



Cuộc thi Trí tuệ kinh doanh
Business Intelligence 7

BÀI DỰ THI VÒNG 1

Nhóm NTS



Tháng 06 năm 2023

DANH SÁCH NHÓM

Họ và tên	Trường	Vai trò
Trần Nhật Nguyên	ĐH Kinh tế - Luật	Trưởng nhóm
Man Đắc Sang	ĐH Kinh tế - Luật	Thành viên
Trần Thị Kiều Trinh	ĐH Kinh tế - Luật	Thành viên

MỤC LỤC

I. TIỀN XỬ LÝ DỮ LIỆU	1
II. TẢI DỮ LIỆU VÀO KHO DỮ LIỆU	3
1. Bus matrix.....	3
2. Master data	3
3. Transaction data.....	3
4. Mô hình lược đồ hình sao	3
5. Quy trình ETL	4
III. GIẢI QUYẾT CÂU HỎI	5
Câu 1:.....	5
Câu 2:.....	6
a) Phân tích độ tuổi	6
c) Giới tính	7
d) Số năm gia nhập	9
Câu 3:.....	9
Câu 4:.....	10
a) Lý thuyết	10
b) Kết quả phân khúc khách hàng	12
c) Phân tích kết quả	13
Câu 5:.....	13
a) Quý I năm 2021	13
b) Quý I năm 2022	14
c) So sánh	16
Câu 6: 3 insights tâm đắc nhất từ câu 3 và câu 4.....	16
a) Khách hàng trung thành mang về doanh thu rất cao dù không nhiều khách hàng	16
b) Những khách hàng chưa xác định được giới tính có chung nhóm độ tuổi và mang lại doanh thu rất cao	16
c) Doanh thu tháng 4 năm 2022 tăng vượt bậc	17

DANH MỤC HÌNH

Hình 1: Chuyển kiểu dữ liệu và gộp dữ liệu	1
Hình 2: Thông tin bảng dữ liệu đã gộp	2
Hình 3: Các dòng dữ liệu trùng	2
Hình 4: Xuất file dữ liệu.....	2
Hình 5: Sơ đồ quan hệ giữa các bảng.....	4
Hình 6: ETL	4
Hình 7: Biểu đồ thực thể - quan hệ	5
Hình 8: Biểu đồ thể hiện phân bố tuổi của khách hàng.....	6
Hình 9: Tỷ lệ các nhóm tuổi so với toàn bộ tệp khách hàng.....	7
Hình 10: Tỷ lệ giới tính trong tệp khách hàng	7
Hình 11: Biểu đồ phân bố giới tính ở các nhóm tuổi	8
Hình 12: Biểu đồ thể hiện số năm trở thành khách hàng	9
Hình 13: Sales overview dashboard	10
Hình 14: Đánh giá điểm M dựa vào tứ phân vị.....	11
Hình 15: Đánh giá điểm M dựa vào tứ phân vị.....	11
Hình 16: Đánh giá điểm M dựa vào tứ phân vị.....	12
Hình 17: Biểu đồ tròn số lượng các phân khúc khách hàng.....	12
Hình 18: Bảng kết quả chi tiết phân tích RFM các phân khúc khách hàng	12
Hình 19: Biểu đồ cột và đường biểu thị số lượng khách hàng và phần trăm doanh thu	16
Hình 20: Các biểu đồ thể hiện doanh số, phần trăm doanh số, số lượng khách hàng theo độ tuổi & giới tính và biểu đồ Hình 13.	17
Hình 21: Doanh thu và biên lợi nhuận	17
Hình 22: Matrix biểu hiện phần trăm số lượng hàng hóa giảm giá ở mỗi cửa hàng.....	18
Hình 23: Matrix biểu hiện doanh thu và lợi nhuận ở mỗi cửa hàng.....	18

DANH MỤC BẢNG

Bảng 1: Mối quan hệ giữa các bảng	5
Bảng 2: Mô tả phân khúc khách hàng	13
Bảng 3: Doanh thu, hoa hồng, phần trăm đóng góp vào doanh thu của mỗi nhân viên trong Quý I - 2021	13
Bảng 4: Doanh thu, hoa hồng, phần trăm đóng góp vào doanh thu của mỗi nhân viên trong Quý I - 2022	14

