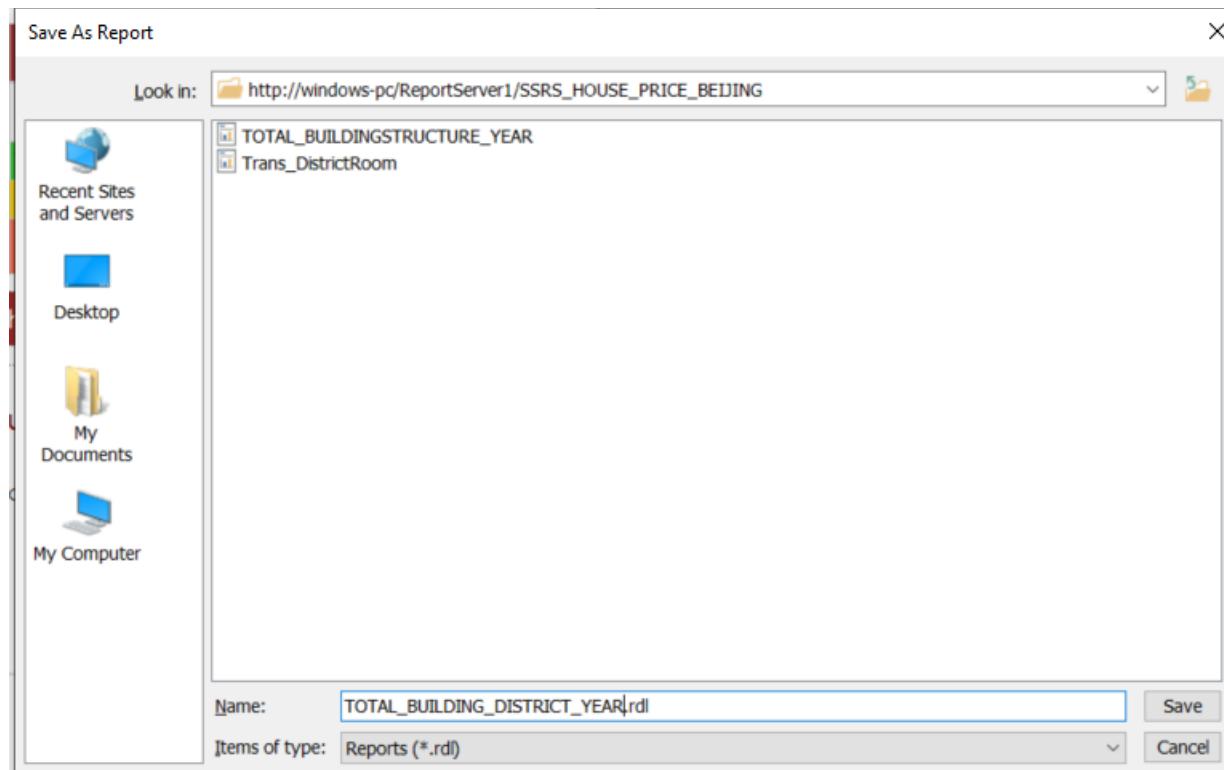
Kết quả câu Report 2 bằng report builder

- Ta xuất file pdf:



Hình 372: Xuất file pdf

- Để deploy lên Report Server:
- Nhấn Save dẫn đến đường dẫn → Recent Sites and Servers → Save



Hình 373: Deploy lên Report Server

TOTAL PRICE OF BUILDING STRUCTURE AND DISTRICT

TRADE YEAR	BUILDINGSTRUCTURE	DISTRICT	SUM OF TOTAL PRICE
2002	6	6	1525000
		7	5110000
		TOTAL BUILDING	6635000
	TOTAL YEAR		6635000
2003	6	2	1607000
		TOTAL BUILDING	1607000
	TOTAL YEAR		1607000
2008	6	6	1260000
		TOTAL BUILDING	1260000
	TOTAL YEAR		1260000
2009	2	10	1500000
		TOTAL BUILDING	1500000
	TOTAL YEAR		1500000
2010	2	10	1565000
		12	129120000
		TOTAL BUILDING	130685000
	4	12	4115000
		TOTAL BUILDING	4115000
	6	12	120702000
		9	966000
		TOTAL BUILDING	121668000

Hình 374: In report 2 trên Report Server

2.3. Tạo report 3

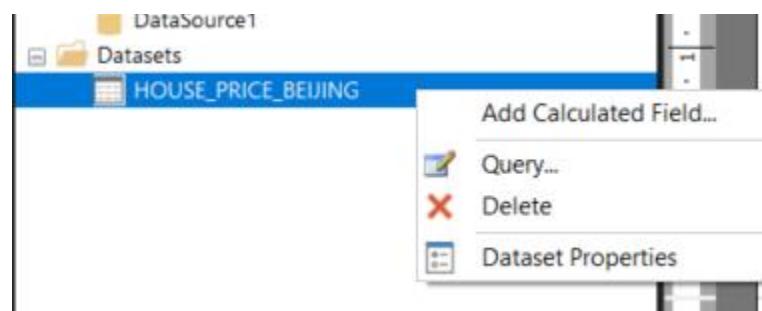
2.3.1. Mục đích, câu truy vấn 3

Xây dựng báo cáo thống kê số lượng giao dịch trên từng loại **renovation condition** theo tháng, năm với nhà có thang máy và từ năm 2015 đến năm 2018

2.3.2 Tạo các dữ liệu cần thiết cho Report



Hình 375: Add Dataset



Hình 376: Chọn Query

- Chọn những thuộc tính cần thiết cho báo biểu giống như chọn trong quá trình SSAS bằng công cụ BI sau đó nhấn **OK**.

The screenshot shows the Microsoft Analysis Services Query Designer interface. On the left, there's a tree view of the model structure under 'HOUSE PRICE BEIJING 1' (Metadata). A 'Calculated Members' section is also visible. On the right, a table is displayed with columns: TRADE YEAR, TRADE MONTH, RENOVATIONCONDITION, and FACT TRANSACTION ... (partially visible). The table contains 31 rows of data. At the top right, there are filter settings for 'DIM HOUSE' (Hierarchy: ELEVATOR, Operator: Equal, Filter Expression: { True }) and 'DIM DATE' (Hierarchy: TRADE YEAR, Operator: Equal, Filter Expression: { 2015, 2016, 2017, 2018 }). Below the table, there are 'OK' and 'Cancel' buttons.

Hình 377: Chọn những thuộc tính cần thiết cho báo biểu

Insert một **Table** vào với tùy chọn là **Insert Table** và kéo **table** qua vùng thiết kế (**Design**), kéo những thuộc tính vào bảng **table** như hình minh họa:

The screenshot shows the Microsoft Visio ribbon and a 'Design' tab. On the left, a 'Data Sources' pane lists 'DataSource1' and 'HOUSE_PRICE_BEIJING' with its fields: TRADE_YEAR, TRADE_MONTH, RENOVATIONCONDITION, and FACT_TRANSACTION_Count. On the right, a 'Click to add title' placeholder is shown with a table structure. The table has four columns: TRADE YEAR, TRADE MONTH, RENOVATIONCON DITION, and FACT TRANSACTION Count. Below each column header is a bracketed placeholder: [TRADE_YEAR], [TRADE_MONTH], [RENOVATIONCON], and [FACT_TRANSACTION_Count].

Hình 378: Kéo những thuộc tính vào bảng table như hình

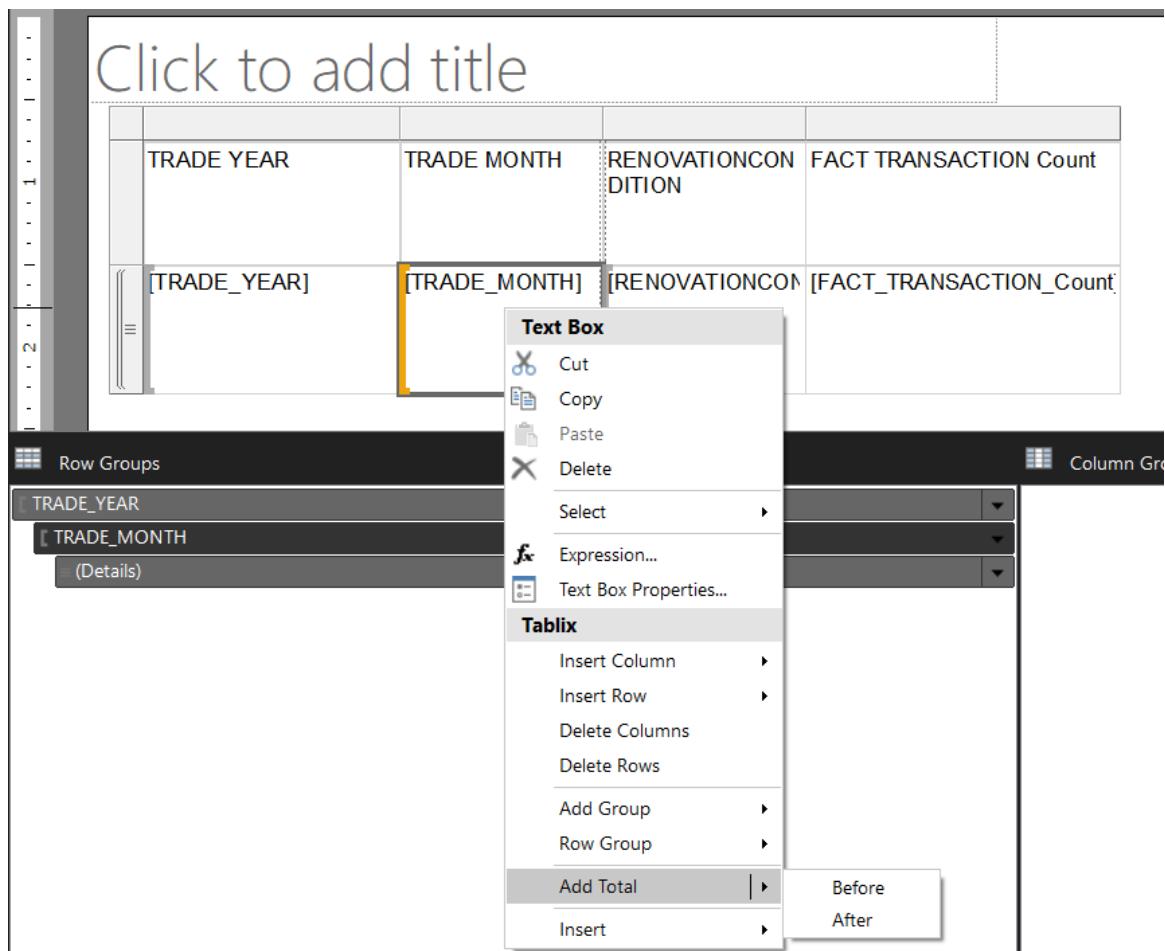
2.3.3. Group dữ liệu

Tạo **Grouping** bằng cách tô đen thuộc các thuộc tính muốn được gom nhóm, chọn **Add Group** để gom nhóm dữ liệu

TRADE YEAR	TRADE MONTH	RENOVATIONCONDITION	FACT TRANSACTION Count
[TRADE_YEAR]	[TRADE_MONTH]	[RENOVATIONCONDITION]	[FACT_TRANSACTION_Count]

Hình 379: Tạo Grouping, Add Group để gom nhóm dữ liệu

- Tạo **Total** bằng cách chọn vào thuộc tính mà ta muốn tính tổng các độ đo theo thuộc tính đó, chọn **Add Total** để tính tổng các giá trị.

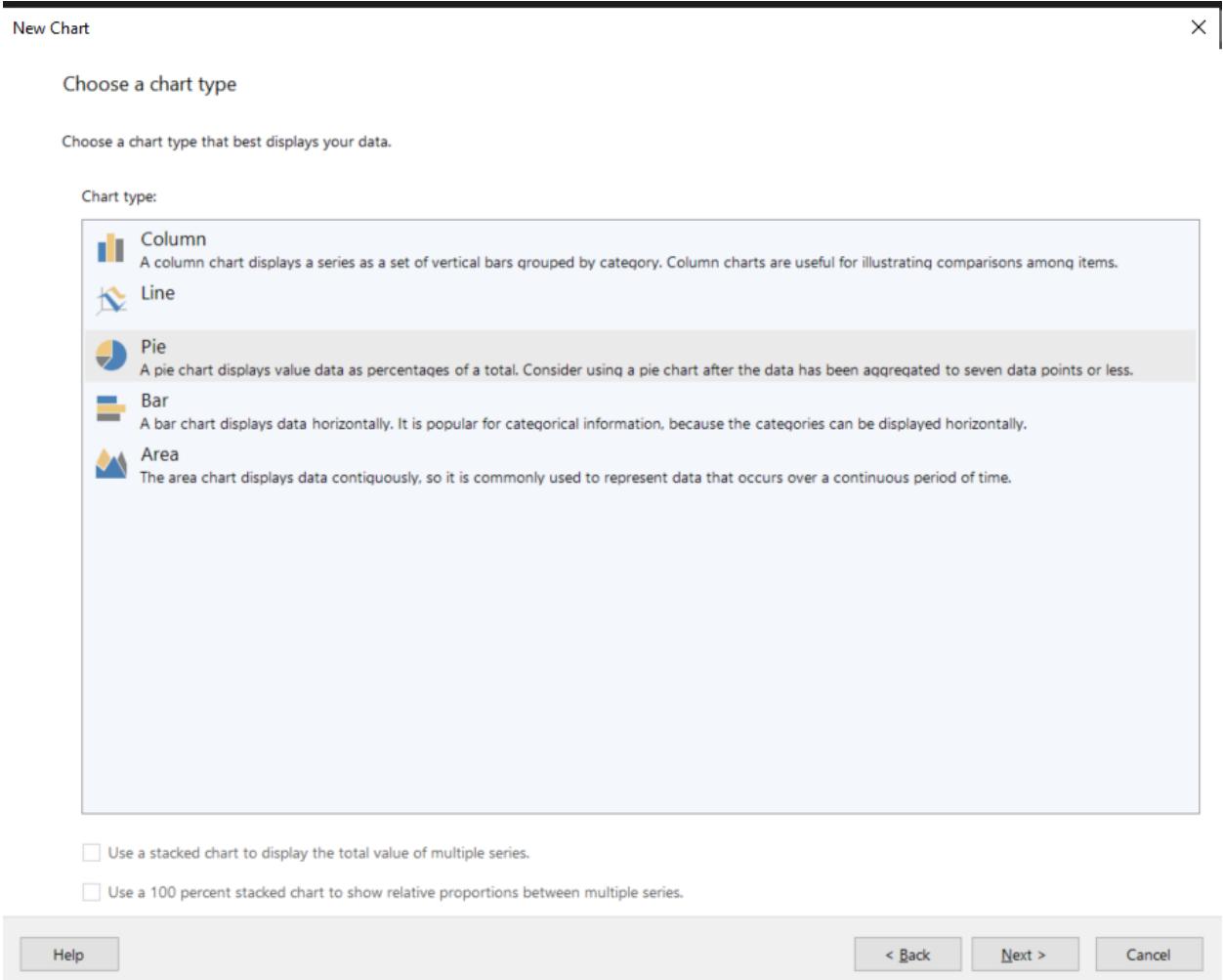


Hình 380: Add Total để tính tổng các giá trị

TRADE YEAR	TRADE MONTH	RENOVATIONCONDITION	FACT TRANSACTION Count
[TRADE_YEAR]	[TRADE_MONTH]	[RENOVATIONCONDITON]	[FACT_TRANSACTION_Count]
Total			[Sum(FACT_TRANSACTION_Count)]
Total			[Sum(FACT_TRANSACTION_Count)]

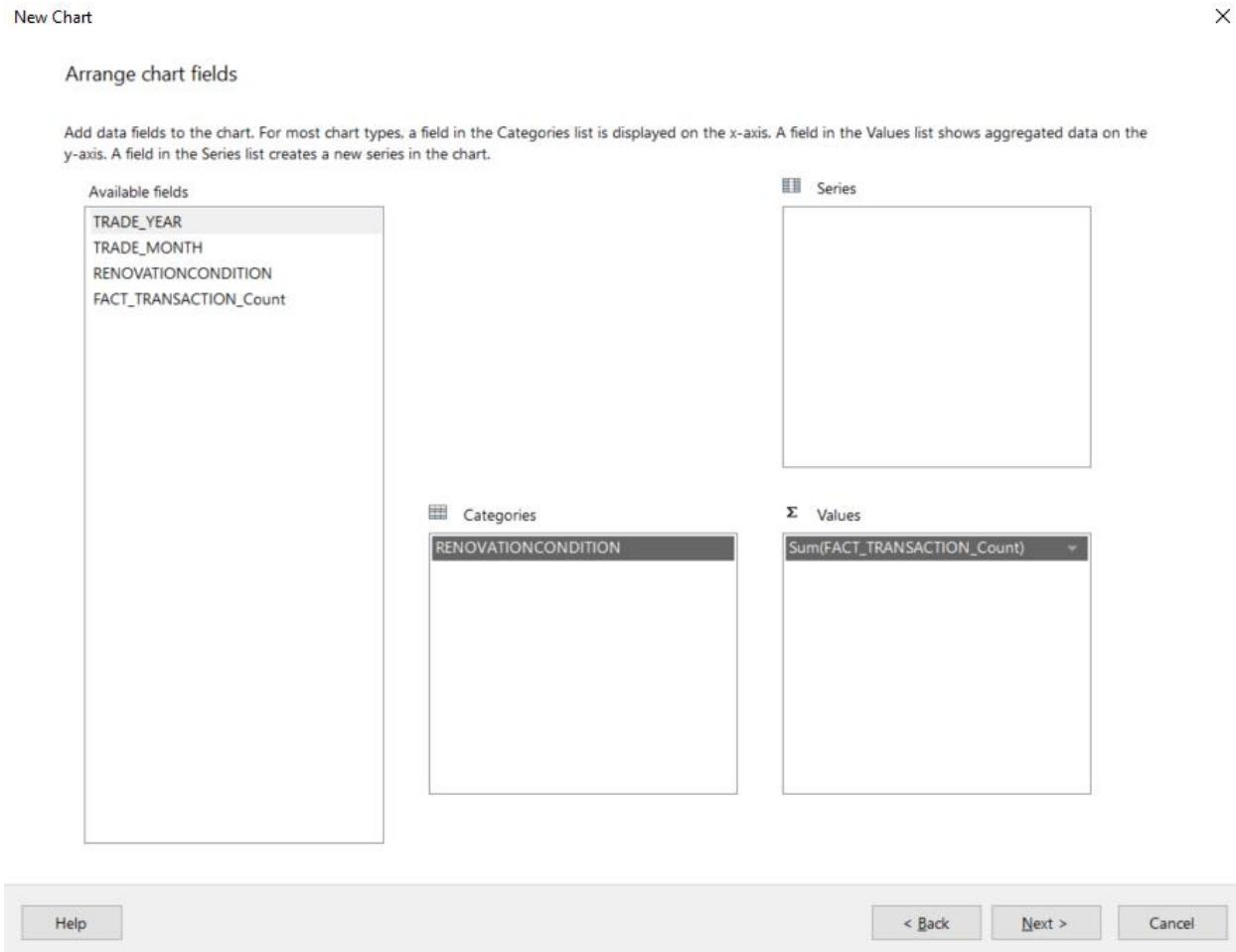
Hình 381: Thêm Title

- Để thống kê doanh thu của từng khách hàng theo từng năm, ta chọn **Insert** một **Chart** → Chọn **Chart Wizard** → Chọn lược đồ các dạng hình tròn, cột và đường.



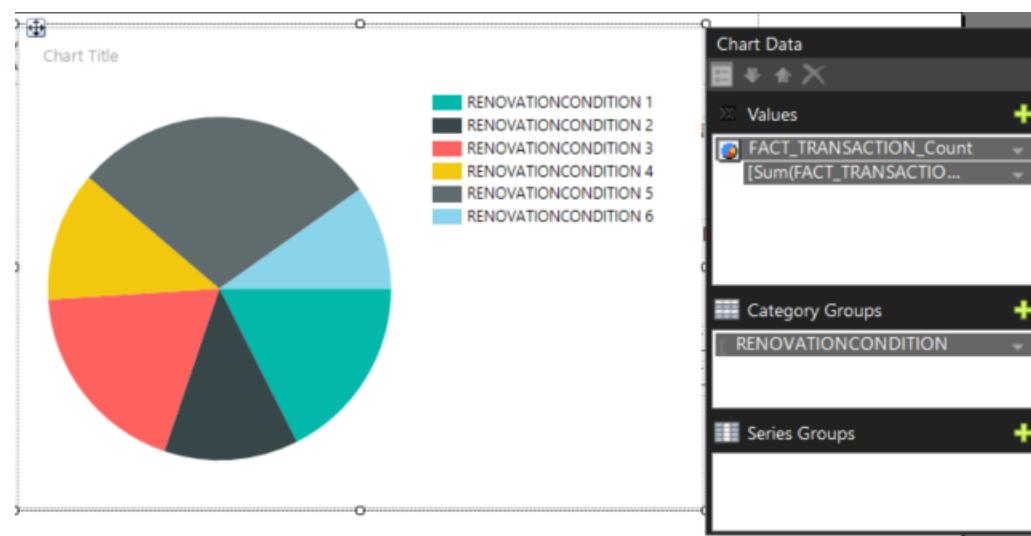
Hình 382: Chọn lược đồ các dạng hình tròn, cột và đường

- Kéo thả độ đo vào ô thích hợp



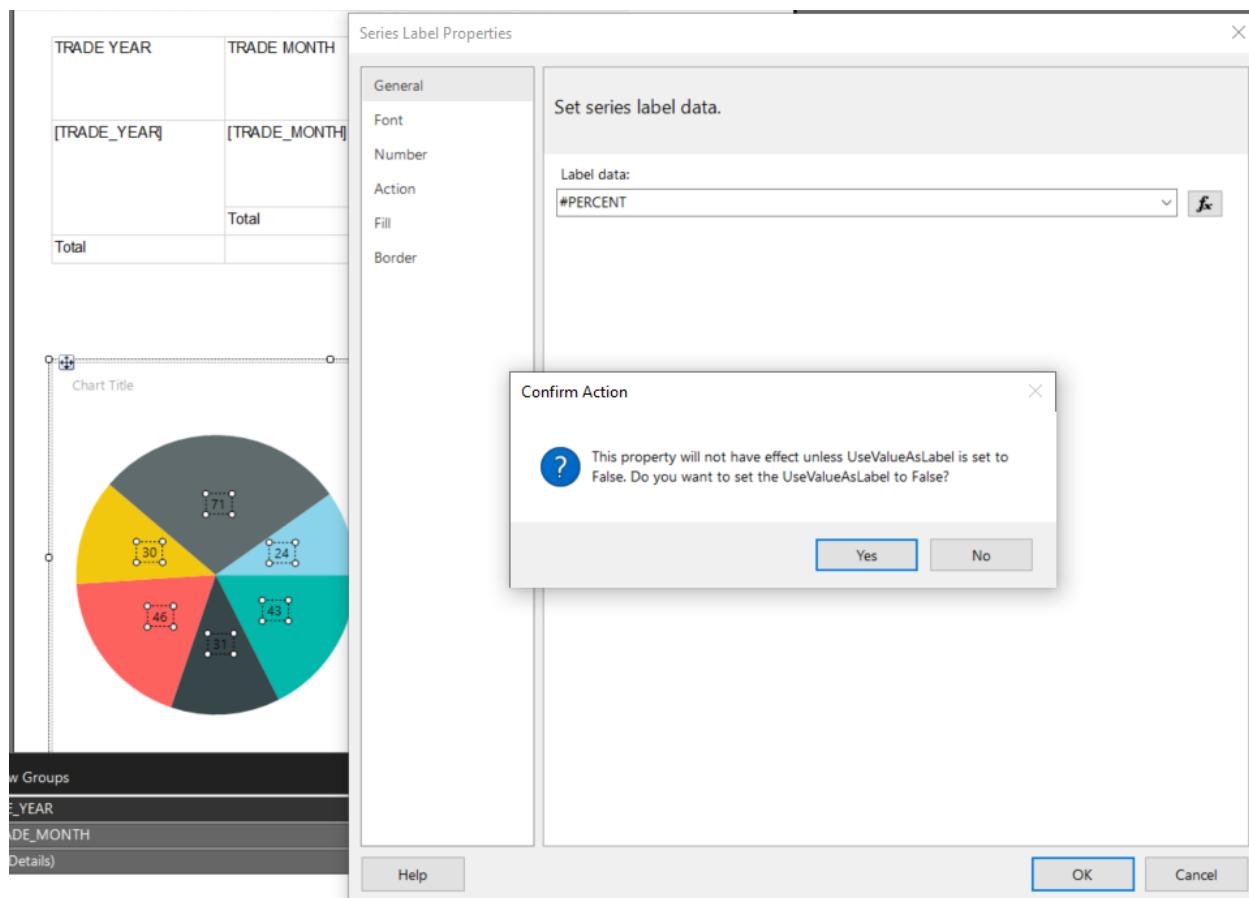
Hình 383: Kéo thả dữ liệu vào ô thích hợp

- Kết quả thu được biểu đồ:



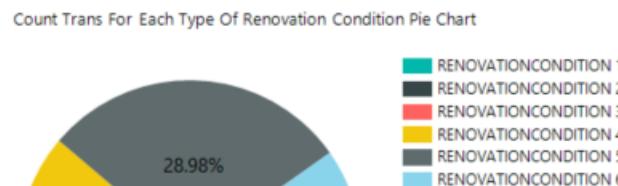
Hình 384: Biểu đồ thu được

- Thêm chỉ số phần trăm Chuyển sang **Design View**, r-click vào biểu đồ, chọn **Show Data Label**. Tiếp tục r-click vào 1 chỉ số bất kì, chọn **Series Label Properties**. Ở phần **General**, mục **Label Data** chọn **#PERCENT**, thông báo mới hiện lên chọn **Yes**.



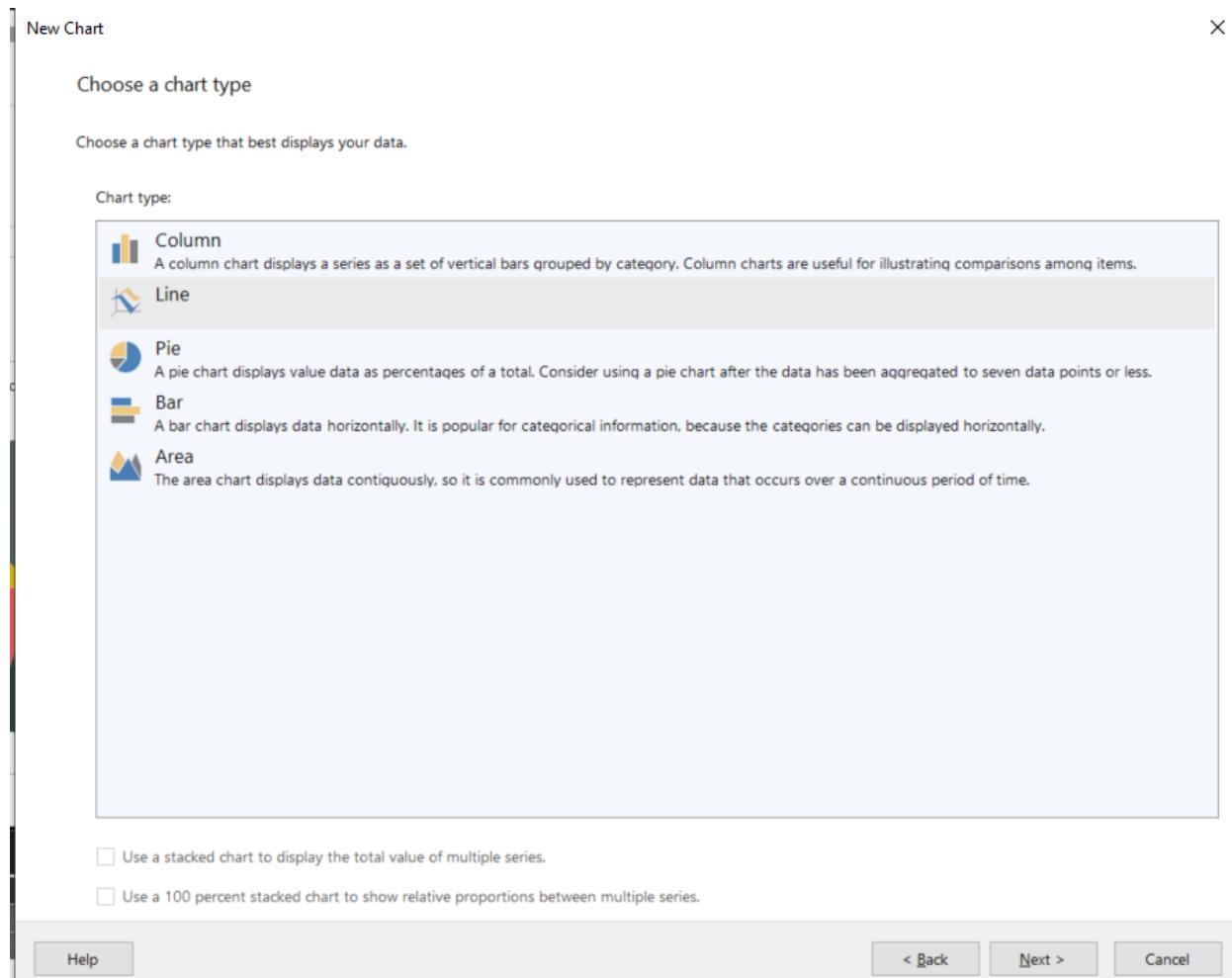
Hình 385: Chọn Yes khi gặp thông báo

- Tiếp tục **Run** để xem những thay đổi hiển thị.



