

HỆ THỐNG THÔNG TIN

PHỤC VỤ TRÍ TUỆ KINH DOANH

ĐỒ ÁN THỰC HÀNH

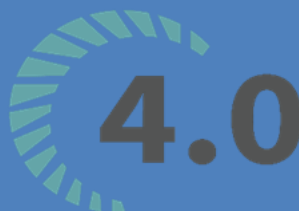
Nhóm 6: ETL

Giáo Viên phụ trách:

- Hồ Thị Hoàng Vy
- Tiết Gia Hồng

Thành viên nhóm:

- 19127615 Nguyễn Anh Tuấn
- 20127401 Quách Đỗ Gia Huy
- 20127549 Nguyễn Chí Linh
- 20127584 Trần Hữu Minh Nhật



HỌC KỲ I NĂM HỌC: 2023-2024

Mục lục

I. Phân công công việc:	2
II. Thiết kế kho dữ liệu và mô tả dữ liệu	3
1. Stage	3
2. NDS	5
3. DDS	7
III. Các phân cấp được triển khai trong hệ thống và mô hình hóa	8
1. Mô hình hóa	8
2. Phân cấp dữ liệu	9
IV. ETL	10
1. Source To Stage	10
2. Stage To NDS	11
3. NDS To DDS	12
V. Report	14
1. Thống kê số thành viên mua hàng theo ngày, tháng, năm	14
2. Thống kê doanh thu của khách hàng (member, normal) theo ngày, tháng, năm và theo chi nhánh	15
3. Thống kê số lượng khách thanh toán theo cash/ debit/... ở từng chi nhánh theo từng tháng trong năm	16
4. Thống kê lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine)	16
5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time / date)	17
6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm	18
7. Cho biết doanh thu của từng chi nhánh của các năm	20
8. Cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit... của mỗi chi nhánh / của tất cả chi nhánh theo từng năm	20
Tổng quan	21
VI. OLAP	23
1. Yêu cầu đề	23
2. Thực hiện bằng Cube và Excel	24
3. MDX	27
VII. Data Mining	29
1. Dự đoán số lượng mua hàng của khách hàng	29



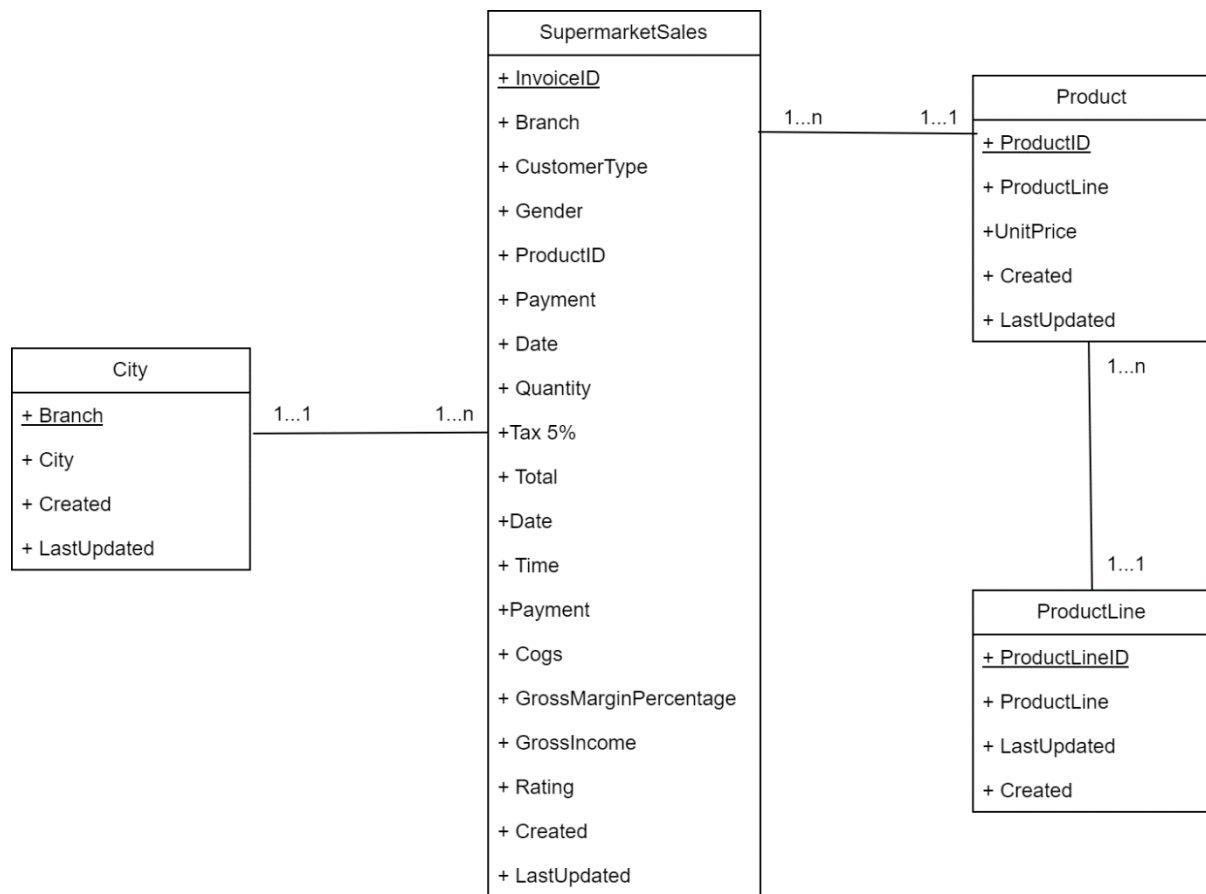
2. Dự đoán đánh giá của khách hàng 33
3. Dự đoán tổng giá tiền mua hàng của khách hàng dựa vào thu nhập 37

I. Phân công công việc:

MSSV	Họ và tên	Công việc	Đánh giá
19127615	Nguyễn Anh Tuấn	Thiết kế DDS, tạo Script DDS Report: thực hiện 4,5,6,8	25
20127401	Quách Đỗ Gia Huy	Thiết kế NDS, tạo Script NDS Data mining	25
20127549	Nguyễn Chí Linh	ETL: StageToNDS, NDSToDDS MDX, OLAP	25
20127584	Trần Hữu Minh Nhật	ETL: SourceToStage Report: thực hiện 1,2,3,7	25

II. Thiết kế kho dữ liệu và mô tả dữ liệu

1. Stage



Vai trò kho dữ liệu Stage:

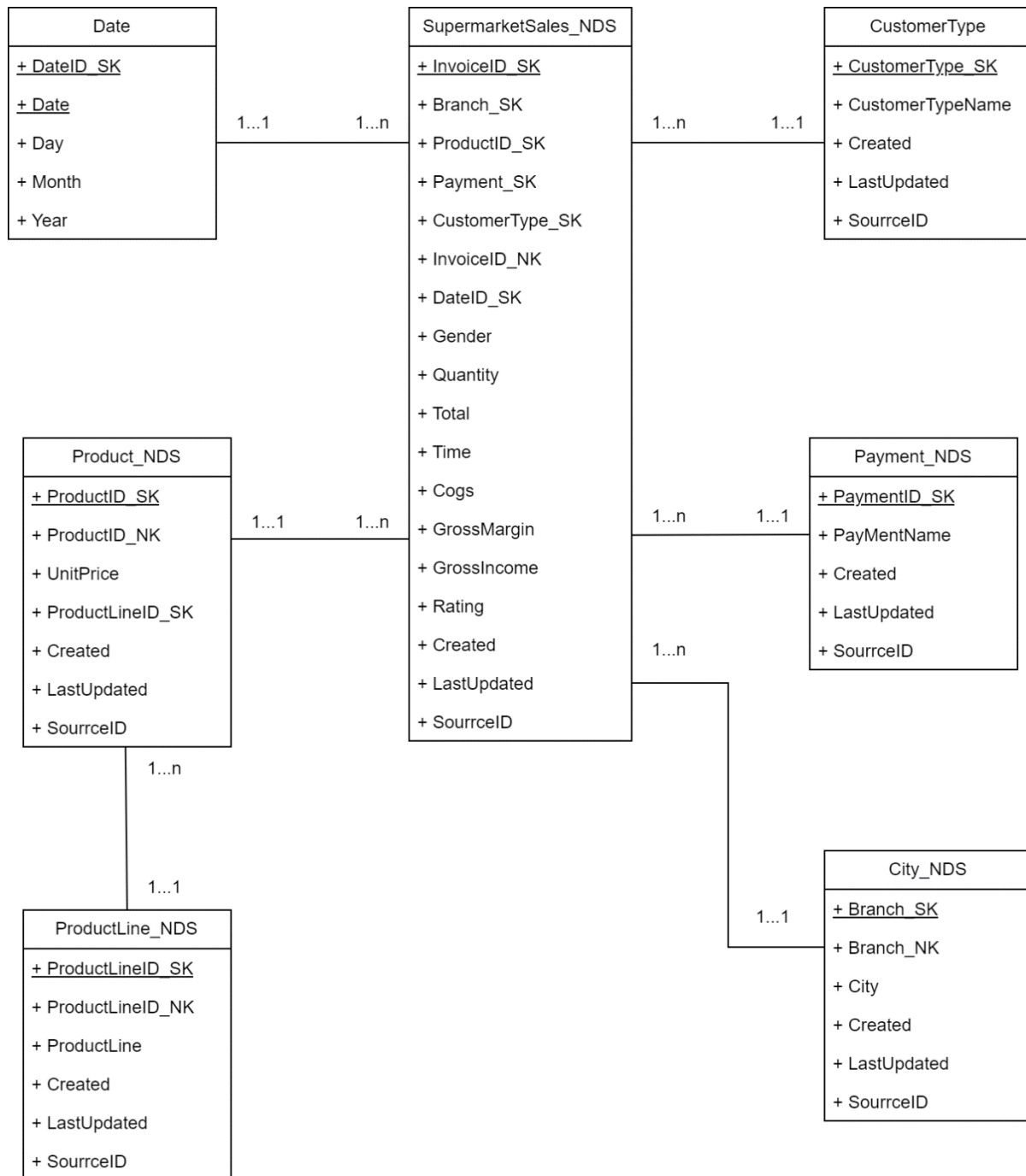
- Kho dữ liệu Stage dùng để chứa thông tin, dữ liệu từ Source Excel đã được giao: Suppermarket_sales
- Lưu trữ thông tin các bảng của dữ liệu nguồn

Mô tả dữ liệu: Gồm 4 bảng

- City:
 - Branch: Chi nhánh của các khu siêu thị (có 3 chi nhánh là A, B và C)
 - City: Tên thành phố nơi các siêu thị được đặt.
- Bảng Product:
 - ProductID: Mã sản phẩm.
 - Unit price: Giá tiền của mỗi đơn vị sản phẩm (tính bằng USD)
 - ProductLline: Loại sản phẩm.
- Bảng ProductLine:

- ProductlineID: Mã số cho mỗi loại sản phẩm.
- Productline: Tên của mỗi loại sản phẩm.
- Bảng supermarket_sale
 - Invoice ID: Số nhận dạng hóa đơn phiếu bán hàng do máy tính tạo ra.
 - Branch: Chi nhánh của các khu siêu thị (có 3 chi nhánh là A, B và C)
 - Customer type: Loại khách hàng, có thể là khách hàng bình thường không có thẻ thành viên hoặc khách hàng thành viên có thẻ thành viên.
 - Gender: Giới tính của khách hàng.
 - ProductID: Mã sản phẩm.
 - Quantity: Số lượng sản phẩm khách hàng mua.
 - Tax 5%: Thuế phải trả, tính dựa trên tổng giá trị đơn hàng.
 - Total: Tổng giá trị đơn hàng bao gồm thuế.
 - Date: Ngày mua hàng.
 - Time: Thời gian mua hàng (từ 10h sáng tới 9h tối)
 - Payment: Phương thức thanh toán. (có 3 phương thức : Tiền, thẻ tín dụng và ví điện tử)
 - COGS: Giá gốc của hàng bán.
 - Gross margin percentage: Tỷ lệ lợi nhuận gộp, số tiền lãi về sau khi bớt đi chi phí hàng hoá.
 - Gross income: Tổng lợi nhuận gộp thu được từ mỗi đơn hàng.
 - Rating: Đánh giá từ khách hàng về sản phẩm hoặc dịch vụ (trên thang điểm từ 1 tới 10)

2. NDS



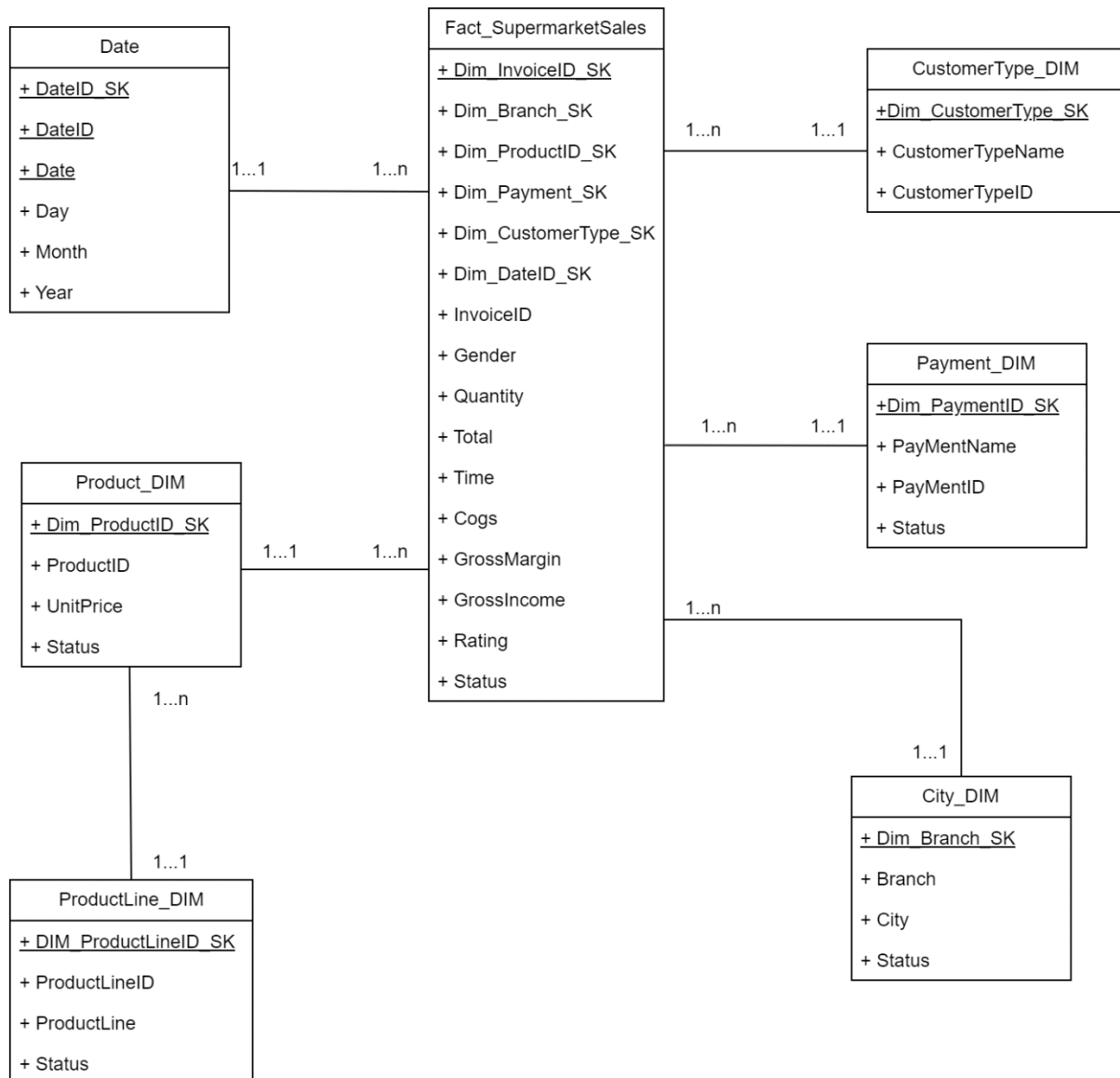
Vai trò kho dữ liệu NDS:

- Dùng để lưu dữ liệu được ETL từ Stage
- Chuẩn hóa dữ liệu trên các bảng về dạng chuẩn 3NF
- Từ bảng SupermaketSales sử dụng các phép biến đổi để chuẩn hóa và tạo ra các bảng mới như: Date, Customertype, Payment_NDS

Các phép biến đổi cần dùng để ETL dữ liệu:

- Data availability:
 - Trong bảng SupermaketSales có 1 file ‘Date’. Trong khi đó ta cần tạo bảng “Date” từ thuộc tính ‘Date’ của bảng SupermaketSales với 3 file (day, month, year). => Để ETL dữ liệu cho bảng “Date” ta cần tách file “Date” của bảng “SupermarketSale” thành các file “Day, Month, Year” để có thể nạp dữ liệu cho bảng “Date” ở kho dữ liệu NDS
- Ngoài ra các bảng mới được tạo thêm như: Customertype, Payment_NDS, các dữ liệu sẽ được ETL từ bảng SupermaketSales theo thứ tự (Customer type, Payment). Và sẽ tự tạo ra khóa cho các bảng này là 1 khóa SK
- Bảng City_NDS được ETL từ bảng City
- Bảng Product_NDS được ETL từ bảng Product
- Bảng ProductLine_NDS được ETL từ bảng ProductLine
- Cuối cùng là bảng SupermaketSales_NDS được ETL từ bảng SupermaketSales.
 - Các dữ liệu của những thuộc tính từ bảng SupermaketSales như Payment, CustomerType, Date khi ETL lên NDS sẽ được thay đổi thành các khóa SK của các bảng đó trên NDS
 - Còn các dữ liệu còn lại thì vẫn được ETL từ bảng SupermaketSales lên

3. DDS



Vai trò kho dữ liệu DDS:

- Đây là kho dữ liệu được dùng để lưu dữ liệu từ NDS
- Các bảng DIM và Fact trong kho dữ liệu nhằm mục đích để phục vụ nhu cầu trực quan hóa và phân tích dữ liệu

Các phép biến đổi:

- Dữ liệu trong bảng Fact_SupermaketSales được nạp từ bảng SupermaketSales_NDS. Ta cần thay đổi các khóa SK trên bảng Fact_SupermaketSales cho phù hợp với các khóa SK mới được sinh ra của các bảng trên DDS

- Ta dùng Slowly Changing Dimension để ETL dữ liệu lên DDS. Và ta sử dụng thuộc tính “Status” để lưu vết dữ liệu cũ. Trong ETL nhóm chọn Historical attribute: nếu dữ liệu nạp vào thì status = 1, nếu dữ liệu đó được update thì sẽ chuyển status = 0

III. Các phân cấp được triển khai trong hệ thống và mô hình hóa

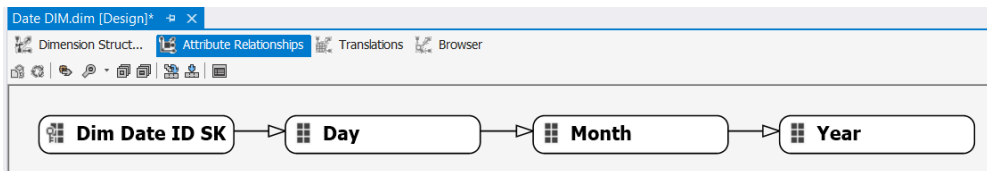
1. Mô hình hóa

- Date Dimension
 - Do nhu cầu phân tích không cần lưu lại giá trị cũ nên không cần Status
- Product Dimension
 - Cần lưu lại giá trị cũ: UnitPrice
 - SCD loại 3: để có thể xem được lịch sử thay đổi giá của sản phẩm
 - Chỉ có 1 nguồn từ SupermarketSales: Chuyển dữ liệu product từ Product_NDS đến Product_DIM
- ProductLine Dimension
 - Cần lưu lại giá trị cũ: ProductLine
 - SCD loại 3: để có thể xem được lịch sử thay đổi của tên loại sản phẩm
 - Chỉ có 1 nguồn từ SupermarketSales: Chuyển dữ liệu product từ ProductLine_NDS đến ProductLine_DIM
- CustomerType Dimension
 - Cần lưu lại giá trị cũ: CustomerTypeName
 - SCD loại 3: để có thể xem được lịch sử thay đổi của tên của các loại khách hàng
 - Chỉ có 1 nguồn từ SupermarketSales: Chuyển dữ liệu product từ CustomerType_NDS đến CustomerType_DIM
- Payment Dimension
 - Cần lưu lại giá trị cũ: PaymentName
 - SCD loại 3: để có thể xem được lịch sử thay đổi thông tin các phương thức thanh toán

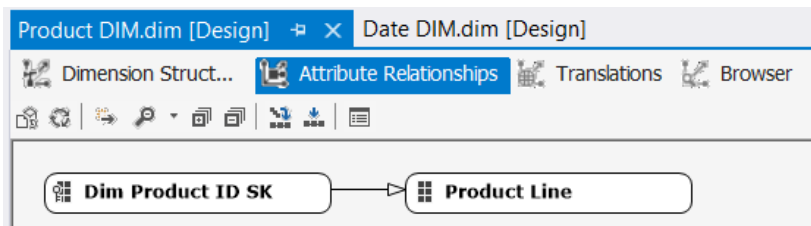
- Chỉ có 1 nguồn từ SupermarketSales: Chuyển dữ liệu product từ CustomerType_NDS đến CustomerType_DIM
- Fact_ SupermarketSales
 - Các giá trị có sẵn từ nguồn: Quantity, Total, Cogs, Rating, GrossIncome, GrossMargin
 - Cấp chi tiết dữ liệu (độ mịn):
 - Đơn vị nhỏ nhất xảy ra sự kiện: Mỗi dòng trong fact tương ứng với 1 đơn hàng bán được của 1 sản phẩm

2. Phân cấp dữ liệu

- Hierarchy Date: Date -> Day -> Month -> Year



- Hierarchy Product: Product -> ProductLine





IV. ETL

1. Source To Stage



Kết quả:

```
use stage
select* from Product_Stage
select* from ProductLine_Stage
select* from City_Stage
select* from SupermarketSales_Stage
```

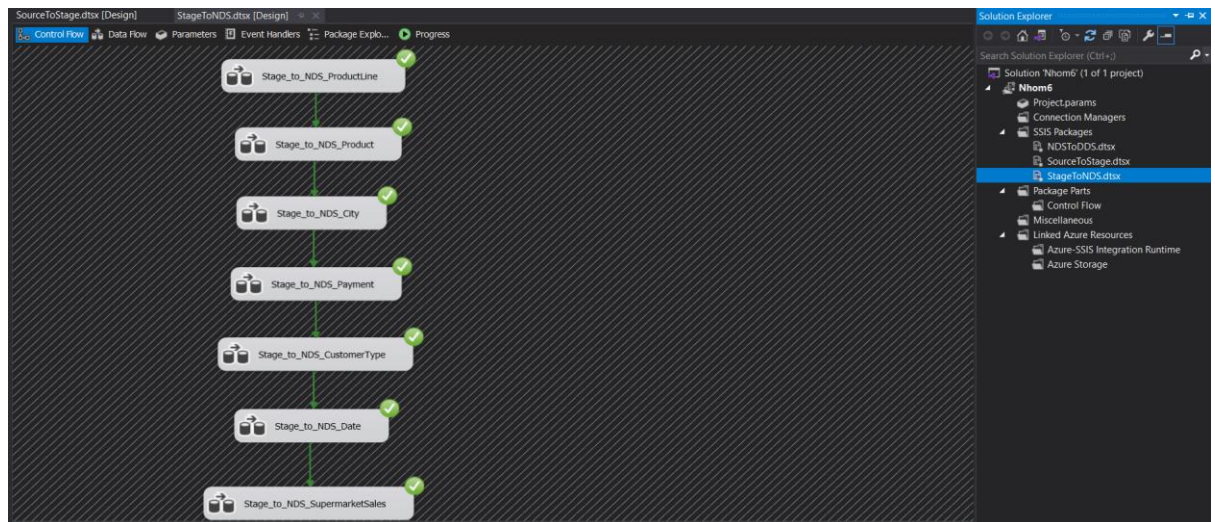
ProductID	Unit price	ProductLine	Created	LastUpdate
PID01	74.69	PD001	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID02	58.22	PD002	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID03	68.84	PD003	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID04	36.26	PD001	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID05	68.93	PD004	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID06	88.63	PD002	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID07	52.59	PD002	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
PID08	62.62	PD003	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000

Product line	ProductLineID	Created	LastUpdate
Health and beauty	PD001	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
Electronic accessories	PD002	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
Home and lifestyle	PD003	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
Sports and travel	PD004	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
Food and beverages	PD005	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
Fashion accessories	PD006	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000

Branch	City	Created	LastUpdate
A	Yangon	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
C	Naypyitaw	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
B	Mandalay	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000

Invoice ID	Branch	Customer type	Gender	ProductID	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating	Created	LastUpdate
750-67-8428	A	Member	Female	PID01	7	26.1415	548.9715	2019-01-05 00:00:00.000	1899-12-30 13:08:00.000	Ewallet	522.83	4.761904762	26.1415	9.1	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
123-19-1176	B	Member	Male	PID02	8	23.288	489.048	2019-01-27 00:00:00.000	1899-12-30 20:33:00.000	Cash	465.76	4.761904762	23.288	8.4	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
355-53-5943	C	Normal	Female	PID03	6	20.652	433.692	2019-02-25 00:00:00.000	1899-12-30 14:36:00.000	Credit ...	413.04	4.761904762	20.652	5.8	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
665-32-9167	A	Member	Female	PID04	2	3.626	76.146	2019-01-10 00:00:00.000	1899-12-30 17:15:00.000	Credit ...	72.52	4.761904762	3.626	7.2	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
656-95-9349	A	Member	Female	PID05	7	24.1255	506.6355	2019-03-11 00:00:00.000	1899-12-30 11:03:00.000	Credit ...	482.51	4.761904762	24.1255	4.6	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
549-59-1358	A	Member	Male	PID06	3	13.2945	279.1845	2019-03-02 00:00:00.000	1899-12-30 17:36:00.000	Ewallet	265.89	4.761904762	13.2945	6	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
227-03-5010	A	Member	Female	PID07	6	21.036	441.756	2019-03-22 00:00:00.000	1899-12-30 19:20:00.000	Credit ...	420.72	4.761904762	21.036	8.5	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000
129-29-8530	A	Member	Male	PID08	5	15.655	328.755	2019-03-10 00:00:00.000	1899-12-30 19:15:00.000	Ewallet	313.1	4.761904762	15.655	7	2023-11-20 00:00:00.000	2023-11-20 00:00:00.000

2. Stage To NDS



Kết quả:

```

use NDS
select* from City_NDS
select* from CustomerType_NDS
select* from Payment_NDS
select* from Product_NDS
select* from ProductLine_NDS
select* from Date_NDS
select* from Supermarket_Sales_NDS

```

Branch_SK	Branch	City	Created	LastUpdated	SourceID
1	A	Yangon	2023-12-22 10:06:43.647	2023-12-22 10:08:01.427	1
2	C	Naypyitaw	2023-12-22 10:06:43.647	2023-12-22 10:08:01.427	1
3	B	Mandalay	2023-12-22 10:06:43.647	2023-12-22 10:08:01.427	1

CustomerType_SK	CustomerTypeName	Created	LastUpdated	SourceID
1	Normal	2023-12-22 10:06:44.193	2023-12-22 10:06:44.193	1
2	Member	2023-12-22 10:06:44.193	2023-12-22 10:06:44.193	1

PaymentID_SK	PaymentName	Created	LastUpdated	SourceID
1	Ewallet	2023-12-22 10:06:43.897	2023-12-22 10:06:43.897	1
2	Cash	2023-12-22 10:06:43.897	2023-12-22 10:06:43.897	1
3	Credit card	2023-12-22 10:06:43.897	2023-12-22 10:06:43.897	1

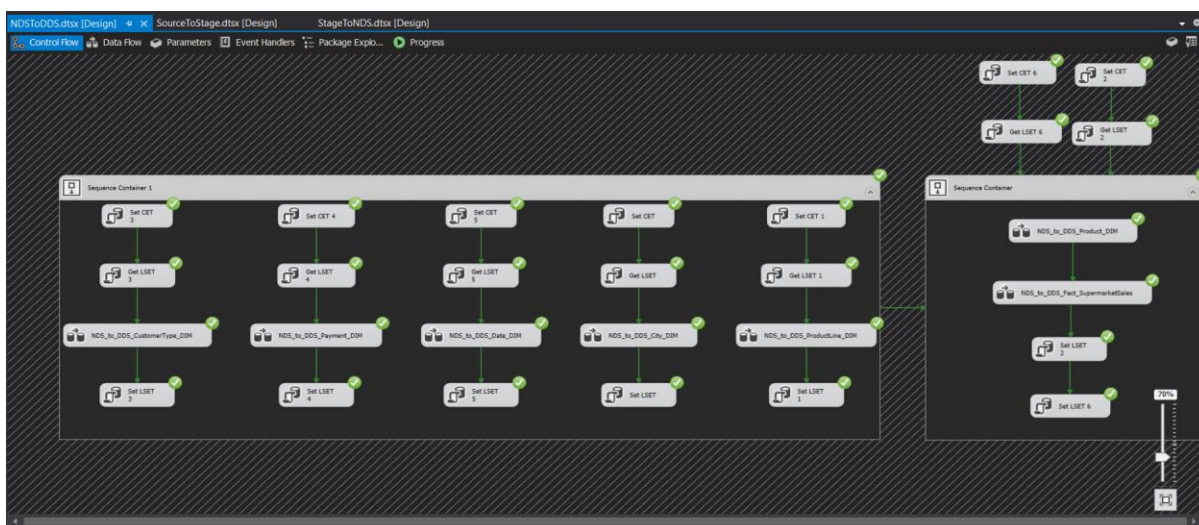
ProductID_SK	ProductID_NK	UnitPrice	ProductLineID_SK	Created	LastUpdated	SourceID
1	PID01	74.89	1	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
2	PID02	58.22	2	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
3	PID03	68.84	3	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
4	PID04	36.26	1	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
5	PID05	68.93	4	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
6	PID06	88.63	2	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
7	PID07	52.59	2	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1
8	PID08	62.62	3	2023-12-22 10:06:43.397	2023-12-22 10:08:01.033	1

ProductLineID_SK	ProductLineID_NK	ProductLine	Created	LastUpdated	SourceID
1	PD001	Health and beauty	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1
2	PD002	Electronic accessories	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1
3	PD003	Home and lifestyle	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1
4	PD004	Sports and travel	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1
5	PD005	Food and beverages	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1
6	PD006	Fashion accessories	2023-12-22 10:06:43.103	2023-12-22 10:08:00.730	1