I. TIỀN XỬ LÝ DỮ LIỆU

Với tập dữ liệu được trích xuất từ hoạt động bán hàng của Meki Group, nhóm K20-NTS nhận thấy các file về Historical_sales được phân tách thành 3 file dữ liệu theo từng năm với hai loại định dạng .xlsx và .csv. Vì vậy, nhóm quyết định thực hiện gộp 3 file dữ liệu tại thành một file dữ liệu duy nhất bao gồm cả ba năm 2020, 2021 và 2022 để dễ dàng thực hiện việc đẩy dữ liệu vào kho dữ liệu.

Trước khi thực hiện việc gộp dữ liệu, nhóm thực hiện việc chuyển kiểu dữ liệu cột “transaction_date” và “transaction_time” ở file dữ liệu năm 2022 sang format “yyyy-mm-dd” và “HH:mm:ss”.

```
# read historical_sales(2020,2021,2022) dataset
df_2020 = pd.read_excel(path_sales2020)
df_2021 = pd.read_excel(path_sales2021)
df_2022 = pd.read_csv(path_sales2022)

# convert string to datetime (transaction_date)
df_2022['transaction_date'] = pd.to_datetime(df_2022['transaction_date'])
# convert string to time (transaction_time)
df_2022['transaction_time'] = pd.to_datetime(df_2022['transaction_time'], format='%H:%M:%S').dt.time

# concat
df_sales = pd.concat([df_2020, df_2021, df_2022], ignore_index=True)
df_sales
```

	transaction_id	transaction_date	transaction_time	store_id	staff_id	customer_id	product_id	quantity_sold	unit_price	promo_item_yn
0	363	2020-01-15	11:17:45	8	6	8600	72	1	3.25	Y
1	363	2020-01-15	11:17:45	8	6	8600	87	1	2.10	Y
2	1812	2020-01-15	17:14:59	8	43	8600	87	1	2.10	Y
3	1812	2020-01-15	17:14:59	8	43	8600	72	1	3.25	Y
4	1597	2020-01-15	19:17:28	8	45	8059	87	1	2.10	Y
...
553310	626	2022-04-29	19:57:00	8	15	8440	22	1	2.00	N
553311	626	2022-04-29	19:57:00	8	15	8440	77	1	3.00	N
553312	3849	2022-04-29	19:57:00	3	14	1000	52	2	2.50	N
553313	1578	2022-04-29	19:57:00	3	14	1000	51	2	3.00	N
553314	330	2022-04-29	19:59:00	8	15	8600	30	1	3.00	N

553315 rows x 10 columns

Hình 1: Chuyển kiểu dữ liệu và gộp dữ liệu

Kết thúc quá trình gộp dữ liệu, ta thu được tập dữ liệu với **553315 dòng** với **10 cột** dữ liệu.

```
df_sales.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 553315 entries, 0 to 553314
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   transaction_id         553315 non-null  int64  
1   transaction_date       553315 non-null  datetime64[ns]
2   transaction_time       553315 non-null  object  
3   store_id              553315 non-null  int64  
4   staff_id              553315 non-null  int64  
5   customer_id           553315 non-null  int64  
6   product_id            553315 non-null  int64  
7   quantity_sold         553315 non-null  int64  
8   unit_price            553315 non-null  float64 
9   promo_item_yn         553315 non-null  object  
dtypes: datetime64[ns](1), float64(1), int64(6), object(2)
memory usage: 42.2+ MB
```

