

Nhóm 12

Kho và khai phá dữ liệu

DATA WAREHOUSE

MIDTERM REPORT

NHÓM 12 - THÀNH VIÊN NHÓM

STT	Họ và tên	Phân công công việc	% đóng góp
1	Vương Thị Mỹ Lê	Tìm data, Giới thiệu tổng quan về data, Thực hiện ETL, Star Schema, OLAP, MDX	20%
2	Phan Thị Thanh Nhàn	Tìm data, Giới thiệu tổng quan về data, Thực hiện ETL, Star Schema, OLAP, làm Slide báo cáo	20%
3	Lâm Hồng Phúc	Tìm data, Thực hiện ETL, Star Schema, OLAP, Dashboard, làm Slide báo cáo	20%
4	Võ Thị Thu Trang	Tìm data, Giới thiệu tổng quan về data, Thực hiện ETL, Star Schema, OLAP, làm Slide báo cáo	20%
5	Nguyễn Ngọc Đan Trâm	Tìm data, Thực hiện ETL, Star Schema, OLAP, MDX, Dashboard	20%

NỘI DUNG

1. Giới thiệu tổng quan về data
2. Mô hình ETL & Star Schema
3. OLAP
4. Truy vấn MDX
5. Dashboard
6. Kết luận

I. GIỚI THIỆU TỔNG QUAN VỀ DATA



MÔ TẢ DỮ LIỆU:

Gồm 1000 dòng, 10 cột

Thông tin các đơn hàng trong vòng 2 năm 2022-2023 của một cửa hàng bán đồ ăn nhanh và nước giải khát

Các trường dữ liệu bao gồm như dưới đây

order_id	date	item_name	item_type	item_price	quantity	transaction_amount	transaction_type	received_by	time_of_sale
1	3/7/2022	Aalopuri	Fastfood	20	13	260		Mr.	Night
2	8/23/2022	Vadapav	Fastfood	20	15	300	Cash	Mr.	Afternoon
3	11/20/2022	Vadapav	Fastfood	20	1	20	Cash	Mr.	Afternoon
4	3/2/2023	Sugarcane juice	Beverages	25	6	150	Online	Mr.	Night
5	2/10/2022	Sugarcane juice	Beverages	25	8	200	Online	Mr.	Evening
6	11/14/2022	Vadapav	Fastfood	20	10	200	Cash	Mr.	Evening
7	3/5/2022	Sugarcane juice	Beverages	25	9	225	Cash	Mr.	Evening
8	12/22/2022	Panipuri	Fastfood	20	14	280	Online	Mr.	Night
9	10/6/2022	Panipuri	Fastfood	20	1	20	Cash	Mrs.	Morning
10	9/16/2022	Panipuri	Fastfood	20	5	100	Online	Mr.	Afternoon
11	1/12/2022	Frankie	Fastfood	50	8	400	Online	Mrs.	Afternoon
12	12/7/2022	Vadapav	Fastfood	20	8	160	Online	Mrs.	Night
13	12/22/2022	Panipuri	Fastfood	20	9	180	Online	Mrs.	Afternoon
14	11/25/2022	Frankie	Fastfood	50	4	200	Online	Mr.	Morning
15	3/2/2023	Aalopuri	Fastfood	20	3	60	Cash	Mrs.	Evening
16	4/14/2022	Sandwich	Fastfood	60	11	660		Mrs.	Midnight
17	10/16/2022	Panipuri	Fastfood	20	11	220	Cash	Mrs.	Morning
18	5/11/2022	Panipuri	Fastfood	20	10	200	Cash	Mrs.	Night
19	8/22/2022	Panipuri	Fastfood	20	11	220	Cash	Mrs.	Night
20	9/15/2022	Cold coffee	Beverages	40	10	400	Online	Mr.	Night
21	12/21/2022	Cold coffee	Beverages	40	15	600	Cash	Mr.	Evening

I. GIỚI THIỆU TỔNG QUAN VỀ DATA



Vấn đề đặt ra : Phân tích hành vi mua hàng của khách hàng

- Ảnh hưởng đến doanh số bán hàng của cửa hàng như thế nào ?

Vấn đề cần giải quyết :

- Xem xét khách hàng có xu hướng mua những sản phẩm nào theo từng khung giờ - mặt hàng nào là bán chạy nhất.
- Khách hàng có thay đổi hành vi mua theo thời tiết hay không.
- Phương thức thanh toán có ảnh hưởng gì đến hành vi mua không.
- Ngày bán cao điểm và ngày bán thấp điểm như thế nào?



Phân tích bảng dữ liệu

order_id	date	item_name	item_type	item_price	quantity	transaction_amount	transaction_type	received_by	time_of_sale
1	3/7/2022	Aalopuri	Fastfood	20	13	260		Mr.	Night
2	8/23/2022	Vadapav	Fastfood	20	15	300	Cash	Mr.	Afternoon
3	11/20/2022	Vadapav	Fastfood	20	1	20	Cash	Mr.	Afternoon
4	3/2/2023	Sugarcane juice	Beverages	25	6	150	Online	Mr.	Night
5	2/10/2022	Sugarcane juice	Beverages	25	8	200	Online	Mr.	Evening
6	11/14/2022	Vadapav	Fastfood	20	10	200	Cash	Mr.	Evening
7	3/5/2022	Sugarcane juice	Beverages	25	9	225	Cash	Mr.	Evening
8	12/22/2022	Panipuri	Fastfood	20	14	280	Online	Mr.	Night
9	10/6/2022	Panipuri	Fastfood	20	1	20	Cash	Mrs.	Morning
10	9/16/2022	Panipuri	Fastfood	20	5	100	Online	Mr.	Afternoon
11	1/12/2022	Frankie	Fastfood	50	8	400	Online	Mrs.	Afternoon
12	12/7/2022	Vadapav	Fastfood	20	8	160	Online	Mrs.	Night
13	12/22/2022	Panipuri	Fastfood	20	9	180	Online	Mrs.	Afternoon
14	11/25/2022	Frankie	Fastfood	50	4	200	Online	Mr.	Morning
15	3/2/2023	Aalopuri	Fastfood	20	3	60	Cash	Mrs.	Evening
16	4/14/2022	Sandwich	Fastfood	60	11	660		Mrs.	Midnight
17	10/16/2022	Panipuri	Fastfood	20	11	220	Cash	Mrs.	Morning
18	5/11/2022	Panipuri	Fastfood	20	10	200	Cash	Mrs.	Night

Dùng dữ liệu chạy ETL

Chỉ có 1 khóa chính (**order_id**)
=> tạo thêm các khóa chính về các chiều khác

Bảng Dimension

Dim_Time

Tên trường	Kiểu dữ liệu	Ý nghĩa
time_id	int	Mã thời gian giao dịch
date	date	Ngày giao dịch
time_of_sale	varchar(20)	Buổi giao dịch

time_id	date	time_of_sale
1	3/7/2022	Night
2	8/23/2022	Afternoon
3	11/20/2022	Afternoon

Dim_Item

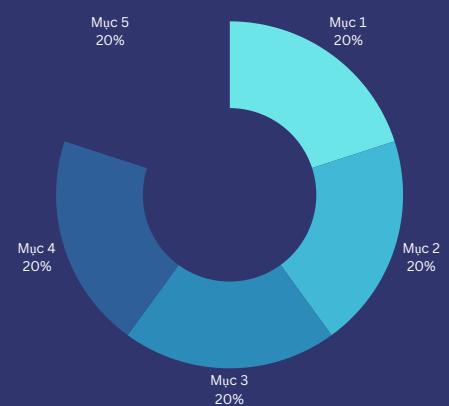
Tên trường	Kiểu dữ liệu	Ý nghĩa
item_id	int	Mã sản phẩm
item_name	varchar(50)	Tên sản phẩm
item_type	varchar(30)	Loại sản phẩm
item_price	int	Giá

item_id	item_name	item_type	item_price
1	Aalopuri	Fastfood	20
2	Vadapav	Fastfood	20
3	Sugarcane juice	Beverages	25

Dim_Transaction

Tên trường	Kiểu dữ liệu	Ý nghĩa
transaction_id	int	Mã giao dịch
transaction_type	varchar (50)	Loại giao dịch

transaction_id	transaction_type
1	Cash
2	Cash
3	Online



Phan Thị Thanh Nhàn

Bảng Fact

quantity	transaction_amount
13	260
15	300
1	20
6	150
8	200
10	200
9	225
14	280
1	20
5	100
8	400
8	160
9	180
4	200
3	60

Mô tả :

Quantity : Số lượng món

Transaction_amount : Số tiền giao dịch cho mỗi đơn hàng

- Các measure:

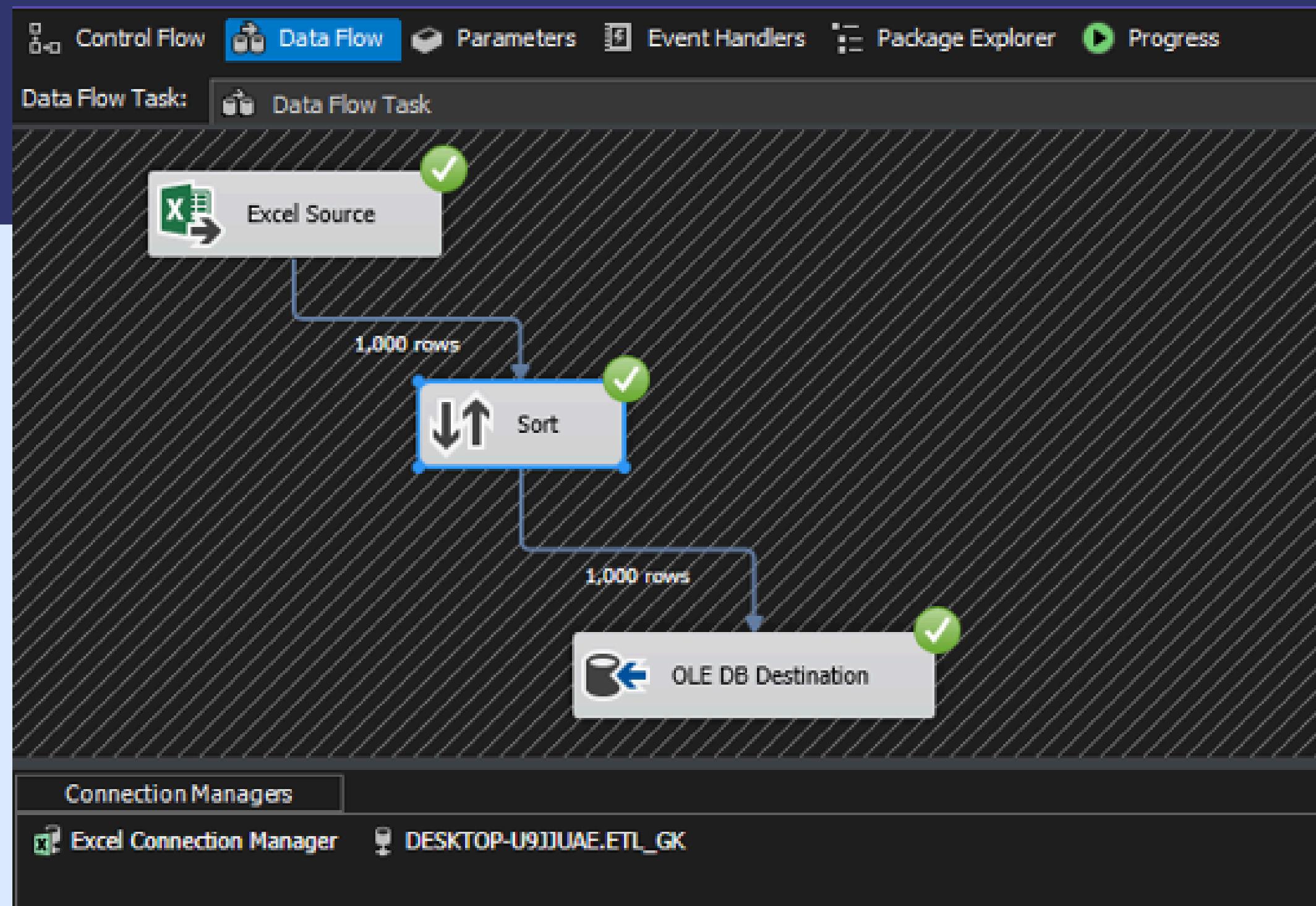
quantity

transaction_amount

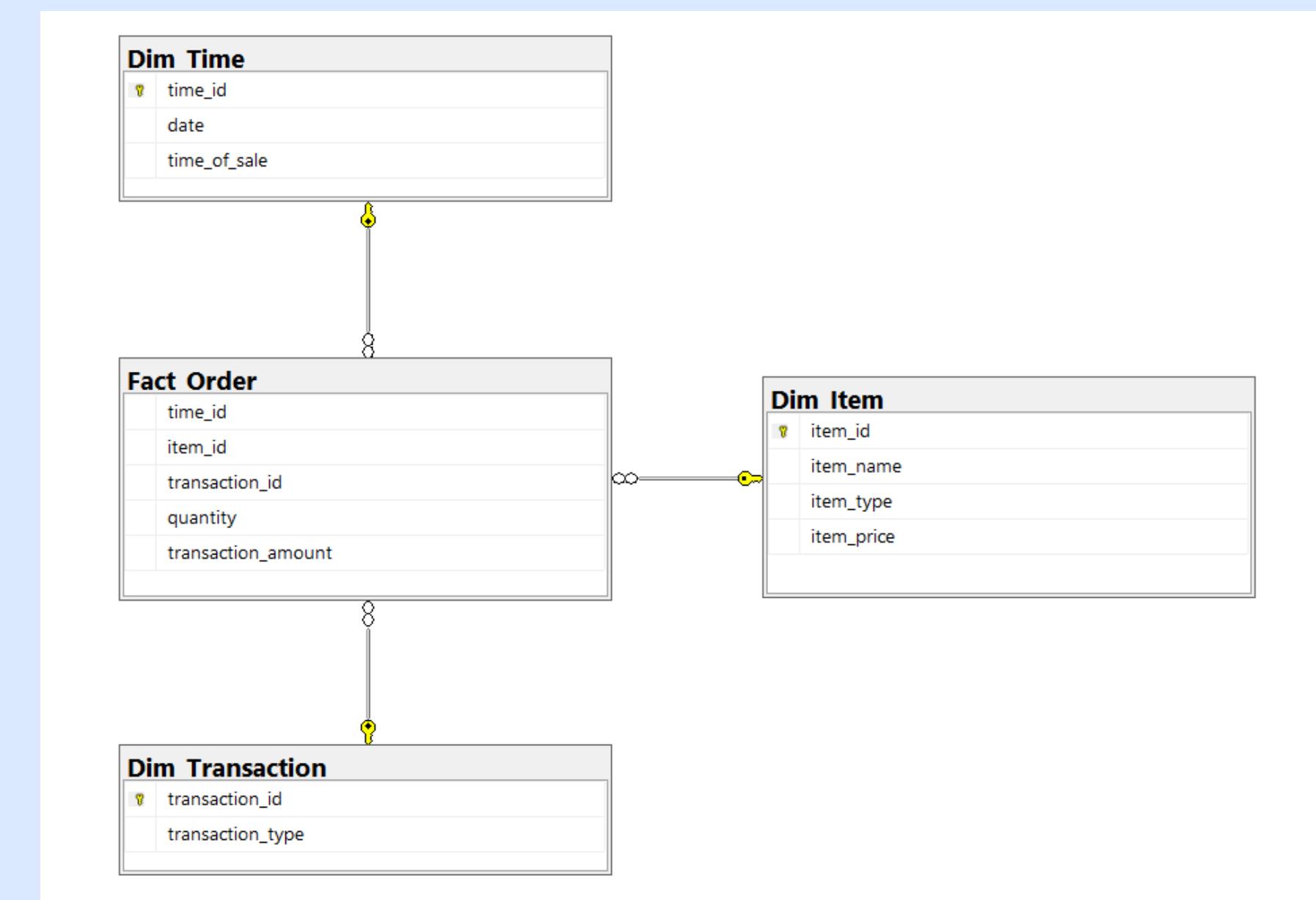
- Các khoá trỏ đến các Dimension:

TimeID, TransactionID, ItemID

2. ETL

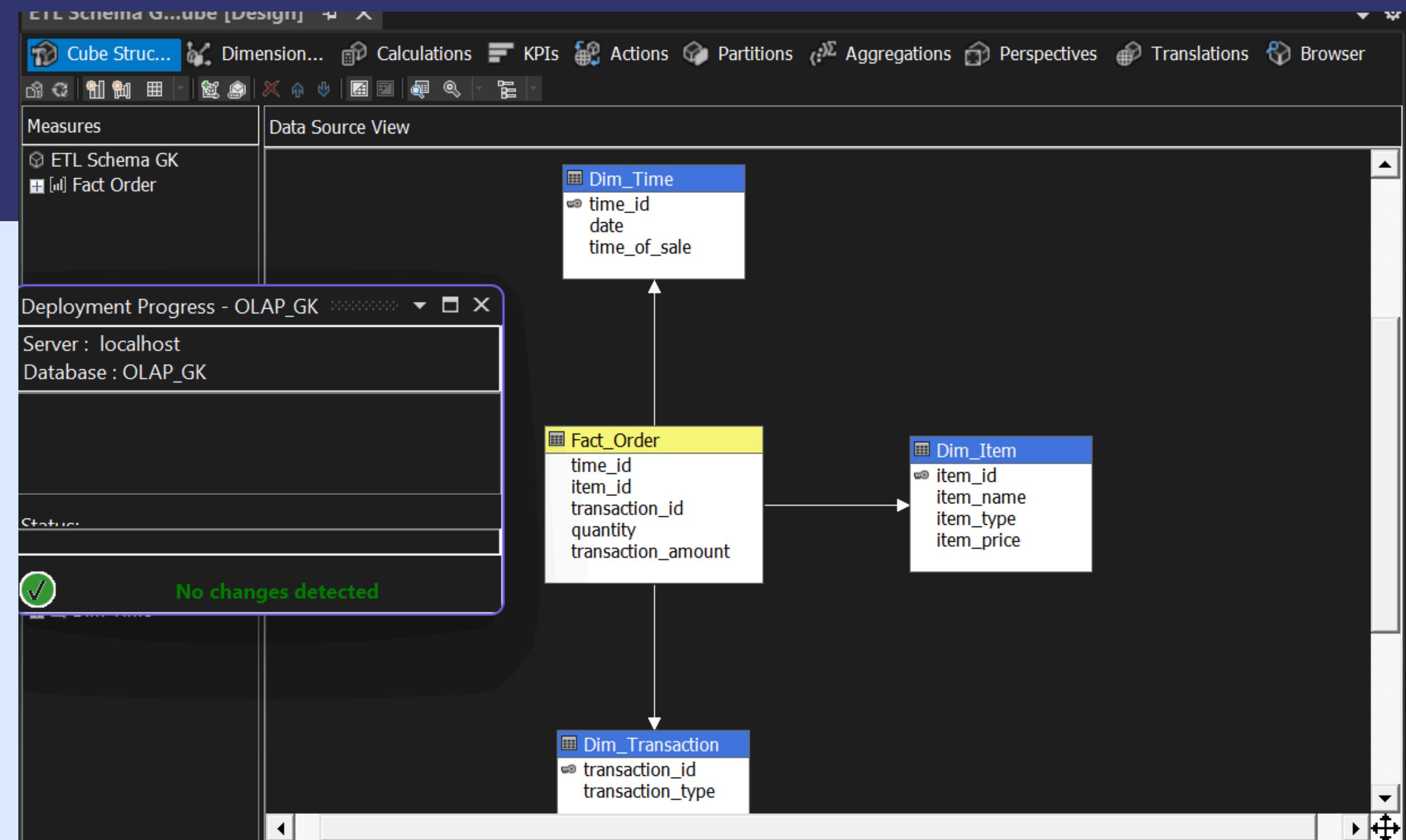


STAR SCHEMA



Phan Thị Thanh Nhàn

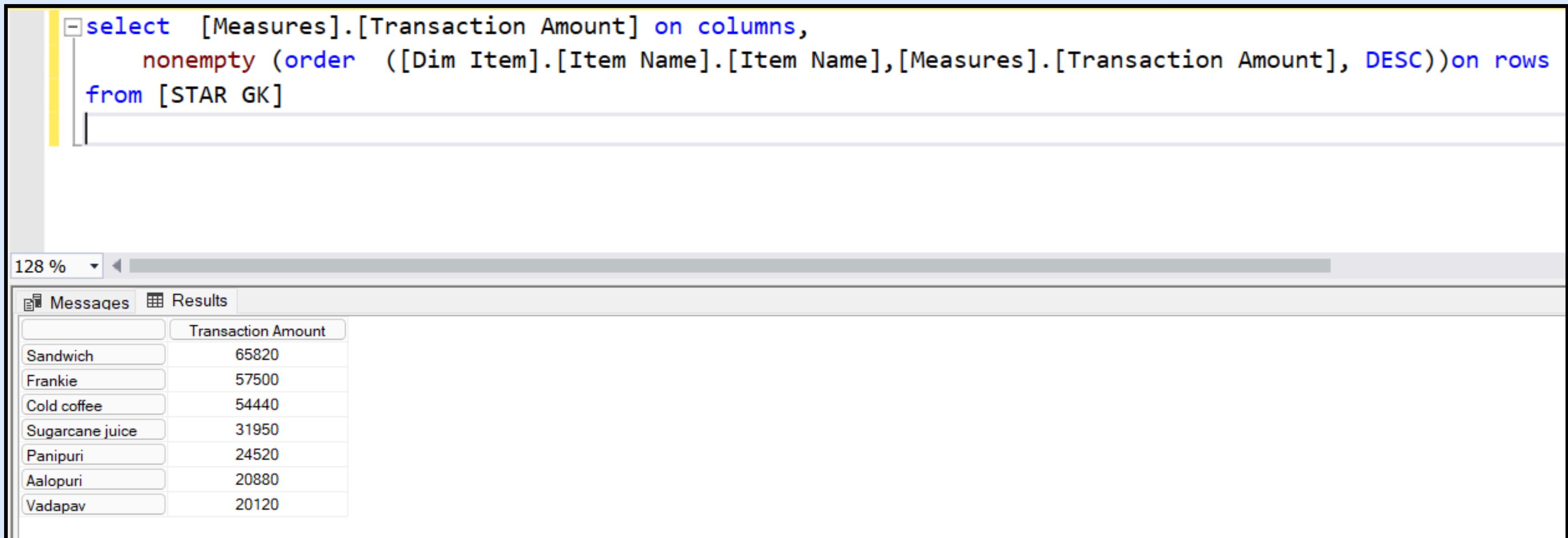
OLAP



4. MDX



Câu 1: Doanh thu bán của từng sản phẩm theo thứ tự giảm dần



The screenshot shows a Microsoft SQL Server Management Studio (SSMS) window. The top pane contains T-SQL code:

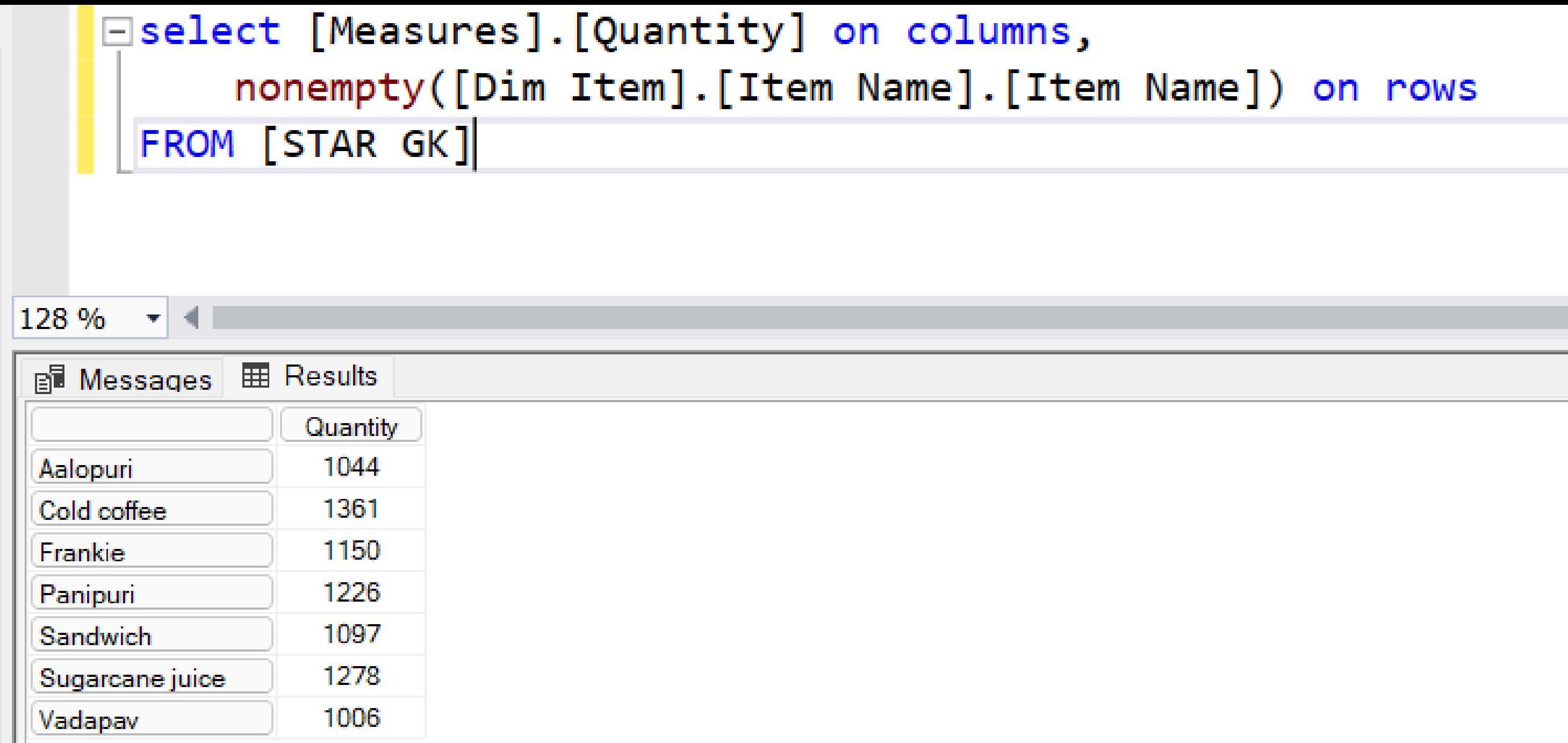
```
select [Measures].[Transaction Amount] on columns,
       nonempty (order ([Dim Item].[Item Name].[Item Name],[Measures].[Transaction Amount], DESC))on rows
  from [STAR GK]
```