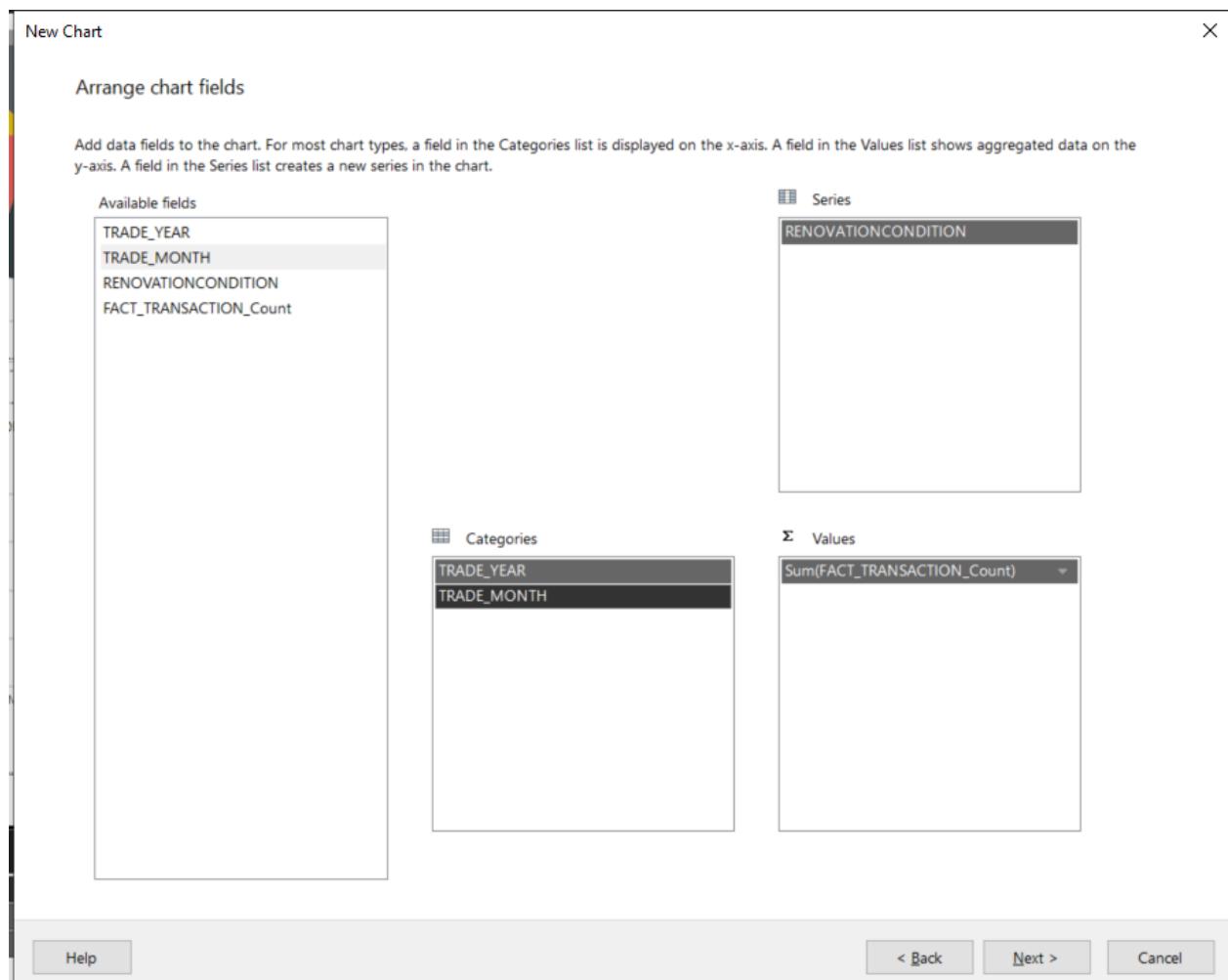
Hình 386: Hiển thị được thay đổi

- Thực hiện tương tự, tạo biểu đồ đường để thống kê số doanh thu của mỗi trạng thái trong từng tháng của từng năm.



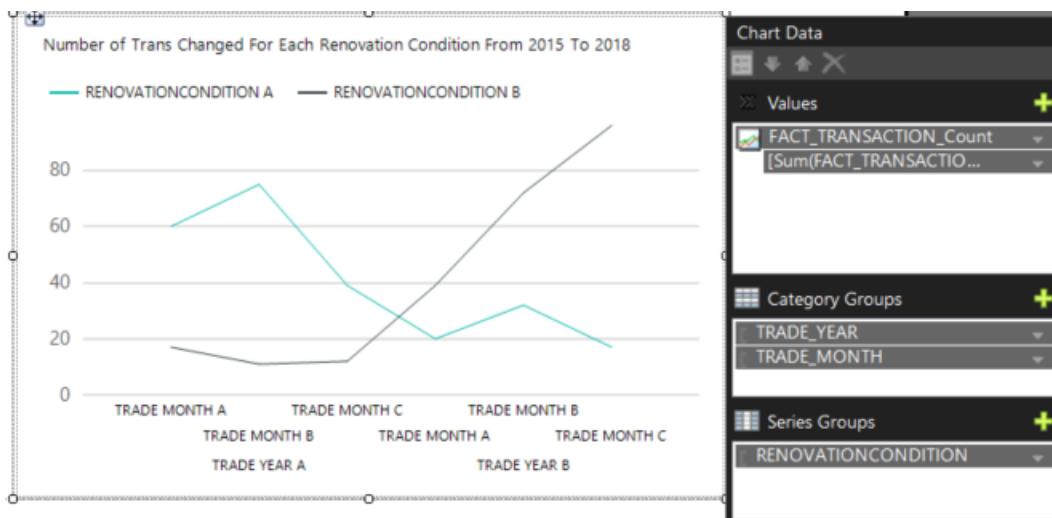
Hình 387: Tạo biểu đồ đường

- Kéo thả độ đo vào ô thích hợp:



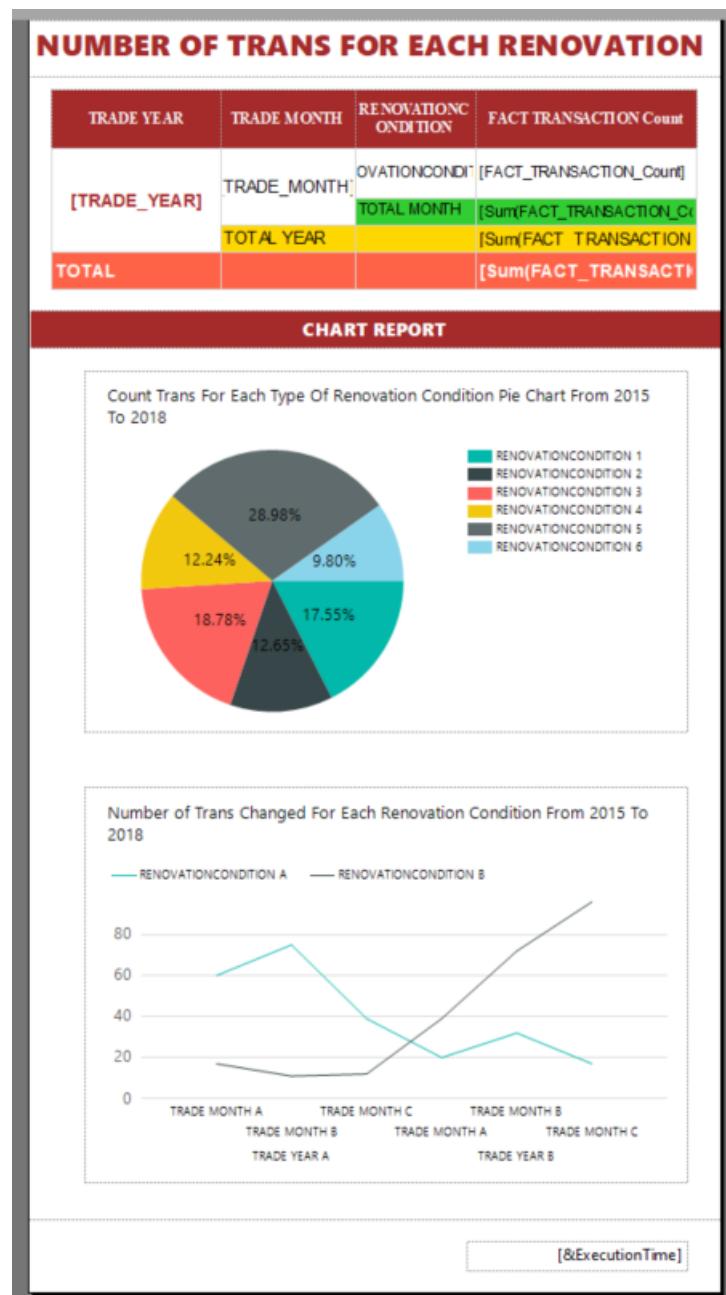
Hình 388: Chọn độ đo thích hợp

- Kết quả trả về của biểu đồ đường



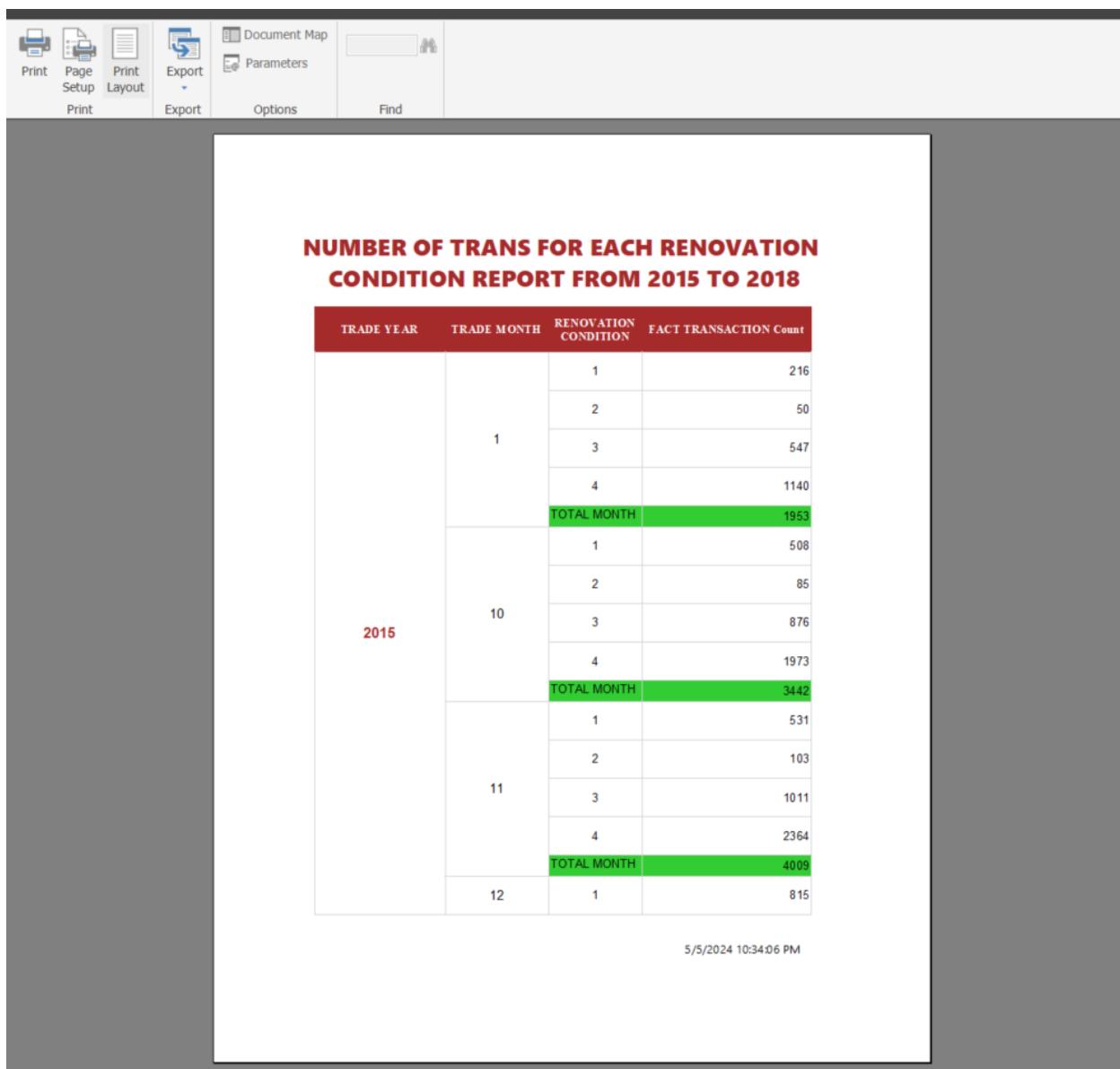
Hình 389: Kết quả trả về

Định dạng report:

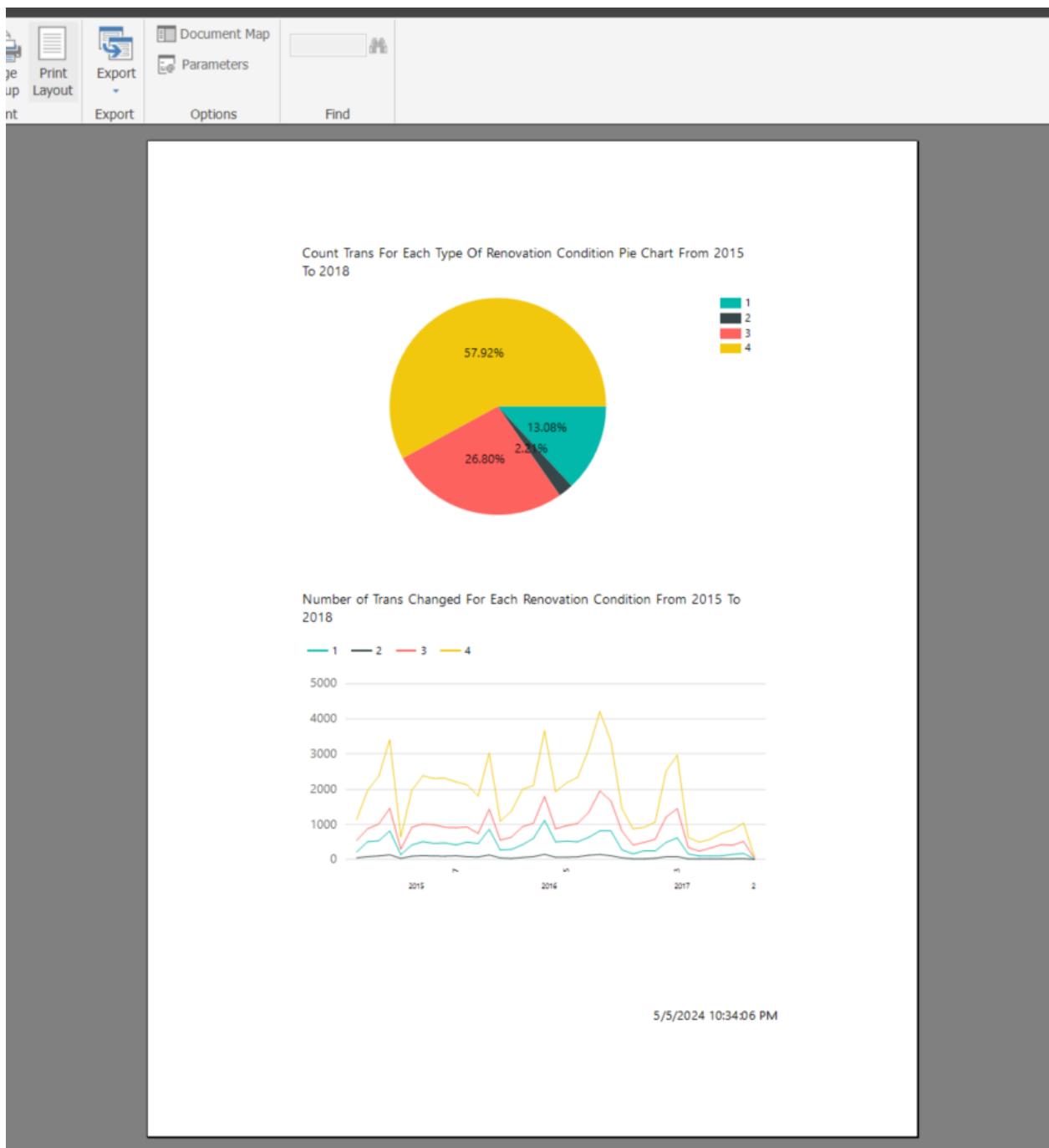


Hình 390: Kết quả Report

Report trong Print Layout

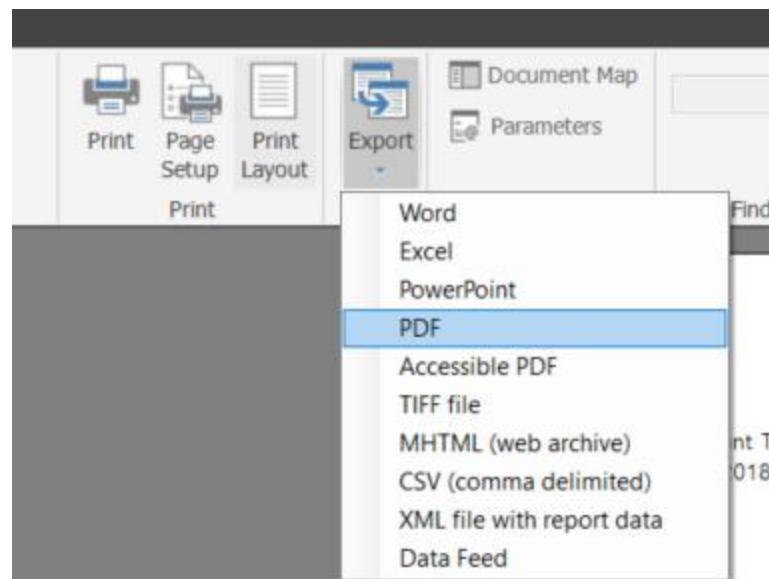


Hình 391: Kết quả trong PrintLayout



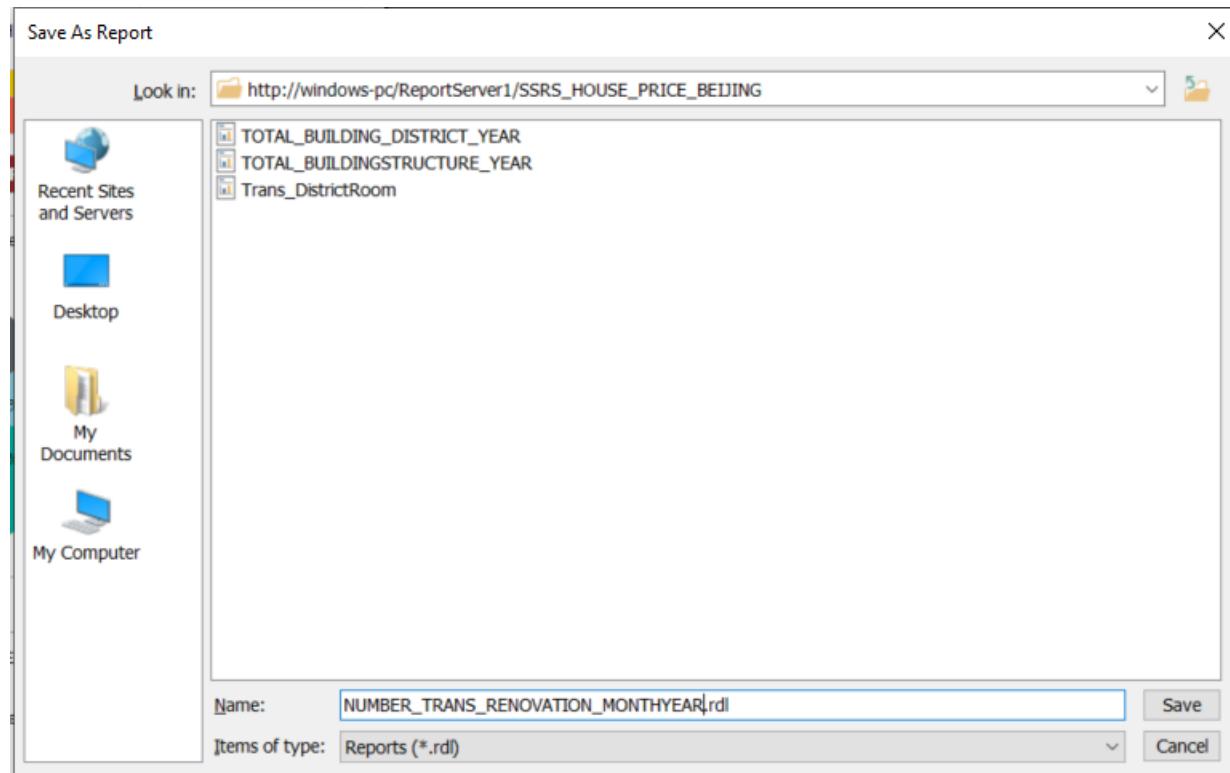
Hình 392: Trang 2 của kết quả

Ta xuất file pdf của report 3:



Hình 393: Xuất file PDF

Nhấn Save dẫn đến đường dẫn → Recent Sites and Servers → Save



Hình 394: Đặt tên và chọn đường dẫn

NUMBER OF TRANS FOR EACH RENOVATION CONDITION REPORT FROM 2015 TO 2018

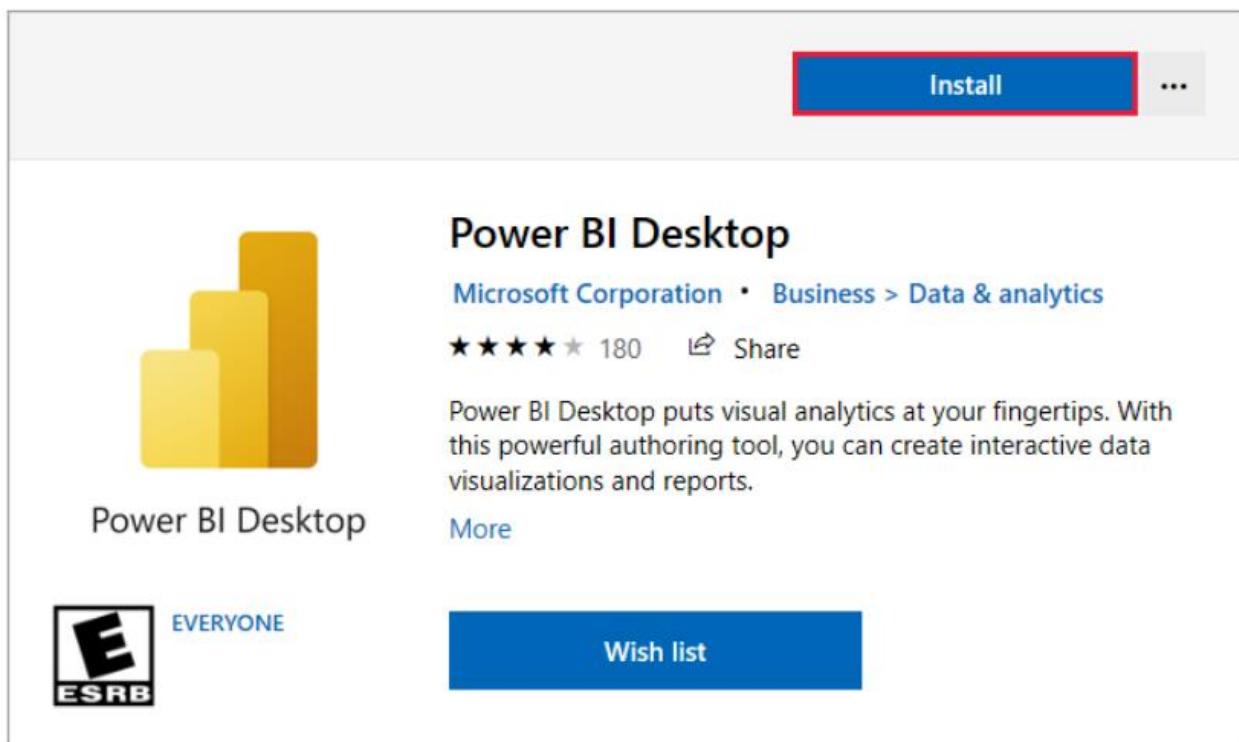
TRADE YEAR	TRADE MONTH	RENOVATION CONDITION	FACT TRANSACTION Count
2015	1	1	216
		2	50
		3	547
		4	1140
		TOTAL MONTH	1953
	10	1	508
		2	85
		3	876
		4	1973
		TOTAL MONTH	3442
	11	1	531
		2	103
		3	1011
		4	2364
		TOTAL MONTH	4009
	12	1	815
		2	136
		3	1455
		4	3408
		TOTAL MONTH	5814
		1	140

Hình 395: In report 3 trên Report Server

3. Quá trình lập báo biểu bằng công cụ Power BI

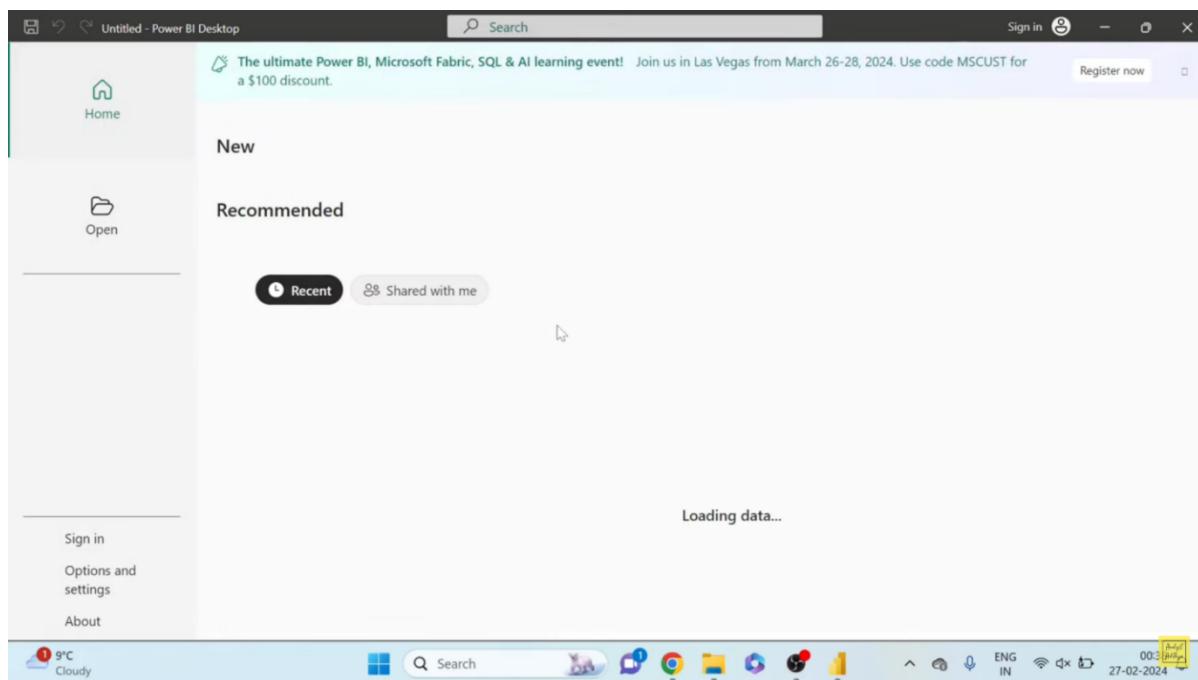
3.1. Cài đặt Power BI và kết nối với SQL Server

Bước 1: Mở Microsoft Store và tìm “Power BI Desktop”, sau đó bấm **install** để cài đặt.