Hình 2: Thông tin bảng dữ liệu đã gộp

Kiểm tra tính trùng lặp dữ liệu được thực nghiệm để đảm bảo tính toàn vẹn dữ liệu. Kết quả trả về có **1662** dòng trùng lặp:

```
# check duplicate
print('Number of duplicate rows: ' + str(df_sales.duplicated().sum()))

# print duplicate
df_sales[df_sales.duplicated()]

Number of duplicate rows: 1662
```

	transaction_id	transaction_date	transaction_time	store_id	staff_id	customer_id	product_id	quantity_sold	unit_price	promo_item_yn
79	557	2020-01-24	09:39:04	8	44	8600	87	1	2.10	Y
82	2089	2020-01-24	12:23:04	8	43	8600	87	1	2.10	Y
159	819	2020-01-31	08:06:45	8	42	8600	87	1	2.10	Y
251	440	2020-02-24	09:20:55	8	42	8600	87	1	2.10	Y
254	1318	2020-02-24	09:31:54	8	42	8113	87	1	2.10	Y
...
544292	2082	2022-04-24	12:30:00	8	43	8138	87	1	2.10	Y
545026	1921	2022-04-24	19:28:00	8	45	8392	87	1	2.10	Y
548823	208	2022-04-27	08:56:00	5	30	6000	71	1	3.75	Y
548872	206	2022-04-27	09:06:00	8	45	8600	71	1	3.75	N
549068	207	2022-04-27	09:49:00	8	45	8600	71	1	3.75	N

1662 rows x 10 columns

Hình 3: Các dòng dữ liệu trùng

Nhóm thực hiện việc loại bỏ các dòng dữ liệu trùng lặp và xuất file csv “Fact_sales.csv” với **551653** dòng dữ liệu.

```
# export file csv
final = df_sales.drop_duplicates(keep='first', ignore_index=True)
final.sort_values(by=['transaction_id'], inplace=True)
final.to_csv('Fact_sales.csv', index=False)
```

Hình 4: Xuất file dữ liệu

II. TẢI DỮ LIỆU VÀO KHO DỮ LIỆU

1. Bus matrix

	Product	Store	Employee	Customer	Date
Nhân khẩu học				X	
Hiệu suất bán hàng	X	X			X
Phân khúc khách hàng				X	X
Phân tích lương thưởng nhân viên			X		X

2. Master data

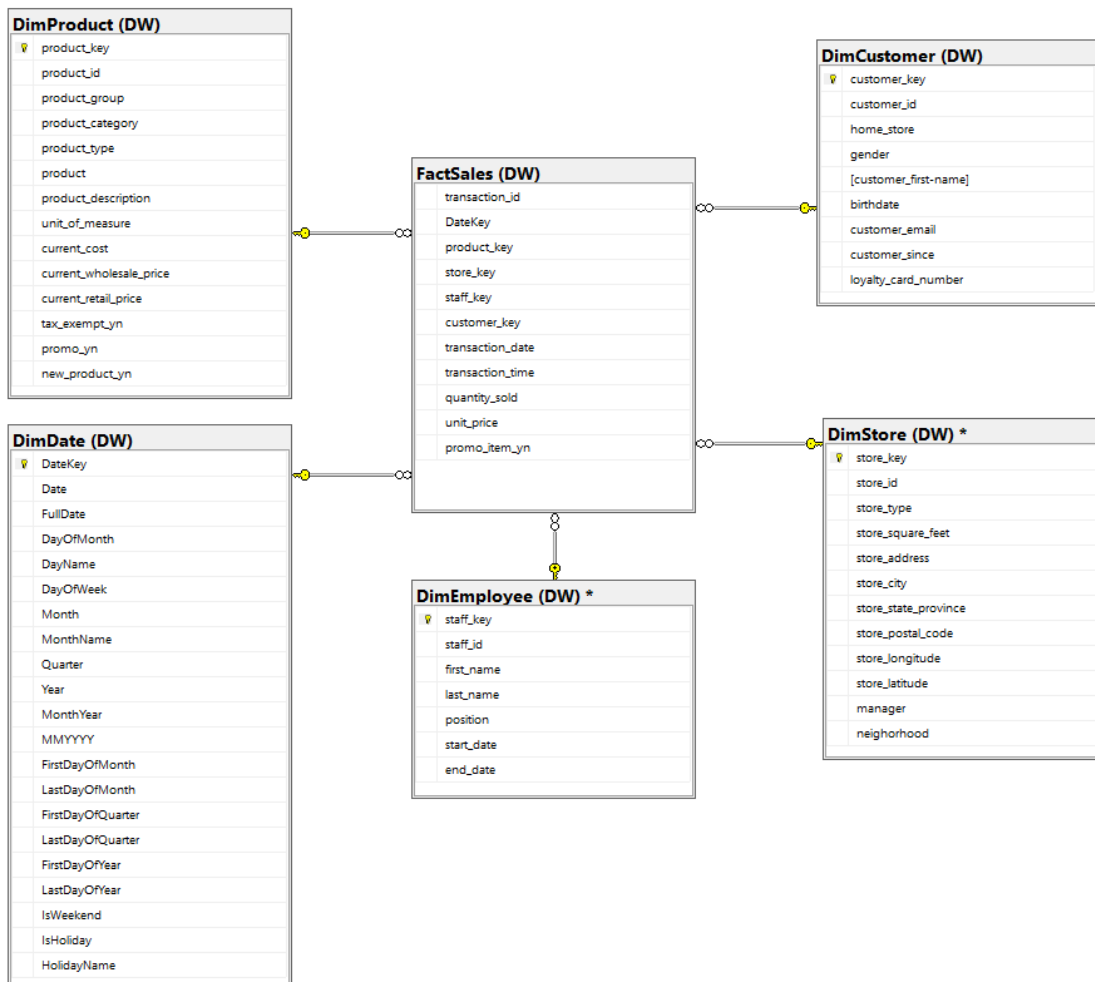
Đối tượng	Miêu tả
Product	Thông tin về sản phẩm (loại sản phẩm, tên sản phẩm, miêu tả,..)
Store	Thông tin về cửa hàng (diện tích, địa chỉ, postal code,..)
Employee	Thông tin về nhân viên (tên, vị trí, ngày bắt đầu,..)
Customer	Thông tin về khách hàng (giới tính, tên, ngày sinh, email,..)