The bottom pane displays the results in a table:

	Transaction Amount
Sandwich	65820
Frankie	57500
Cold coffee	54440
Sugarcane juice	31950
Panipuri	24520
Aalopuri	20880
Vadapav	20120

=> Sản phẩm có doanh thu cao nhất là Sandwich (65820)
và sản phẩm có doanh thu thấp nhất Vadapav (20120)

Câu 2: Truy vấn số lượng bán theo sản phẩm



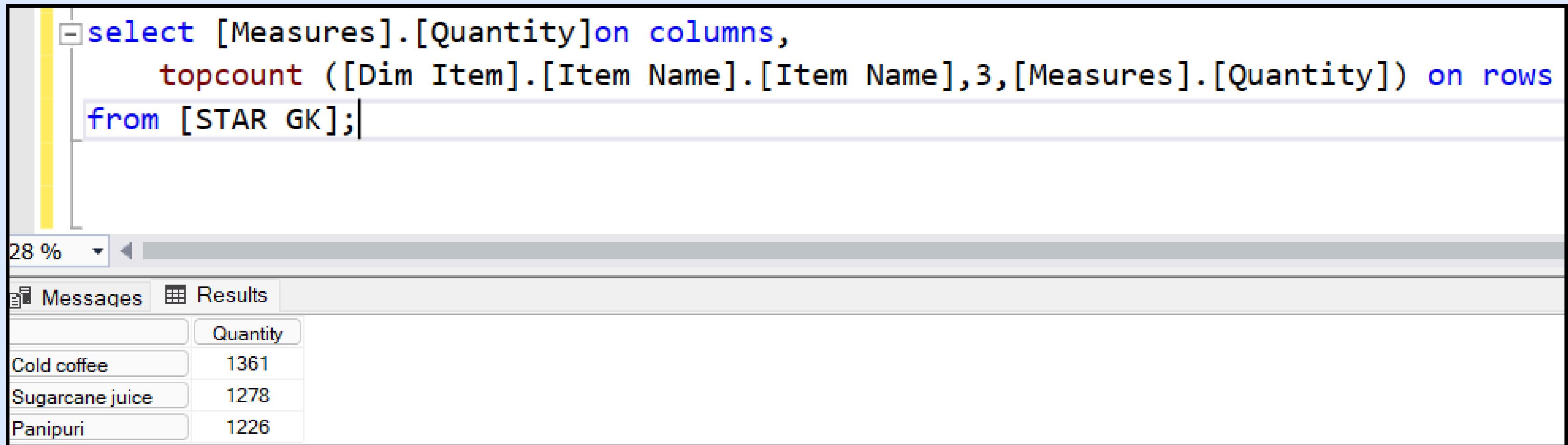
```
select [Measures].[Quantity] on columns,
nonempty([Dim Item].[Item Name].[Item Name]) on rows
FROM [STAR GK]
```

The screenshot shows the SSMS Results grid with the following data:

	Quantity
Aalopuri	1044
Cold coffee	1361
Frankie	1150
Panipuri	1226
Sandwich	1097
Sugarcane juice	1278
Vadapav	1006

=> Số lượng sản phẩm được bán nhiều nhất là Cold coffee
và bán ít nhất là Vadapav

Câu 3: Top 3 sản phẩm bán chạy nhất



The screenshot shows a Microsoft SQL Server Management Studio (SSMS) interface. In the top pane, there is a query editor window containing the following T-SQL code:

```
select [Measures].[Quantity] on columns,
       topcount ([Dim Item].[Item Name].[Item Name], 3, [Measures].[Quantity]) on rows
  from [STAR GK];
```

In the bottom pane, there is a results grid titled "Results" with the following data:

	Quantity
Cold coffee	1361
Sugarcane juice	1278
Panipuri	1226

=> Top 3 sản phẩm được bán chạy nhất là Cold coffee,
Sugarcane juice và Panipuri

Câu 4: Doanh số theo khung giờ

The screenshot shows a SQL query in the top pane and its execution results in the bottom pane.

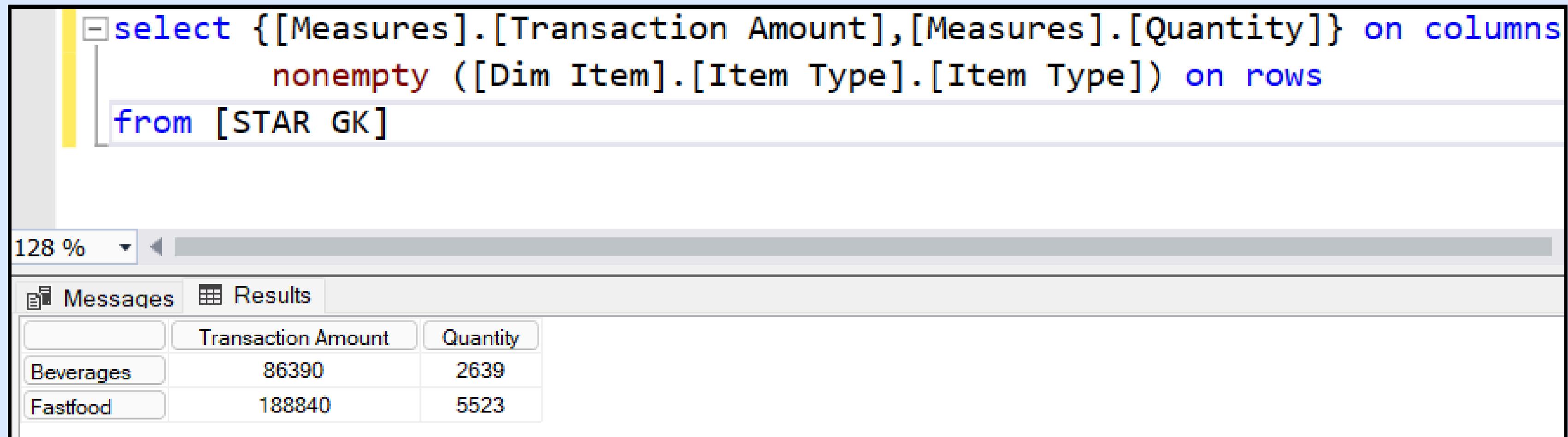
```
select [Measures].[Transaction Amount] on columns,
       nonempty ([Dim Time].[Time Of Sale].[Time Of Sale]) on rows
  from [STAR GK];
```

The Results pane displays a table with the following data:

	Transaction Amount
Afternoon	55685
Evening	77380
Midnight	36860
Morning	49750
Night	55555

=> Số tiền giao dịch vào các thời điểm trong ngày sẽ khác nhau và cao nhất vào buổi Evening (77380) và thấp nhất vào buổi Midnight (36860)

Câu 5: Truy vấn doanh thu và số lượng theo từng loại sản phẩm



The screenshot shows a Microsoft SQL Server Management Studio (SSMS) interface. In the top pane, there is a T-SQL query:

```
select {[Measures].[Transaction Amount],[Measures].[Quantity]} on columns  
nonempty ([Dim Item].[Item Type].[Item Type]) on rows  
from [STAR GK]
```