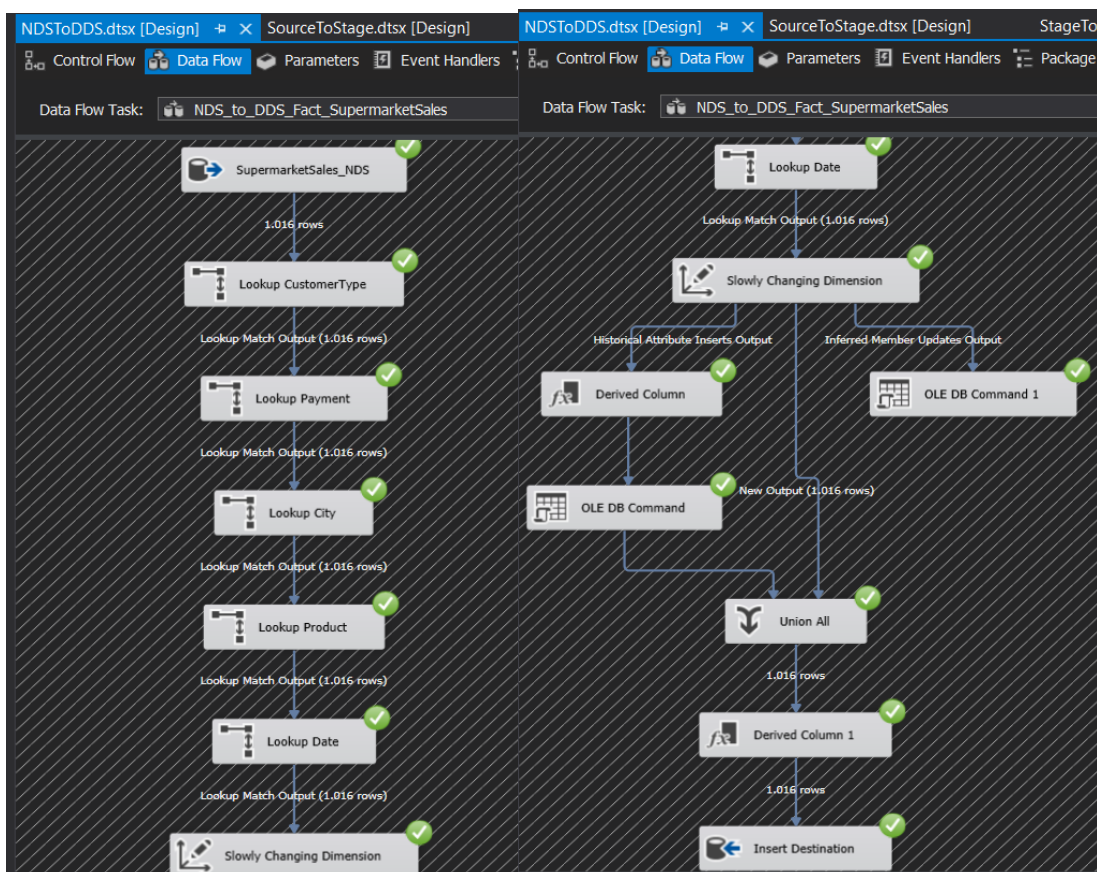
DateID_SK	Date	Day	Month	Year
1	2019-03-22 00:00:00.000	22	3	2019
2	2019-03-16 00:00:00.000	16	3	2019
3	2019-01-04 00:00:00.000	4	1	2019
4	2019-02-21 00:00:00.000	21	2	2019
5	2019-02-04 00:00:00.000	4	2	2019
6	2019-03-28 00:00:00.000	28	3	2019
7	2020-03-05 00:00:00.000	5	3	2020
8	2020-03-11 00:00:00.000	11	3	2020

InvoiceID_SK	Branch_SK	ProductID_SK	Payment_SK	CustomerType_SK	InvoiceID_NK	DateID_SK	Time	Gender	Quantity	Total	cogs	GrossMargin	GrossIncome	Rating	Created	LastUpdated	SourceID
1	1	1	1	2	750-67-8428	69	13:08:00.00000000	Female	7	548.9715	522.83	4.761904762	26.1415	9.1	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
2	2	3	2	2	123-19-1176	33	20:33:00.00000000	Male	8	489.048	465.76	4.761904762	23.288	8.4	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
3	3	2	3	3	355-53-5943	30	14:36:00.00000000	Female	6	433.692	413.04	4.761904762	20.652	5.8	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
4	4	1	4	3	665-32-9167	37	17:15:00.00000000	Female	2	76.146	72.52	4.761904762	3.629	7.2	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
5	5	1	5	3	656-95-9349	49	11:03:00.00000000	Female	7	506.8355	482.51	4.761904762	24.3255	4.6	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
6	6	1	6	1	549-59-1358	12	17:36:00.00000000	Male	3	279.1845	265.89	4.761904762	13.2945	6	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
7	7	1	7	3	227-03-5010	1	19:20:00.00000000	Female	8	441.756	420.72	4.761904762	21.036	8.5	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1
8	8	1	8	1	129-29-8530	86	19:15:00.00000000	Male	5	328.755	313.1	4.761904762	15.655	7	2023-12-22 10:06:44.837	2023-12-22 10:08:02.617	1

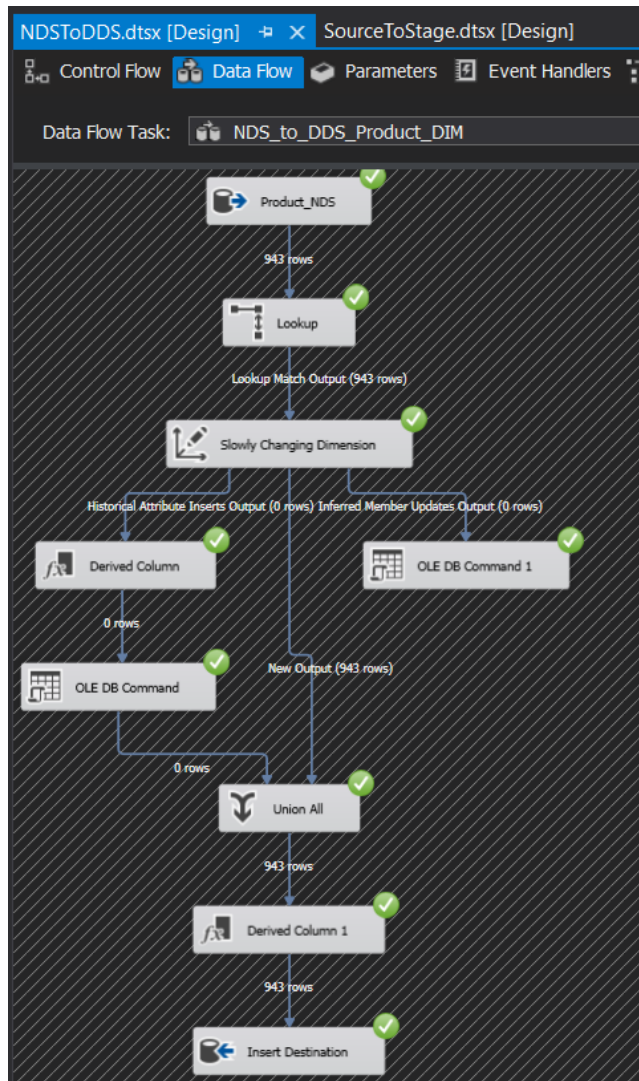
3. NDS To DDS



Flow bảng Fact:



Flow bảng Product_Dim:



Kết quả:

```
use DDS
select* from City_DIM
select* from CustomerType_DIM
select* from Payment_DIM
select* from Product_DIM
select* from ProductLine_DIM
select* from Date_Dim
select* from Fact_SupermarketSales
```

Dim_Branch_SK	BranchID	Branch	City	Status
1	1	A	Yangon	1
2	2	C	Naypyitaw	1
3	3	B	Mandalay	1

Dim_CustomerType_SK	CustomerTypeID	CustomerTypeName	Status
1	1	Normal	1
2	2	Member	1

Dim_PaymentID_SK	PaymentID	PaymentName	Status
1	1	Ewallet	1
2	2	Cash	1
3	3	Credit card	1

Dim_ProductID_SK	ProductID	UnitPrice	Dim_ProductLineID_SK	Status
1	1	74.99	1	1
2	2	58.22	2	1
3	3	68.84	3	1
4	4	36.26	1	1
5	5	68.93	4	1
6	6	88.63	2	1
7	7	52.59	2	1
8	8	62.62	3	1



Dim_ProductLineID_SK	ProductLineID	ProductLine	Status
1	1	Health and beauty	1
2	2	Electronic accessories	1
3	3	Home and lifestyle	1
4	4	Sports and travel	1
5	5	Food and beverages	1
6	6	Fashion accessories	1

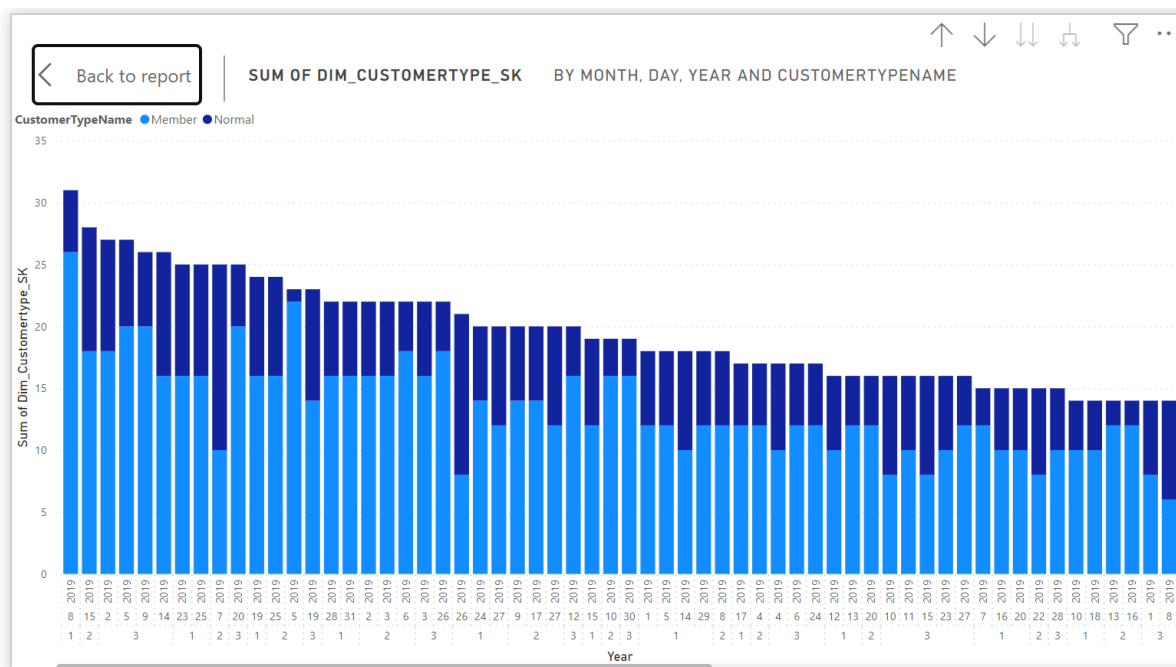
Dim_DateID_SK	DateID	Date	Day	Month	Year
1	1	2019-03-22 00:00:00.000	22	3	2019
2	2	2019-03-16 00:00:00.000	16	3	2019
3	3	2019-01-04 00:00:00.000	4	1	2019
4	4	2019-02-21 00:00:00.000	21	2	2019
5	5	2019-02-04 00:00:00.000	4	2	2019
6	6	2019-03-28 00:00:00.000	28	3	2019
7	7	2020-03-05 00:00:00.000	5	3	2020
8	8	2020-03-11 00:00:00.000	11	3	2020

Dim_InvoiceID_SK	InvoiceID	Dim_Branch_SK	Dim_ProductID_SK	Dim_Payment_SK	Dim_Customertype_SK	Dim_DateID_SK	Gender	Quantity	Total	Time	cogs	GrossMargin	GrossIncome	Rating	Status
1	1	1	1	1	2	69	Female	7	548.9715	13:08:00.00000000	522.83	4,761904762	26,1415	9,1	1
2	2	2	2	2	2	33	Male	8	489.048	20:33:00.00000000	465.76	4,761904762	23,288	8,4	1
3	3	3	2	3	1	30	Female	6	433.692	14:36:00.00000000	413.04	4,761904762	20,652	5,8	1
4	4	4	1	4	3	37	Female	2	76.146	17:15:00.00000000	72.52	4,761904762	3,626	7,2	1
5	5	5	1	5	3	49	Female	7	506.6355	11:03:00.00000000	482.51	4,761904762	24,1255	4,6	1
6	6	6	1	6	1	12	Male	3	278.1845	17:36:00.00000000	265.99	4,761904762	13,2845	6	1
7	7	7	1	7	3	1	Female	8	441.756	19:20:00.00000000	420.72	4,761904762	21,036	8,5	1
8	8	8	1	8	2	86	Male	5	328.755	19:15:00.00000000	313.1	4,761904762	15,655	7	1

V. Report

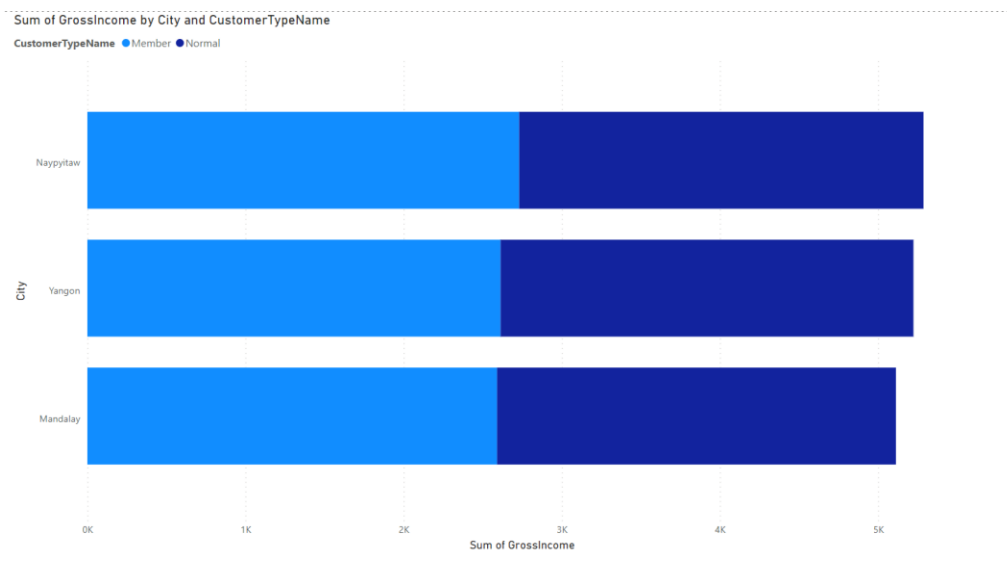
1. Thống kê số thành viên mua hàng theo ngày, tháng, năm

- Sự kiện: Khi khách hàng mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng
 - Cái gì: Sản phẩm
 - Khi nào: ngày mua hàng
- Đo lường: Số lượng

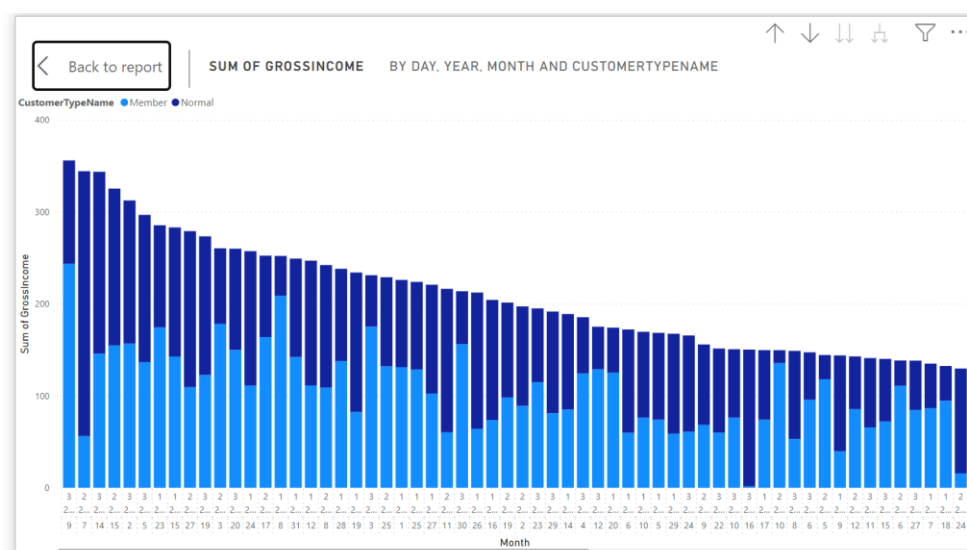


2. Thống kê doanh thu của khách hàng (member, normal) theo ngày, tháng, năm và theo chi nhánh.

- Sự kiện: Khi khách hàng mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng
 - Cái gì: Sản phẩm
 - Khi nào: ngày mua hàng
- Đo lường: Đơn giá, giá trị