3. Transaction data

Đối tượng	Miêu tả
Historical Sales	Thông tin về thời gian giao dịch, số lượng, đơn giá, khuyến mãi,..

4. Mô hình lược đồ hình sao

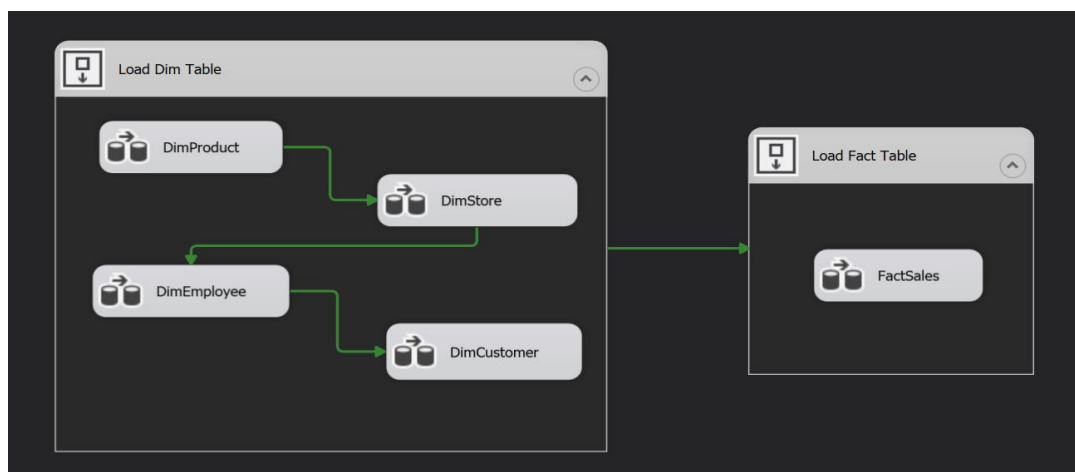
Mô hình Star Schema là một trong những mô hình thiết kế cơ sở dữ liệu phổ biến nhất cho các hệ thống Data Warehouse và mang lại nhiều lợi ích cho các doanh nghiệp trong việc quản lý và phân tích dữ liệu.