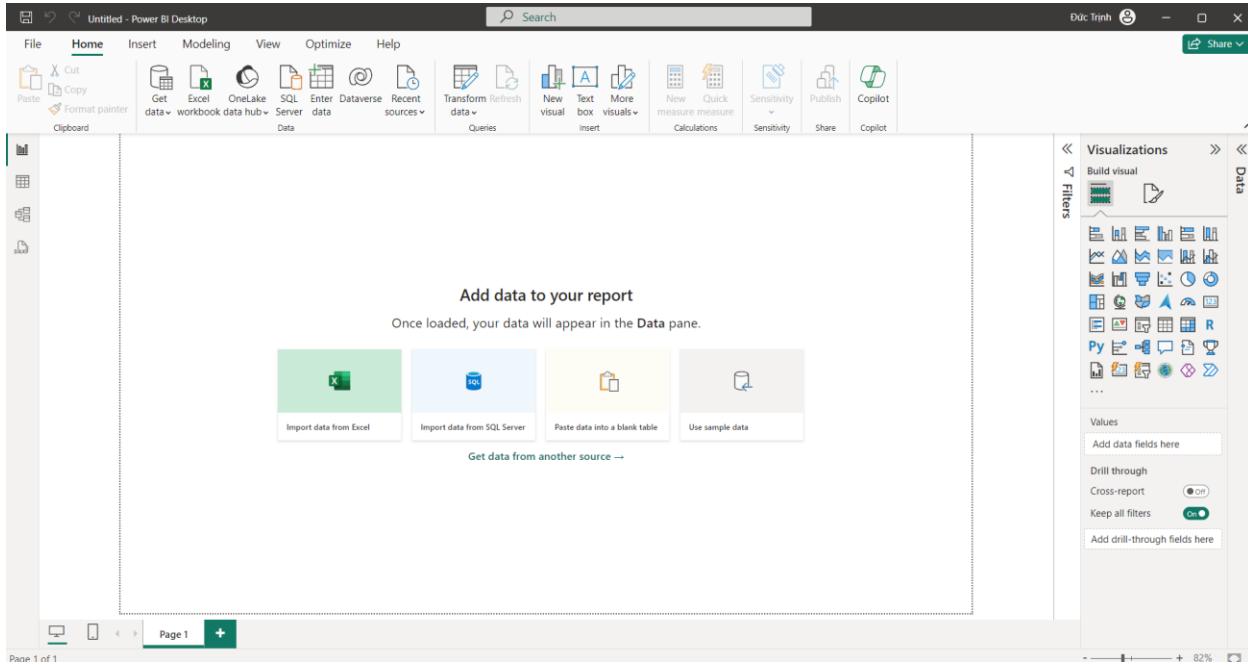
Hình 396: Bấm Install

Bước 2: Mở Power BI sau khi cài đặt thành công, bấm vào nút **Sign in** ở góc dưới bên trái màn hình để tiến hành đăng nhập



Hình 397: Bấm nút Sign in

Bước 3: Sau khi đăng nhập thành công, bấm **New Report** giao diện Power BI sẽ nhìn như sau:



Hình 398: Giao diện Power BI

Bước 4: Bấm vào nút **SQL Server** trên thanh công cụ, sau đó nhập tên server Sql và tên Database tương ứng



Hình 399: Nhập tên server Sql và tên Database tương ứng

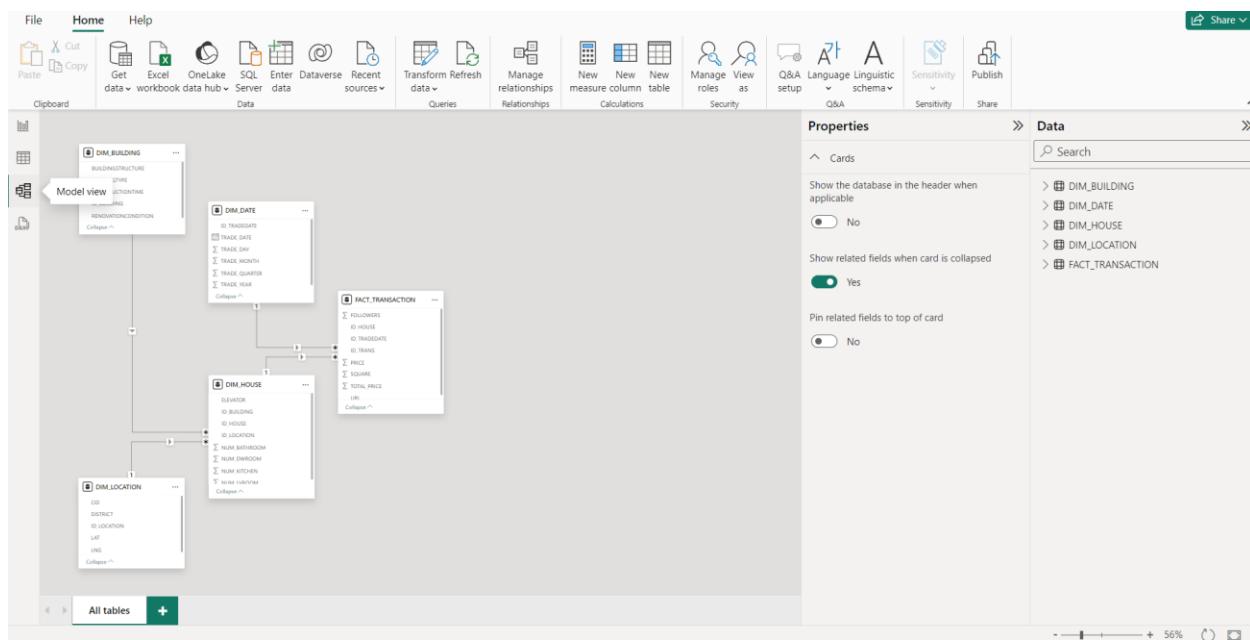
Bước 5: Chọn các bảng Fact và Dim để import, sau khi hoàn thành bấm Load

The screenshot shows the Power BI Navigator window. On the left, there is a tree view of tables under 'DESIRE2U: HOUSE_PRICE_BEIJING [11]'. Several tables are selected: DIM_BUILDING, DIM_DATE, DIM_HOUSE, FACT_TRANSACTION, and DIM_LOCATION. The table 'DIM_LOCATION' is currently highlighted. To the right of the tree view is a large table preview titled 'DIM_LOCATION' with columns ID_LOCATION, LNG, LAT, CID, DISTRICT, and DIM_HC. The preview shows 23 rows of data. At the bottom of the window are buttons for 'Select Related Tables', 'Load' (which is green), 'Transform Data', and 'Cancel'.

ID_LOCATION	LNG	LAT	CID	DISTRICT	DIM_HC
1	116.072514	39.678855	1.11106E+12	5	Tabl
2	116.073082	39.660414	1.11106E+12	5	Tabl
3	116.077958	39.656478	1.11103E+12	5	Tabl
4	116.081646	39.720928	1.11106E+12	5	Tabl
5	116.111318	39.949921	1.11103E+12	12	Tabl
6	116.115806	40.235586	1.11106E+12	6	Tabl
7	116.116904	39.910865	1.11105E+12	12	Tabl
8	116.117019	39.93386	1.11105E+12	12	Tabl
9	116.117724	39.941152	1.11105E+12	12	Tabl
10	116.118409	39.937888	1.11104E+12	12	Tabl
11	116.119651	39.934504	1.11103E+12	12	Tabl
12	116.121949	40.215899	1.11103E+12	6	Tabl
13	116.121964	39.939762	1.11103E+12	12	Tabl
14	116.122115	39.932268	1.11103E+12	12	Tabl
15	116.122354	39.937384	1.11103E+12	12	Tabl
16	116.122491	39.939735	1.11103E+12	12	Tabl
17	116.128533	39.723864	1.11103E+12	5	Tabl
18	116.128779	40.242216	1.11106E+12	6	Tabl
19	116.129677	40.252758	1.1111E+12	6	Tabl
20	116.131543	40.25064	1.11106E+12	6	Tabl
21	116.132633	40.240934	1.11106E+12	6	Tabl
22	116.132908	40.248933	1.11106E+12	6	Tabl
23	116.134232	40.251337	1.11106E+12	6	Tabl

Hình 400: Chọn các bảng Fact và Dim để import

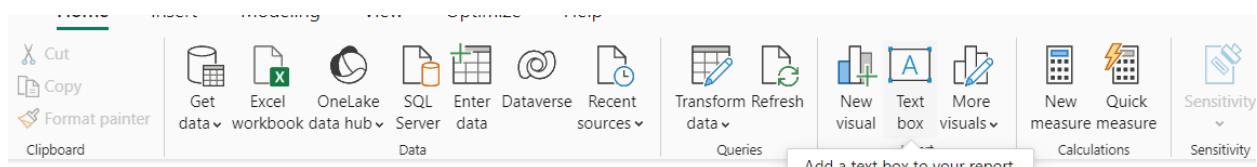
Bước 6: Sau khi load thành công, có thể kiểm tra lại bằng cách bấm vào mục Model View để đảm bảo đã import đúng



Hình 401: Giao diện Model View

3.2. Báo biểu 1: Thống kê tổng số tiền giao dịch qua từng năm từ 2011 đến 2017

Bước 1: Tạo text box để ghi tiêu đề cho báo biểu



Hình 402: Add Textbox

Bước 2: Từ mục **Visualizations**, chọn tạo mới 1 Clustered column chart. Sau đó kéo **TRADE_YEAR** từ **Data** vào cột X, và **Sum of TOTAL_PRICE** vào cột Y.

The screenshot shows the Power BI interface with the 'Visualizations' pane on the left and the 'Data' pane on the right.

Visualizations pane:

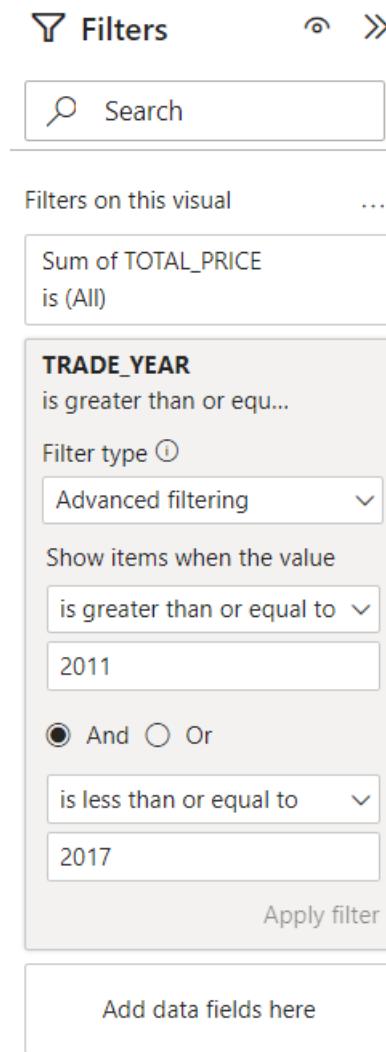
- Build visual:** Shows a preview of a 'Clustered column chart'.
- Visualizations gallery:** A grid of icons representing various chart types, with 'Clustered column chart' highlighted.
- X-axis:** Set to 'TRADE_YEAR'.
- Y-axis:** Set to 'Sum of TOTAL_PRICE'.
- Legend:** Placeholder 'Add data fields here'.
- Small multiples:** Placeholder 'Add data fields here'.
- Toolips:** Placeholder 'Add data fields here'.

Data pane:

- Search:** Search bar.
- DIM_DATE:** Expanded node.
 - ID_TRADEDATE
 - TRADE_DATE
 - Σ TRADE_DAY
 - Σ TRADE_MONTH
 - Σ TRADE_QUAR...
 - Σ TRADE_YEAR
- DIM_LOCATION:** Expanded node.
 - CID
 - DISTRICT
 - ID_LOCATION
 - LAT
 - LNG
- FACT_TRANSACTION:** Expanded node.
 - Σ FOLLOWERS
 - ID_HOUSE
 - ID_TRADEDATE
 - ID_TRANS
 - Σ PRICE
 - Σ SQUARE
 - Σ TOTAL_PRICE
 - URL

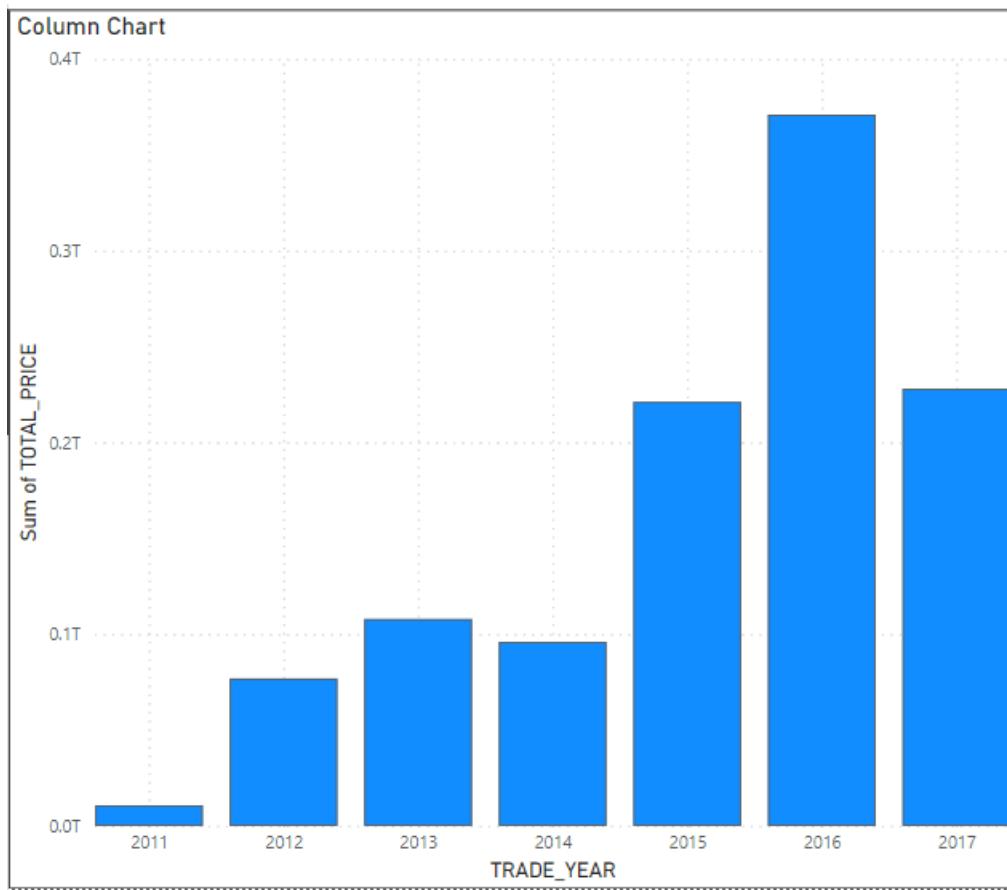
Hình 403: Chọn Clusterd column chart và kéo thuộc tính tương ứng

Bước 3: Nhìn sang phần **Filter**, bấm chọn **TRADE_YEAR** để thêm điều kiện, vì chúng ta muốn thống kê từ năm 2011 đến 2017 nên sẽ cài chọn thông tin có giá trị lớn hơn hoặc bằng 2011 và thông tin có giá trị nhỏ hơn hoặc bằng 2017.



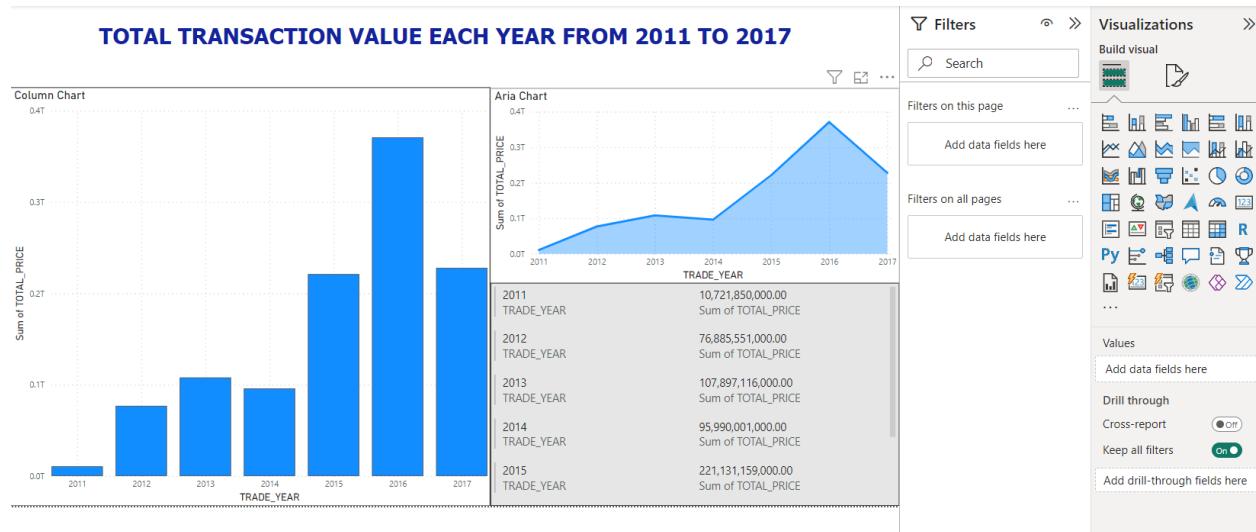
Hình 404: Thêm điều kiện

- Chúng ta sẽ có kết quả như sau:



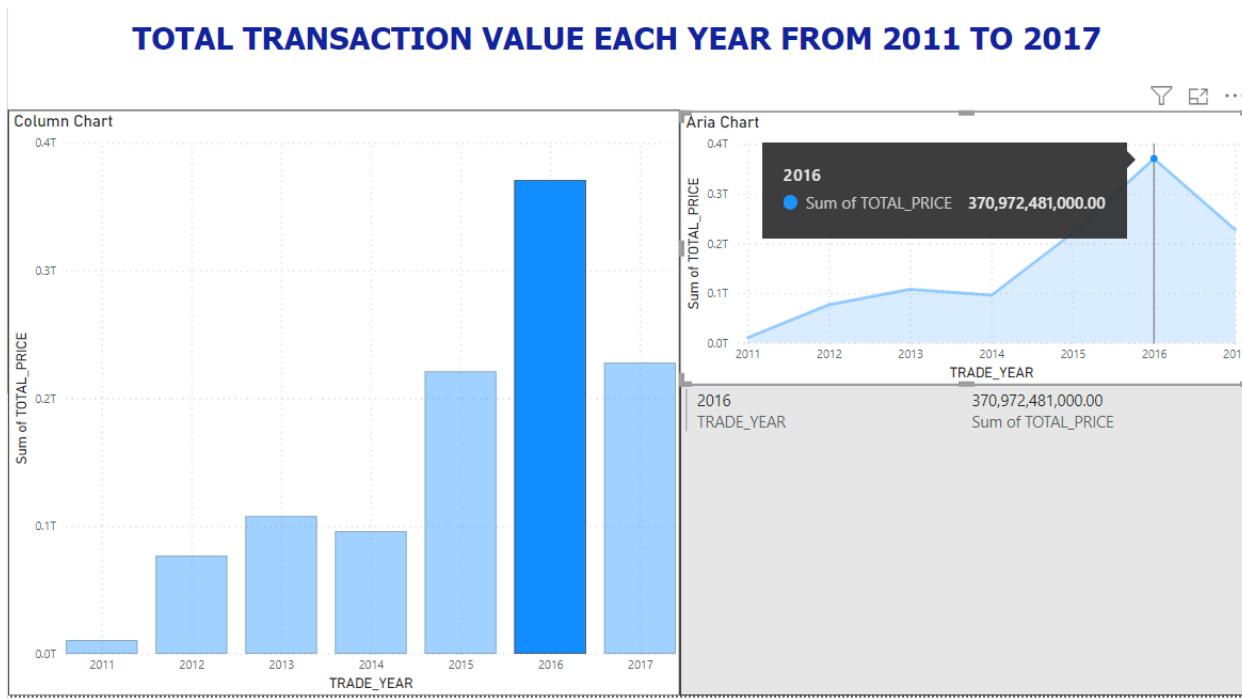
Hình 405: Kết quả biểu đồ

Bước 4: Làm tương tự với các **Visualizations Area Chart** và **Multi-row Card**, sau đó chúng ta sẽ có kết quả như sau:



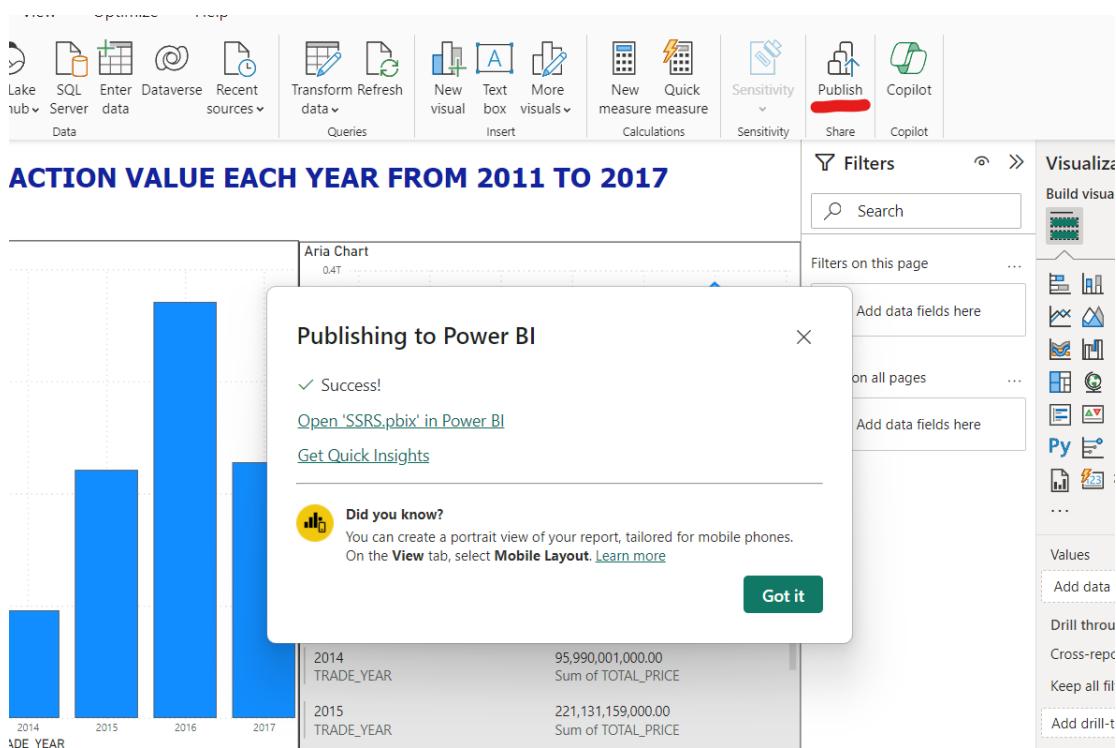
Hình 406: Kết quả sau khi thêm Visualizations Area Chart và Multi-row Card

- Có thể chọn các thời điểm khác nhau để thấy kết quả cụ thể hơn như hình vẽ



Hình 407: Hiển thị chi tiết hơn

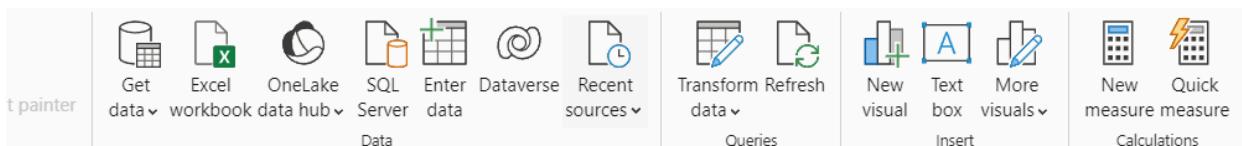
Bước 5: Publish report, sau khi thành công sẽ như sau



Hình 408: Publish thành công

3.3. Báo biểu 2: Thống kê diện tích bất động sản của từng khu vực được giao dịch qua từng năm, chỉ tính những mảnh đất có loại hình kiến trúc là “Bungalow” và “Tower”

Bước 1: Tạo text box ghi tiêu đề cho bảng biểu



TOTAL SQUARE OF BUNGALOW BUILDING IN DISTRICT

Hình 409: Tạo Textbox

Bước 2: Từ mục **Visualizations**, chọn tạo mới 1 **Matrix**. Sau đó thêm **TRADE_YEAR** vào **rows**, **DISTRICT** vào **Columns**, **Sum of SQUARE** vào **Values**.

The screenshot shows the Power BI interface with the 'Visualizations' pane open on the left and the 'Data' pane on the right.

Visualizations pane:

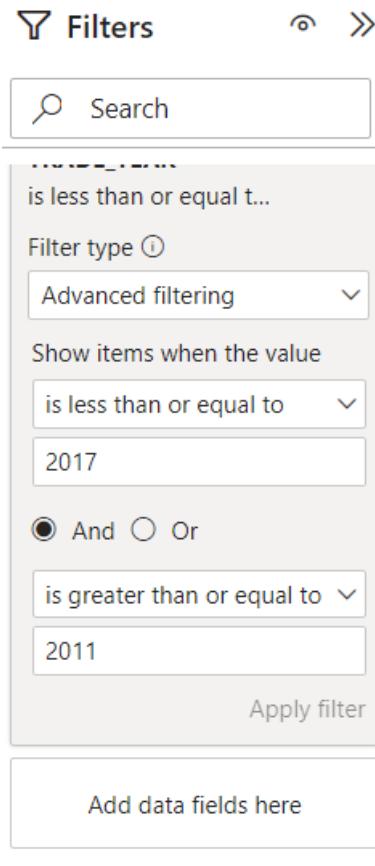
- Build visual:** Shows various chart and matrix icons.
- Rows:** Set to 'TRADE_YEAR'.
- Columns:** Set to 'DISTRICT'.
- Values:** Set to 'Sum of SQUARE'.
- Drill through:** Off.
- Cross-report:** Off.
- Keep all filters:** On.

Data pane:

- Search:** Search bar.
- DIM_BUILDING:** Contains fields: BUILDINGSTR..., BUILDINGTYPE, CONSTRUCTI..., ID_BUILDING, RENOVATION... (none selected).
- DIM_DATE:** Contains fields: ID_TRADEDATE, TRADE_DATE, TRADE_DAY, TRADE_MONTH, TRADE_QUAR... (TRADE_YEAR is selected).
- DIM_HOUSE:** Contains fields: ELEVATOR, ID_BUILDING, ID_HOUSE, ID_LOCATION, NUM_BATHR..., NUM_DWRO..., NUM_KITCHEN, NUM_LVROOM (none selected).

Hình 410: Tạo mới 1 Matrix

Bước 3: Sử dụng Filter để đặt điều kiện cần thiết cho TRADE_YEAR, BUILDINGTYPE. Với BUILDINGTYPE chỉ nhận chỉ số 1, 2 tương ứng với 2 loại nhà “Tower” và “Bungalow”



Filters on this page ...