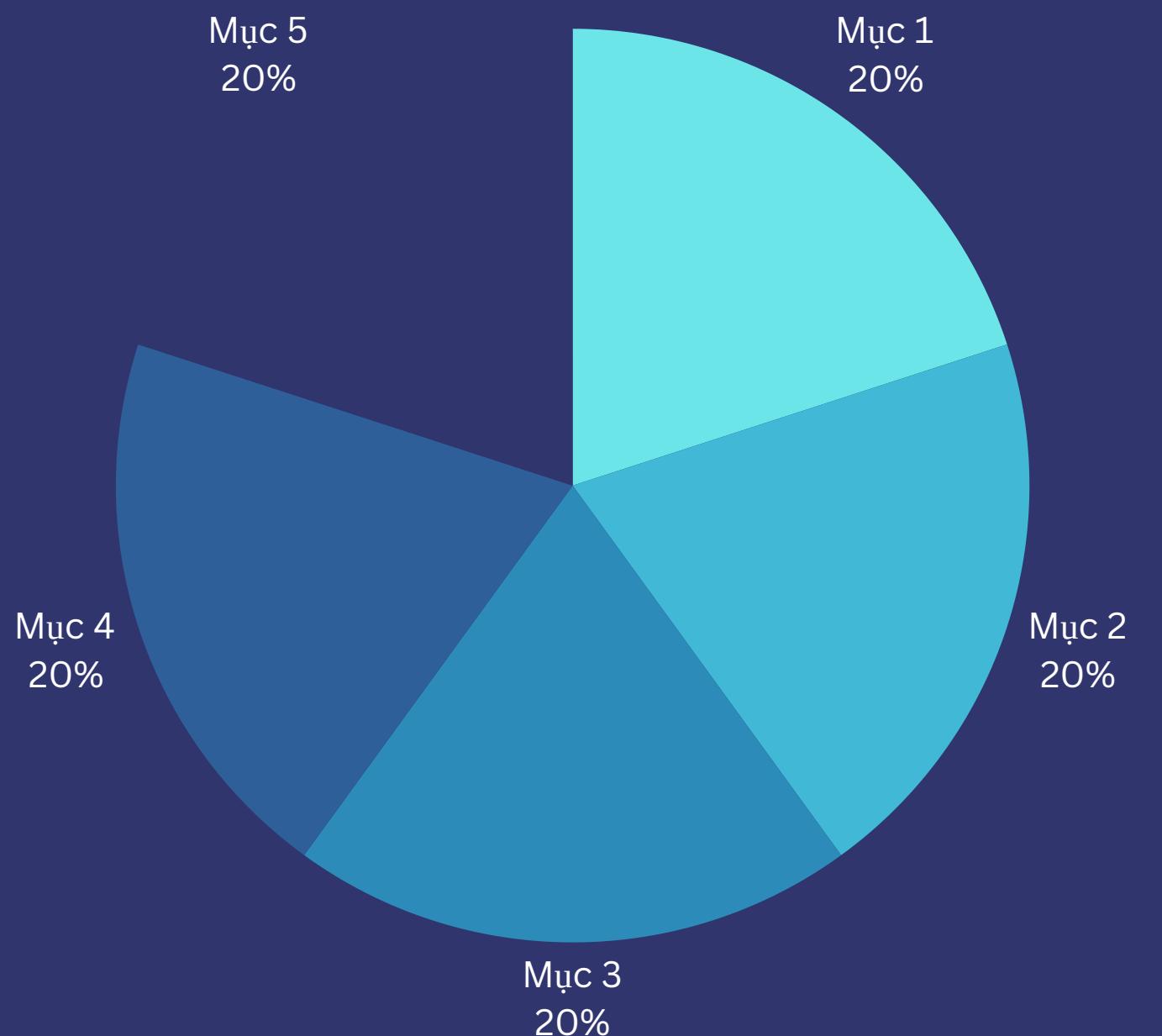
In the bottom pane, under the "Results" tab, the output is displayed in a table:

	Transaction Amount	Quantity
Beverages	86390	2639
Fastfood	188840	5523

=> Doanh thu và số lượng của Fastfood sẽ cao hơn so với Beverages

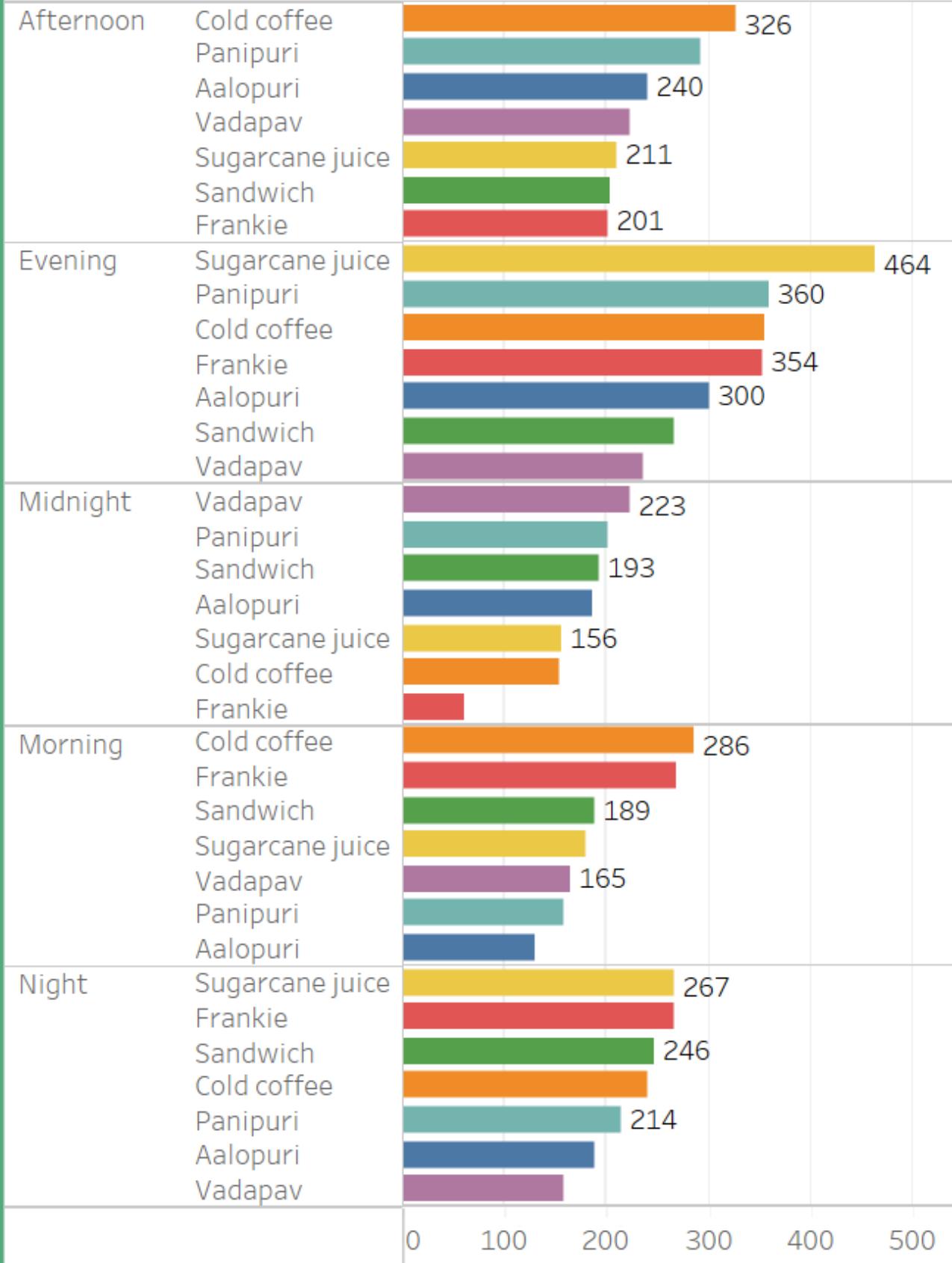
DASHBOARD

Trực quan dữ liệu bằng công cụ Tableau

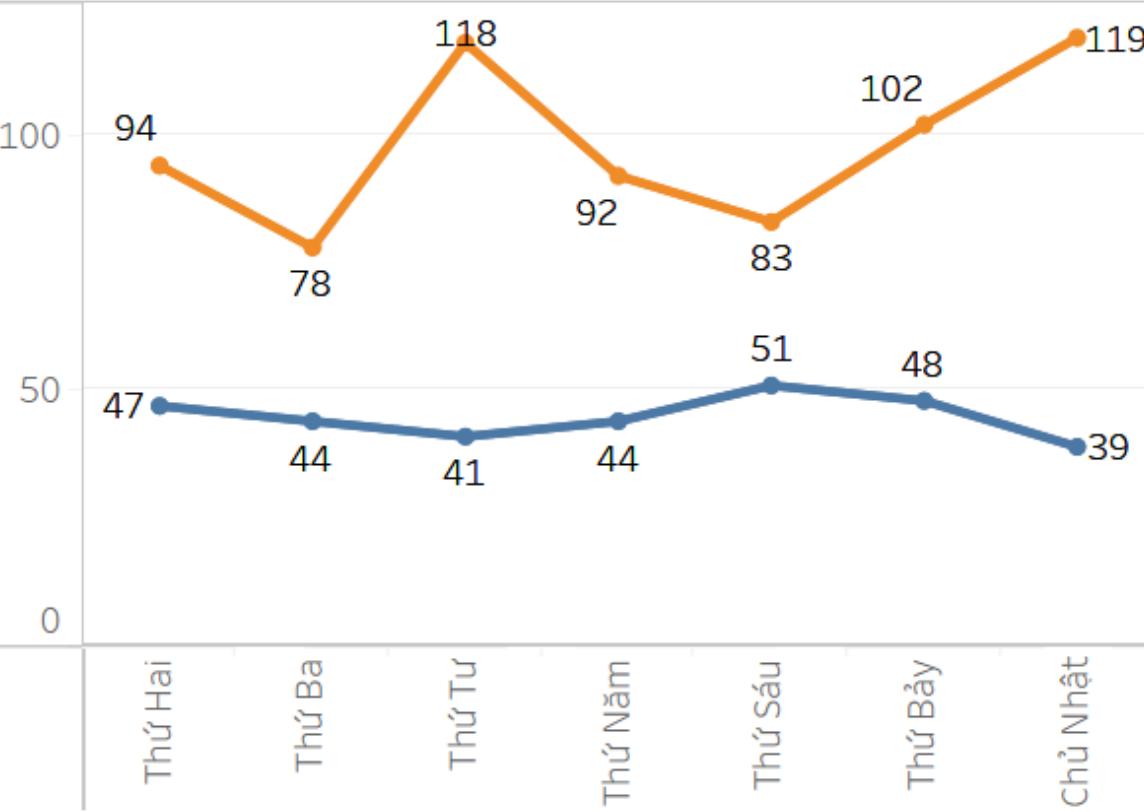


PHÂN TÍCH THỜI GIAN VÀ TẦN SUẤT MUA HÀNG

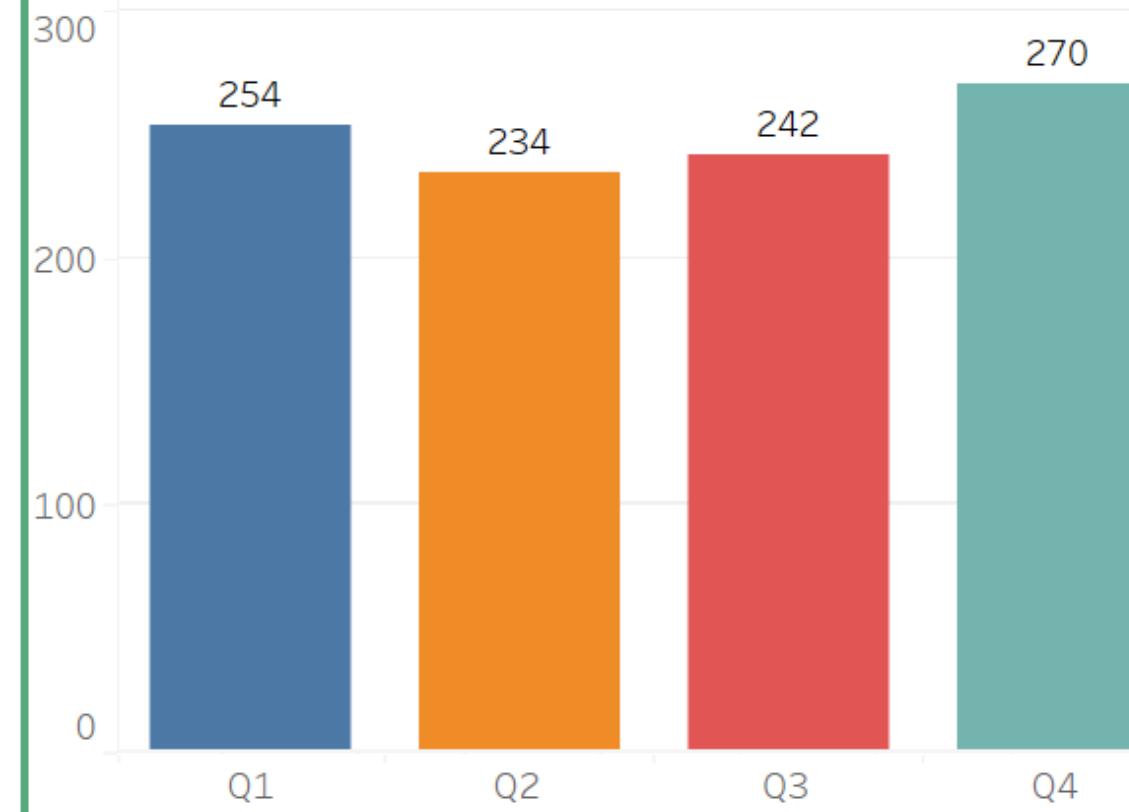
Đồ ăn bán chạy theo Khung giờ



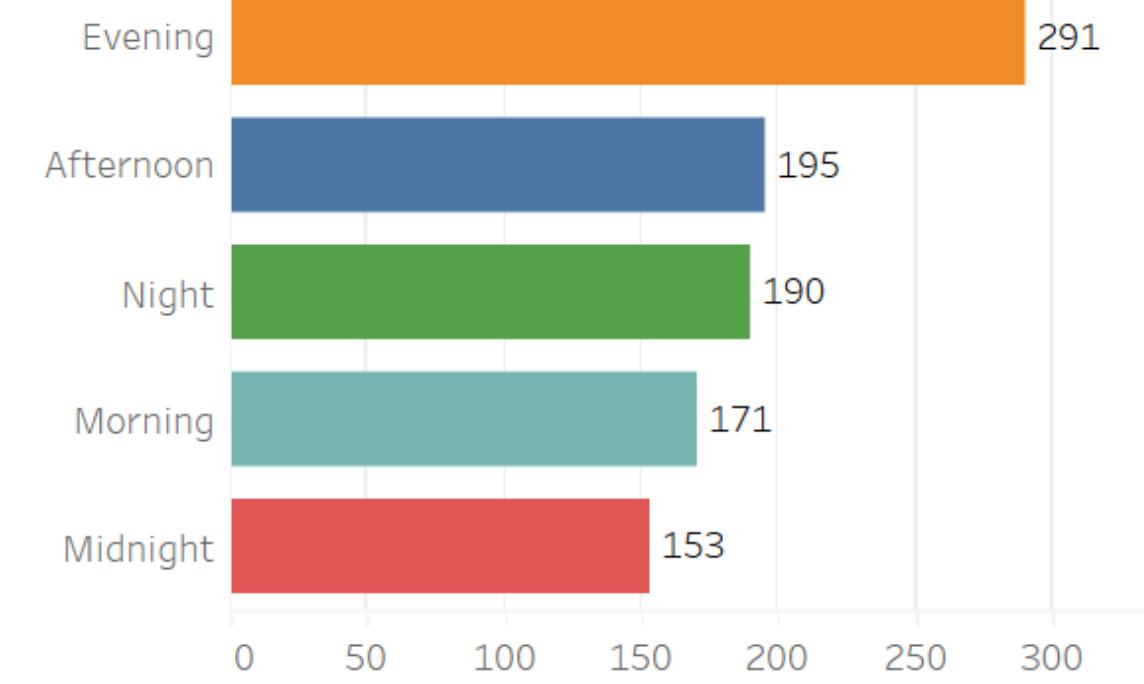
Tần suất giao dịch theo Ngày trong Tuần



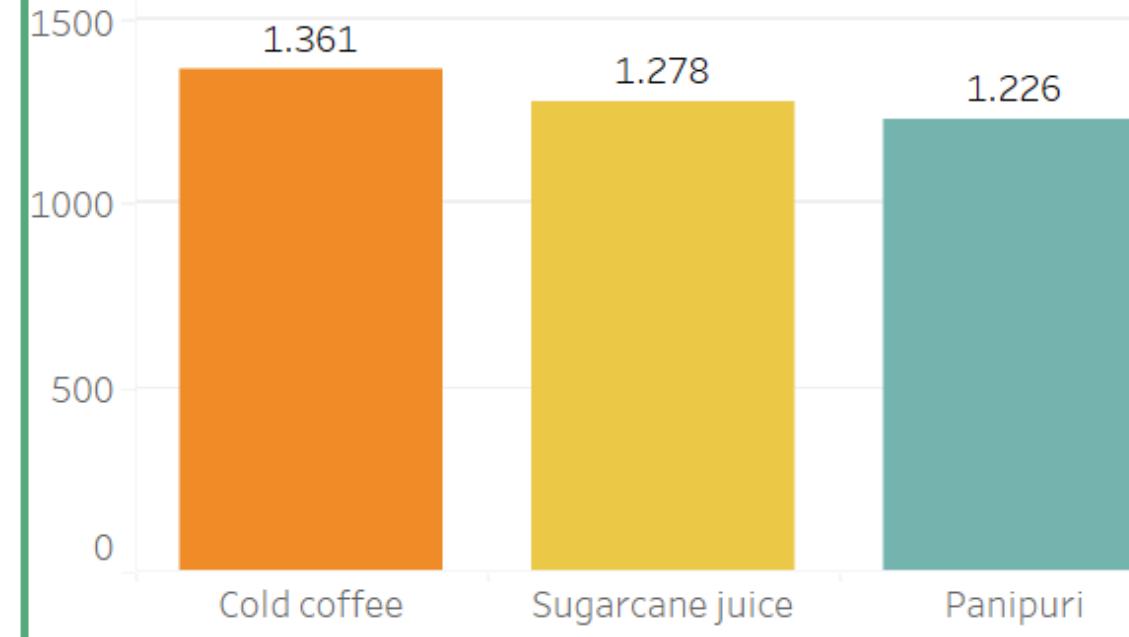
Xu hướng mua hàng theo Quý