Hình 5: Sơ đồ quan hệ giữa các bảng

5. Quy trình ETL

ETL là quá trình trích xuất, biến đổi và tải dữ liệu vào hệ thống data warehouse. Công cụ nhóm sử dụng để thực hiện quy trình ETL là SSIS. Dưới đây là package ETL bao gồm 2 pipeline: Load Dimension table và Load Fact Table:

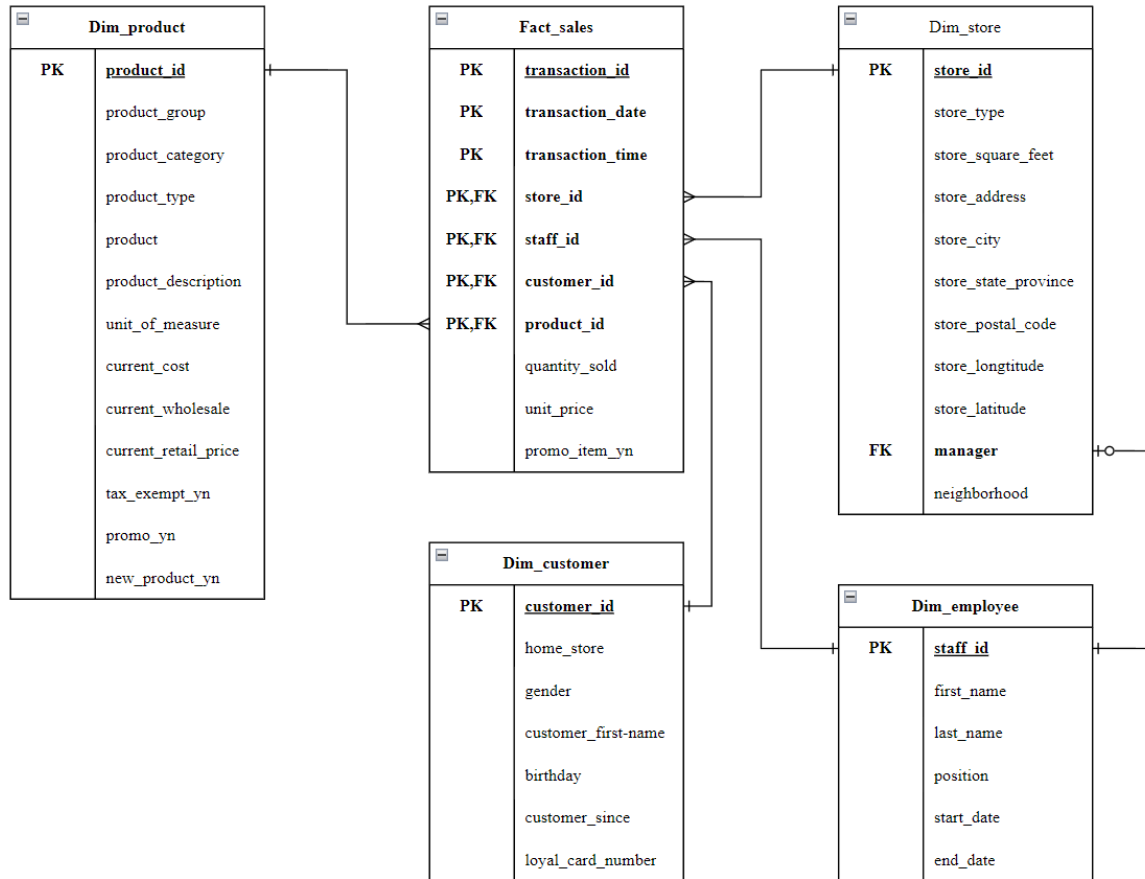


Hình 6: ETL

III. GIẢI QUYẾT CÂU HỎI

Câu 1:

Mối quan hệ giữa các bảng được thể hiện một cách tổng quát qua biểu đồ thực thể - quan hệ ở Hình 1.