BUILDINGTYPE

is 1 or is 2

Add data fields here

Hình 411: Sử dụng filter

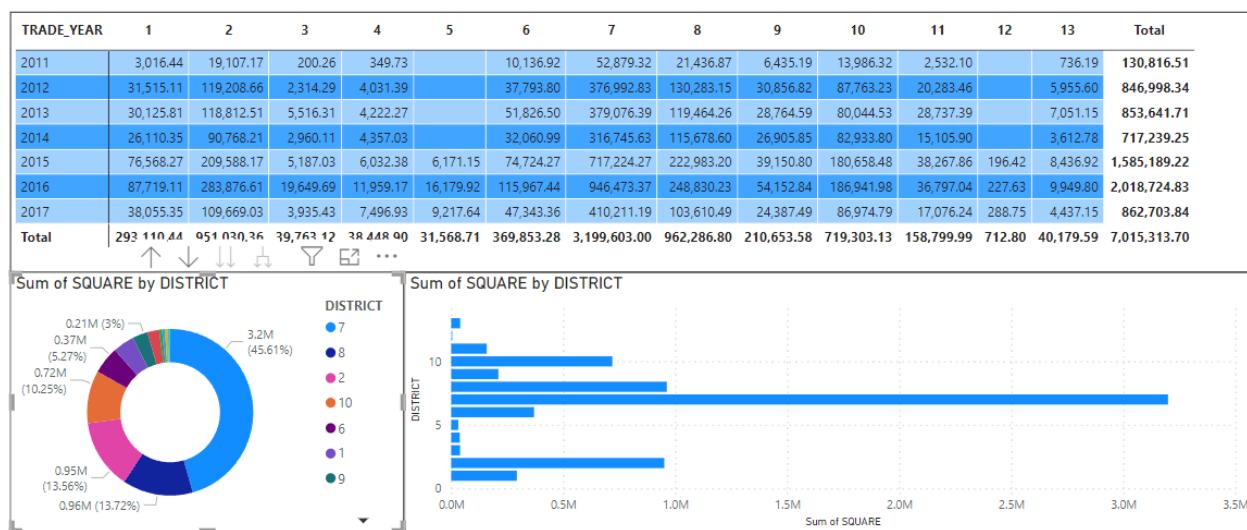
- Ta được kết quả như sau

TRADE_YEAR	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
2011	3,016.44	19,107.17	200.26	349.73		10,136.92	52,879.32	21,436.87	6,435.19	13,986.32	2,532.10		736.19	130,816.51
2012	31,515.11	119,208.66	2,314.29	4,031.39		37,793.80	376,992.83	130,283.15	30,856.82	87,763.23	20,283.46		5,955.60	846,998.34
2013	30,125.81	118,812.51	5,516.31	4,222.27		51,826.50	379,076.39	119,464.26	28,764.59	80,044.53	28,737.39		7,051.15	853,641.71
2014	26,110.35	90,768.21	2,960.11	4,357.03		32,060.99	316,745.63	115,678.60	26,905.85	82,933.80	15,105.90		3,612.78	717,239.25
2015	76,568.27	209,588.17	5,187.03	6,032.38	6,171.15	74,724.27	717,242.27	222,983.20	39,150.80	180,658.48	38,267.86	196.42	8,436.92	1,585,189.22
2016	87,719.11	283,876.61	19,649.69	11,959.17	16,179.92	115,967.44	946,473.37	248,830.23	54,152.84	186,941.98	36,797.04	227.63	9,949.80	2,018,724.83
2017	38,055.35	109,669.03	3,935.43	7,496.93	9,217.64	47,343.36	410,211.19	103,610.49	24,387.49	86,974.79	17,076.24	288.75	4,437.15	862,703.84
Total	293,110.44	951,030.36	39,763.12	38,448.90	31,568.71	369,853.28	3,199,603.00	962,286.80	210,653.58	719,303.13	158,799.99	712.80	40,179.59	7,015,313.70

Hình 412: Kết quả sau khi dùng filter

Bước 4: Thực hiện tương tự để tạo ra **Visualizations Donut Chart** và **Stacked Bar Chart**, ta được kết quả như sau

TOTAL SQUARE OF TOWER AND BUNGALOW BUILDING IN DISTRICT



Hình 413: Kết quả sau khi thêm Visualizations Donut Chart và Stacked Bar Chart

Bước 5: Thực hiện Publish report

TOTAL SQUARE OF TOWER AND BUNGALOW BUILDING IN DISTRICT

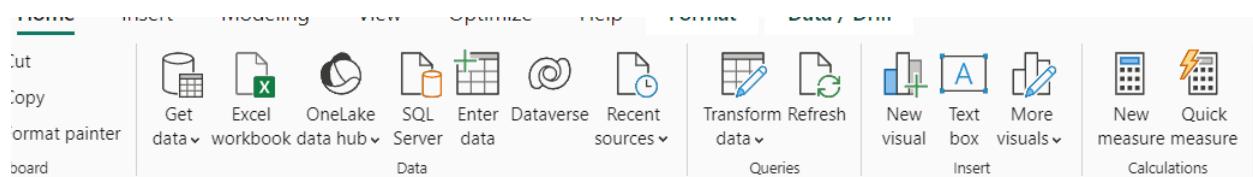
TRADE_YEAR	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
2011	3,016.44	19,107.17	200.26	349.73		10,136.92	52,879.32	21,436.87	6,435.19	13,986.32	2,532.10		736.19	130,816.51
2012	31,515.11	119,208.66	2,314.29	4,031.39		37,793.80	376,992.83	130,283.15	30,856.82	87,763.23	20,283.46		5,955.60	846,998.34
2013	30,125.81	118,812.51	5,516.31	4,222.27		51,826.50	379,076.39	119,464.26	28,764.59	80,044.53	28,737.39		7,051.15	853,641.71
2014	26,110.35	90,768.21	2,960.11	4,357.03		32,060.99	316,745.63	115,678.60	26,905.85	82,933.80	15,105.90		3,612.78	717,239.25
2015	76,568.27	209,588.17	5,187.03	6,032.38	6,171.15	74,724.27	717,224.27	222,983.20	39,150.80	180,658.48	38,267.86	196.42	8,436.92	1,585,189.22
2016	87,719.11	283,876.61	19,649.69	11,959.17	16,179.92	115,967.44	946,473.37	248,830.23	54,152.84	186,941.98	36,797.04	227.63	9,949.80	2,018,724.83
2017	38,055.35	109,669.03	3,935.43	7,496.93	9,217.64	47,343.36	410,211.19	103,610.49	24,387.49	86,974.79	17,076.24	288.75	4,437.15	862,703.84
Total	293,110.44	951,030.36	39,763.12	38,448.90	31,568.71	369,853.28	3,199,603.00	962,286.80	210,653.58	719,303.13	158,799.99	712.80	40,179.59	7,015,313.70

Hình 414: Publish thành công

3.4. Báo biểu 3: Thống kê số lượng và tổng số tiền giao dịch đạt chuẩn chất lượng cao qua từng năm

Note: Được biết giao dịch đạt chuẩn chất lượng cao chỉ bàn giao những ngôi nhà có tình trạng cải tạo đạt mức “**Hardcover**”, được làm bởi các vật liệu có chất lượng cao như thép, bê tông, kết hợp thép và bê tông (**Brick and Concrete, Steel, Steel-Concrete Composite**)

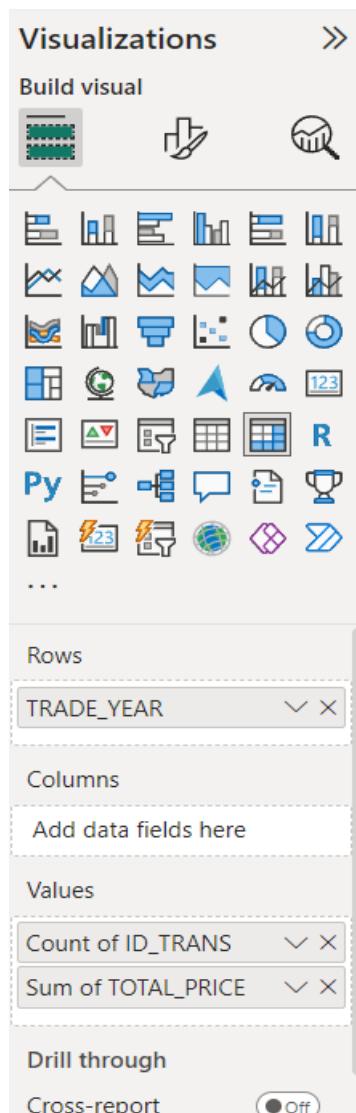
Bước 1: Tạo **text box** ghi tiêu đề cho bảng biểu



STATISTIC OF HIGH QUALITY TRANSACTION

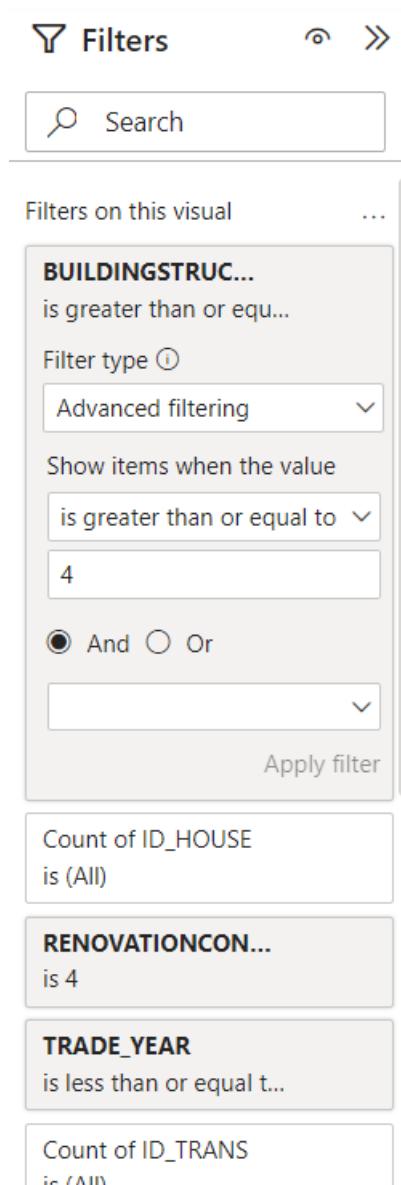
Hình 415: Tạo Textbox

Bước 2: Từ mục **Visualizations**, chọn tạo mới 1 **Matrix**. Sau đó thêm **TRADE_YEAR** vào **rows**, **Count of ID_TRANS**, **Sum of TOTAL_PRICE** vào **Values**.



Hình 416: Tạo mới Matrix và thêm hàng cột tương ứng

Bước 3: Sử dụng **filter** để lọc các điều kiện để đạt được giao dịch đạt tiêu chuẩn cao



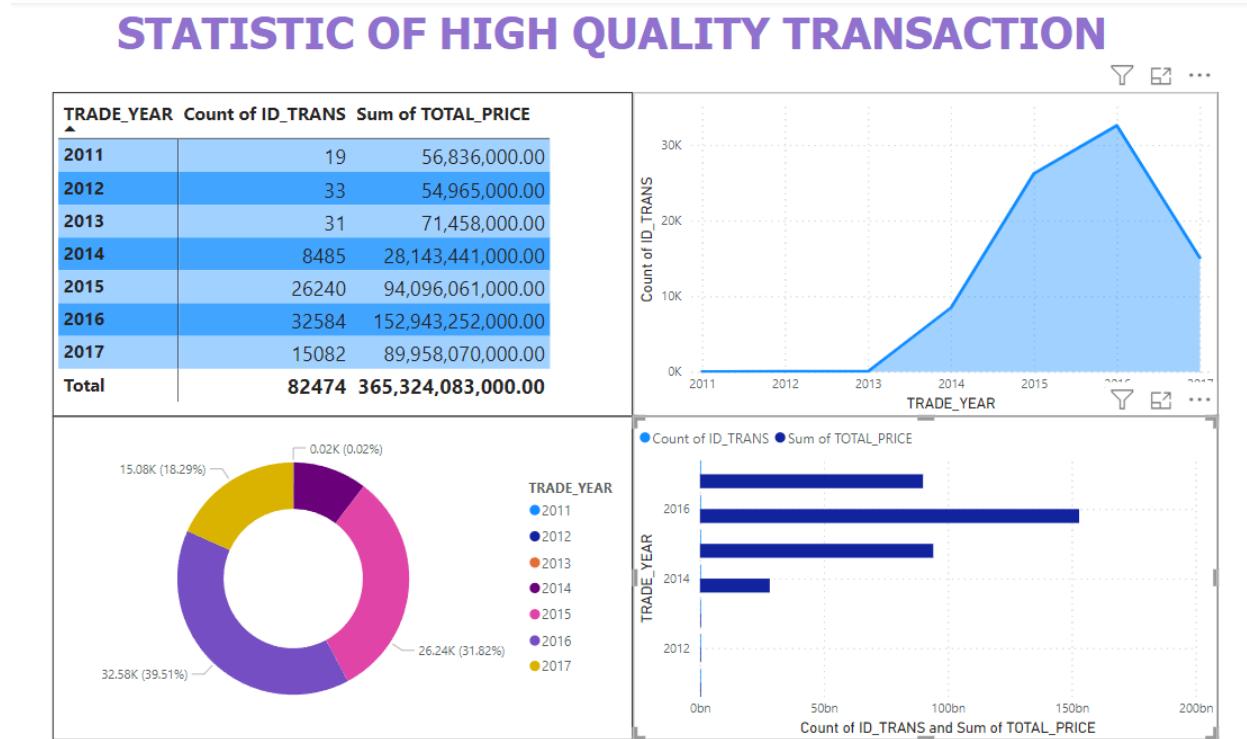
Hình 417: Sử dụng filter

- Chúng ta có kết quả như sau:

TRADE_YEAR	Count of ID_TRANS	Sum of TOTAL_PRICE
2011	19	56,836,000.00
2012	33	54,965,000.00
2013	31	71,458,000.00
2014	8485	28,143,441,000.00
2015	26240	94,096,061,000.00
2016	32584	152,943,252,000.00
2017	15082	89,958,070,000.00
Total	82474	365,324,083,000.00

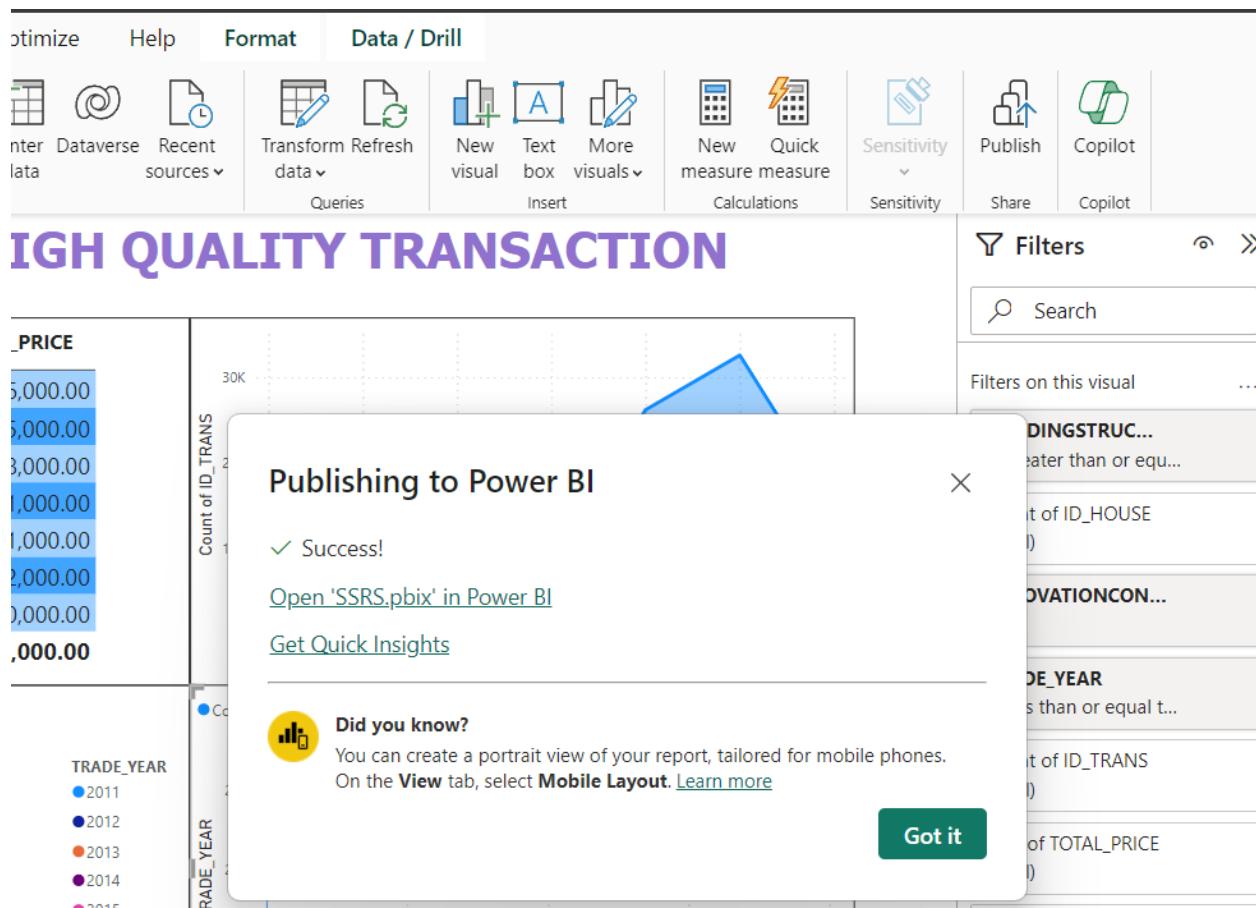
Hình 418: Kết quả sau khi áp Filter

Bước 4: Thực hiện tương tự để tạo ra **Visualizations Donut Chart, Clustered Bar Chart** và **Stacked Bar Chart**, ta được kết quả như sau



Hình 419: Thêm Visualizations Donut Chart, Clustered Bar Chart và Stacked Bar Chart

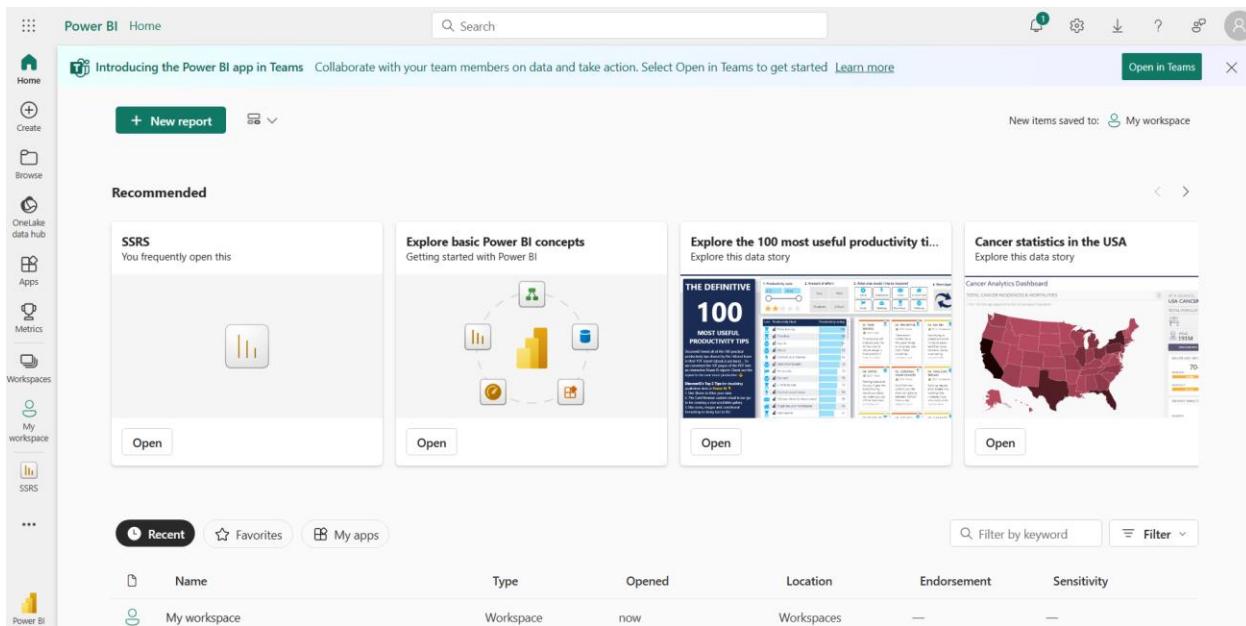
Bước 5: Thực hiện **Publish report**



Hình 420: Publish thành công

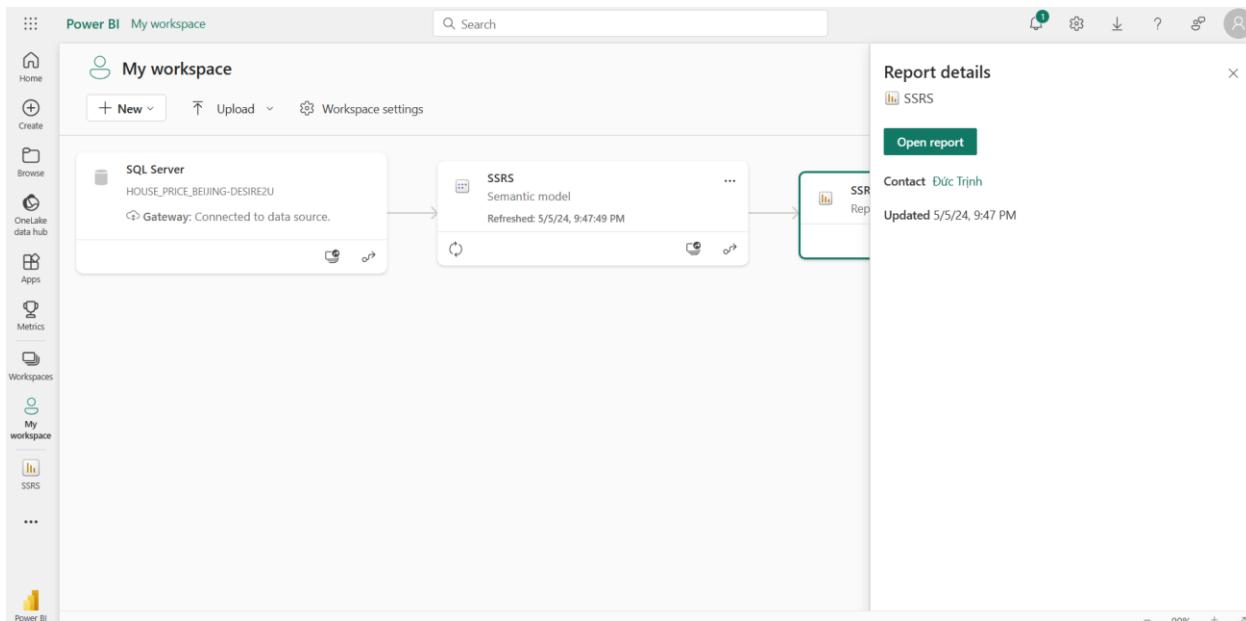
3.5. Cách kiểm tra và xem lại report đã được Publish

Bước 1: Truy cập trang web <https://app.powerbi.com/home?experience=power-bi>



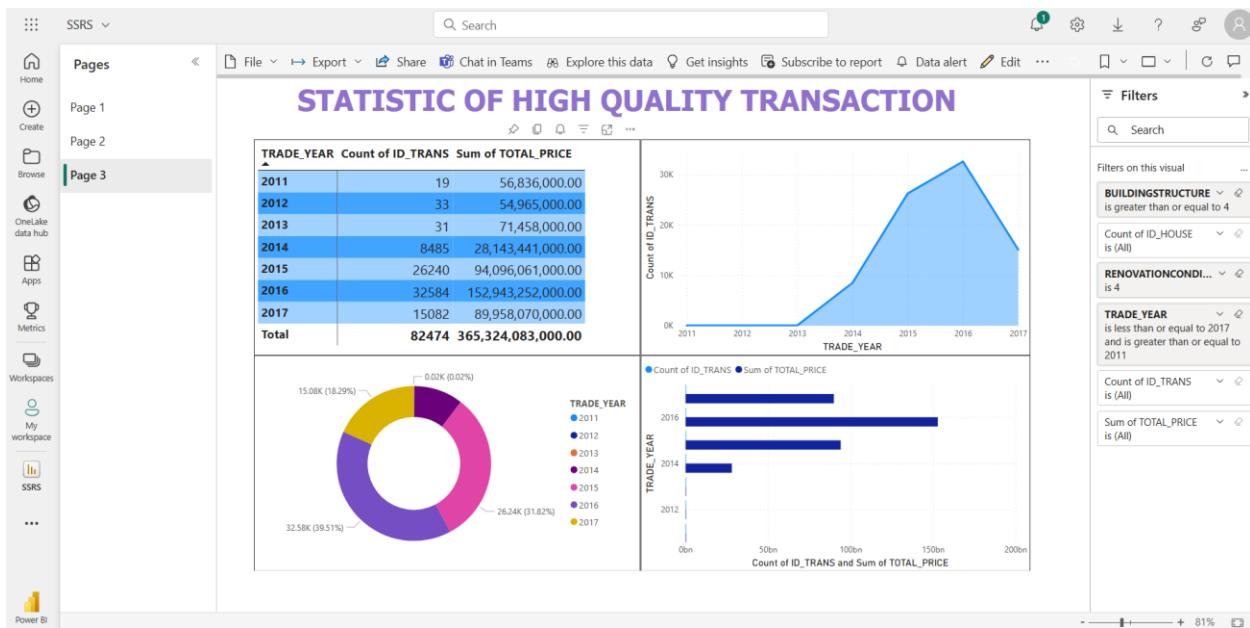
Hình 421: Giao diện web

Bước 2: Lúc trước mình đã lưu report được publish tại **My Workspace** nên lúc này chọn **My Workspace**, sau đó bấm vào xem **report**.



Hình 422: Nhấn Open report

Bước 3: Lúc này bạn có thể xem lại các **report** cũng như **filter** mà mình đã sử dụng

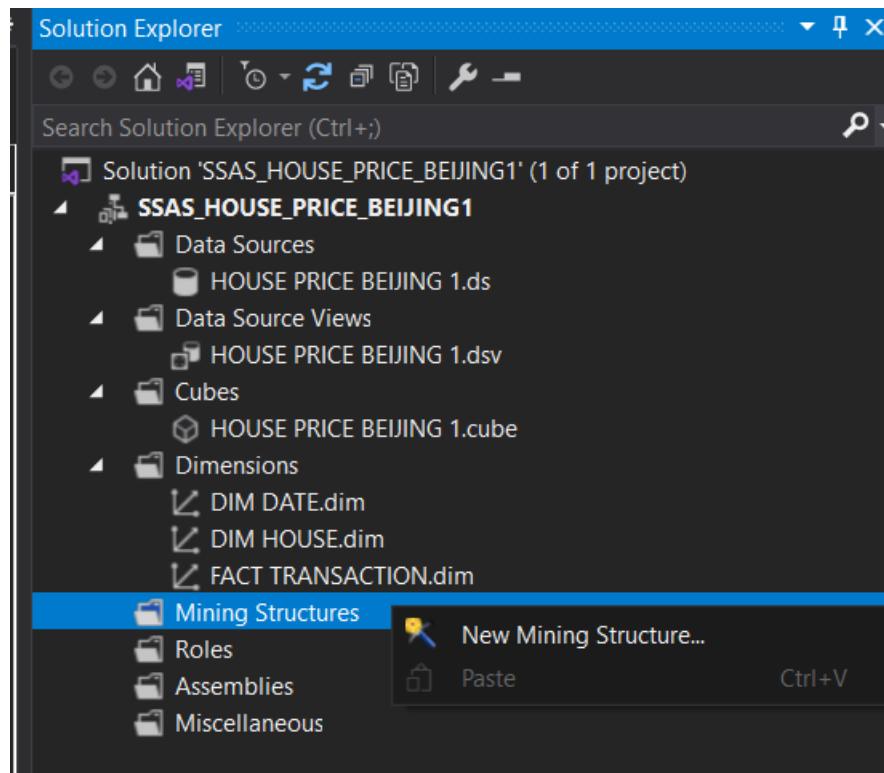


Hình 423: Đang xem lại Báo cáo 3

CHƯƠNG V: DATA MINING

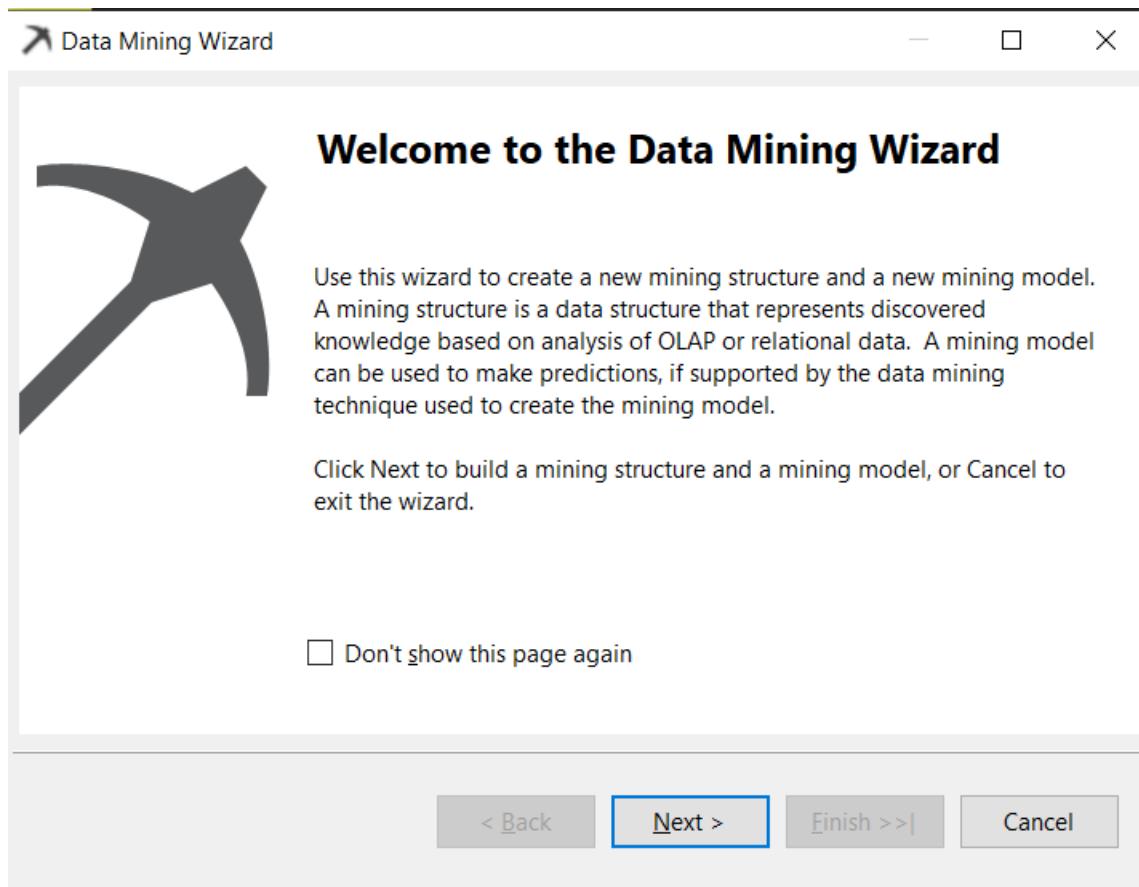
1. Tạo Mining Structures

- Kích chuột phải vào **Mining Tructures** của dự án SSAS và chọn **new Mining Struture**