Tần suất giao dịch theo Khung giờ



Top 3 Mặt hàng bán chạy nhất



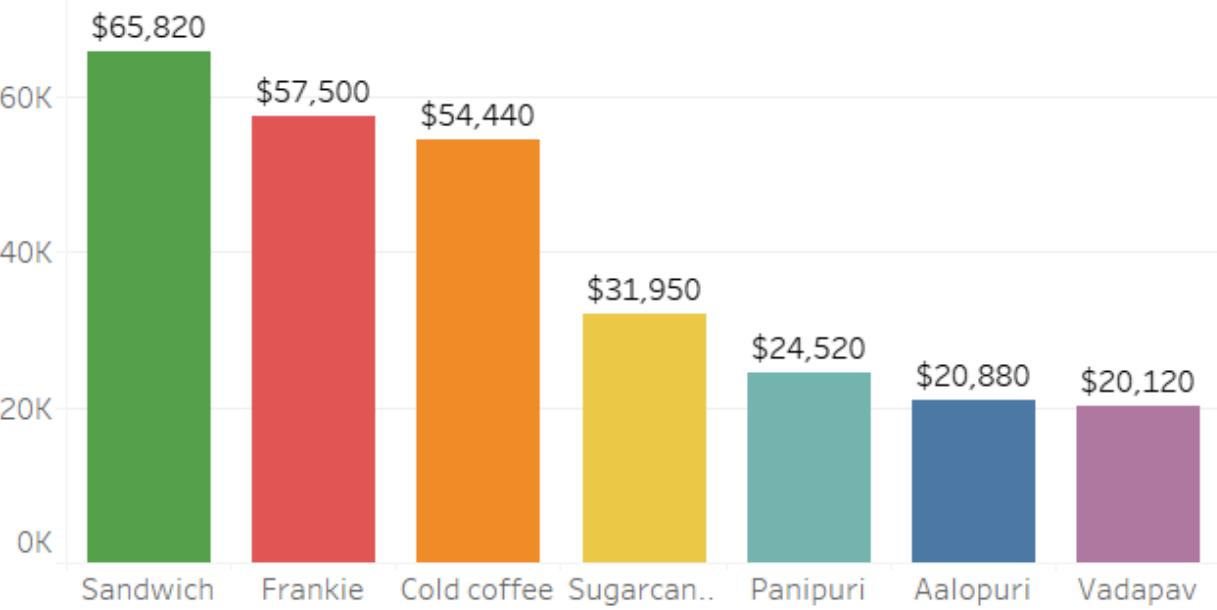
Phân tích doanh số bán của tên sản phẩm

Tổng số lượng và doanh số bán

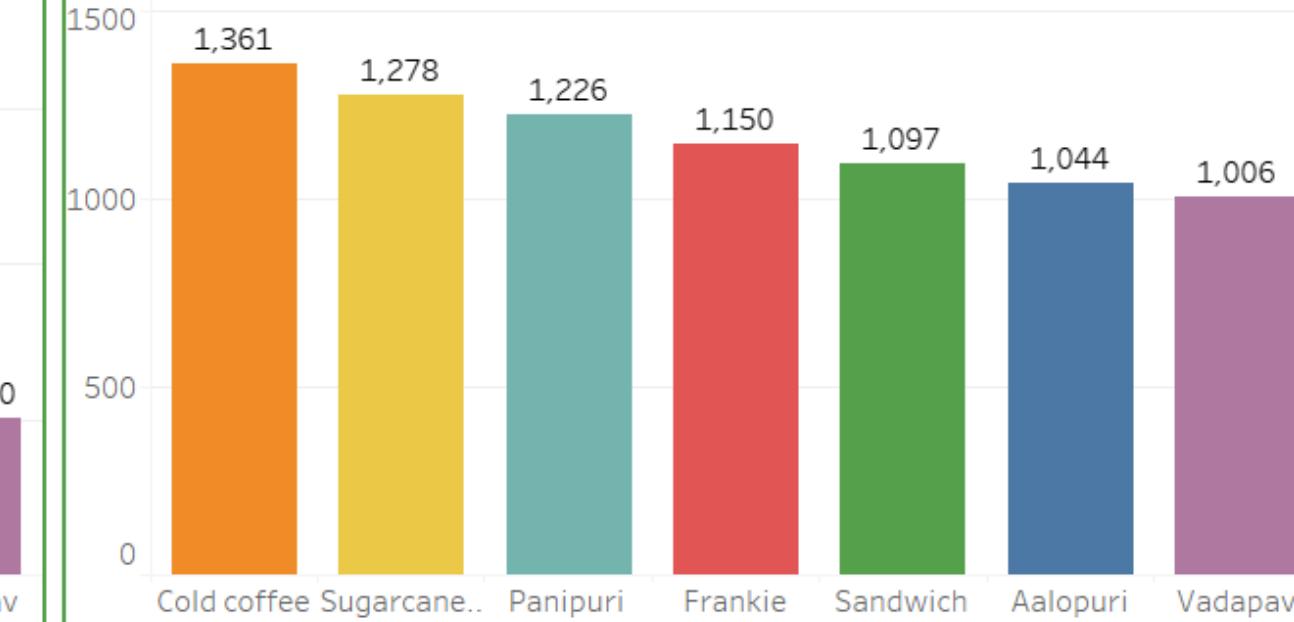
Quantity Transaction Amount

8.162K \$275,230

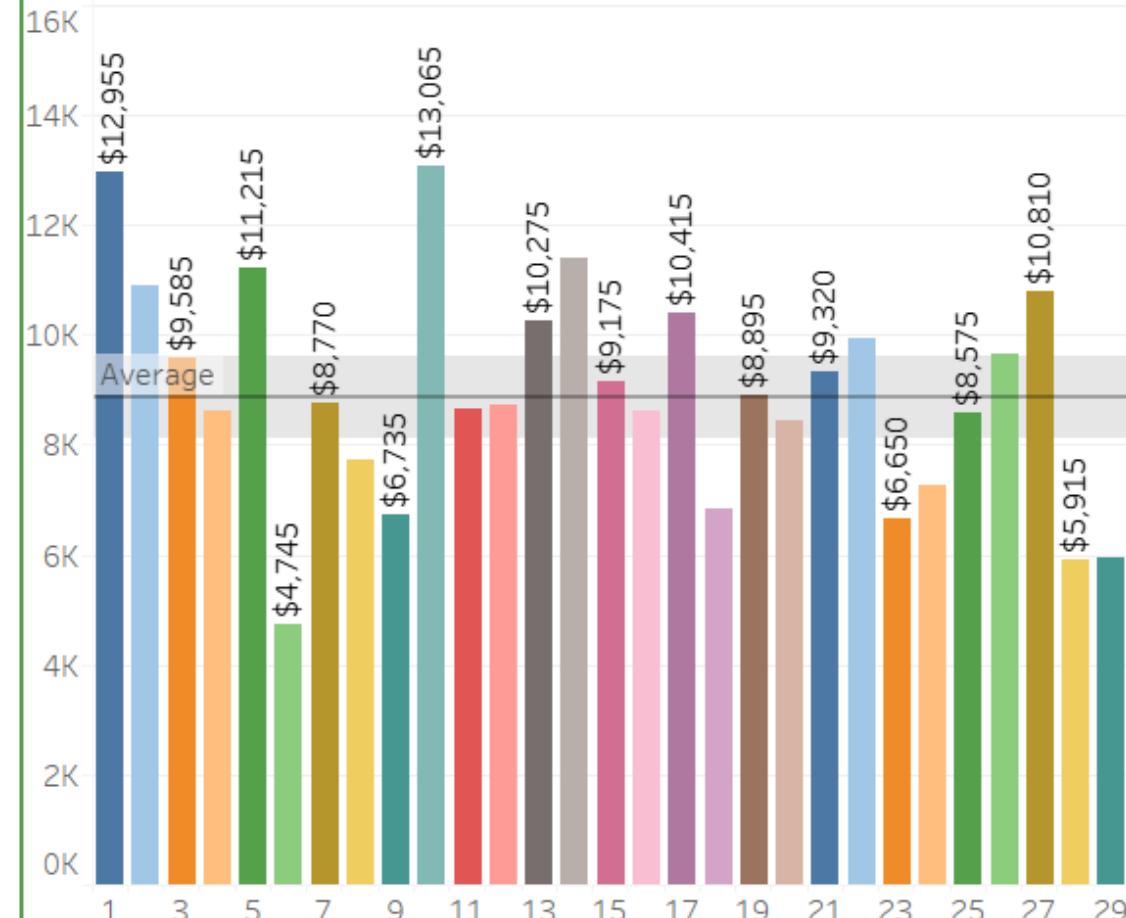
Doanh số bán theo Từng sản phẩm



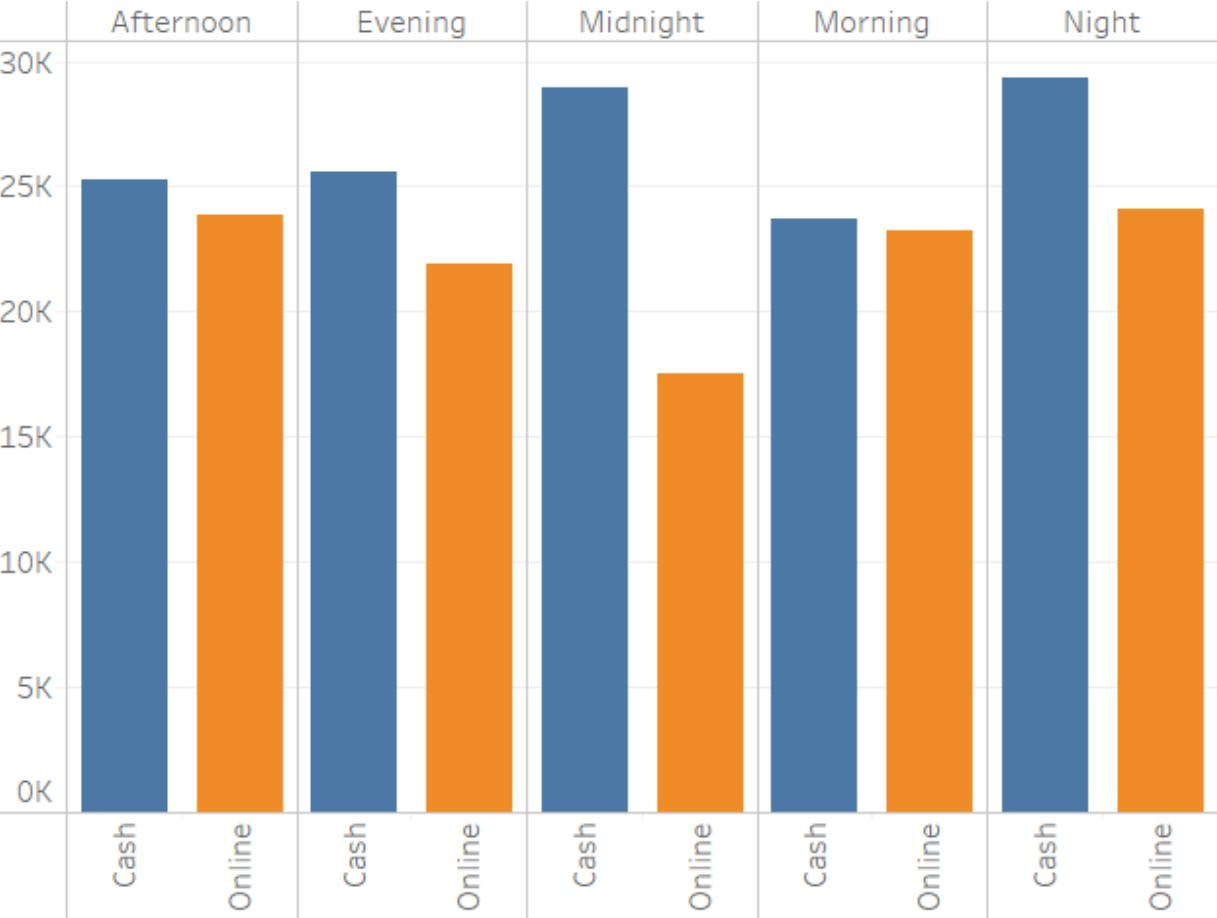
Số lượng bán theo sản phẩm



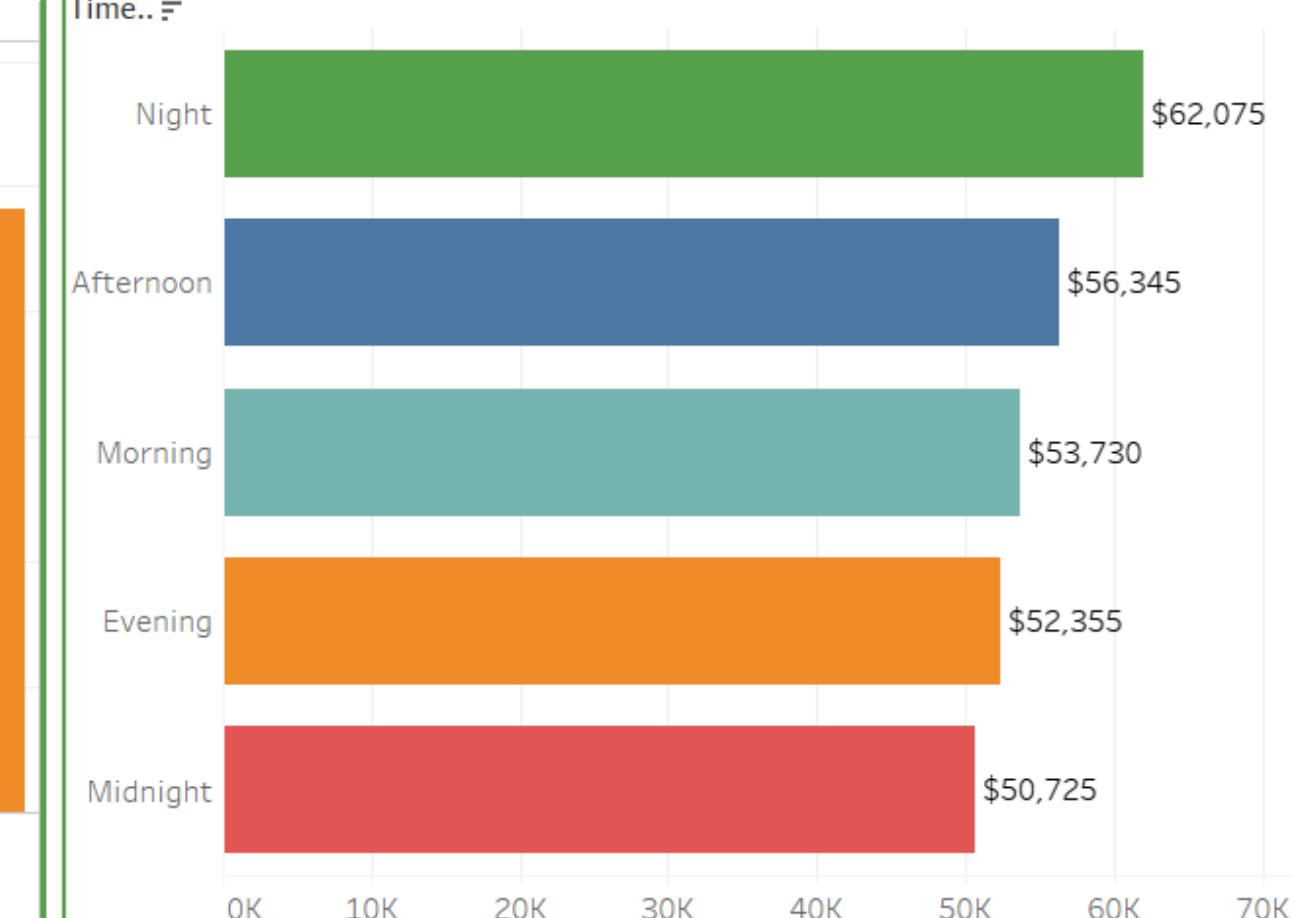
Doanh số bán hàng theo Ngày trong Tháng



Doanh số bán theo các phương thức thanh toán



Doanh số theo Khung giờ



KẾT LUẬN



Data
Warehouse

1. Khách hàng có xu hướng mua sản phẩm theo từng khung giờ khác nhau, khung giờ cao điểm nhất là vào buổi TỐI
2. Việc thay đổi mùa hay thời tiết không ảnh hưởng đến hành vi mua của khách hàng
3. Thanh toán Cash (tiền mặt) được ưu tiên hơn
4. Ngày bán cao điểm và ngày bán thấp điểm ở đây cũng có chênh lệch và đồ ăn sẽ chênh lệch nhiều hơn.
=> Cửa hàng có thể hiểu rõ khách hàng hơn, sẽ thay đổi kế hoạch bán hàng theo hành vi mua của khách hàng để hỗ trợ khách hàng tốt hơn cũng như nâng cao lợi nhuận.