Hình 7: Biểu đồ thực thể - quan hệ

Cụ thể, mối quan hệ giữa các bảng được xác định như sau

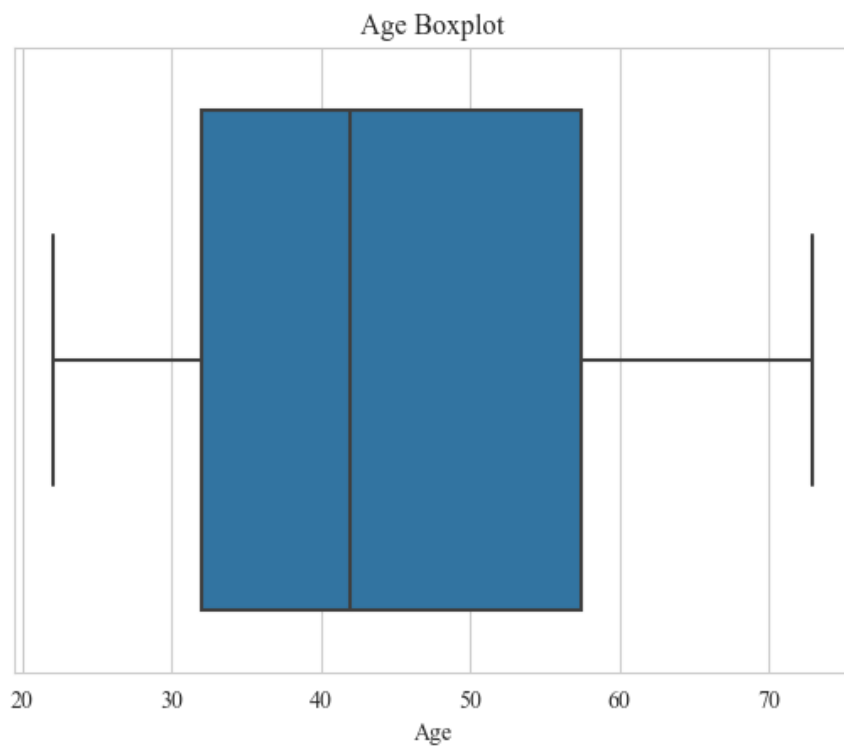
Bảng 1: Mối quan hệ giữa các bảng

Bảng	Mối quan hệ
Dim_product - Fact_sales	1-n
Dim_dim_store - Fact_sales	1-n
Dim_dim_employee - Fact_sales	1-n
Dim_dim_customer - Fact_sales	1-n
Dim_employee - Dim_dim_store	1-1

Câu 2:

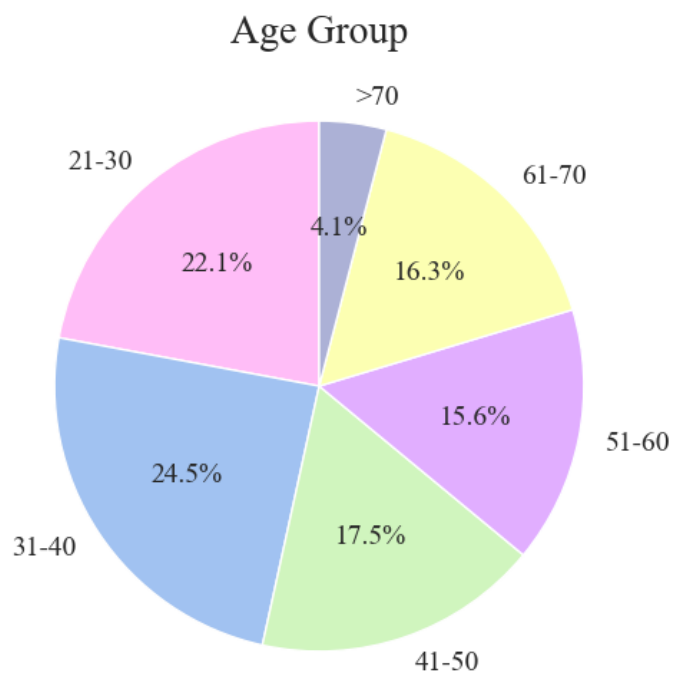
a) Phân tích độ tuổi

- Độ tuổi trung bình của khách hàng là 45.
- Độ tuổi của khách hàng nằm trong khoảng từ 22-73 tuổi.