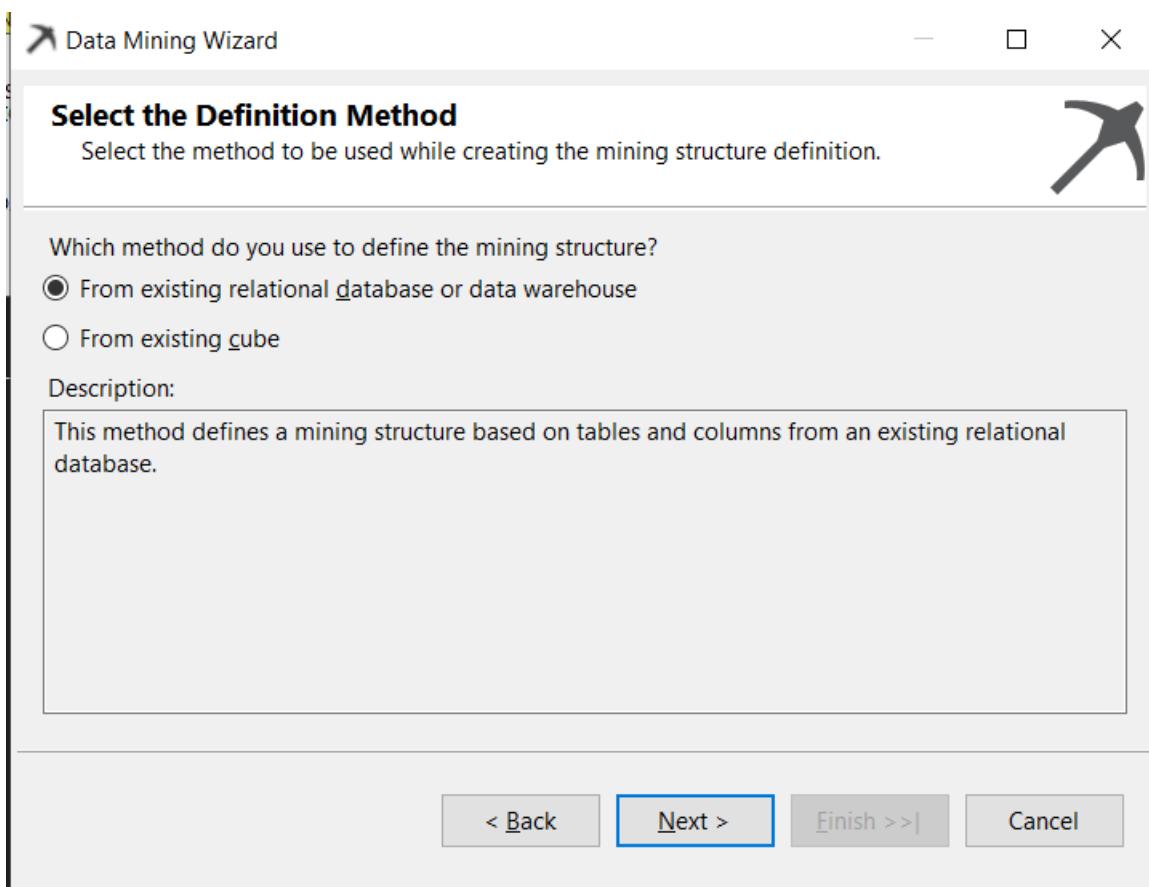
Hình 424: Chọn New Mining Structure

- Nhấn Next:



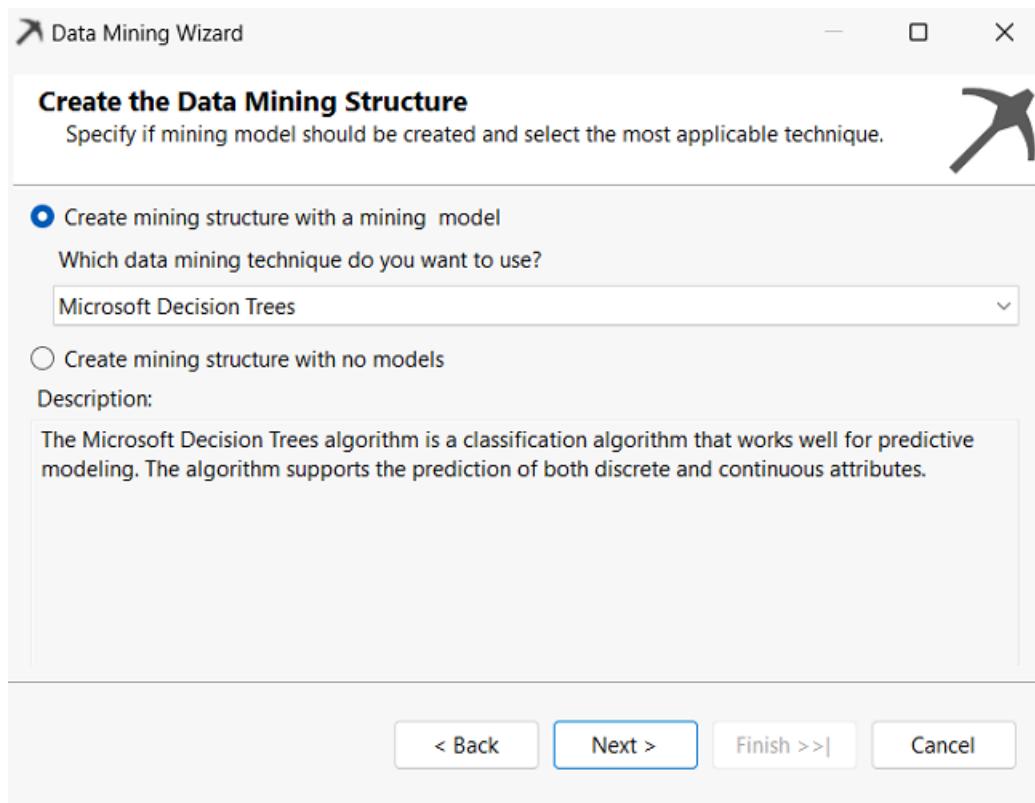
Hình 425: Chọn Next

- Nhấn Next:



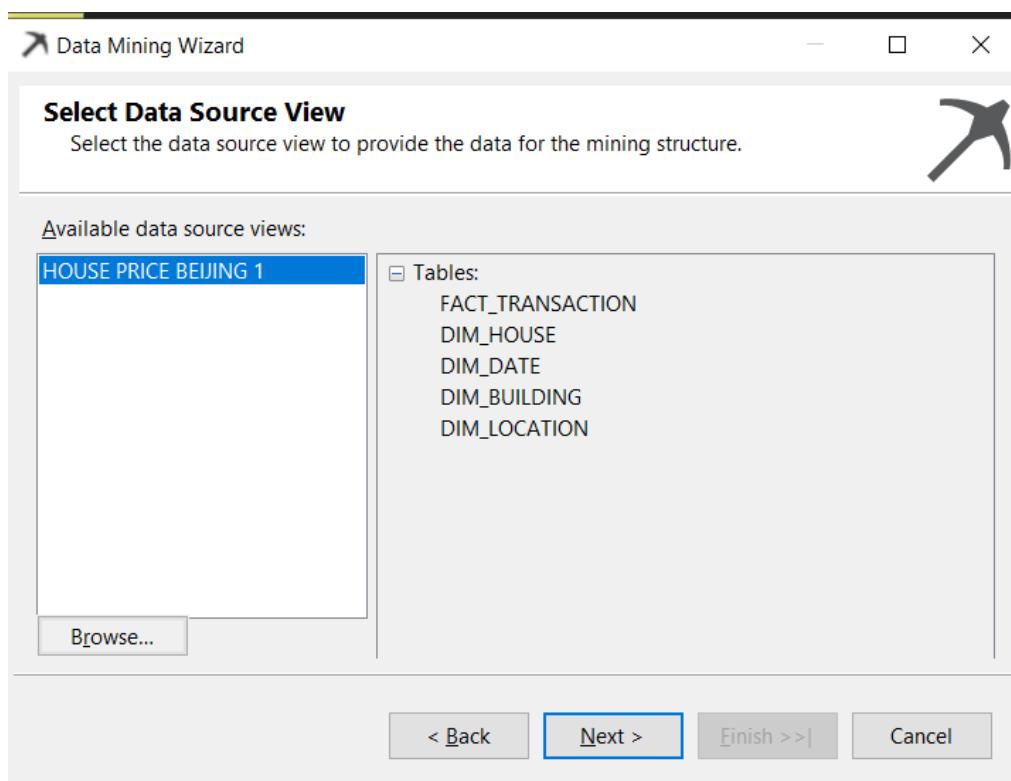
Hình 426: Tích From existing relational or data warehouse

- Chọn thuật toán để mining là **Microsoft Decision Trees** (Thuật toán cây quyết định).



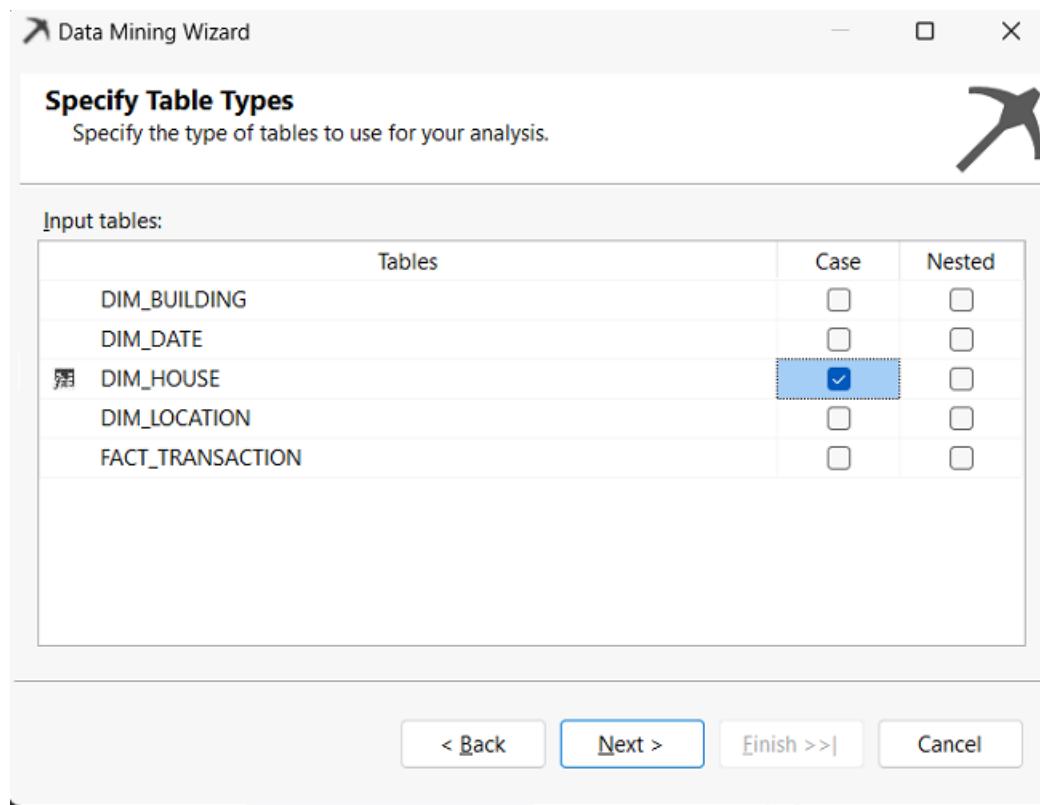
Hình 427: Chọn thuật toán Decision Trees

- Chọn data source **HOUSE PRICE BEIJING 1**



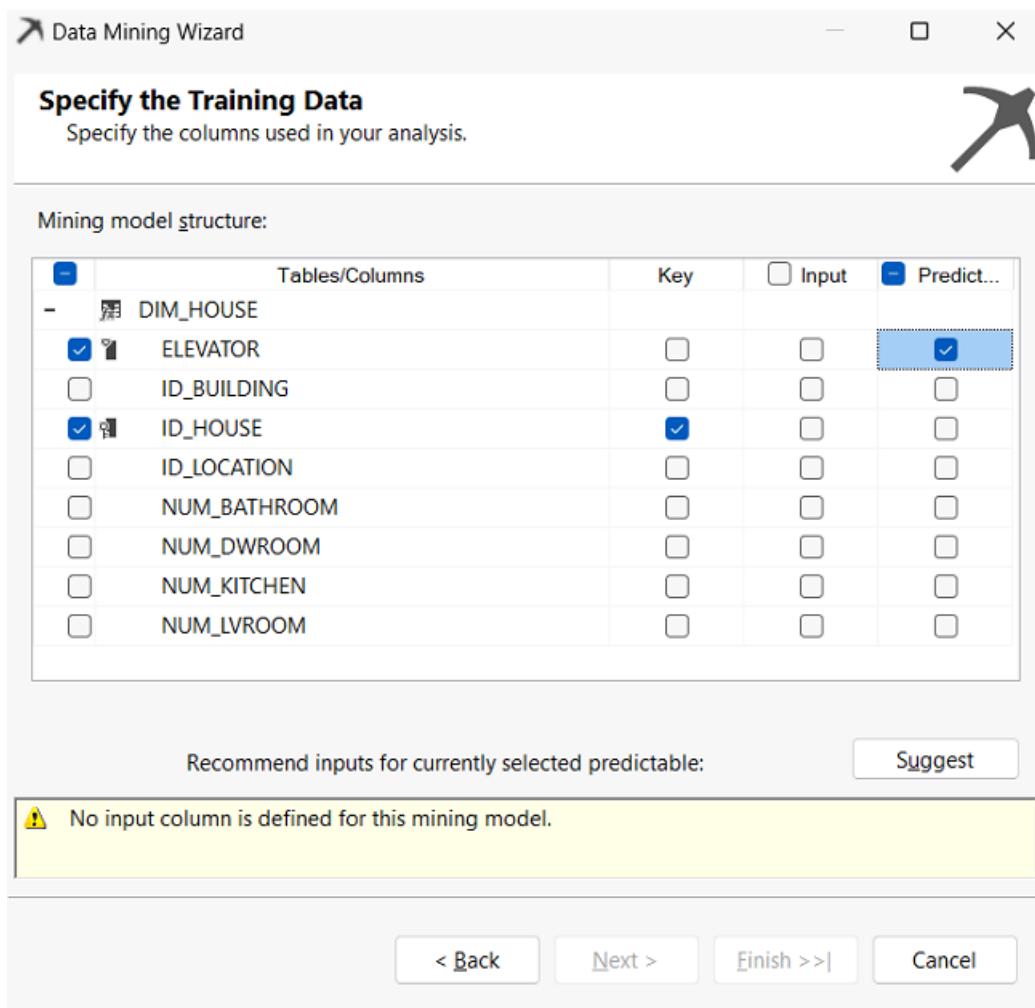
Hình 428: Chọn Data source

- Chọn **DIM_HOUSE** làm case để chạy thuật toán mining



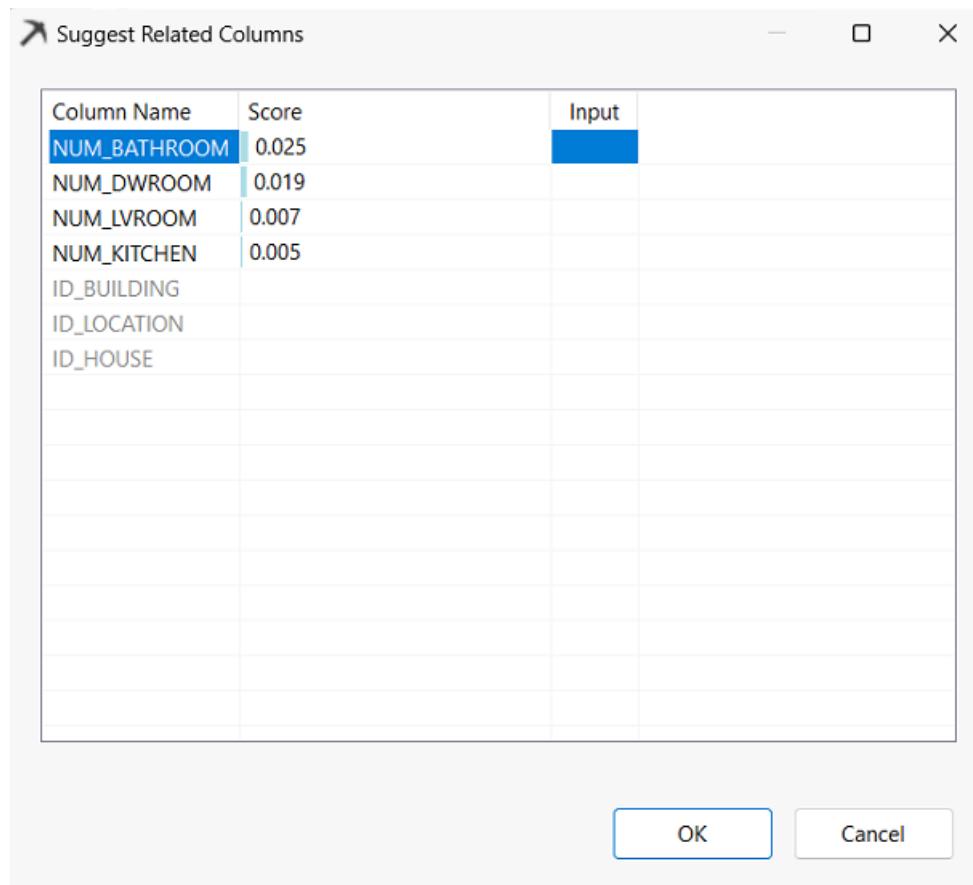
Hình 429: Chọn DIM_HOUSE làm case để chạy thuật toán mining

- Chọn thuộc tính key là **ID_HOUSE**, và thuộc tính **Predict** là **ELAVATOR** (có thang máy hay không)



Hình 430: Chọn thuộc tính key và thuộc tính Predict

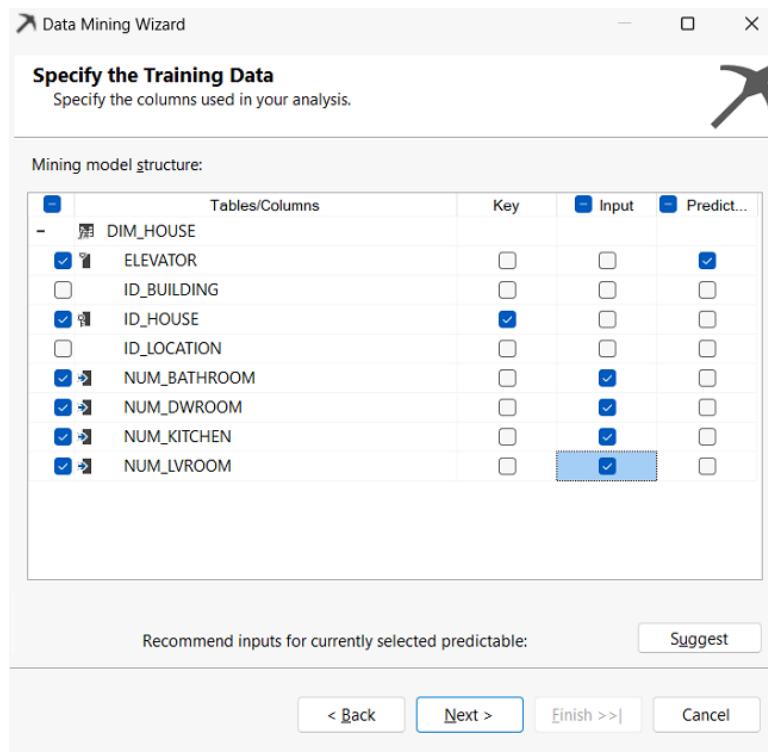
- Xem các thuộc tính được suggest



Hình 431: Xem các thuộc tính được suggest

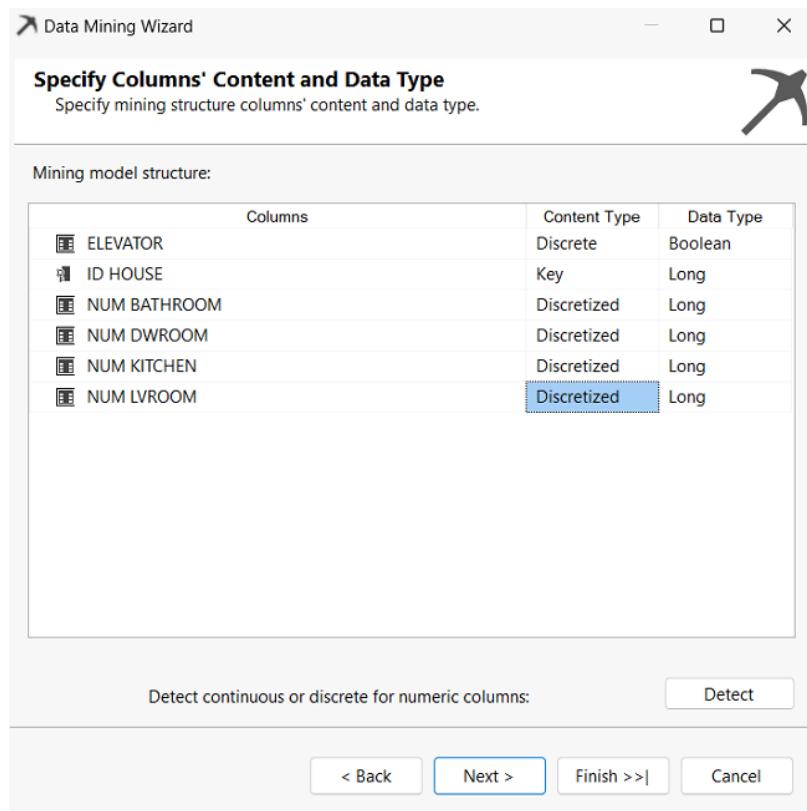
- Chọn các thuộc tính để làm input:

NUM_BATHROOM, NUM_DWROOM, NUM_KITCHEN, NUM_LVROOM

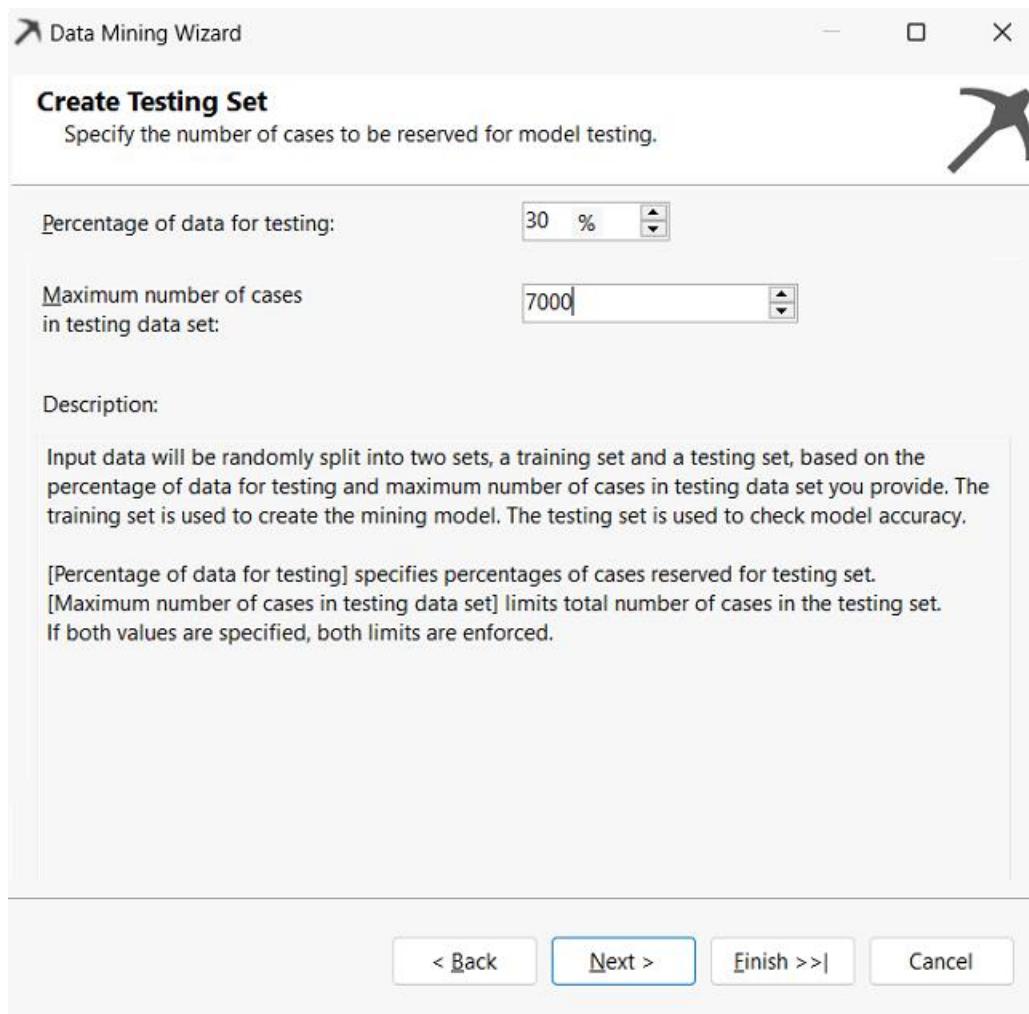


Hình 432: Chọn các thuộc tính để làm input

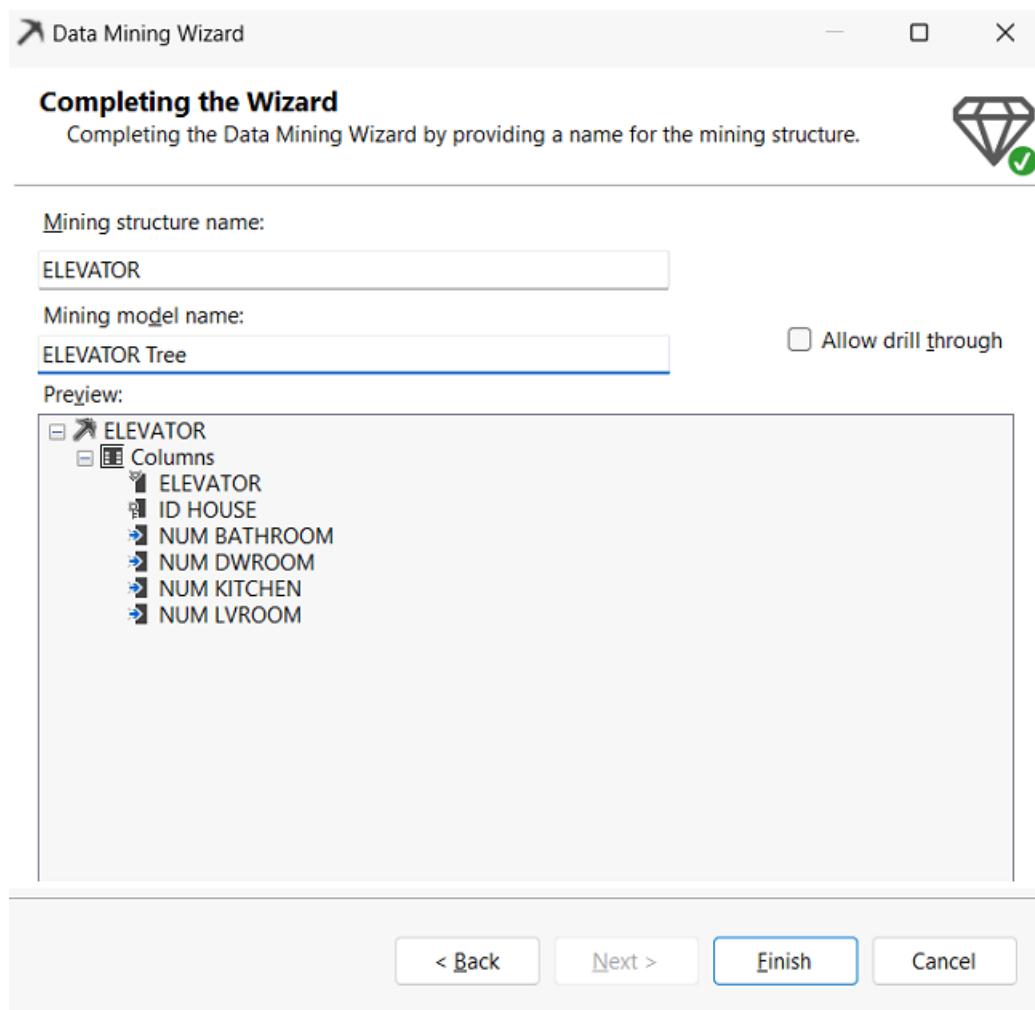
- Detect các trường dữ liệu: chuyển Content Type của các trường **NUM BATHROOM**, **NUM DWROOM**, **NUM KITCHEN**, **NUM LVROOM** từ **Continuous** sang **Discretized**