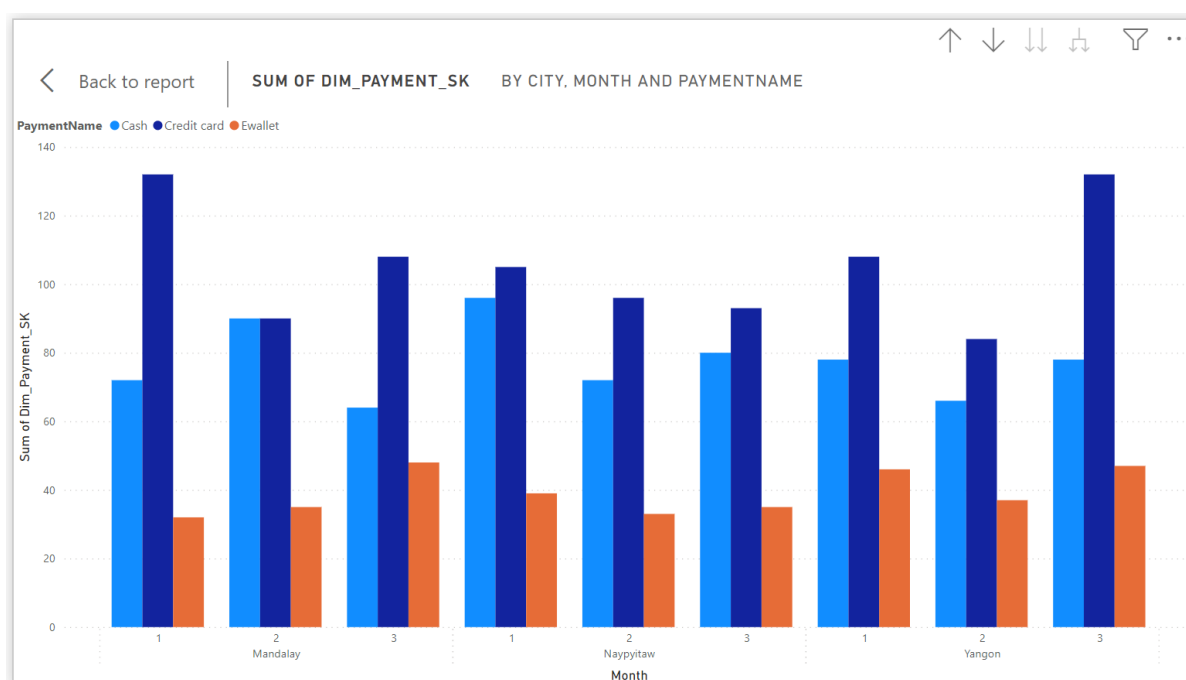
Thống kê doanh thu của khách hàng theo chi nhánh



Thống kê doanh thu của khách hàng theo năm

3. Thống kê số lượng khách thanh toán theo cash/ debit/... ở từng chi nhánh theo từng tháng trong năm

- Sự kiện: Khi khách hàng mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
 - Cái gì: Phương thức thanh toán
 - Khi nào: ngày mua hàng
- Đo lường: Số lượng

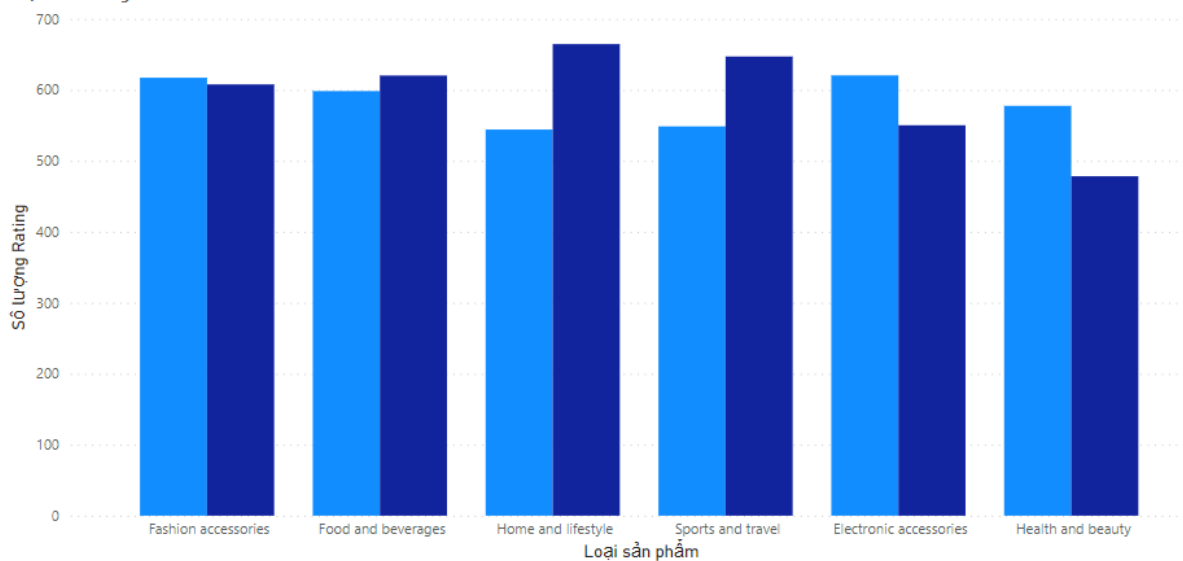


4. Thống kê lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine)

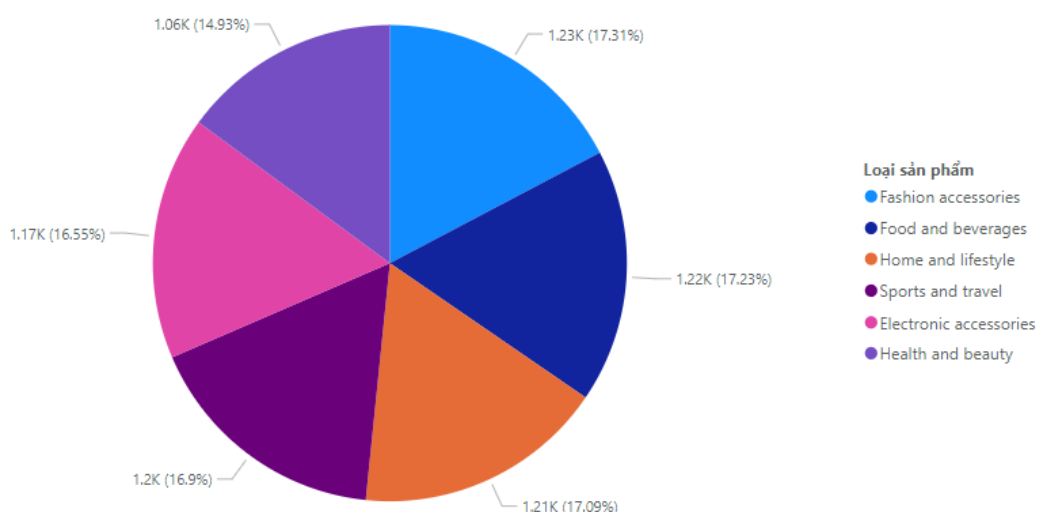
- Sự kiện: Khi khách hàng đánh giá 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
 - Cái gì: Rating của khách hàng
 - Khi nào: ngày mua hàng
- Đo lường: Số lượng

Số lượng Rating của khách hàng theo từng loại sản phẩm

Loại khách hàng ● Member ● Normal



Số lượng rating của khách hàng theo từng loại sản phẩm



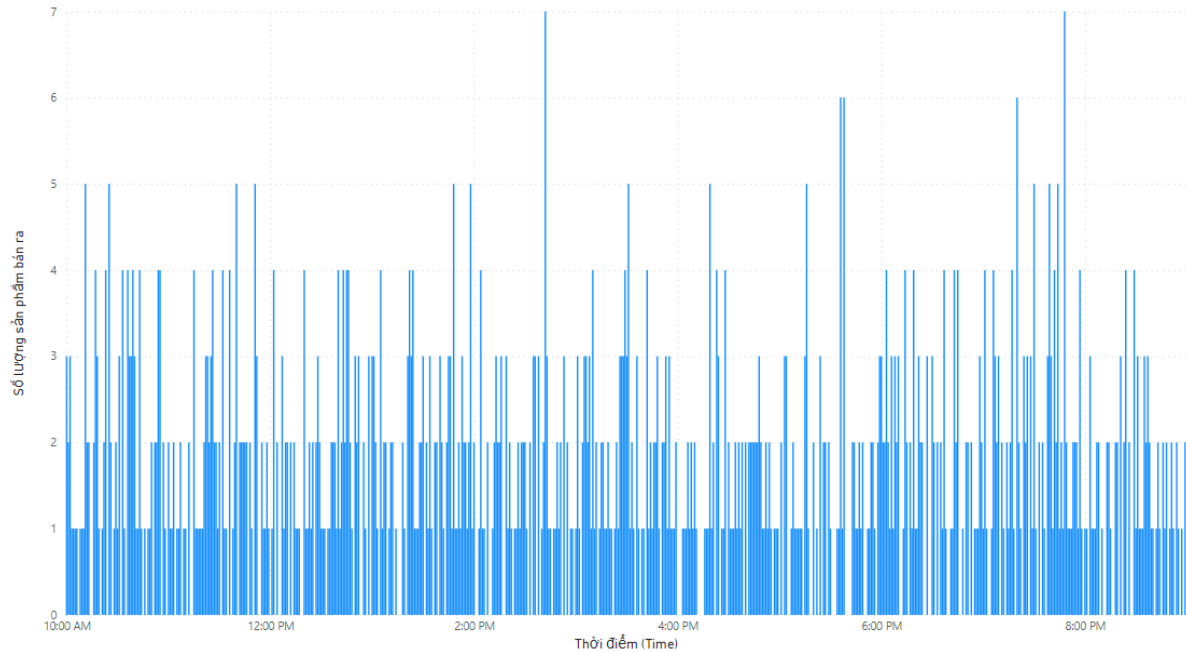
5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time / date)

- Sự kiện: Khi khách hàng mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
 - Cái gì: Sản phẩm

- Khi nào: ngày mua hàng
- Đo lường: Thời gian

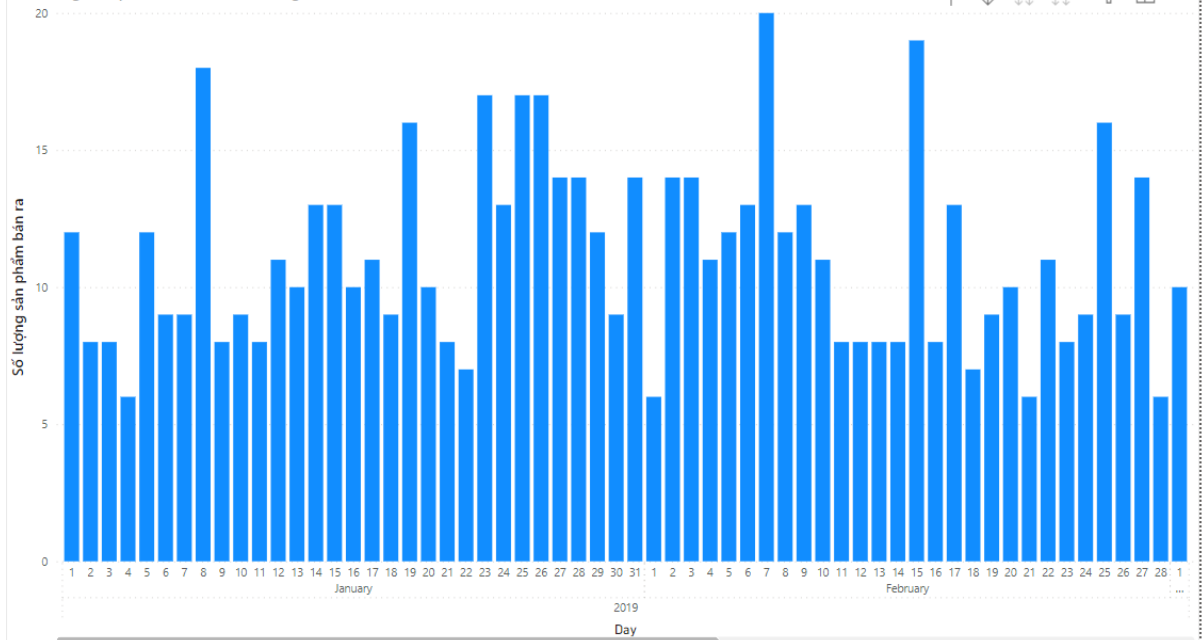
Thống kê số lượng sản phẩm bán được theo từng thời điểm (Time)

Số lượng theo sản phẩm bán ra theo từng thời điểm (Time)



Thống kê số lượng sản phẩm bán được theo từng thời điểm (Date)

Số lượng sản phẩm bán ra theo từng thời điểm (Date)

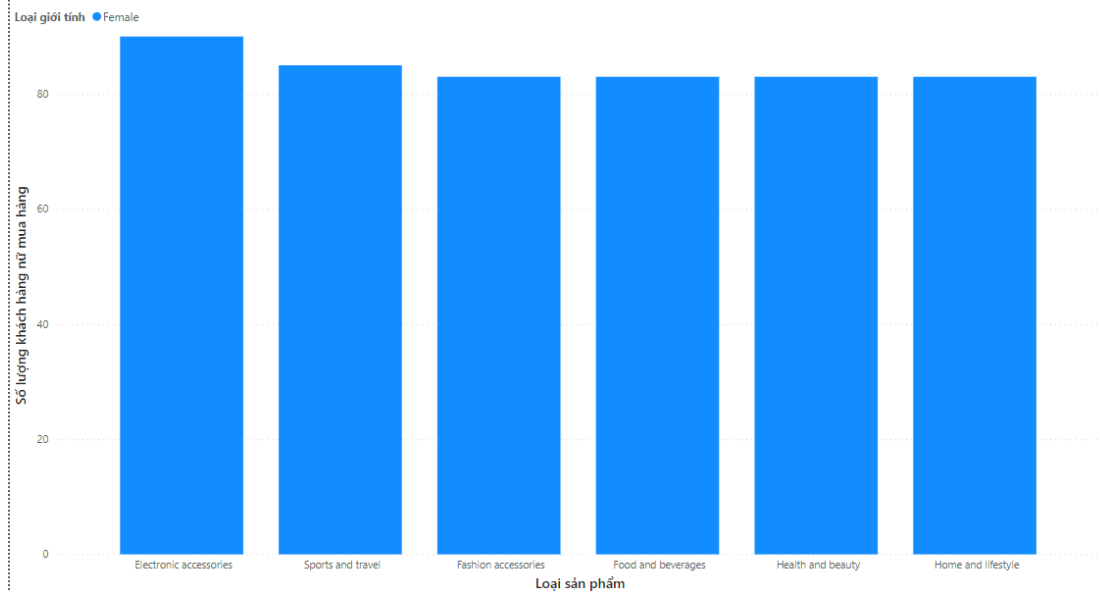


6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm

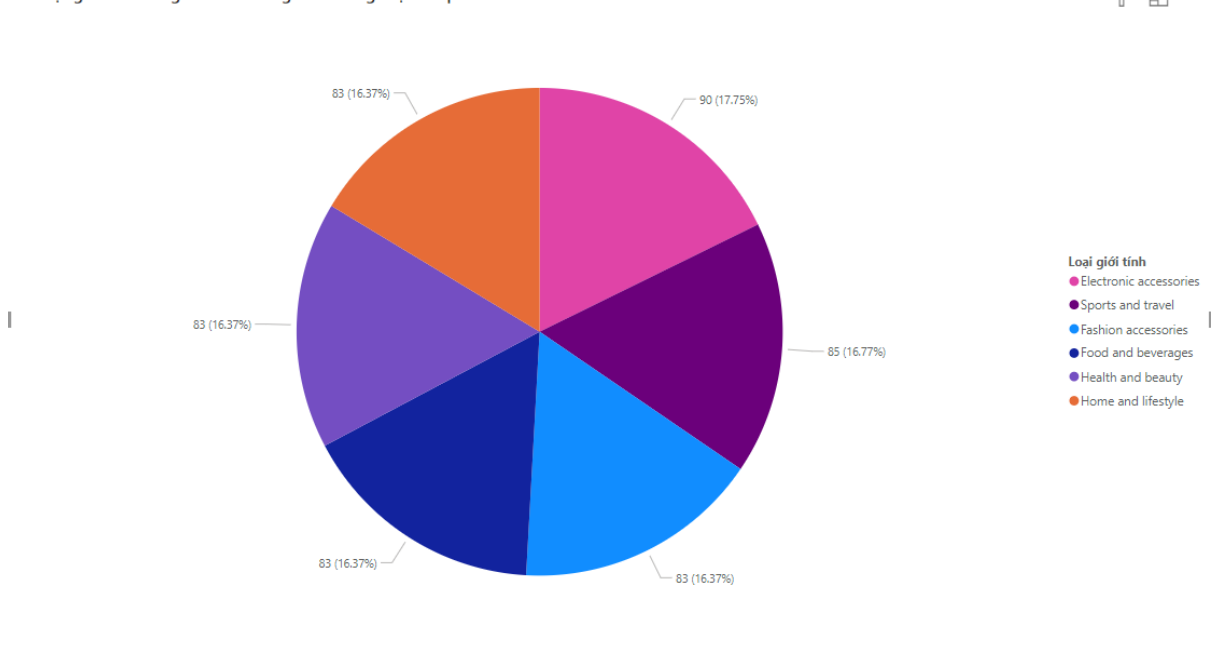
- Sự kiện: Khi khách hàng nữ mua 1 sản phẩm
- Bối cảnh sự kiện:

- Ai: Khách hàng
- Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
- Cái gì: Sản phẩm
- Khi nào: ngày mua hàng
- Đo lường: Số lượng

Số lượng khách hàng nữ mua hàng theo từng loại sản phẩm



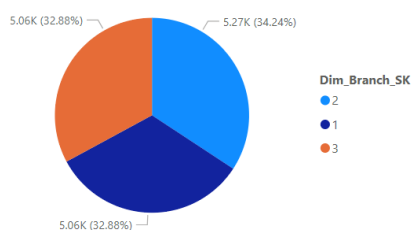
Số lượng khách hàng nữ mua hàng theo từng loại sản phẩm



7. Cho biết doanh thu của từng chi nhánh của các năm

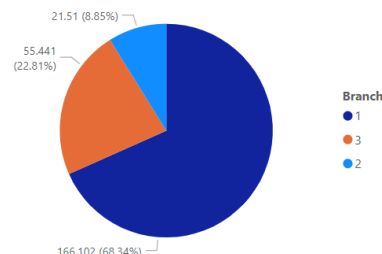
- Sự kiện: Khi khách hàng nữ mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
 - Cái gì: Sản phẩm
 - Khi nào: Năm mua hàng
- Đo lường: Đơn giá

Sum of GrossIncome by Dim_Branch_SK and Year



2019

Sum of GrossIncome by Branch and Year

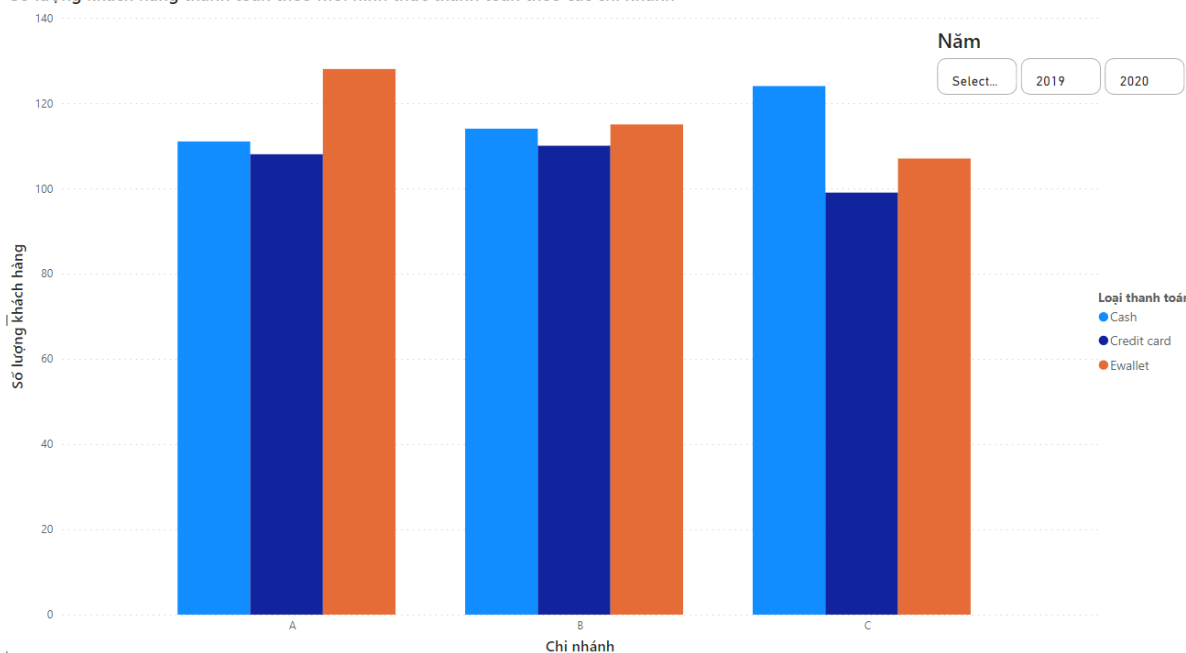


2020

8. Cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit... của mỗi chi nhánh / của tất cả chi nhánh theo từng năm...

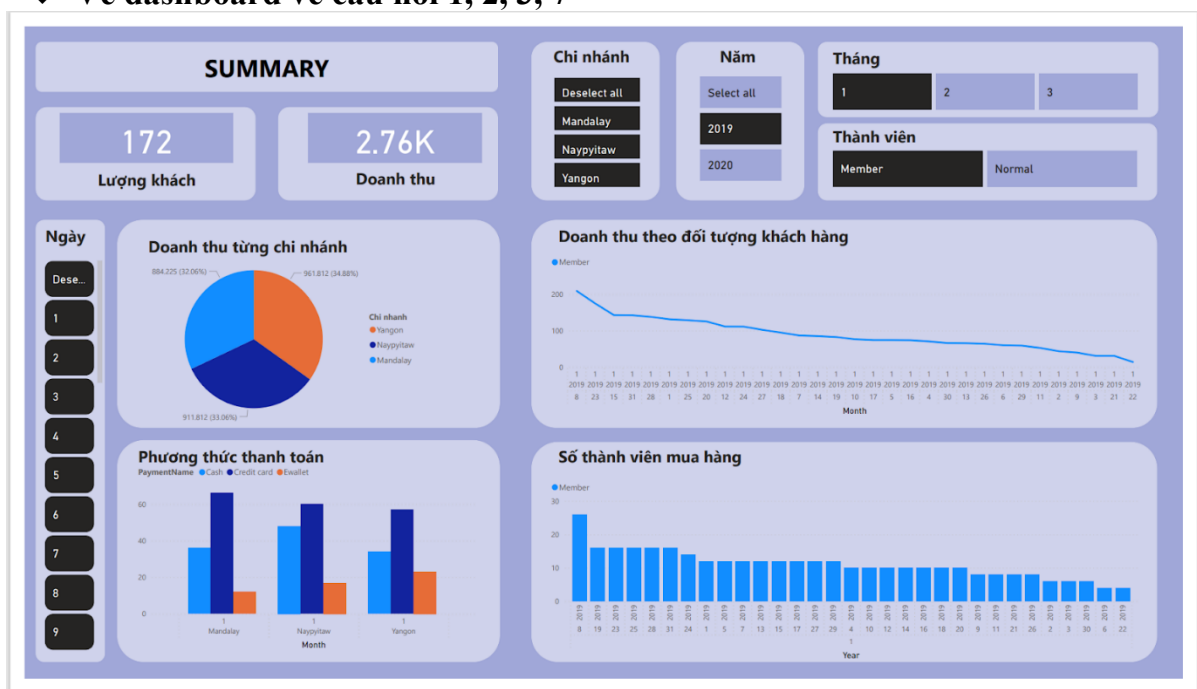
- Sự kiện: Khi khách hàng nữ mua 1 sản phẩm
- Bối cảnh sự kiện:
 - Ai: Khách hàng
 - Ở đâu: Cửa hàng, các chi nhánh thuộc quản lý
 - Cái gì: Phương thức thanh toán
 - Khi nào: Năm mua hàng
- Đo lường: Số lượng

Số lượng khách hàng thanh toán theo mỗi hình thức thanh toán theo các chi nhánh



Tổng quan

❖ Vẽ dashboard về câu hỏi 1, 2, 3, 7



Dựa vào bảng thống kê, ta thấy rằng:

- Doanh thu bán hàng vào mỗi đầu tháng đều cao nhưng lại có xu hướng giảm dần vào cuối tháng.
- Doanh thu giữa các chi nhánh không có sự chênh lệch nhiều
- Lượng khách hàng là thành viên có nhu cầu mua hàng nhỉnh hơn so với lượng khách hàng thông thường. Và khách hàng cũng có xu hướng ít mua hàng hơn so với những ngày đầu tháng.

- Phương thức thanh toán bằng thẻ tín dụng được ưa chuộng nhất ở cả 3 chi nhánh.

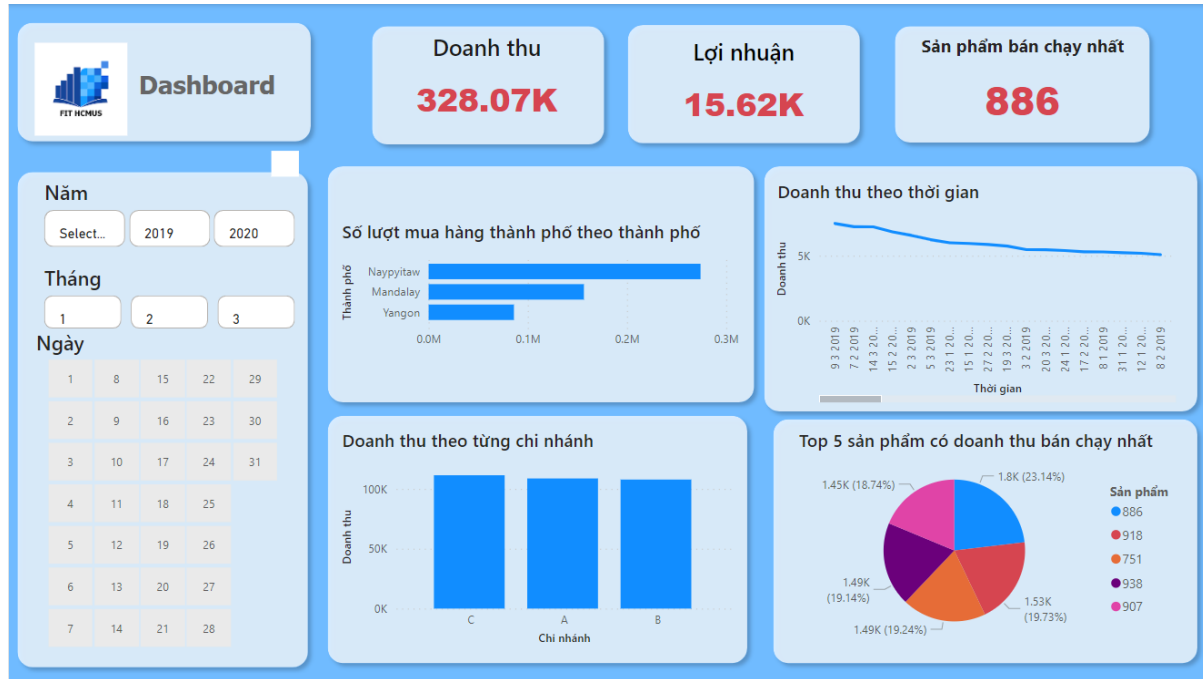
❖ Vẽ dashboard về câu hỏi 4, 5, 6, 8



Dựa vào bảng thống kê, ta thấy rằng:

- Hầu hết không có sự khác biệt về số lượng rating của khách hàng về từng loại sản phẩm
- Thời gian bán hàng từ 10AM tới 9PM, sản phẩm được bán ra nhiều nhất vào khung giờ chiều (3PM) và khung giờ tối (8PM)
- Trong năm 2019, 2020, phương thức thanh toán bằng Ewallet là phổ biến nhất, không có sự chênh lệch nhiều về số lượng khách hàng khi thanh toán bằng 3 phương thức
- Không có sự chênh lệch nhiều về số lượng khách hàng nữ mua hàng theo từng loại sản phẩm

❖ Vẽ dashboard theo dõi doanh thu, top sản phẩm bán chạy, chi nhánh đạt doanh thu cao nhất, city có đông lượt mua hàng nhất, theo dõi doanh thu của 1 chi nhánh bất kỳ



Dựa vào bảng thống kê, ta thấy rằng:

- Tổng doanh thu của cả 3 chi nhánh là 328.07K và lợi nhuận là 15.62K
- Sản phẩm có mã 886 là sản phẩm có doanh thu bán chạy nhất
- Thành phố Naypyitaw là thành phố có đông lượt mua hàng nhất
- Doanh thu giảm dần theo thời gian
- Chi nhánh C có doanh thu cao nhất và doanh thu của 3 chi nhánh không chênh lệch nhiều

VI. OLAP

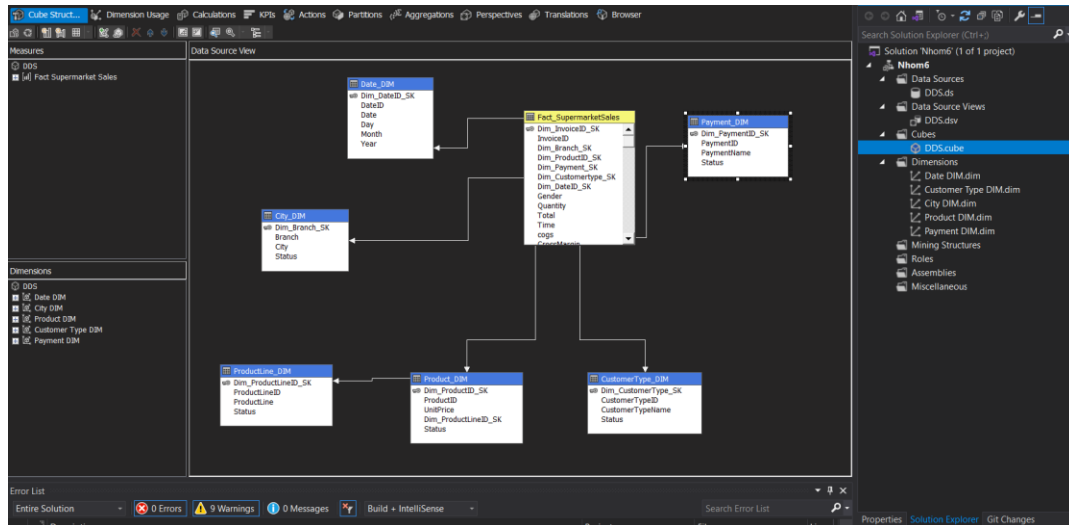
1. Yêu cầu đề

Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm, theo thời gian, hình thức thanh toán