Hình 8: Biểu đồ thể hiện phân bố tuổi của khách hàng

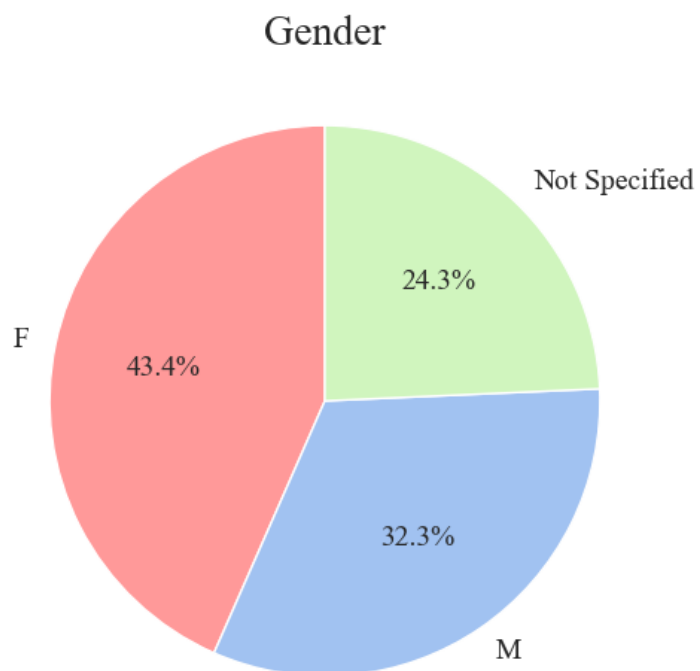
- Phân phối tuổi lệch phải.



Hình 9: Tỷ lệ các nhóm tuổi so với toàn bộ tệp khách hàng

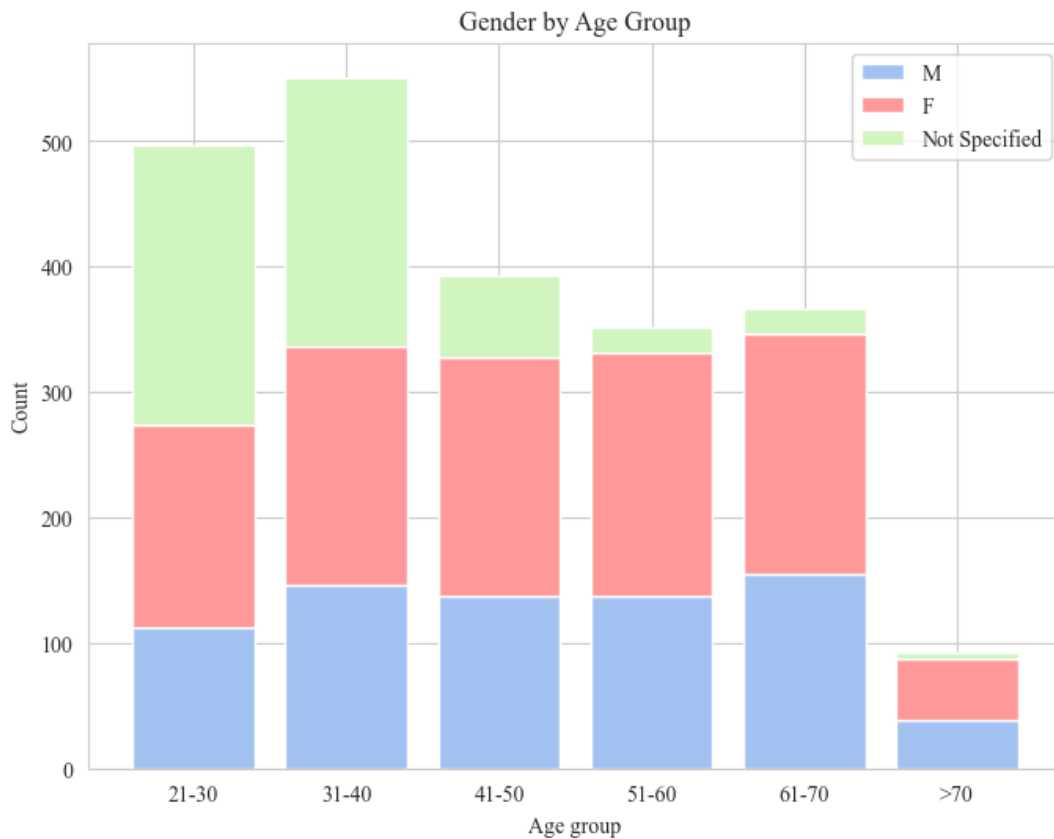
- Có thể thấy, khách hàng từ 31-40 tuổi chiếm tỷ lệ lớn nhất, với 24.5%, tiếp theo đó là nhóm khách hàng từ 21-30 tuổi, chiếm 22.1%.