Hình 433: Detect các trường dữ liệu

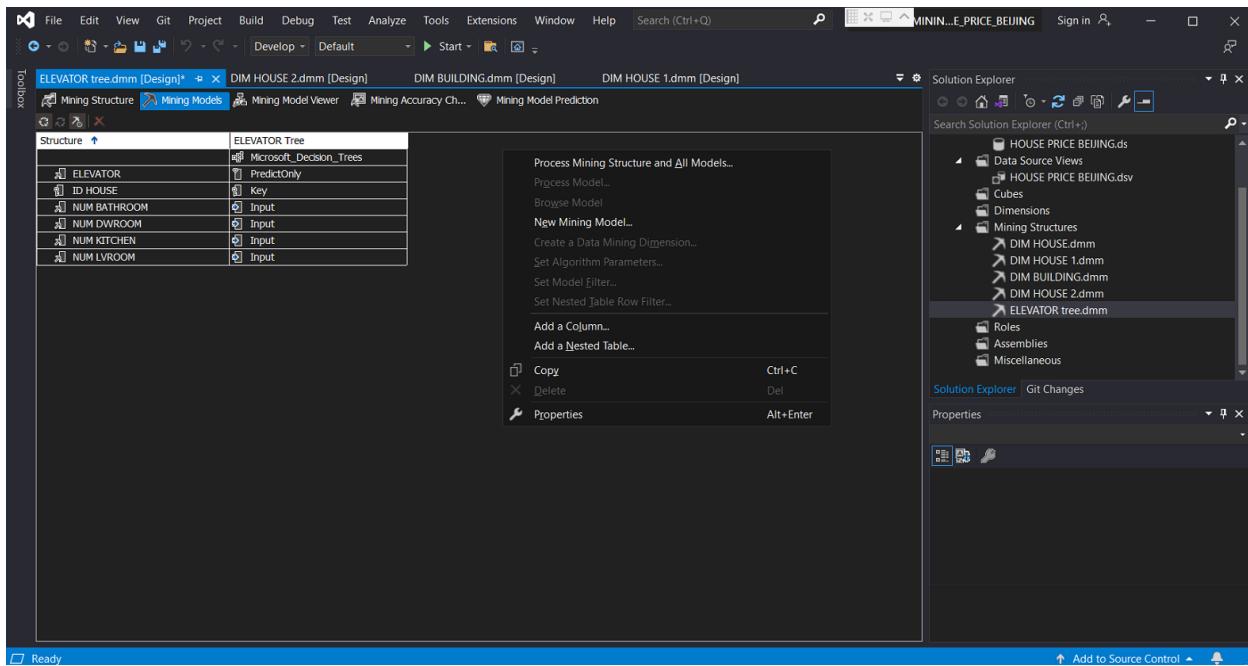


Hình 434: Chọn các thông số thích hợp



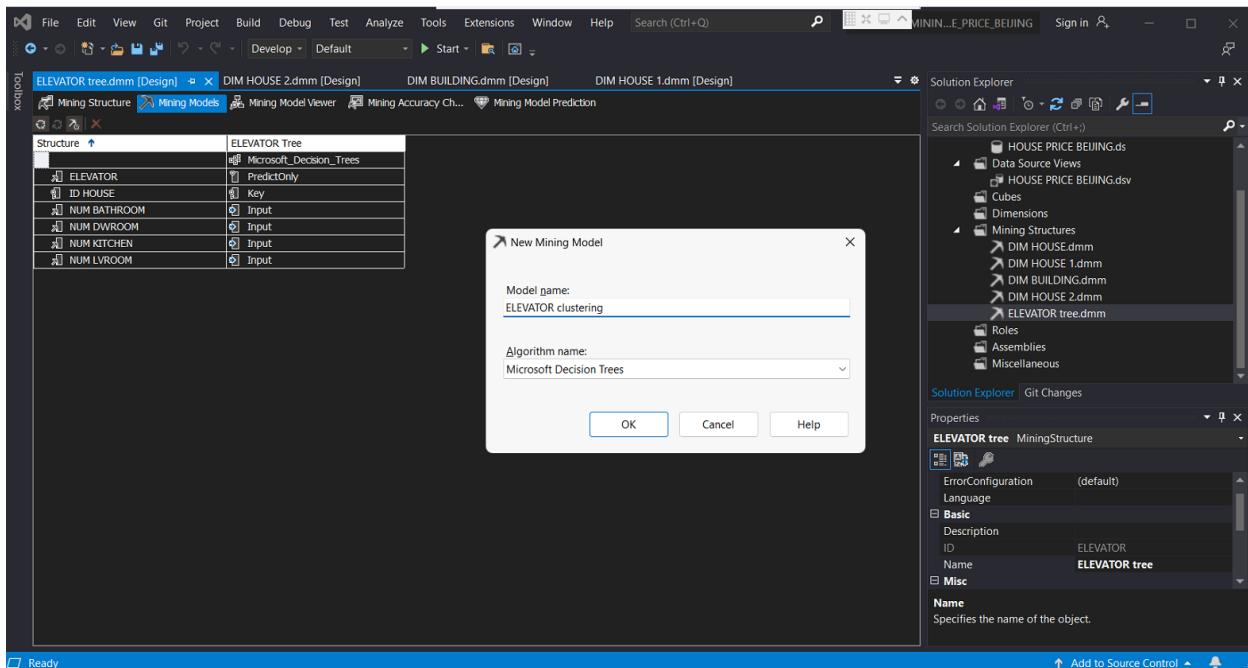
Hình 435: Nhấn Finish

- Thêm thuật toán mới

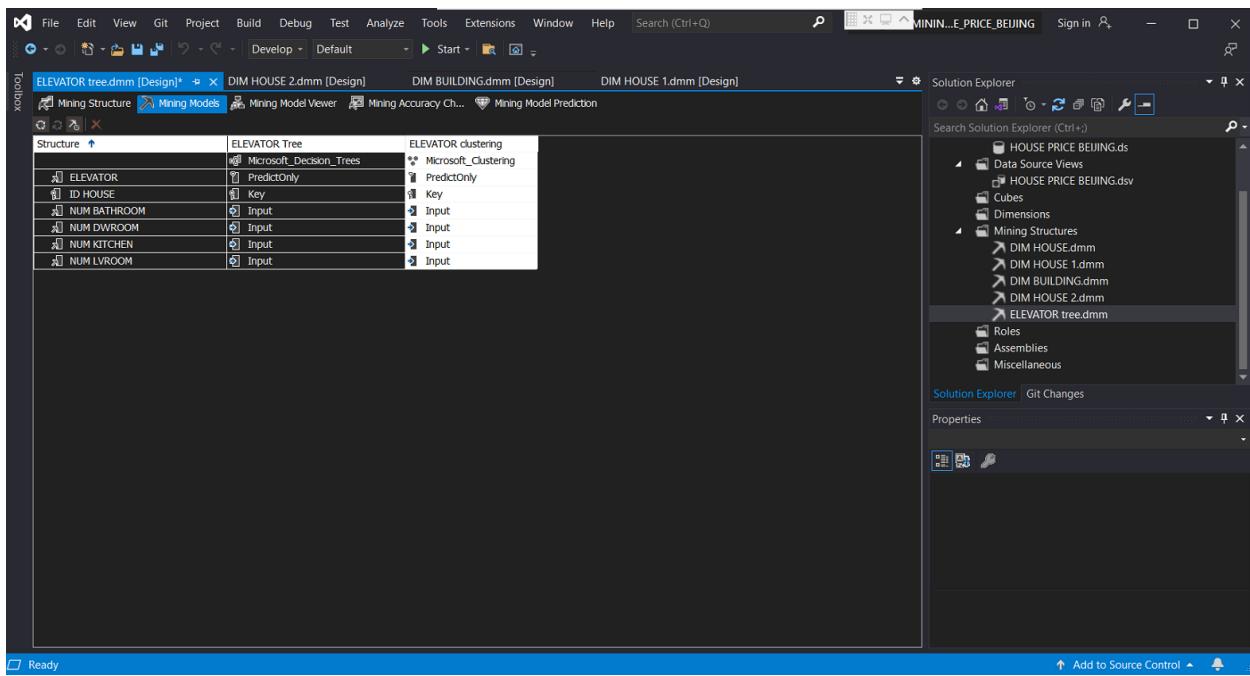


Hình 436: Thêm Mining Model

- Thêm thuật toán Microsoft Clustering

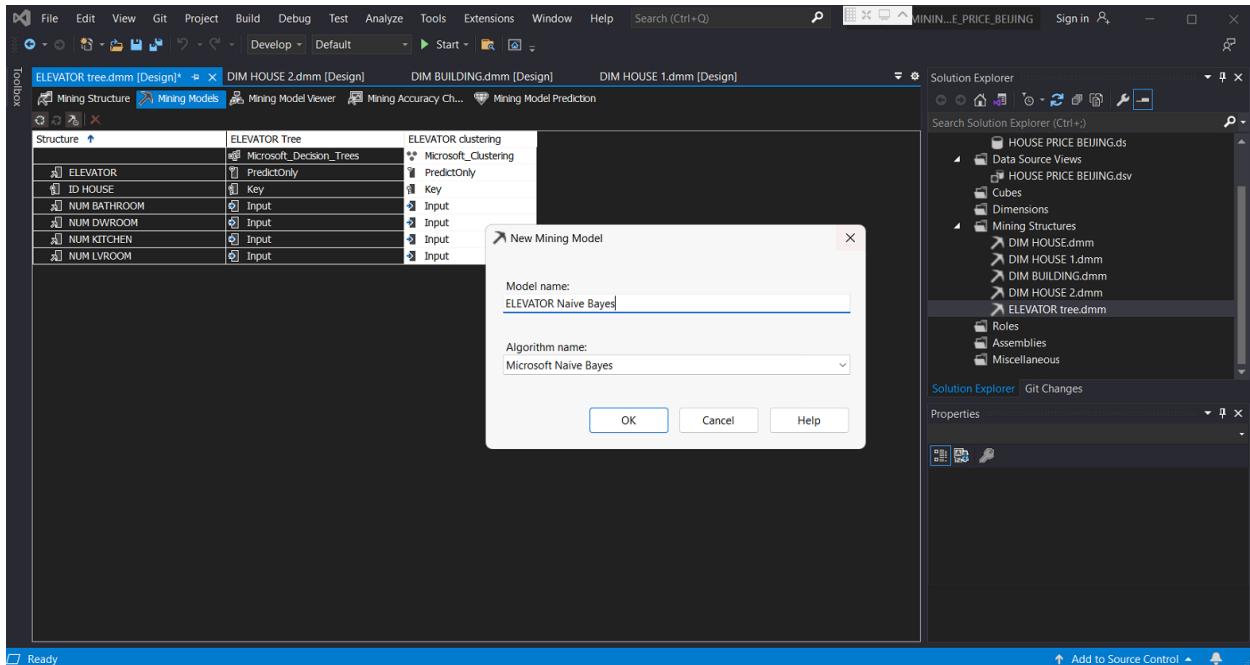


Hình 437: Thêm thuật toán Microsoft Clustering

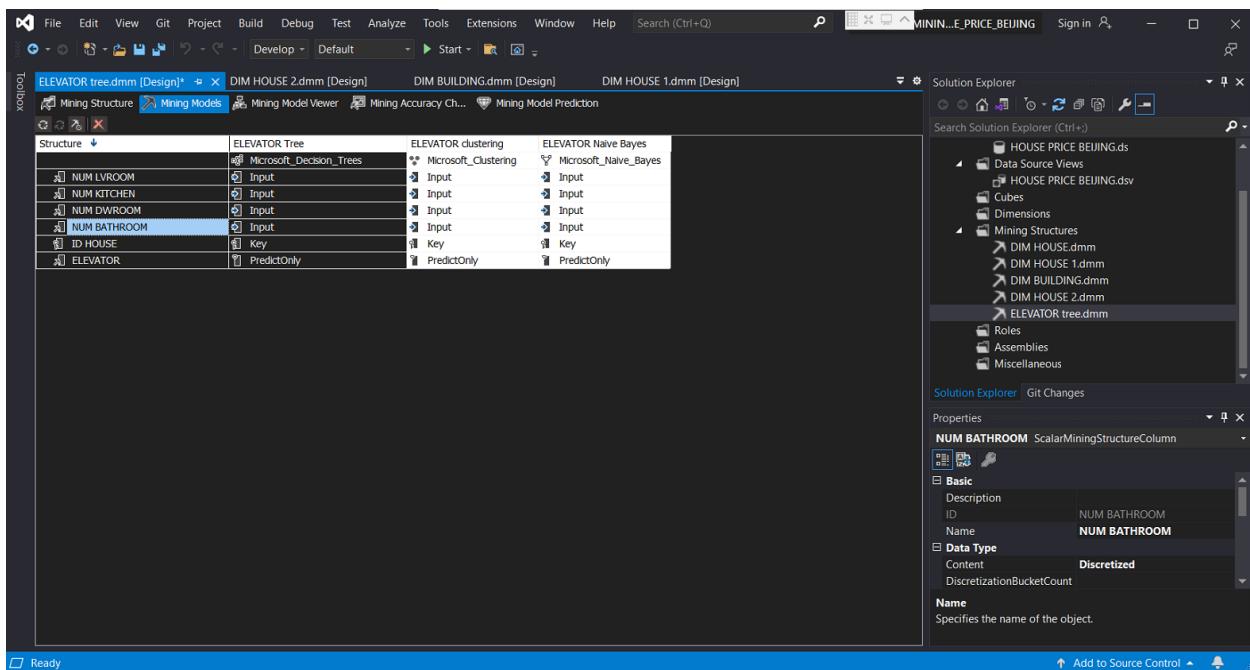


Hình 438: Sau khi thêm thành công

- Thêm thuật toán Microsoft Naive Bayes

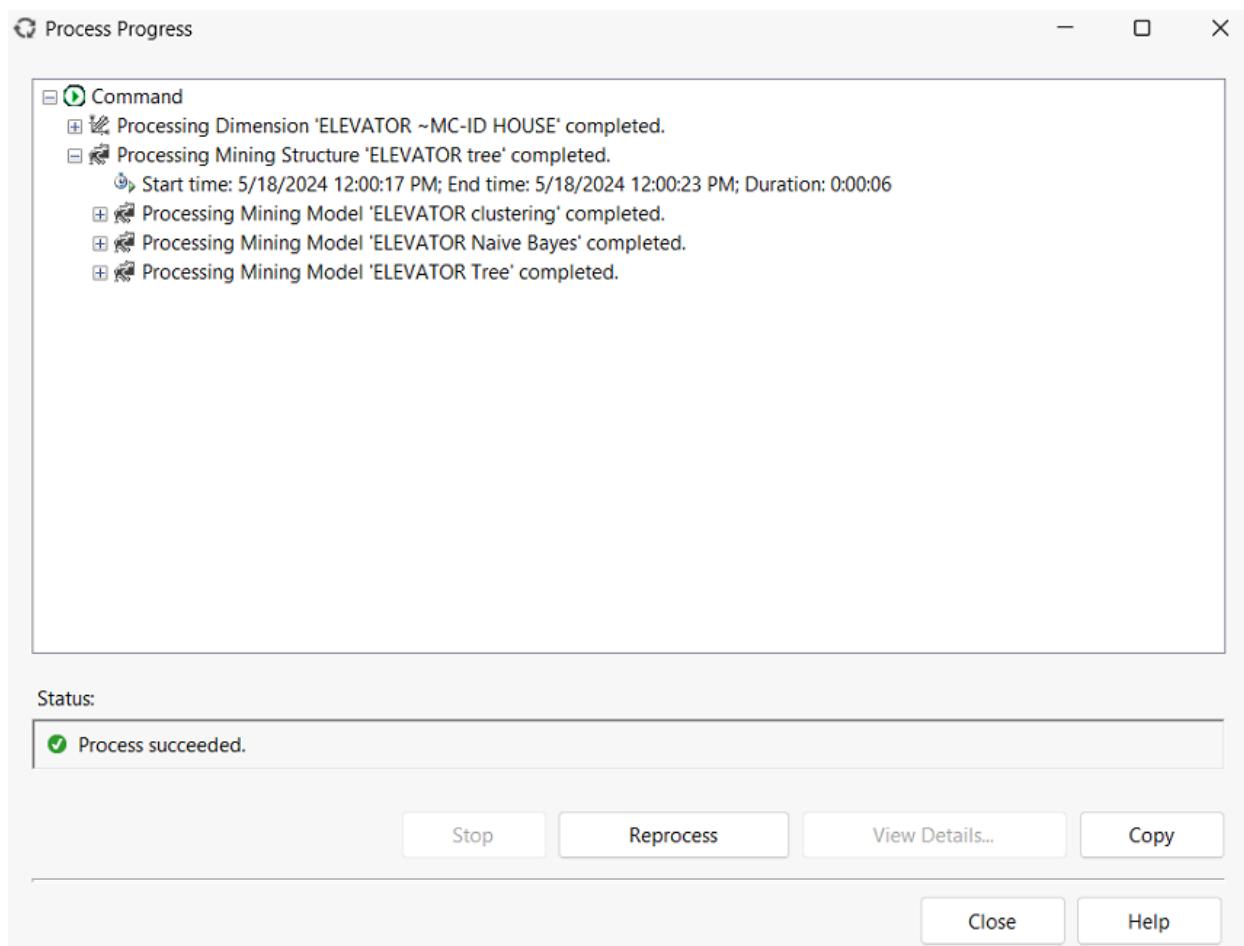


Hình 439: Thêm thuật toán Microsoft Naive Bayes



Hình 440: Sau khi thêm thành công

- Run Process để tiến hành mining



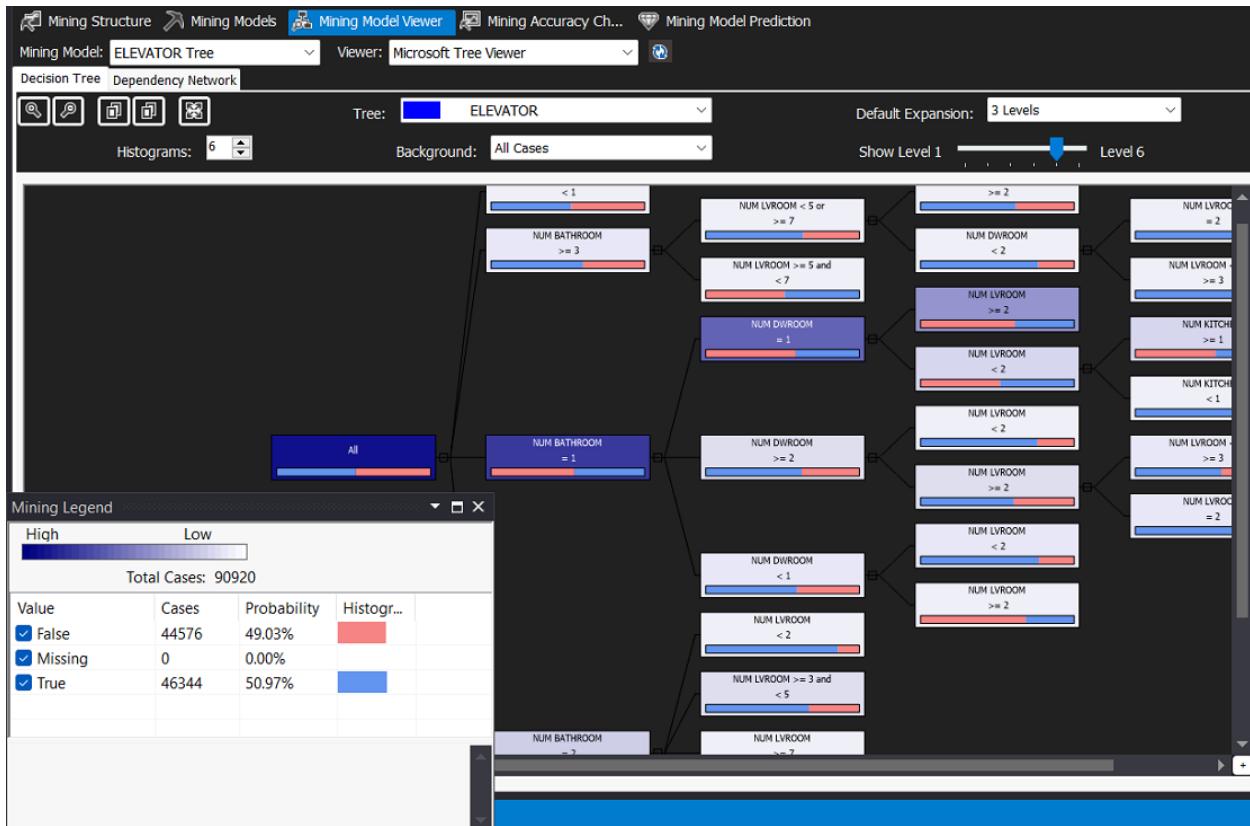
Hình 441: Tiến hành Mining

2. Nhận xét kết quả sau khi Mining dữ liệu

2.1. Thuật toán Microsoft Decision Trees

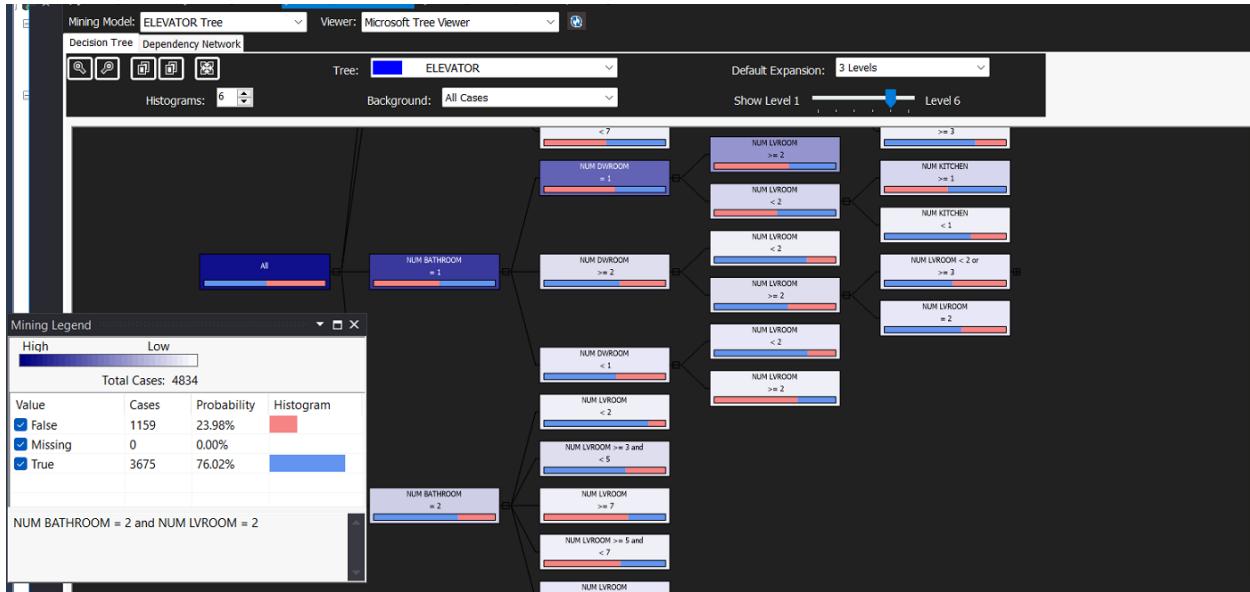
Xem **Mining Model Viewer** của thuật toán **Microsoft Decision Trees** với thuộc tính dự đoán là **ELEVATOR**

- Tỷ lệ nhà có thang máy (**True**) trong bộ dữ liệu là **49.03%**
- Tỷ lệ nhà không có thang máy (**False**) trong bộ dữ liệu là **50.97%**



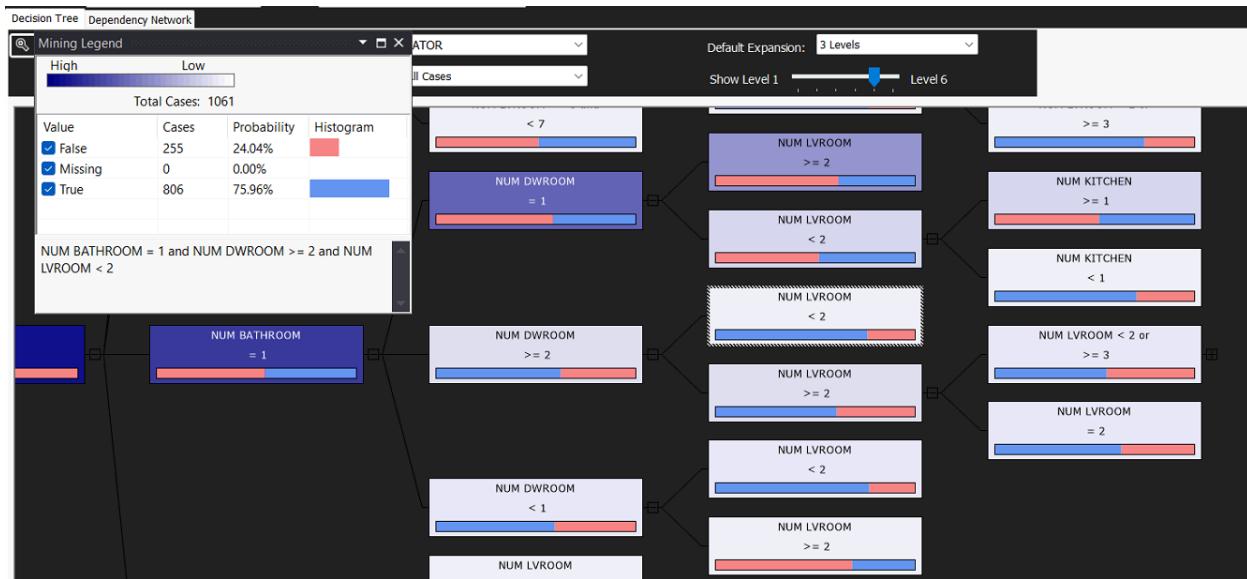
Hình 442: Xem Mining Model Viewer của thuật toán Microsoft Decision Tree với thuộc tính dự đoán là ELEVATOR

- Với những nhà có thang máy có tỷ lệ cao 76.02% là những nhà có 2 phòng tắm và 2 phòng khách



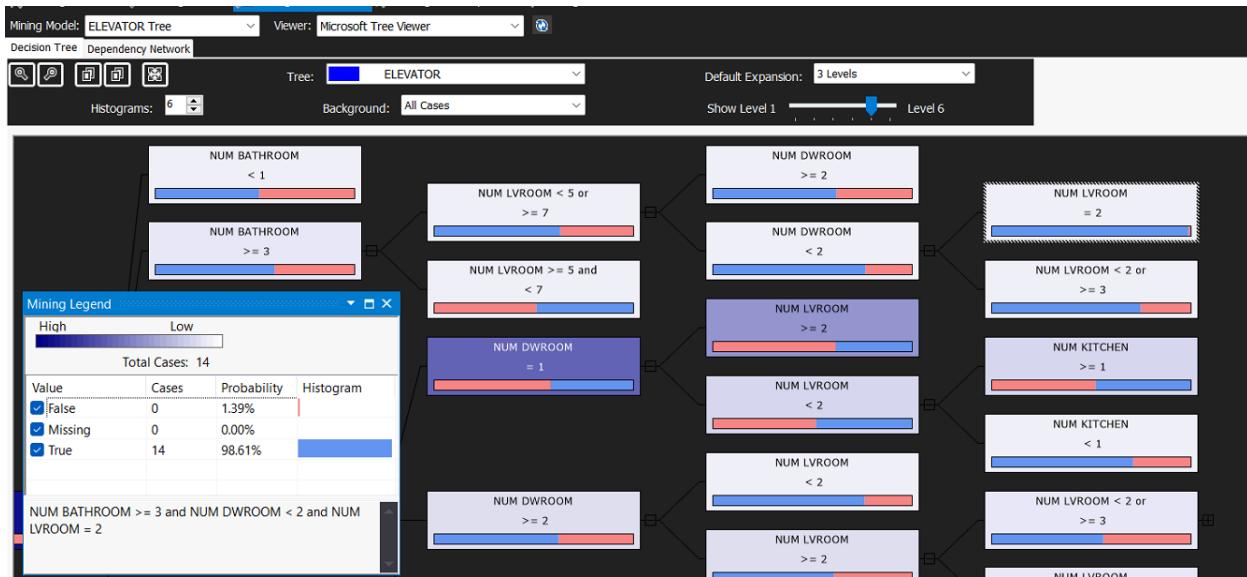
Hình 443: Tỷ lệ cao 76.02% là những nhà có 2 phòng tắm và 2 phòng khách