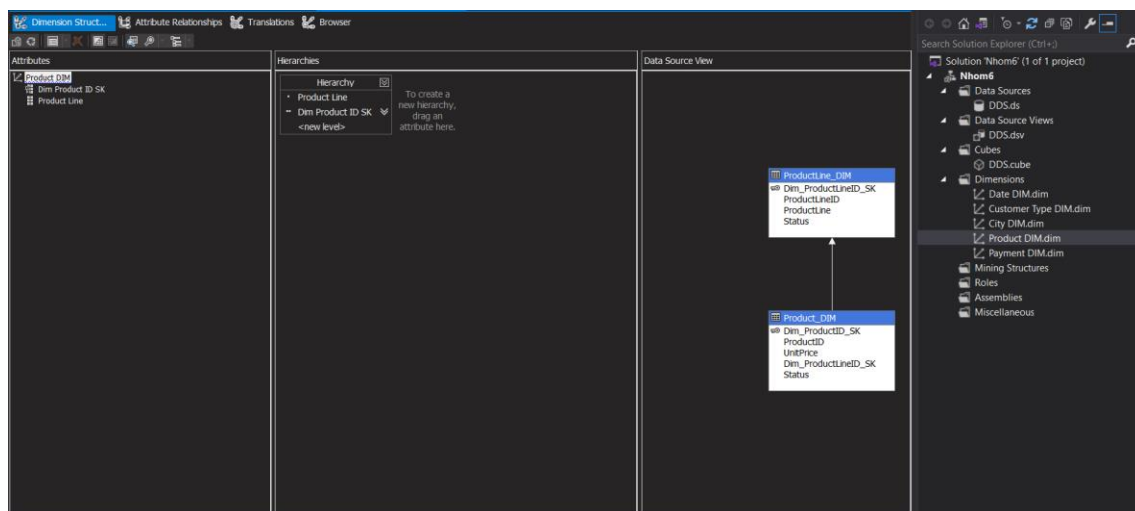
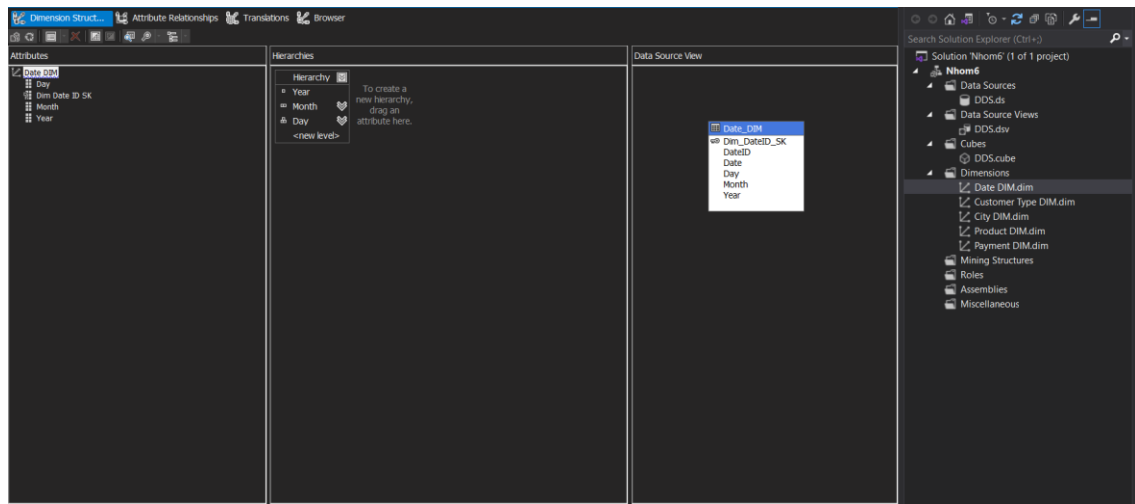
- Cần Phân cấp chiều cho Date: để có thể phân tích dữ liệu theo thời gian
- Cần phân cấp chiều cho ProductLine: để có thể xác định được các sản phẩm nào thuộc loại sản phẩm nào

2. Thực hiện bằng Cube và Excel

1. Thực hiện tạo 1 Cube

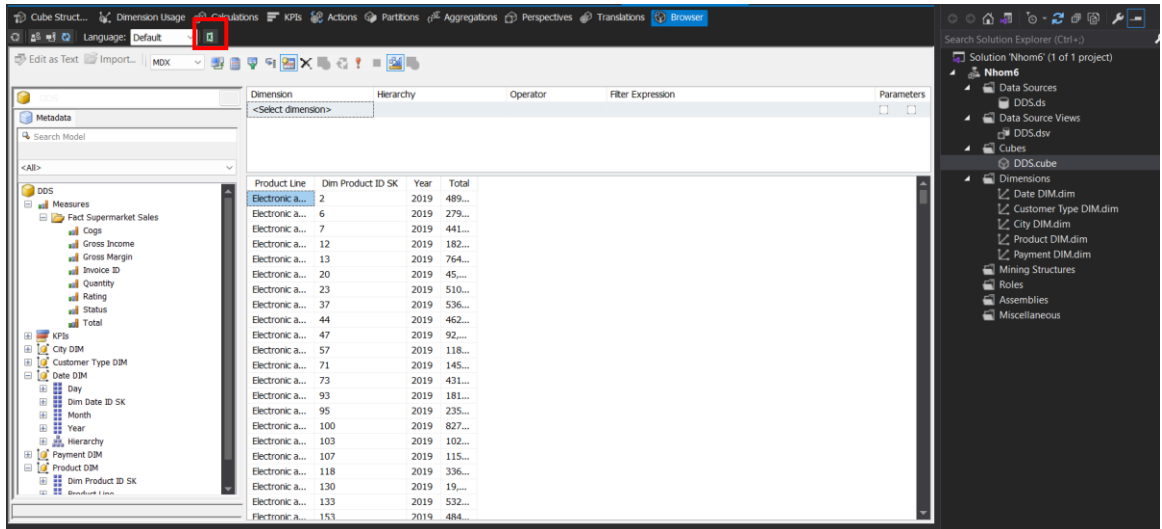


2. Tạo Hierarchy cho Date Dim và Product Dim

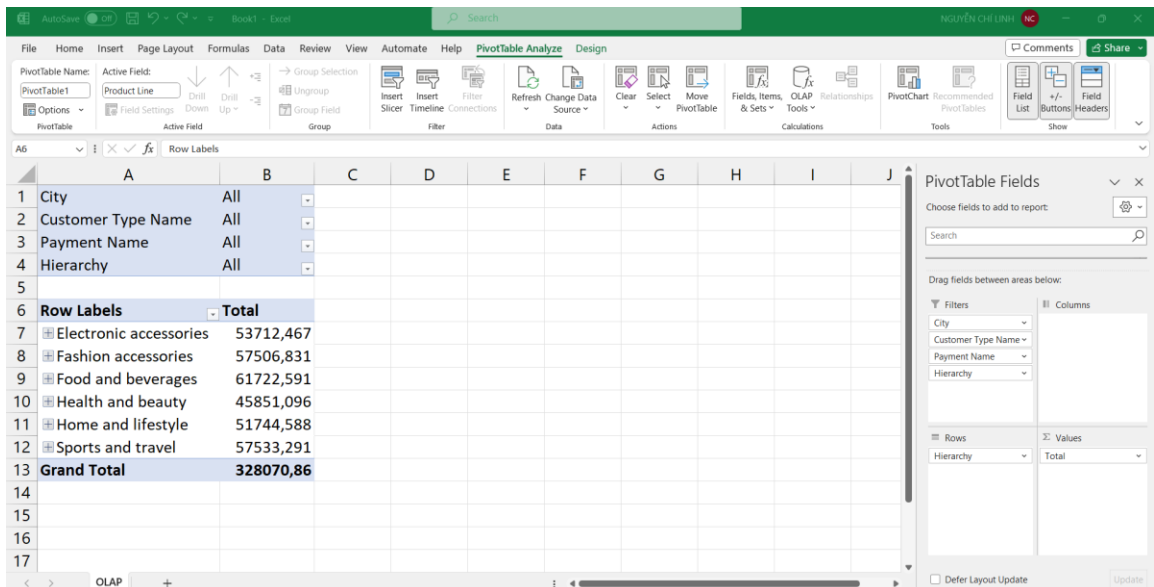


3. Ta thực hiện Process lại để có các Hierarchy vừa tạo

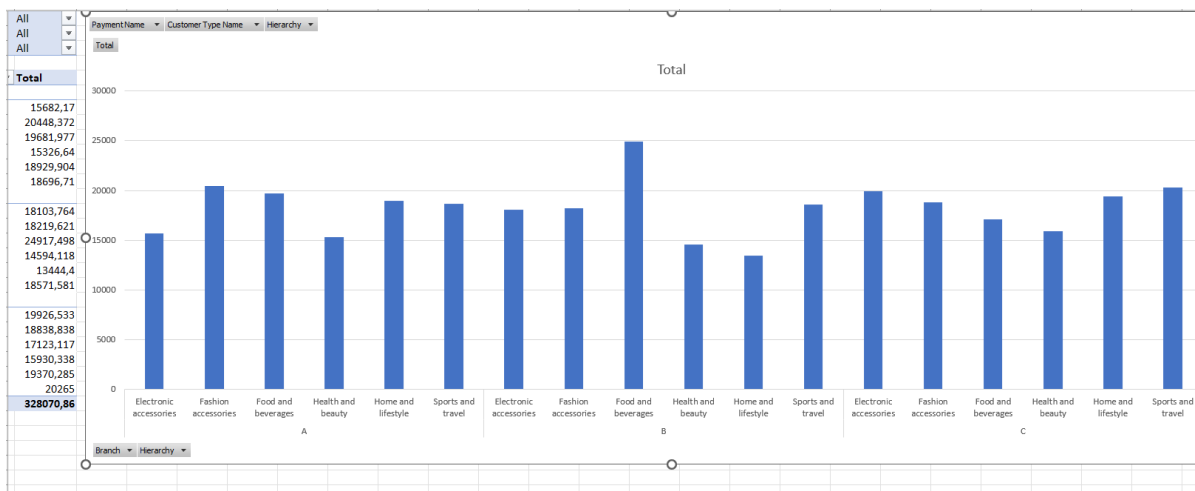
4. Vào phần Browser kiểm tra xem đã có các Dim và Hierarchy chưa
5. Sau khi kiểm tra xong ta sẽ chọn vào excel và chuyển sang file Excel để có thể trực quan



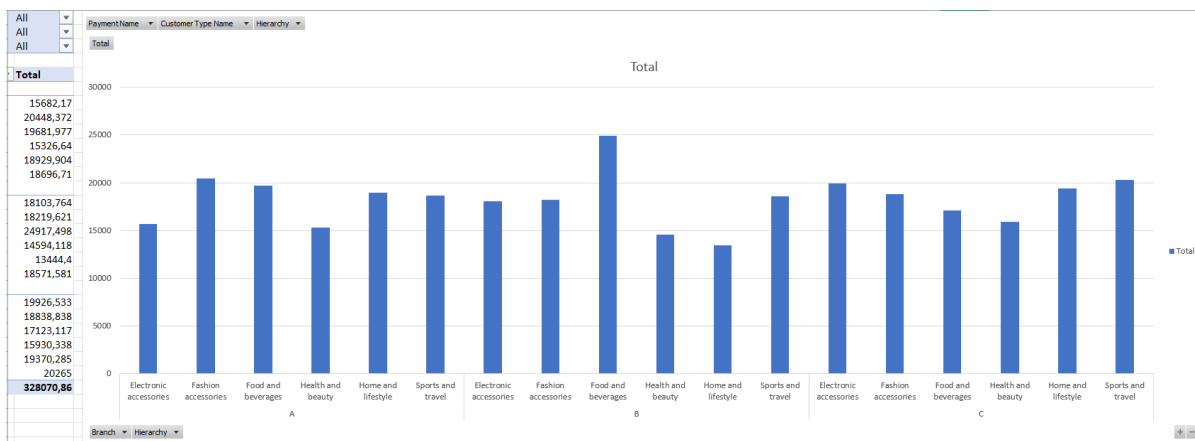
6. Dựa theo yêu cầu đề bài ta setting các thông tin như sau:
 - a. Filters: City, Customer Type Name, Payment Name, Hierarchy của Date DIM
 - b. Row: Hierarchy của bảng Product DIM
 - c. Values: Total tổng danh thu được lấy từ bảng fact



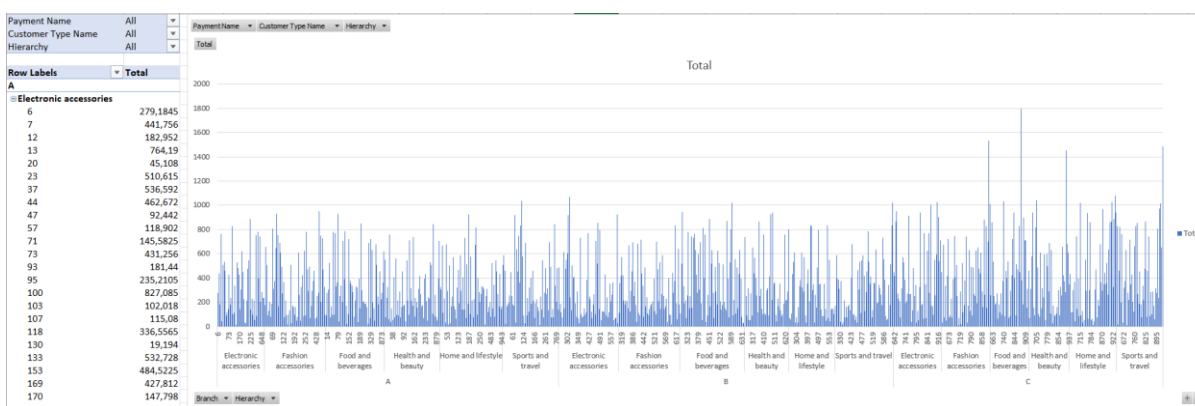
7. Tiếp đến ta thực hiện tạo 1 dashboard từ các dữ kiện mà ta đã tạo



- Ta có thể filter theo nhu cầu mà ta cần phân tích



- Ta có thể xem chi tiết cụ thể từng sản phẩm của loại sản phẩm



3. MDX

- Tình hình mua hàng của từng loại khách hàng theo từng chi nhánh

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

- Measures
- KPIs
- City DIM
- Customer Type DIM
- Date DIM
- Payment DIM
- Product DIM

```
// Tình hình mua hàng của từng loại khách hàng theo từng chi nhánh
SELECT NON EMPTY Crossjoin([Customer Type DIM].[Customer Type Name].MEMBERS,
    {[Measures].[total]}) ON COLUMNS,
    NON EMPTY [City DIM].[Branch].MEMBERS ON ROWS
FROM [DDS];
```

114 %

	All	Member	Normal
	Total	Total	Total
All	328070.862	166166.616	161904.246
A	108765.7725	53942.3115	54823.461
B	107850.981	54891.312	52959.669
C	111454.1085	57332.9925	54121.116

- Tình hình mua hàng của khách hàng theo từng chi nhánh, theo từng loại sản phẩm

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

- Measures
- KPIs
- City DIM
- Customer Type DIM
- Date DIM
- Payment DIM
- Product DIM

```
// Tình hình mua hàng của khách hàng theo từng chi nhánh, theo từng loại sản phẩm
SELECT NON EMPTY Crossjoin([Product DIM].[Hierarchy].[Product Line].Members,
    {[Measures].[Total]}) ON COLUMNS,
    NON EMPTY [City DIM].[Branch].MEMBERS ON ROWS
FROM [DDS];
```

114 %

	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
	Total	Total	Total	Total	Total	Total
All	53712.4665	57506.8305	61722.591	45851.0955	51744.588	57533.2905
A	15682.17	20448.372	19681.977	15326.64	18929.904	18696.7095
B	18103.764	18219.621	24917.4975	14594.118	13444.3995	18571.581
C	19926.5325	18838.8375	17123.1165	15930.3375	19370.2845	20265