- Với những nhà có **thang máy** có tỷ lệ cao **75.96%** là những nhà có **1 phòng tắm**, **có số phòng vách lớn hơn hoặc bằng 2 và có ít hơn 2 phòng khách**



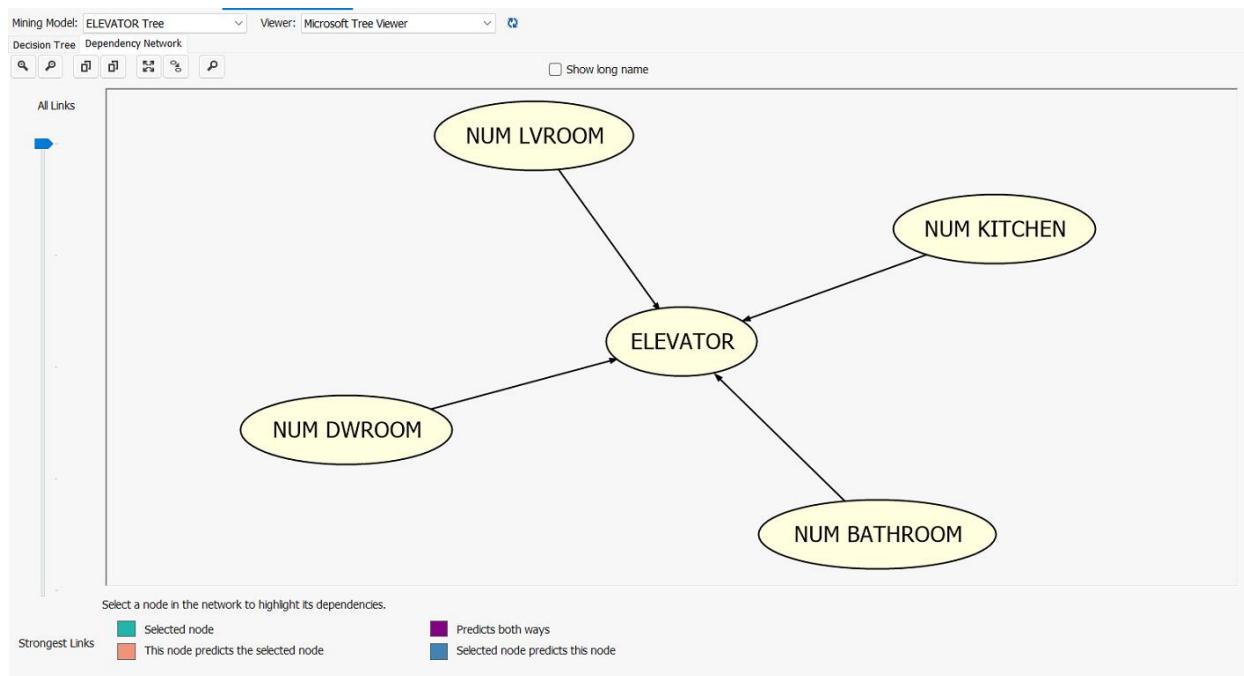
Hình 444: Tỷ lệ cao 75.96% là những nhà có 1 phòng tắm, có số phòng vách lớn hơn hoặc bằng 2 và có ít hơn 2 phòng khách

- Với những **nhà có thang máy** có tỷ lệ cao **98.61%** là những nhà có **số phòng tắm lớn hơn hoặc bằng 3, có số phòng vách bé hơn 2 và có 2 phòng khách**

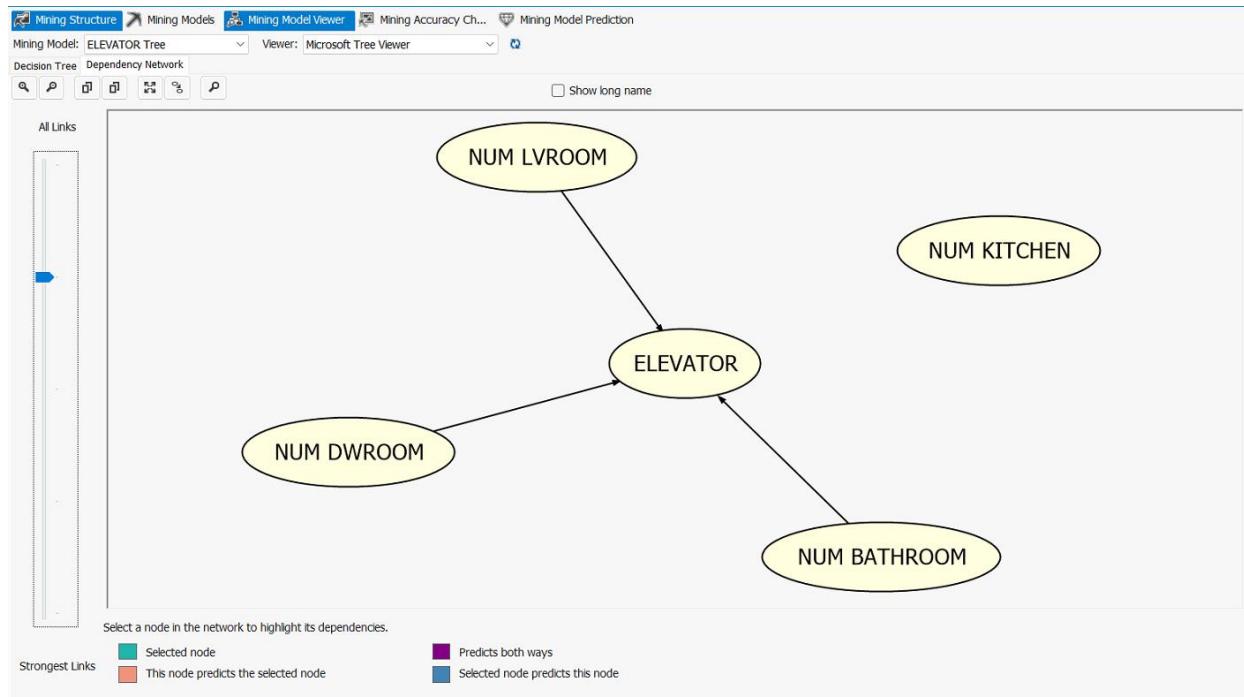


Hình 445: Tỷ lệ cao 98.61% là những nhà có **số phòng tắm lớn hơn hoặc bằng 3, có số phòng vách bé hơn 2 và có 2 phòng khách**

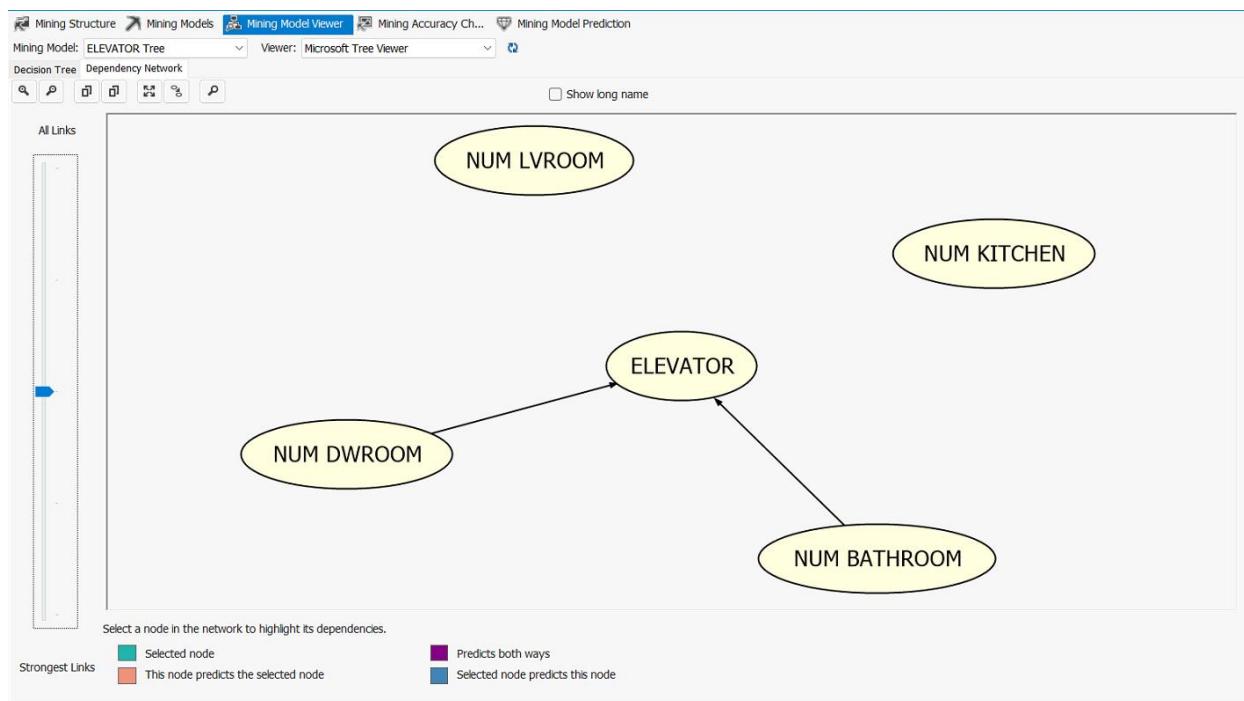
- **Pdependency Network** của thuật toán **Microsoft Decision Tree** cho thấy những thuộc tính có ảnh hưởng đến việc nhà có thang máy.
- Những thuộc tính có độ ảnh hưởng giảm dần là: **NUM BATHROOM -> NUM DWROOM -> NUM LVROOM -> NUM KITCHEN**



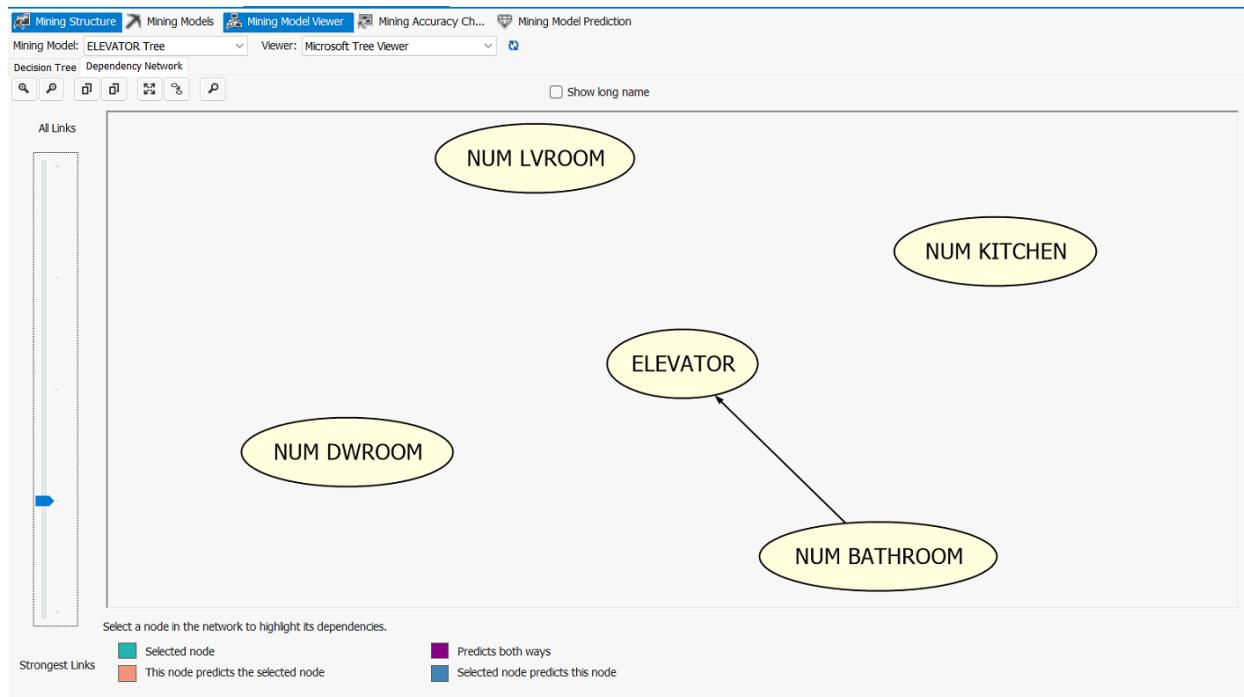
Hình 446: Pendency Network của thuật toán (1)



Hình 447: Pendency Network của thuật toán (2)



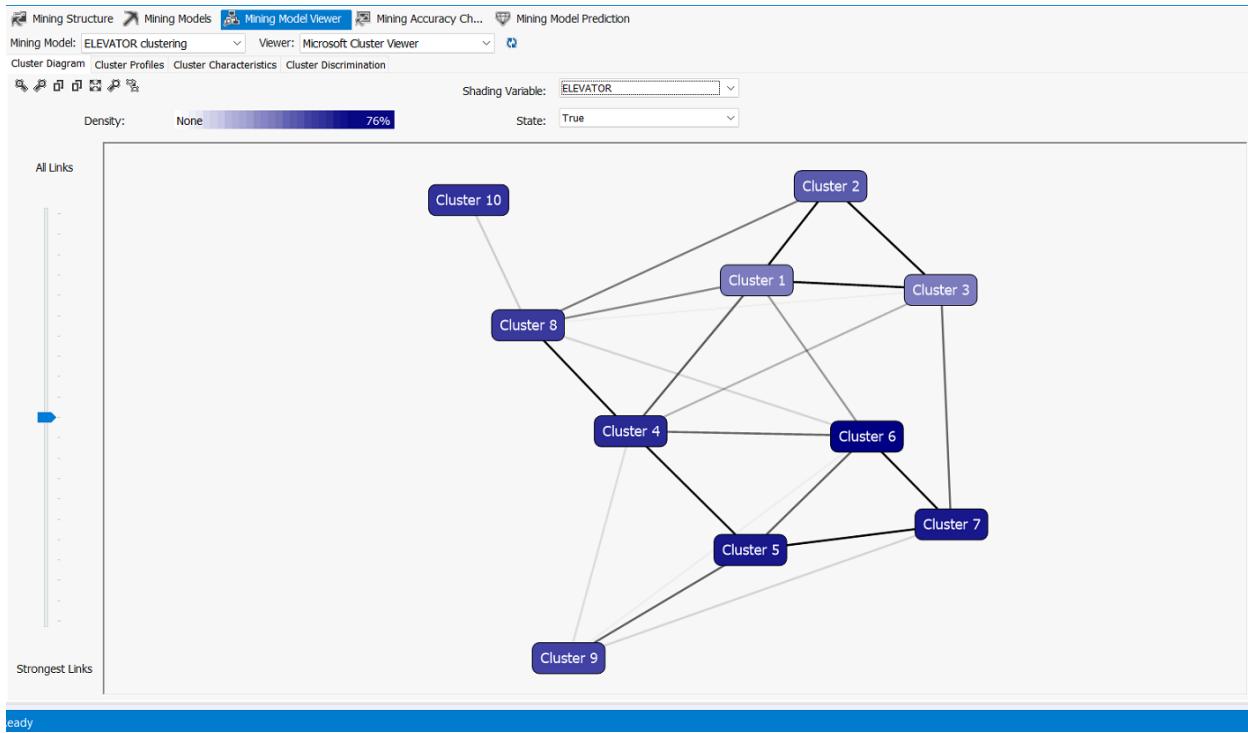
Hình 448: Pependency Network của thuật toán (3)



Hình 449: Pependency Network của thuật toán (4)

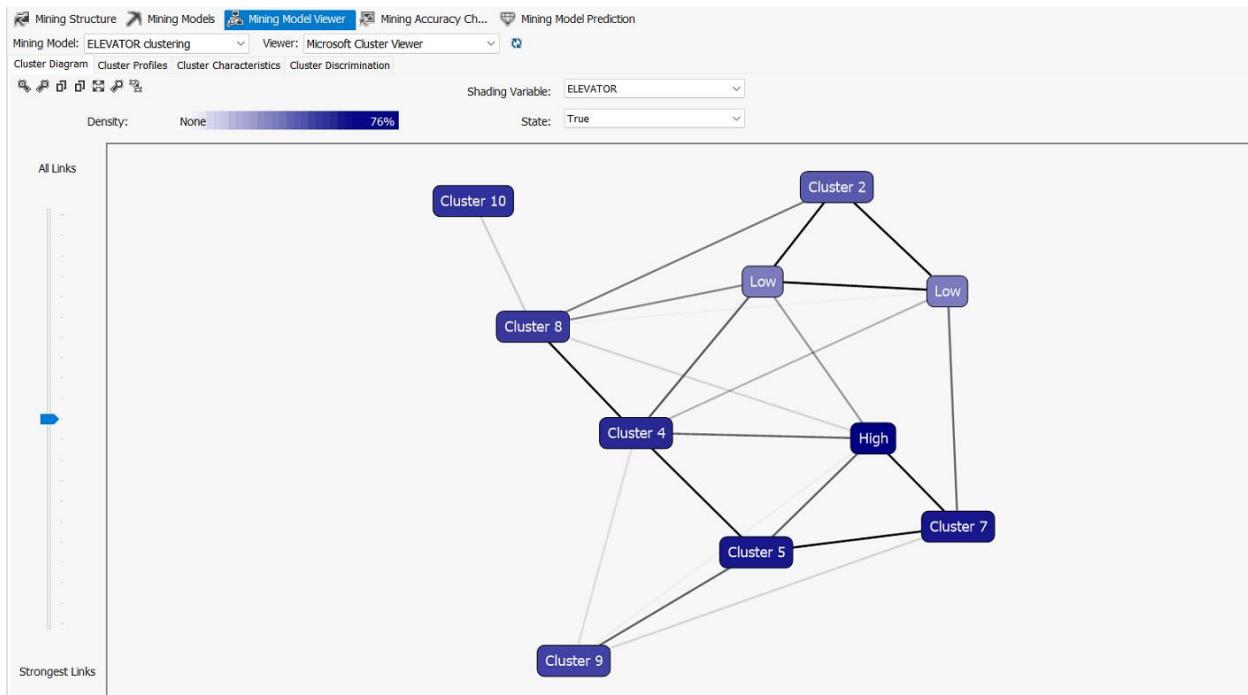
2.2. Thuật toán Microsoft Clustering

Cluster Diagram của thuật toán Shading Variable là **ELEVATOR**, value là **True** (có thang máy)



Hình 450: Cluster Diagram của thuật toán Shading Variable là ELEVATOR

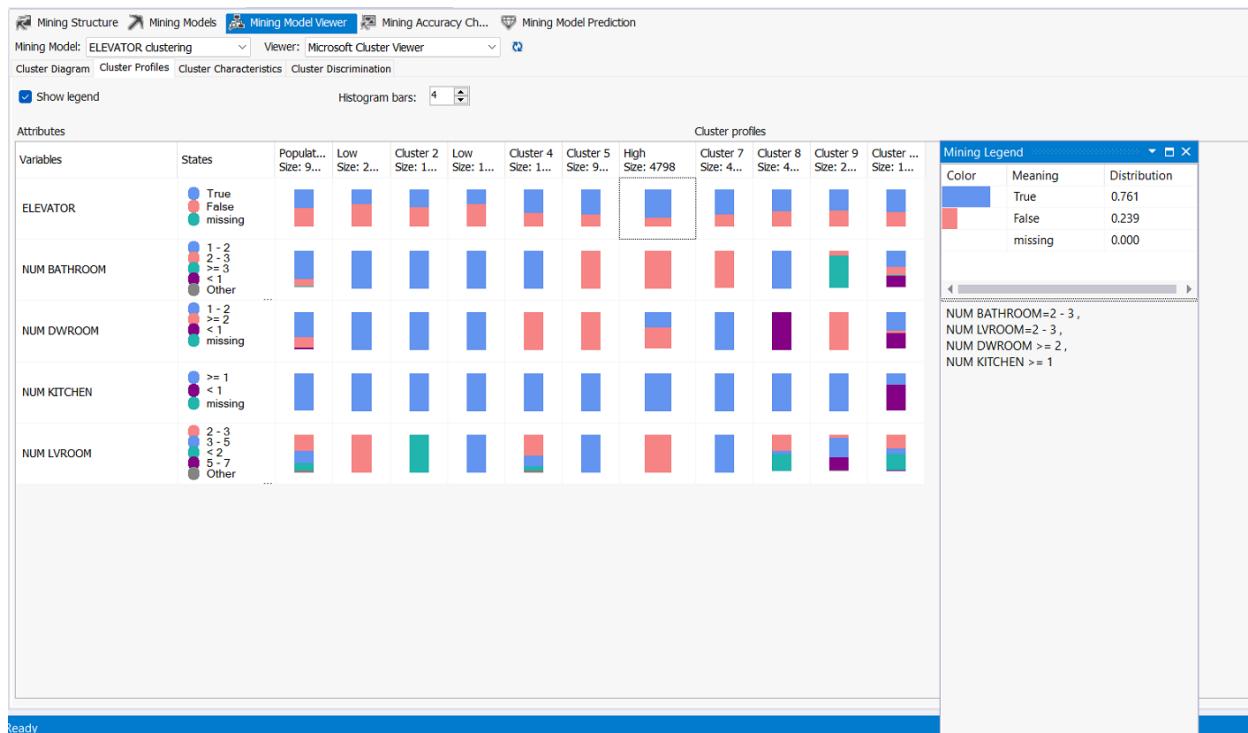
- Đổi tên Cluster đậm nhất thành **High** (khả năng nhà có thang máy cao nhất)
- Đổi tên Cluster nhạt màu nhất thành **Low** (khả năng nhà có thang máy thấp nhất)



Hình 451: Tập lục của những nhà có thang máy chiếm 76.1%

- Với **Cluster Profiles** của **Cluster High** cho ta thấy tập lục của những nhà **có thang máy chiếm 76.1%**

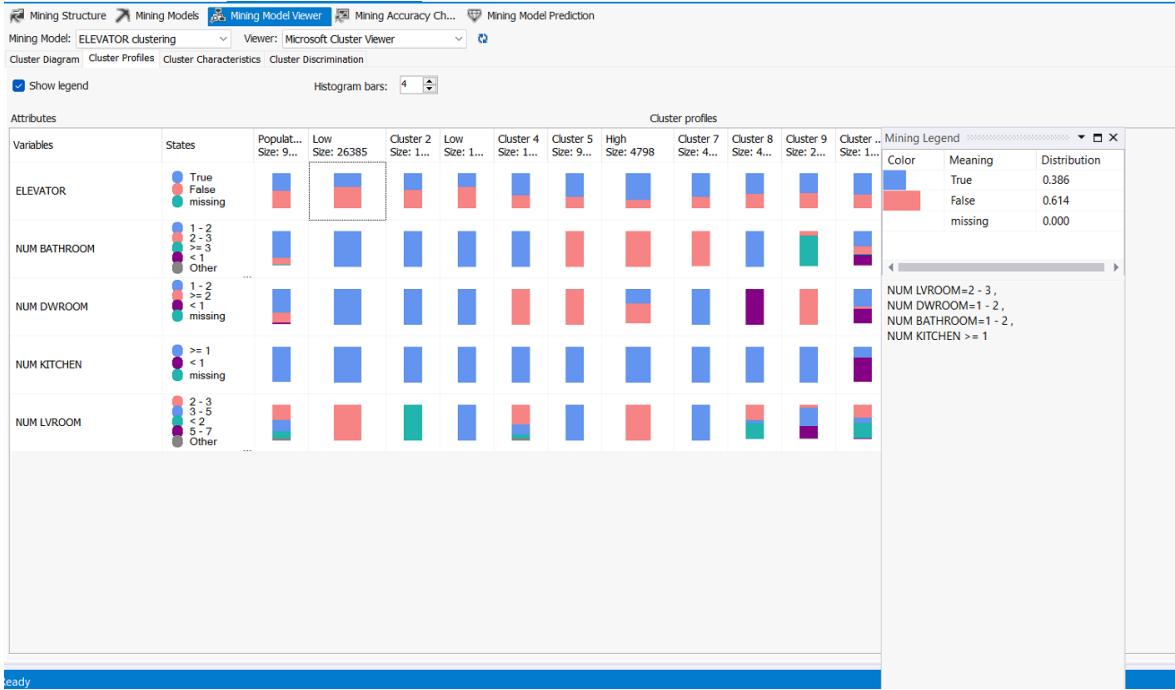
Tập lục: Những nhà có **số phòng tắm (NUM BATHROOM)** từ 2 đến 3, có **số phòng khách (NUM LVROOM)** từ 2 đến 3, có **số phòng vệ (NUM DWROOM)** lớn hơn hoặc bằng 2, có **số phòng bếp (NUM KITCHEN)** lớn hơn hoặc bằng 1



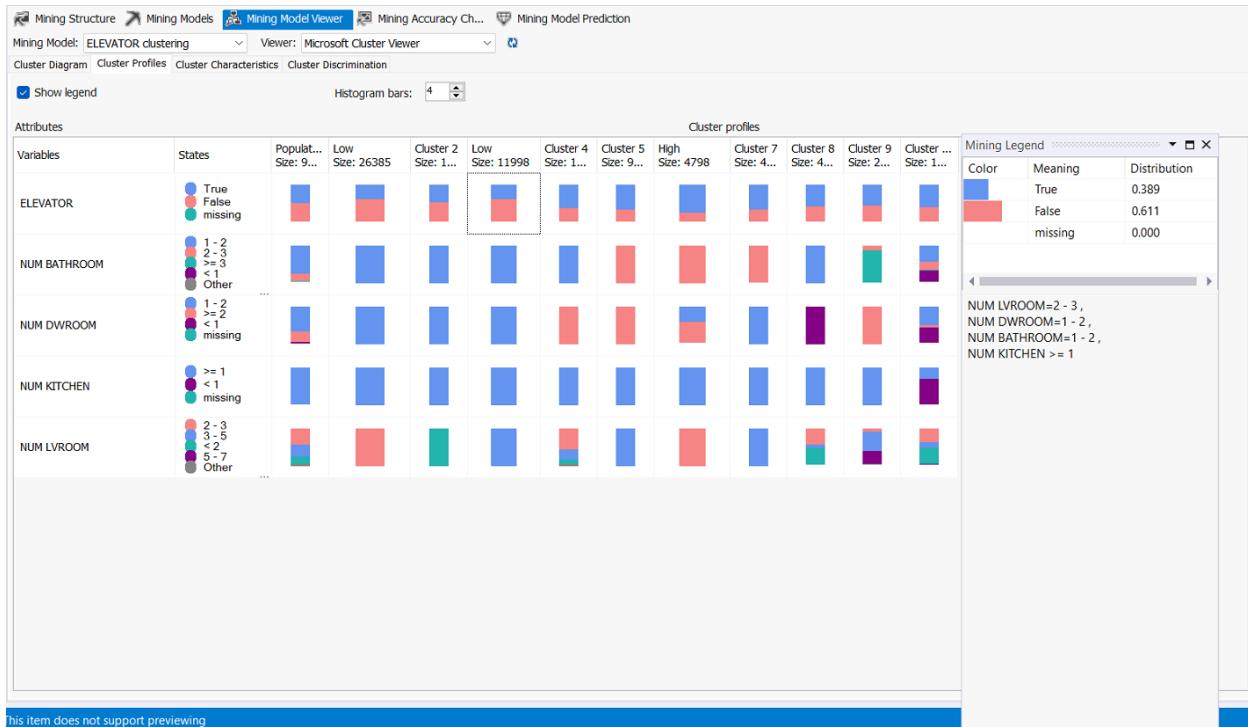
Hình 452: Tập lục của những nhà không có thang máy chiếm 61.4%

- Với **Cluster Profiles** của **Cluster Low** cho ta thấy tập lục của những nhà **không có thang máy chiếm 61.4%**

Tập lục: Những nhà có **số phòng khách (NUM LVROOM)** từ 2 đến 3, có **số phòng vệ (NUM DWROOM)** từ 1 đến 2, có **số phòng tắm (NUM BATHROOM)** từ 1 đến 2, có **số phòng bếp (NUM KITCHEN)** lớn hơn hoặc bằng 1



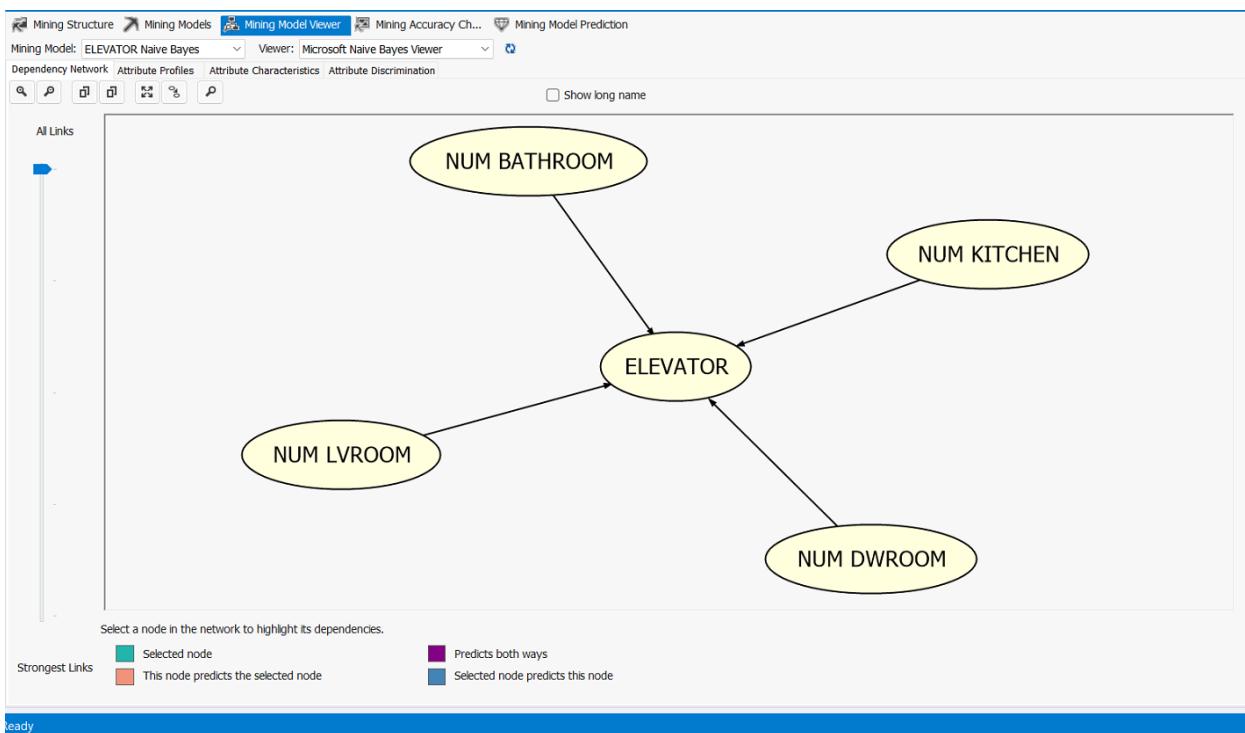
Hình 453: Cluster Profiles của Cluster Low của tập lục trên (1)



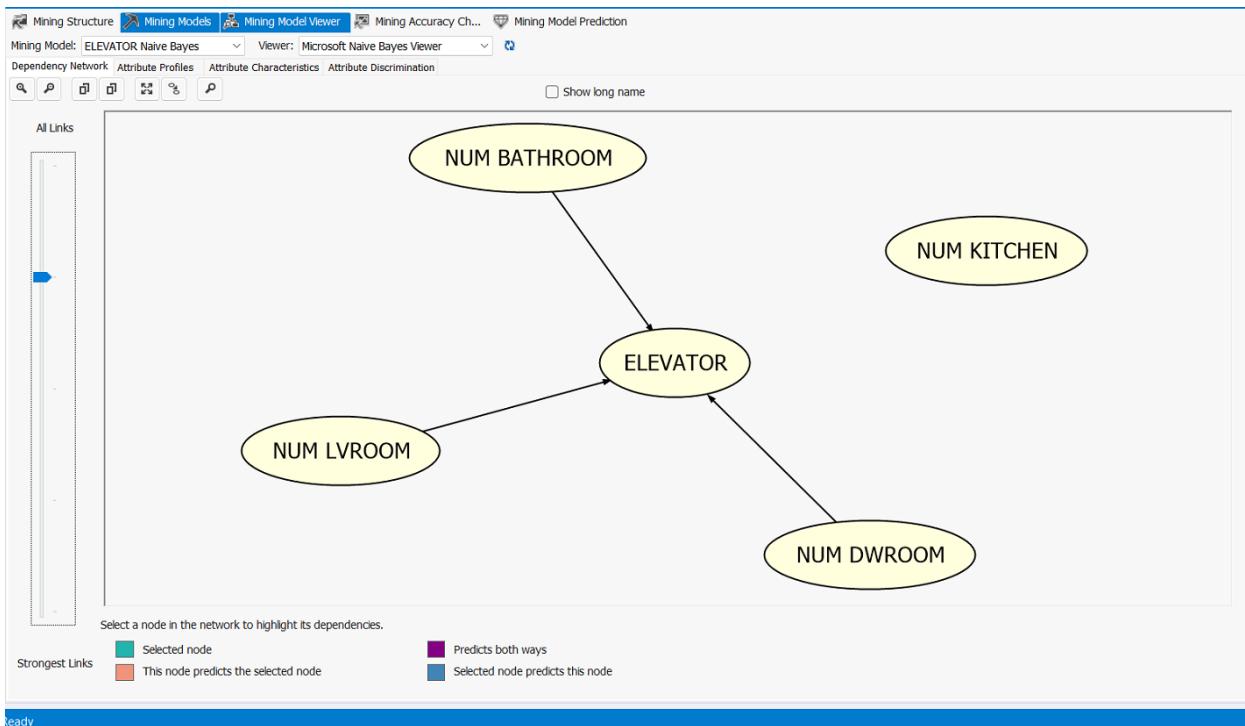
Hình 454: Cluster Profiles của Cluster Low của tập lục trên (2)

2.3. Thuật toán Microsoft Naive Bayes

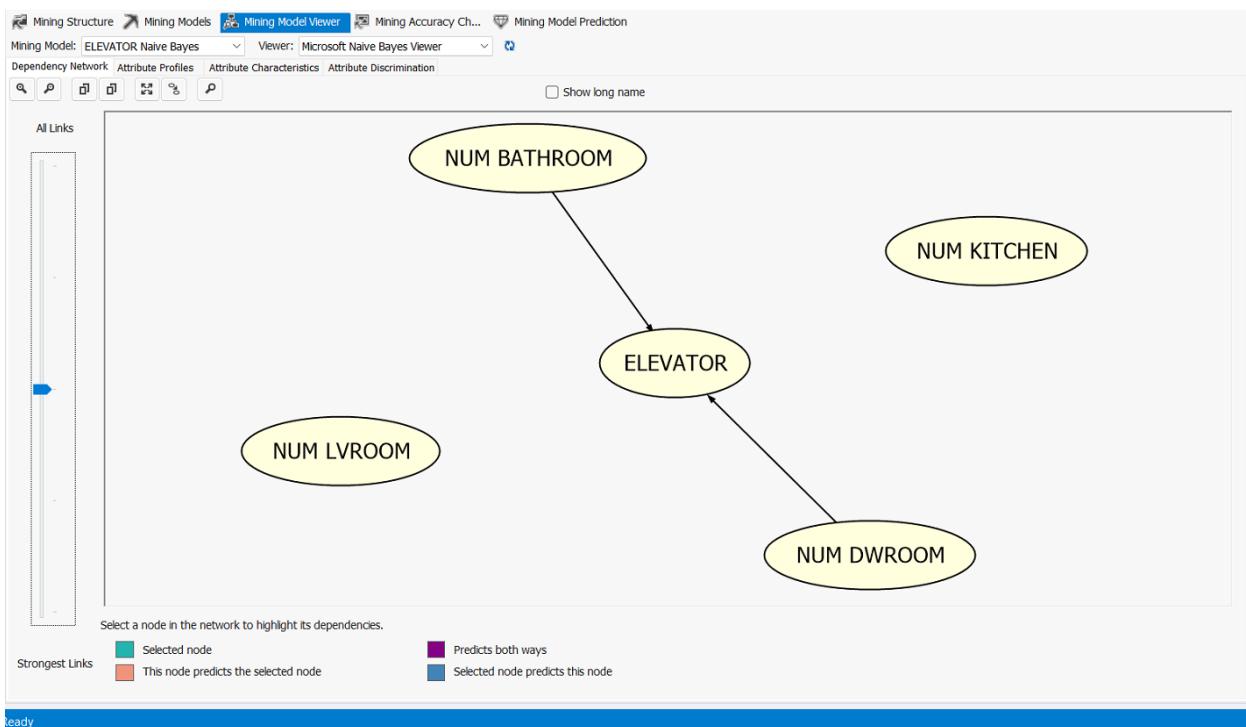
Dependency Network của thuật toán **Microsoft Naive Bayes** cho ta thấy các thuộc tính ảnh hưởng đến **nhà có thang máy**



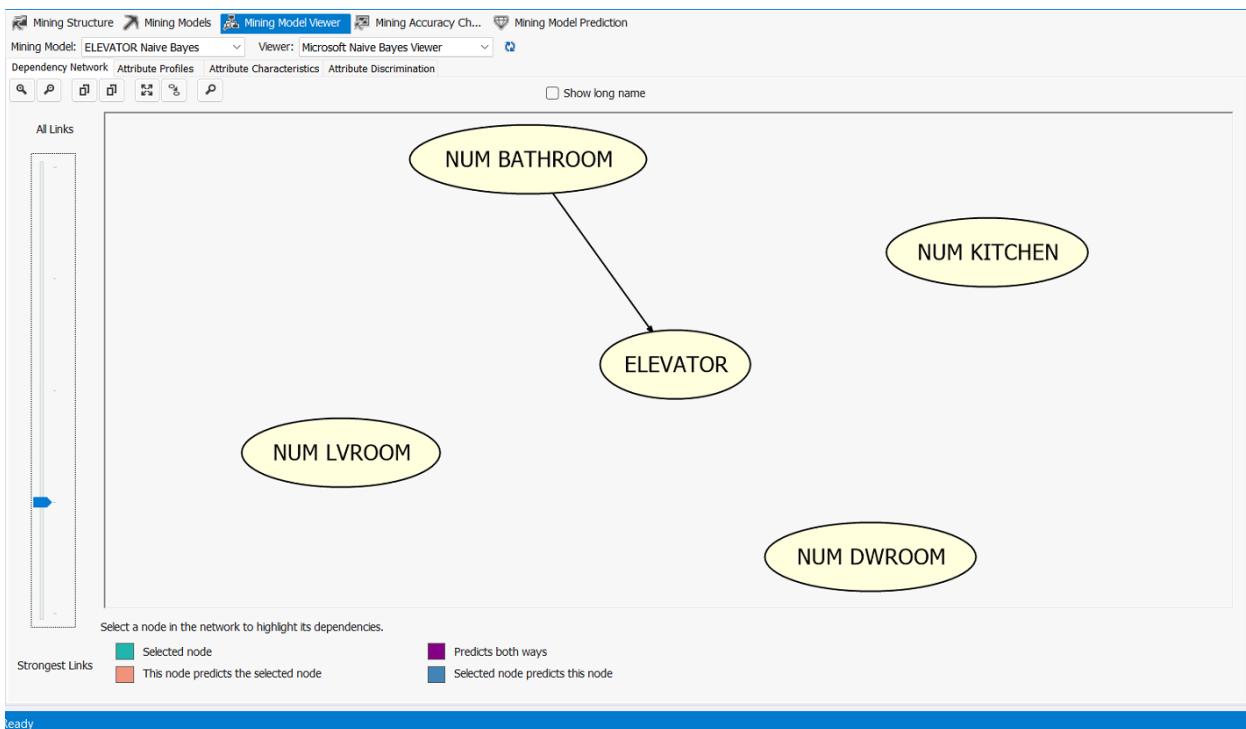
Hình 455: Dependency Network của thuật toán Microsoft Naive Bayes (1)



Hình 456: Dependency Network của thuật toán Microsoft Naive Bayes (2)

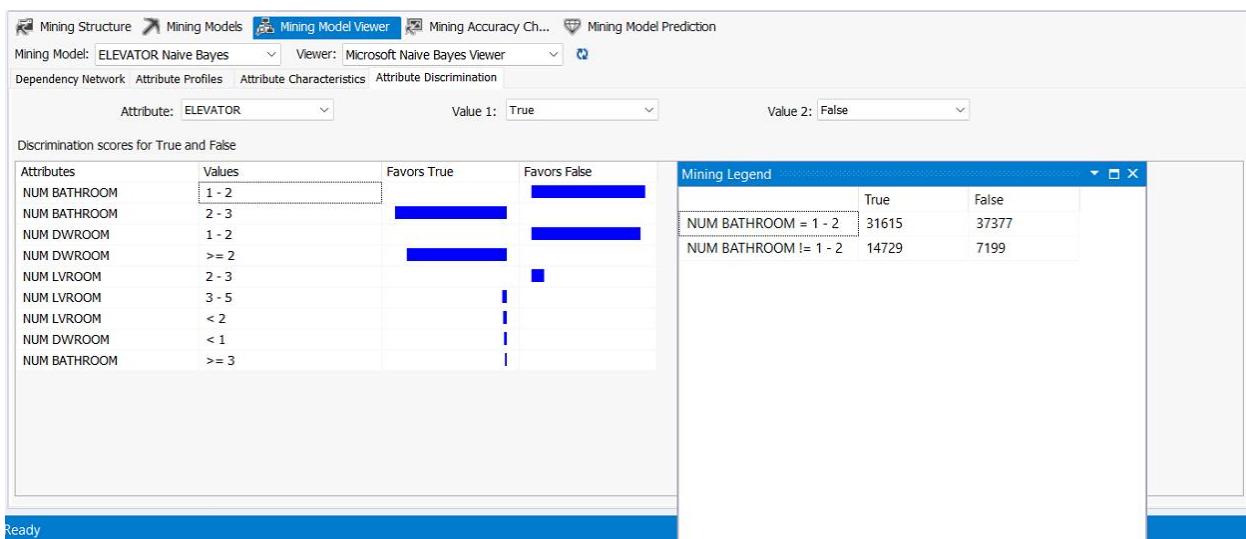


Hình 457: Dependency Network của thuật toán Microsoft Naive Bayes (3)



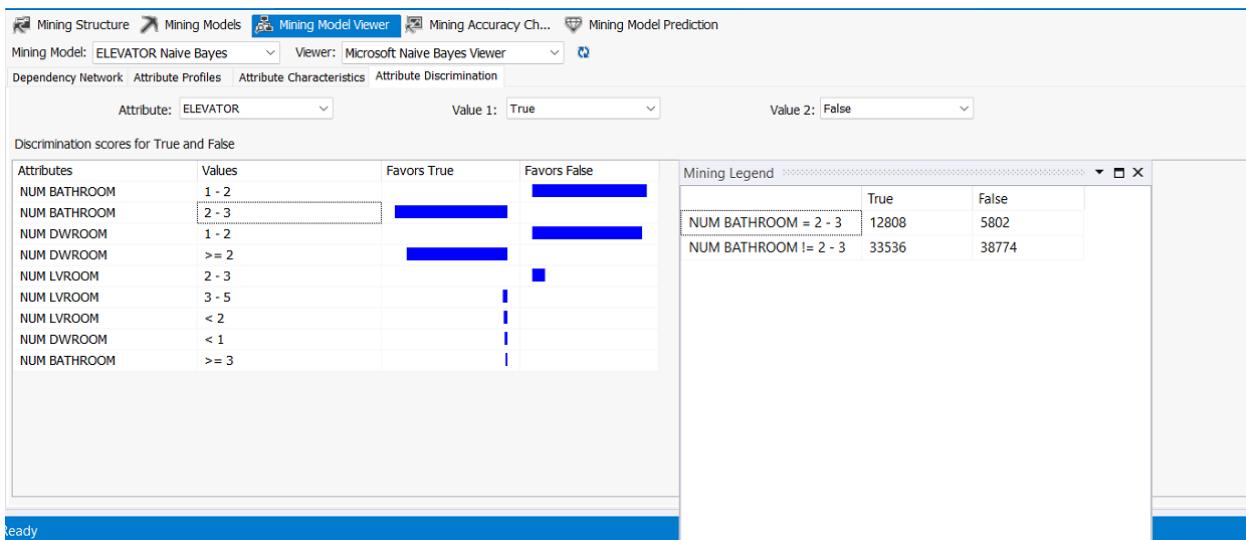
Hình 458: Dependency Network của thuật toán Microsoft Naive Bayes (4)

- Những nhà có **số phòng tắm (NUM BATHROOM)** từ **1 đến 2** thì khả năng cao sẽ **không có thang máy**



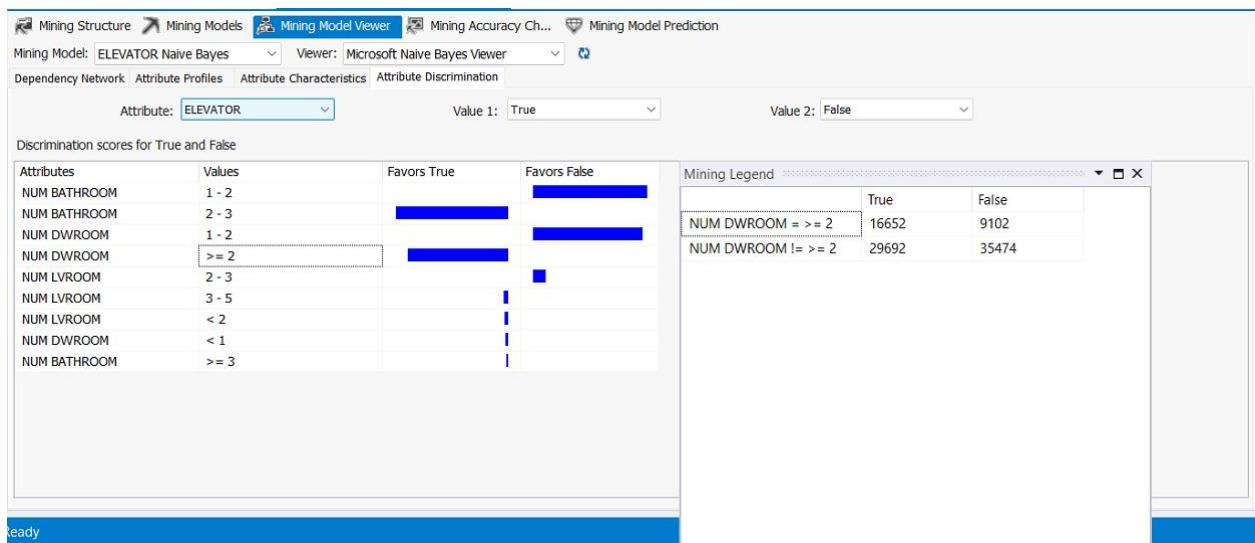
Hình 459: Nhà có số phòng tắm từ 1-2 khả năng cao sẽ không có thang máy

- Những nhà có **số phòng tắm (NUM BATHROOM)** từ **2 đến 3** thì khả năng cao sẽ **có thang máy**



Hình 460: Nhà có số phòng tắm từ 2-3 khả năng cao sẽ có thang máy

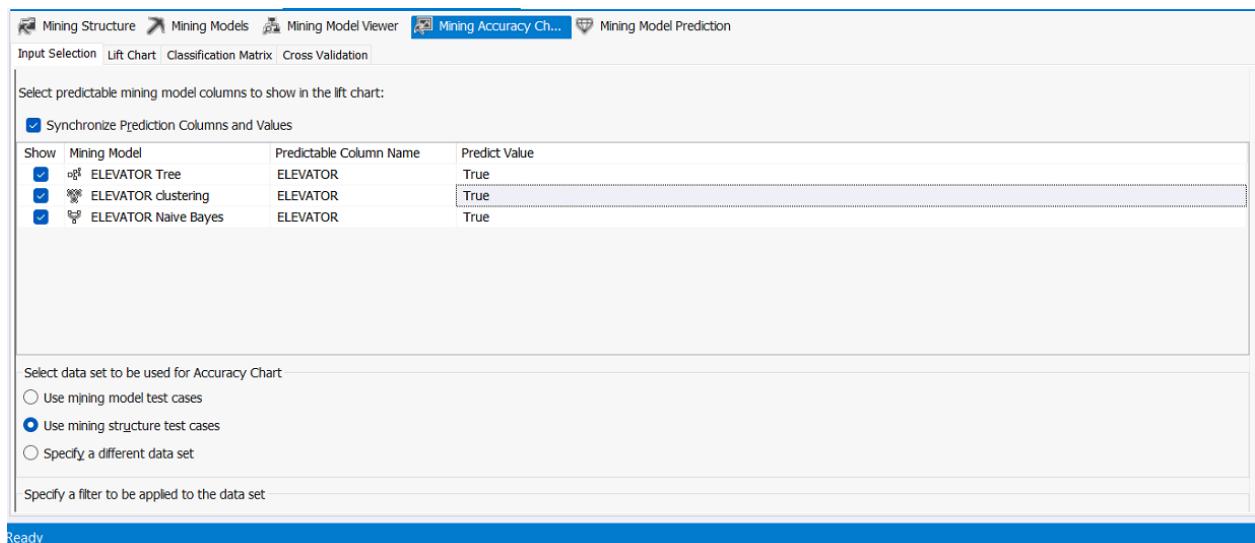
- Những nhà có **số phòng vệ (NUM DWROOM)** lớn hơn hoặc bằng 2 thì khả năng cao sẽ **có thang máy**



Hình 461: Nhà có số phòng vệ lớn hơn hoặc bằng 2 khả năng cao sẽ có thang máy

2.4. Đánh giá các thuật toán bằng Mining Accuracy Chart

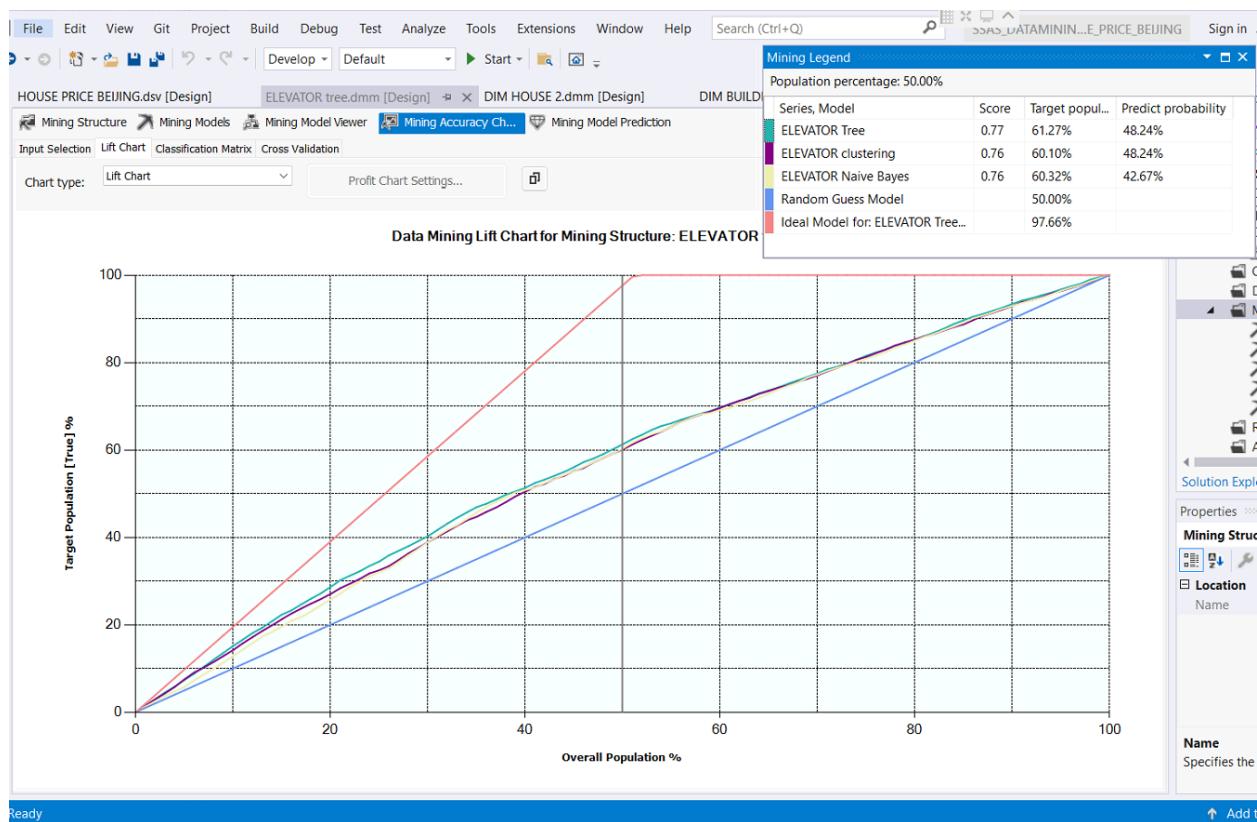
Thiết lập thông số **Input Selection** với các nhà có thang máy



Hình 462: Thiết lập thông số Input Selection với các nhà có thang máy

Line chart cho ta thấy tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà có thang máy

- Microsoft Decision Tree: 0.77 Score.
- Microsoft Clustering: 0.76 Score.
- Microsoft Naive Bayes: 0.76 Score.



Hình 463: Tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà có thang máy

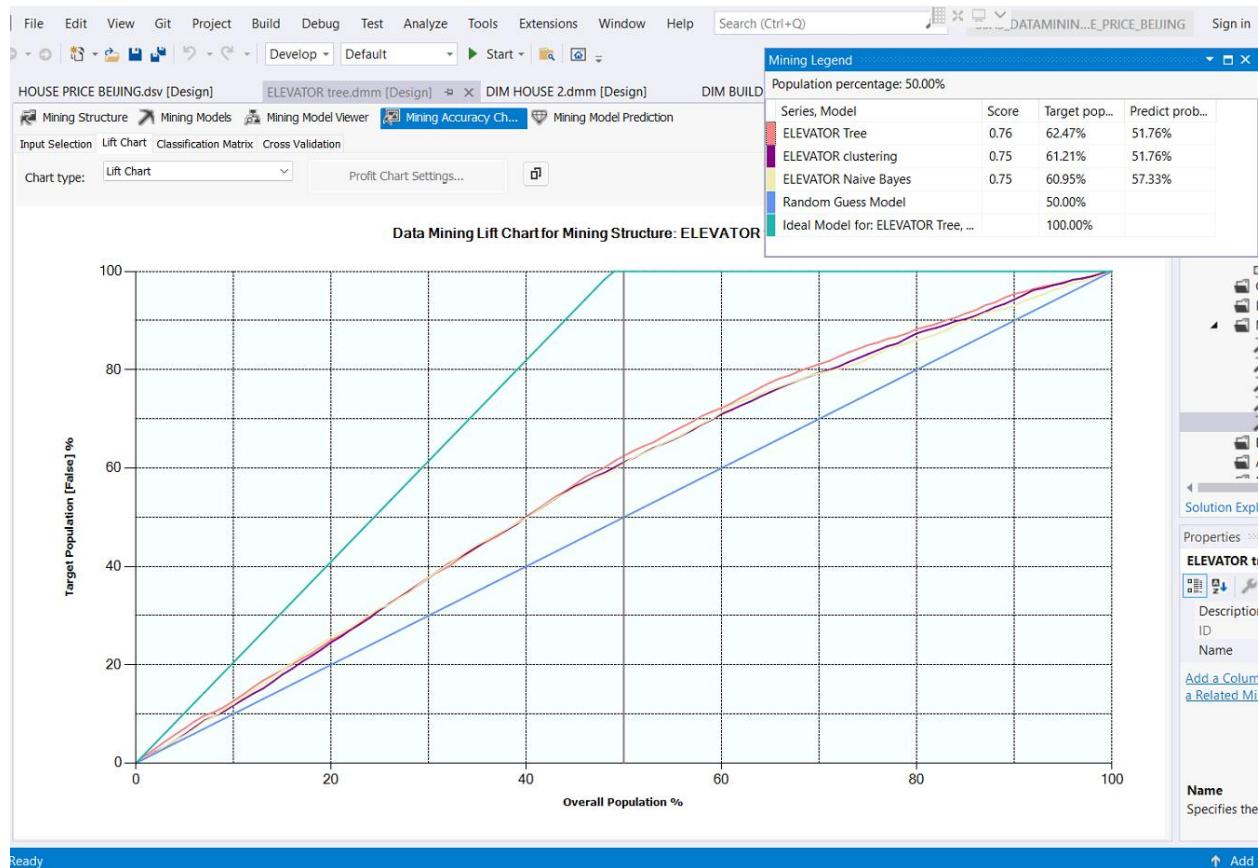
- Thiết lập thông số **Input Selection** với các nhà **không** có thang máy

Show	Mining Model	Predictable Column Name	Predict Value
<input checked="" type="checkbox"/>	ELEVATOR Tree	ELEVATOR	False
<input checked="" type="checkbox"/>	ELEVATOR clustering	ELEVATOR	False
<input checked="" type="checkbox"/>	ELEVATOR Naive Bayes	ELEVATOR	False

Hình 464: Thiết lập Input Selection với các nhà không có thang máy

Line chart cho ta thấy tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà không có thang máy

- Microsoft Decision Tree: 0.76 Score.
- Microsoft Clustering: 0.75 Score.
- Microsoft Naive Bayes: 0.75 Score.



Hình 465: Tỷ lệ chính xác của 3 thuật toán khi mining trường hợp nhà không có thang máy