- Tình hình mua hàng của khách hàng theo từng chi nhánh, theo từng loại sản phẩm. Theo thời gian(Year, month)

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

- Measures
- KPIs
- City DIM
- Customer Type DIM
- Date DIM
- Payment DIM
- Product DIM

```
// Tình hình mua hàng của khách hàng theo từng chi nhánh, theo từng loại sản phẩm. Theo thời gian(Year, month)
SELECT NON EMPTY Crossjoin([Product DIM].[Hierarchy].[Product Line].Members,
    {[Measures].[Total]}) ON COLUMNS,
    NON EMPTY Crossjoin([City DIM].[Branch].MEMBERS,
        [Date DIM].[Hierarchy].[Year].MEMBERS, [Month].MEMBERS) ON ROWS
FROM [DDS];
```

114 %

		Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
		Total	Total	Total	Total	Total	Total
All	2019	52393.1835	56585.13	61054.581	45163.02	51132.795	56638.0395
All	2019	14382.4485	23588.5965	21058.338	18623.7975	17676.12	20962.5675
All	2019	17472.1575	16313.115	17735.3295	14279.454	15626.793	15792.525
All	2019	20538.5775	16683.4185	22260.9135	12259.7685	17829.882	19882.947
All	2019	1319.283	921.7005	668.01	688.0755	611.793	895.251
All	2020	590.436	172.2105	(null)	(null)	457.443	(null)
All	2020	451.71	749.49	281.8535	181.44	(null)	453.495
All	2020	277.137	(null)	388.3265	506.6355	154.35	441.756
A	2019	15682.17	19626.6715	19156.305	14638.5645	18472.461	17801.4585
A	2019	1	3238.746	7381.4265	6968.107	8674.3545	4949.2695
A	2019	2	4395.5205	5555.9595	4716.999	4258.5165	3536.3535
A	2019	3	8047.9035	6589.2855	7479.129	3391.941	5061.753
A	2020	All	(null)	921.7005	525.672	688.0755	457.443
A	2020	1	(null)	172.2105	(null)	(null)	457.443
A	2020	2	(null)	749.49	246.4875	181.44	453.495
A	2020	3	(null)	(null)	279.1845	506.6355	441.756
B	2019	All	17236.191	18219.621	24775.1595	14594.118	13290.0495
B	2019	1	4878.384	9423.162	8337.7035	5440.029	2925.447
B	2019	2	7051.9995	4831.6065	7632.9225	4641.189	5492.5185
B	2019	3	5305.8075	3964.8525	8804.5335	4512.9	4872.084
B	2020	All	867.573	(null)	142.338	(null)	154.35
B	2020	1	590.436	(null)	(null)	(null)	(null)
B	2020	2	(null)	(null)	35.196	(null)	(null)
B	2020	3	277.137	(null)	107.142	(null)	154.35
C	2019	All	19474.8225	18838.8375	17123.1165	15930.3375	19370.2845
C	2019	1	6265.3185	6784.008	5760.4575	6195.6615	6076.3185
C	2019	2	6024.6375	5925.549	5385.408	5379.7485	6297.921

- Tình hình mua hàng của khách hàng theo từng loại sản phẩm. Theo Hình thức thanh toán

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)* ✕

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

- Measures
- KPIs
- City DIM
- Customer Type DIM
- Date DIM
- Payment DIM
- Product DIM

```
// Tình hình mua hàng của khách hàng theo từng loại sản phẩm. Theo Hình thức thanh toán
SELECT NON EMPTY Crossjoin([Product DIM].[Hierarchy].[Product Line].Members,
{[Measures].[Total]}) ON COLUMNS,
NON EMPTY Crossjoin([City DIM].[Branch].MEMBERS,
[Payment DIM].[Payment Name].[Payment Name]) ON ROWS
FROM [DDS];
```

114 %

		Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
		Total	Total	Total	Total	Total	Total
All	Cash	17819.025	21452.2455	19895.043	16551.3705	18354.0525	20106.1455
All	Credit card	13113.3555	18986.8035	24783.9795	13173.447	17231.5605	15559.383
All	Ewallet	22780.086	17067.7815	17043.5685	16126.278	16158.975	21867.762
A	Cash	6709.6785	6364.0185	3574.9035	4913.9265	6396.2115	6572.0025
A	Credit card	4717.4295	6127.9575	8530.2105	2853.543	7495.3515	4948.3035
A	Ewallet	4255.062	7956.396	7576.863	7559.1705	5038.341	7176.4035
B	Cash	6243.72	6507.921	6498.1245	6055.5705	6042.498	5213.4495
B	Credit card	3023.412	8434.9545	12350.352	3780.6825	3861.0495	5964.126
B	Ewallet	8836.632	3276.7455	6069.021	4757.865	3540.852	7394.0055
C	Cash	4865.6265	8580.306	9822.015	5581.8735	5915.343	8320.6935
C	Credit card	5372.514	4423.8915	3903.417	6539.2215	5875.1595	4646.9535
C	Ewallet	9688.392	5834.64	3397.6845	3809.2425	7579.782	7297.353

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)* ✕

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

- Measures
- KPIs
- City DIM
- Customer Type DIM
- Date DIM
- Payment DIM
- Product DIM

```
SELECT
NON EMPTY Crossjoin([Measures].[total]), {[Product DIM].[Product Line].[Product Line].Members}) ON COLUMNS,
NON EMPTY CrossJoin(
Except([City DIM].[Branch].Members, {[City DIM].[Branch].[ALL]}),
Except([Payment DIM].[Payment Name].Members, {[Payment DIM].[Payment Name].[ALL]}),
Except([Date DIM].[Year].Members, {[Date DIM].[Year].[ALL]})
) ON ROWS
FROM [DDS];
```

114 %

		Total	Total	Total	Total	Total	Total
		Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
A	Cash	6709.6785	5614.5285	3574.9035	4913.9265	6396.2115	6572.0025
A	Cash	2020	749.49	(null)	(null)	(null)	(null)
A	Credit card	2019	5955.747	8530.2105	2346.9075	7037.9085	4506.5475
A	Credit card	2020	172.2105	(null)	506.6355	457.443	441.756
A	Ewallet	2019	7956.396	7051.191	7377.7305	5038.341	6722.9085
A	Ewallet	2020	(null)	525.672	181.44	(null)	453.495
B	Cash	2019	5653.284	6355.7865	6055.5705	6042.498	5213.4495
B	Cash	2020	590.436	142.338	(null)	(null)	(null)
B	Credit card	2019	3023.412	8434.9545	12350.352	3780.6825	3791.3295
B	Credit card	2020	(null)	(null)	(null)	69.72	(null)
B	Ewallet	2019	8559.495	3276.7455	6069.021	4757.865	3456.222
B	Ewallet	2020	277.137	(null)	(null)	(null)	84.63
C	Cash	2019	4865.6265	8580.306	9822.015	5581.8735	5915.343
C	Cash	2020	(null)	(null)	(null)	(null)	(null)
C	Credit card	2019	5372.514	4423.8915	3903.417	6539.2215	5875.1595
C	Credit card	2020	9236.682	5834.64	3397.6845	3809.2425	7579.782
C	Ewallet	2019	9688.392	5834.64	3397.6845	3809.2425	7579.782
C	Ewallet	2020	451.71	(null)	(null)	(null)	(null)

- Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm. Theo thời gian (Year = 2019). Hình thức thanh toán (Cash)

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

Measures

KPIs

City DIM

Customer Type DIM

Date DIM

Payment DIM

Product DIM

```
//Tình hình mua hàng của khách hàng theo từng chi nhánh, từng loại sản phẩm,
//theo thời gian (Year = 2019)
//hình thức thanh toán (Cash)
SELECT NON EMPTY Crossjoin([Product DIM].[Hierarchy].[Product Line].Members,
                             {[Measures].[Total]}) ON COLUMNS,
                             NON EMPTY [City DIM].[Branch].MEMBERS ON ROWS
FROM [DDS]
WHERE ([Date DIM].[Hierarchy].[Year].[2019],
       [Payment DIM].[Payment Name].[Payment Name].[Cash]);
```

114 %

Messages Results

	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel
Total	17228.589	20702.7555	19752.705	16551.3705	18354.0525	20106.1455
All	6709.6785	5614.5285	3574.9035	4913.9265	6396.2115	6572.0025
A	5653.284	6507.921	6355.7865	6055.5705	6042.498	5213.4495
B	4865.6265	8580.306	9822.015	5581.8735	5915.343	8320.6935

- Câu hỏi: Cho biết doanh thu của từng chi nhánh của các năm

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

Measures

KPIs

City DIM

Customer Type DIM

Date DIM

Payment DIM

Product DIM

```
// Câu hỏi: Cho biết doanh thu của từng chi nhánh của các năm
SELECT NON EMPTY Crossjoin([Date DIM].[Hierarchy].[Year],
                             {[Measures].[Total]}) ON COLUMNS,
                             NON EMPTY [City DIM].[Branch].MEMBERS ON ROWS
FROM [DDS];
```

115 %

Messages Results

	2019	2020
Total	322966.749	5104.113
All	105277.6305	3488.142
B	106686.72	1164.261
C	111002.3985	451.71

- Câu hỏi: cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit... của mỗi chi nhánh / của tất cả chi nhánh theo từng năm...

20HTTT2.BI.06.mdx...CGEPSB65\ChiLinh)*

Cube: DDS

Metadata Functions

Search Model

Measure Group: <All>

DDS

Measures

KPIs

City DIM

Customer Type DIM

Date DIM

Payment DIM

Product DIM

```
//Câu hỏi: cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit...
//của mỗi chi nhánh / của tất cả chi nhánh theo từng năm...
SELECT NON EMPTY Crossjoin([Payment DIM].[Payment Name].MEMBERS,
                             [Date DIM].[Hierarchy].[Year]) ON COLUMNS,
                             NON EMPTY [City DIM].[Branch].MEMBERS ON ROWS
FROM DDS;
```

115 %

Messages Results

	Cash	Cash	Credit card	Credit card	Ewallet	Ewallet
	2019	2020	2019	2020	2019	2020
All	181939	2366	153075	1842	173434	3980
A	27689	764	23923	1786	29550	2275
B	50460	1602	49685	56	53890	839
C	103790	(null)	79467	(null)	89994	866

VII. Data Mining

1. Dự đoán số lượng mua hàng của khách hàng

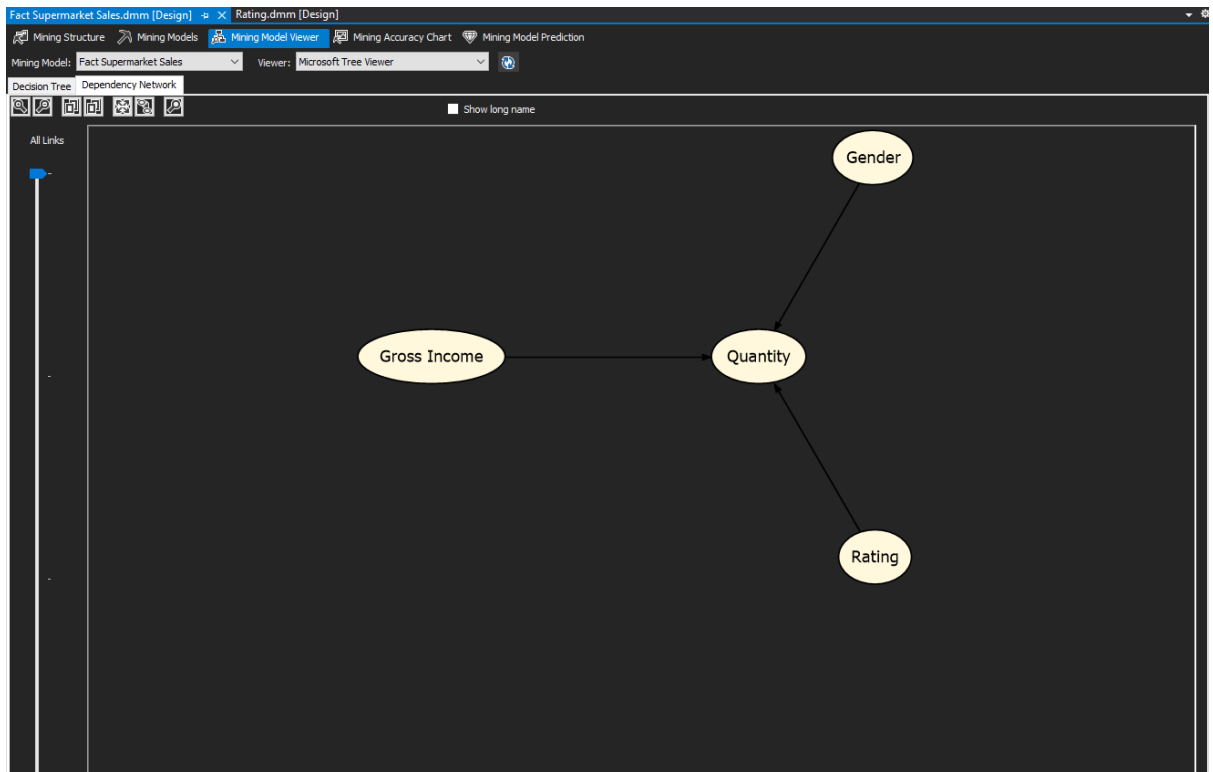
Để dự đoán số lượng sản phẩm mà khách hàng đã mua trong mỗi giao dịch.-> Các thuộc tính ảnh hưởng từ bảng Fact_SupermarketSales:

- Gender : Giới tính của khách hàng có thể ảnh hưởng đến số lượng sản phẩm họ mua.
- Time: Thời gian mua hàng có thể ảnh hưởng đến số lượng sản phẩm mà khách hàng mua.
- Gross Income: Thu nhập có thể ảnh hưởng đến sản phẩm khách hàng mua
- Rating: Đánh giá của khách hàng có thể ảnh hưởng đến số lượng sản phẩm họ mua trong tương lai.
- Sử dụng thuật toán decision tree .

Decision Tree (Cây Quyết Định) là một phương pháp phân loại và hồi quy phổ biến trong data mining. Nó tạo ra một cấu trúc cây quyết định, trong đó mỗi nút của cây đại diện cho một quyết định dựa trên các thuộc tính của dữ liệu đầu vào.

Cấu Trúc Cơ Bản của Decision Tree:

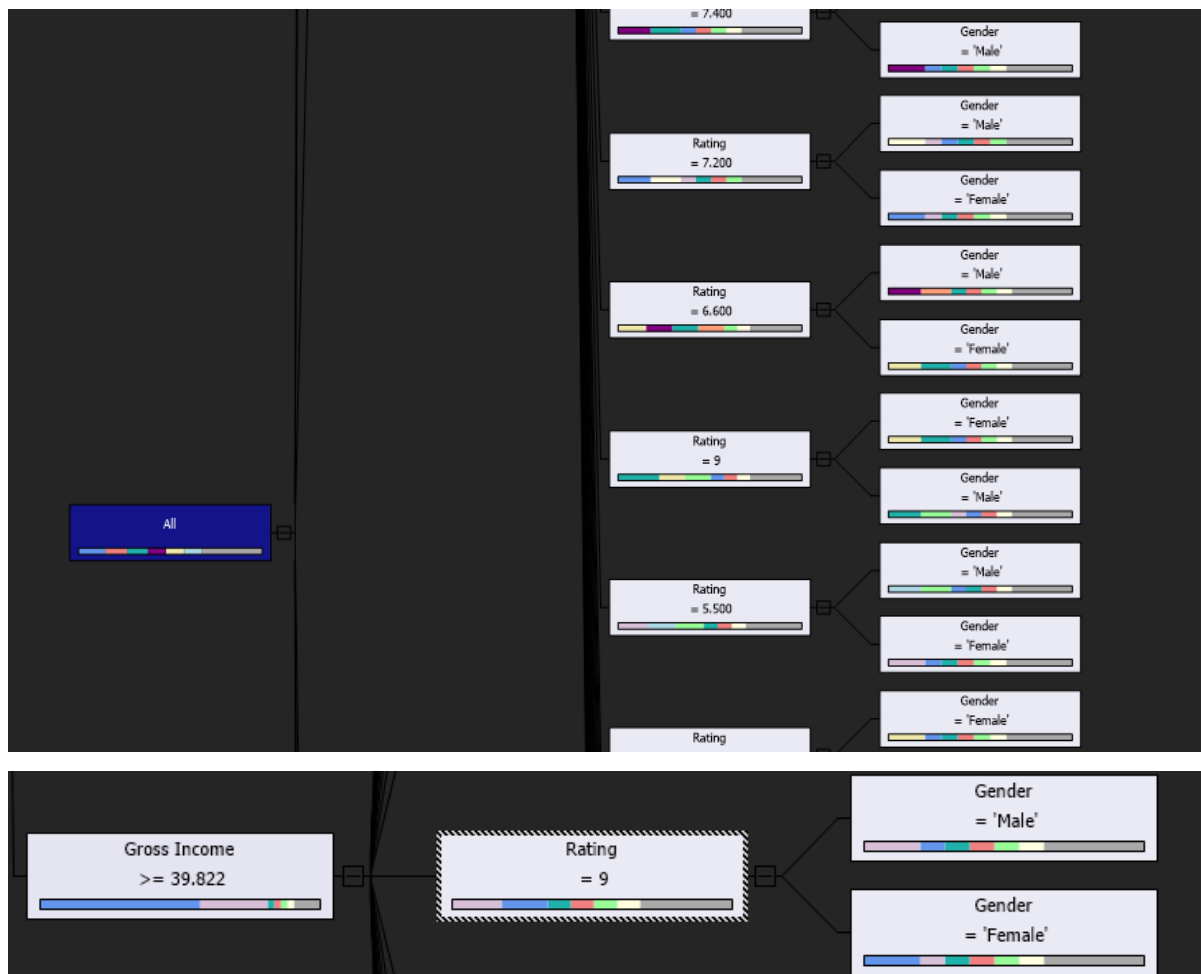
- **Nút Gốc (Root Node):** Nút đầu tiên của cây, đại diện cho toàn bộ dữ liệu và quyết định dựa trên một thuộc tính.
- **Nút Quyết Định (Decision Node):** Các nút nằm giữa cây, đại diện cho một quyết định dựa trên một thuộc tính cụ thể.
- **Nút Lá (Leaf Node):** Các nút ở cuối cây, chứa kết quả dự đoán hoặc phân loại.
- Decision Tree có thể xử lý cả dữ liệu phân loại và dữ liệu số, và không yêu cầu dữ liệu phải được chuẩn hóa trước khi đưa vào mô hình



- Qua biểu đồ có thể thấy có 3 thuộc tính ảnh hưởng đến số lượng mua hàng của khách hàng là Gross Income , Gender , Rating . Và mức độ ảnh hưởng theo thứ tự : Gross Income > Rating > Gender
- Set up lại mô hình với input là các thuộc tính có độ ảnh hưởng cao

Structure	Quantity Tree
	Microsoft_Decision_Trees
Dim Invoice ID SK	Key
Gender	Input
Gross Income	Input
Quantity	PredictOnly
Rating	Input
Time	Ignore
Total	Ignore

❖ Mô hình decision tree dự đoán số lượng sản phẩm khách hàng mua :



- Ví dụ :Khi có Gross Income ≥ 39.822 and Rating = 9 and Gender = 'Male' -> dự đoán khách hàng sẽ mua 9 sản phẩm .

High Low			
Total Cases: 1			
Value	Cases	Probabi...	Histogram
<input checked="" type="checkbox"/> 1	0	9.09%	
<input checked="" type="checkbox"/> 10	0	9.09%	
<input checked="" type="checkbox"/> 2	0	9.09%	
<input checked="" type="checkbox"/> 3	0	9.09%	
<input checked="" type="checkbox"/> 4	0	9.09%	
<input checked="" type="checkbox"/> 5	0	9.09%	
<input checked="" type="checkbox"/> 6	0	9.09%	
<input checked="" type="checkbox"/> 7	0	9.09%	
<input checked="" type="checkbox"/> 8	0	9.09%	
<input checked="" type="checkbox"/> 9	1	18.18%	
<input checked="" type="checkbox"/> Missing	0	0.00%	

❖ Dự đoán số lượng mua hàng :

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Quantity Tree	Quantity		<input checked="" type="checkbox"/>			
Prediction Function	PredictProbability	PredictProbability	<input checked="" type="checkbox"/>			[Quantity Tree].[Quantity]

- Nhập các số liệu như ; Gender là 'Male' , Gross Income : 20 . rating : 9.1 ,

Quantity	Expression
4	0.1818181818...

- Qua đây có thấy là dự đoán khách hàng sẽ mua 4 sản phẩm với tỉ lệ là 18%

2. Dự đoán đánh giá của khách hàng

Để dự đoán **đánh giá của khách hàng** đã mua trong mỗi giao dịch.-> Các thuộc tính ảnh hưởng từ bảng Fact_SupermarketSales:

- Gender : Giới tính của khách hàng có thể ảnh hưởng đến số lượng sản phẩm họ mua.
- Time : Thời gian mua hàng có thể ảnh hưởng đến đánh giá của khách hàng
- Gross Income : Thu nhập có thể ảnh hưởng đến đánh giá của khách hàng
- Rating: Đánh giá của khách hàng có thể ảnh hưởng đến đánh giá của khách hàng
- Total : Tổng tiền khách hàng trả có thể ảnh hưởng đến đánh giá của khách hàng

- Sử dụng thuật toán decision tree .



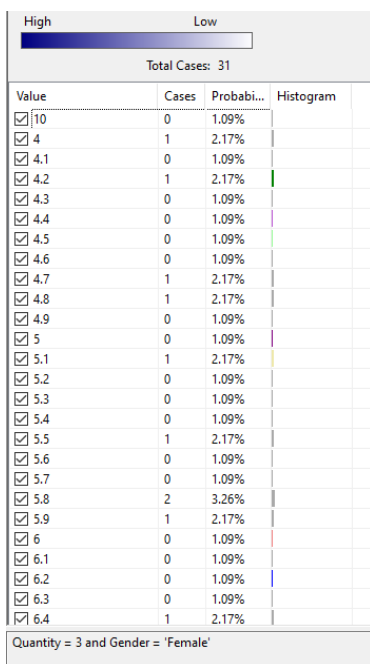
- Qua biểu đồ có thể thấy có 2 thuộc tính ảnh hưởng đến đánh giá của khách hàng là Gender và Quantity . Và mức độ ảnh hưởng : Quantity > Gender.
- Set up lại mô hình với input là các thuộc tính có độ ảnh hưởng cao

Structure	Rating Tree
Dim Invoice ID SK	Microsoft_Decision_Trees
Gender	Key
Gross Income	Input
Quantity	Ignore
Rating	Input
Time	PredictOnly
Total	Ignore

Mô hình decision tree dự đoán đánh giá của khách hàng:



- Ví dụ : Khi khách hàng mua 3 sản phẩm (Quantity = 3) và giới tính : 'Female' .
Có 31 cases dự đoán rating của khách hàng :



❖ Dự đoán rating của khách hàng :

Mining Model

- Rating Tree
 - Dim Invoice ID SK
 - Gender
 - Quantity
 - Rating

Select Model...

Singleton Query Input

Mining Model Column	Value
Gender	Male
Quantity	8
Rating	Missing

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Rating Tree	Rating		<input checked="" type="checkbox"/>			
Prediction Function	PredictProbability		<input checked="" type="checkbox"/>			[Rating Tree].[Rating]

- Nhập các số liệu như ; Gender là 'Male' . quantity : 8

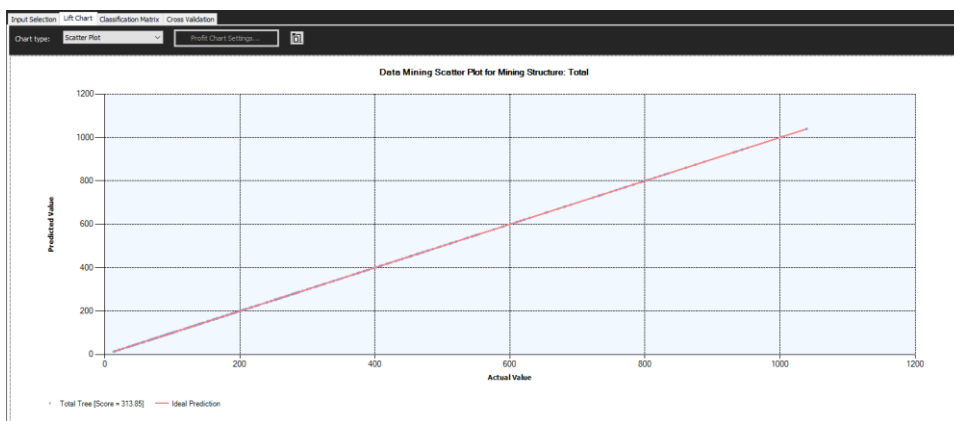
Rating	Expression
6.2	0.0444444444...

⇒ Dự đoán khách hàng sẽ rating 9.4

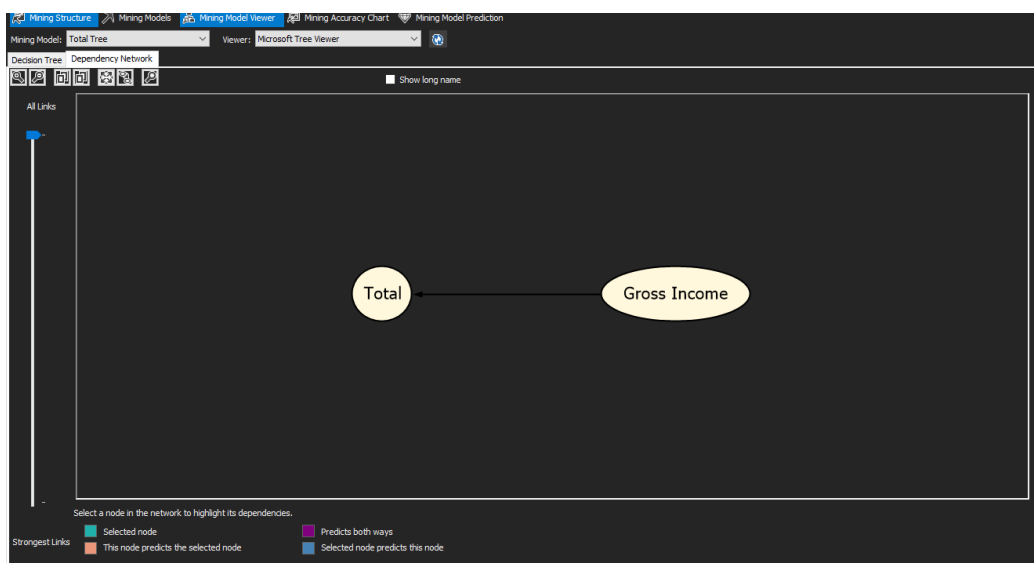
3. Dự đoán tổng giá tiền mua hàng của khách hàng dựa vào thu nhập

Để dự đoán số lượng sản phẩm mà khách hàng đã mua trong mỗi giao dịch.-> Các thuộc tính ảnh hưởng từ bảng Fact_SupermarketSales:

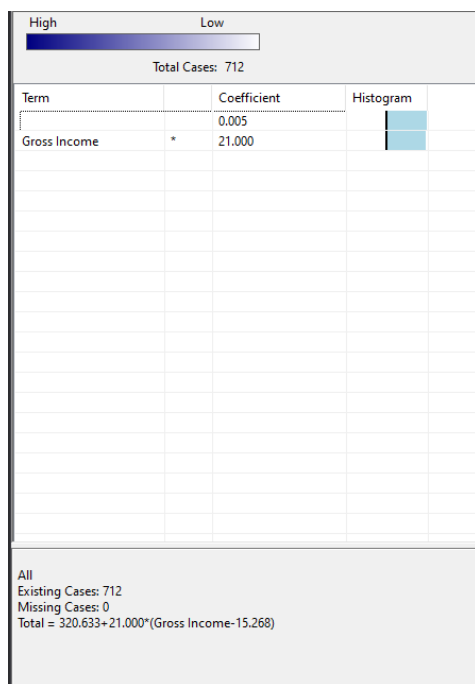
- Gross Income : Thu nhập có thể ảnh hưởng đến đánh giá của khách hàng
- Sử dụng thuật toán linear regression để dự đoán giá
- "Hồi quy tuyến tính" là một phương pháp thống kê để hồi quy dữ liệu với biến phụ thuộc có giá trị liên tục trong khi các biến độc lập có thể có một trong hai giá trị liên tục hoặc là giá trị phân loại. Nói cách khác "Hồi quy tuyến tính" là một phương pháp để dự đoán biến phụ thuộc (Y) dựa trên giá trị của biến độc lập (X). Nó có thể được sử dụng cho các trường hợp chúng ta muốn dự đoán một số lượng liên tục



Qua biểu Đồ Phân Tán (Scatter Plot) : dữ liệu có hình dạng tuyến tính (theo dạng đường thẳng) -> Có thể sử dụng thuật toán Linear Regression



❖ Mô hình linear regression có hệ số tương quan



❖ Dự đoán giá tiền khách hàng mua

Mining Model

- Total Tree
- Dim Invoice ID SK
- Gross Income
- Total

Select Model...

Singleton Query Input

Mining Model Column	Value
Gross Income	30
Total	

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Total Tree	Total	Predict Total	<input checked="" type="checkbox"/>			
Prediction Function	PredictProbability	Predict Probability	<input checked="" type="checkbox"/>			[Total Tree].[Total]

- Nhập vào số liệu : Gross Income : 30

Predict Total	Predict Probability
629.99565509...	0.9985994397...

- Dự đoán giá tiền khách hàng mua hàng là : 629.99 .

- **Dự Đoán Số Lượng Mua Hàng của Khách Hàng:**

- Các thuộc tính ảnh hưởng: Gender, Time, Gross Income, Rating.
- Sử dụng thuật toán Decision Tree để dự đoán số lượng sản phẩm mua.
- Mức độ ảnh hưởng: Gross Income > Rating > Gender.
- **Dự Đoán Đánh Giá Của Khách Hàng:**
 - Các thuộc tính ảnh hưởng: Gender, Time, Gross Income, Rating, Total.
 - Sử dụng thuật toán Decision Tree để dự đoán đánh giá của khách hàng.
 - Mức độ ảnh hưởng: Quantity > Gender.
- **Dự Đoán Tổng Giá Tiền Mua Hàng::**
 - Sử dụng Linear Regression với thuộc tính ảnh hưởng là Gross Income..
 - Mô hình có thể dự đoán tổng giá tiền mua hàng của khách hàng dựa trên thu nhập của họ.

⇒ **Kết Luận**

- Cả ba mô hình có thể cung cấp thông tin chi tiết và dự đoán về hành vi mua sắm và đánh giá của khách hàng trong tương lai .
- Gender, Gross Income và Rating đều là các yếu tố quan trọng trong việc dự đoán hành vi của khách hàng.
- Trong cả ba mô hình, Gross Income được xác định là yếu tố quan trọng nhất, cho thấy sự quan trọng của thu nhập cá nhân đối với hành vi mua hàng và đánh giá
- Các mô hình cung cấp khả năng tùy chỉnh dựa trên giới tính, thời gian mua, đánh giá, và số lượng sản phẩm mua,... -> Sử dụng mô hình có thể giúp doanh nghiệp hiểu rõ hơn về nhóm khách hàng, đồng thời tối ưu hóa chiến lược tiếp thị và dịch vụ khách hàng.