c) Giới tính



Hình 10: Tỷ lệ giới tính trong tệp khách hàng

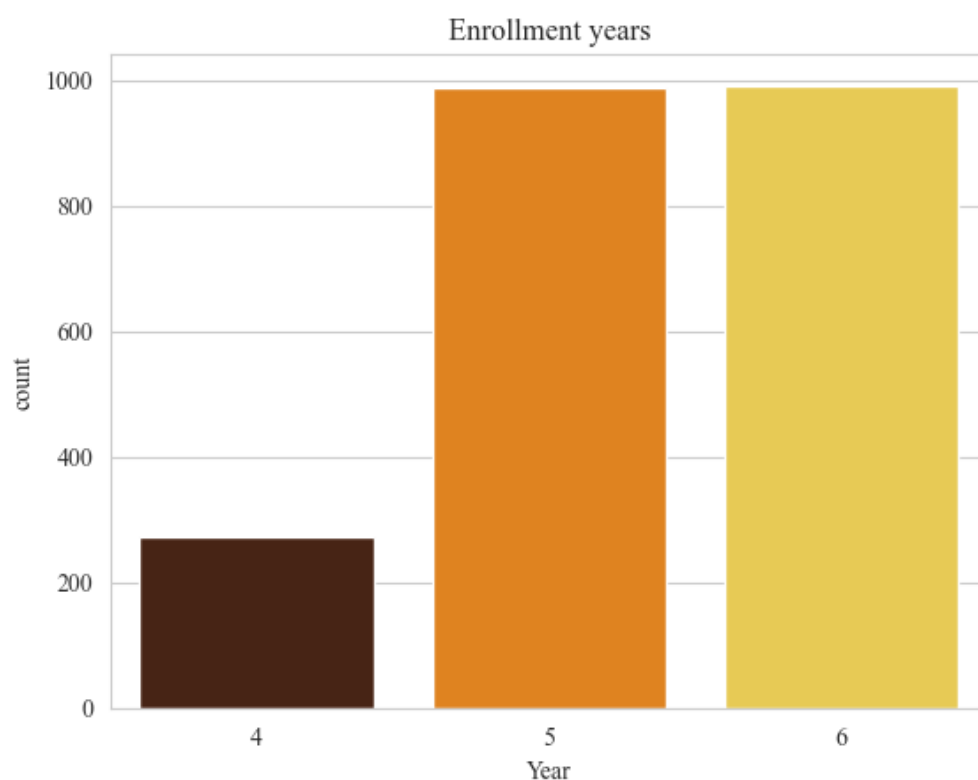
Với dữ liệu được xác định, tỉ lệ khách hàng nữ cao hơn tỉ lệ khách hàng nam, số khách hàng nữ là 977 và con số đối với khách hàng nam là 727, với tỉ lệ lần lượt là 43.4% và 32.3%. Bên cạnh đó, nhóm chưa xác định chiếm tỉ lệ đáng kể, với 24.3% (547 khách hàng), do đó cần tìm xác định thêm thông tin về đối tượng này để có phương án tiếp thị/bán hàng phù hợp.



Hình 11: Biểu đồ phân bố giới tính ở các nhóm tuổi

- Với nhóm khách hàng đã xác định thông tin về giới tính (nam/nữ), ở tất cả các nhóm tuổi, số lượng khách hàng nữ đều cao hơn khách hàng nam.
- Một cách tổng quát, các nhóm khách hàng lớn tuổi hơn có tỉ lệ nam nhiều hơn.
- Thông tin phân bố giới tính ở từng nhóm tuổi là cần thiết trong trường hợp công ty phát triển sản phẩm mới hoặc ra chiến lược tiếp thị tập trung vào một nhóm tuổi cụ thể.

d) Số năm gia nhập



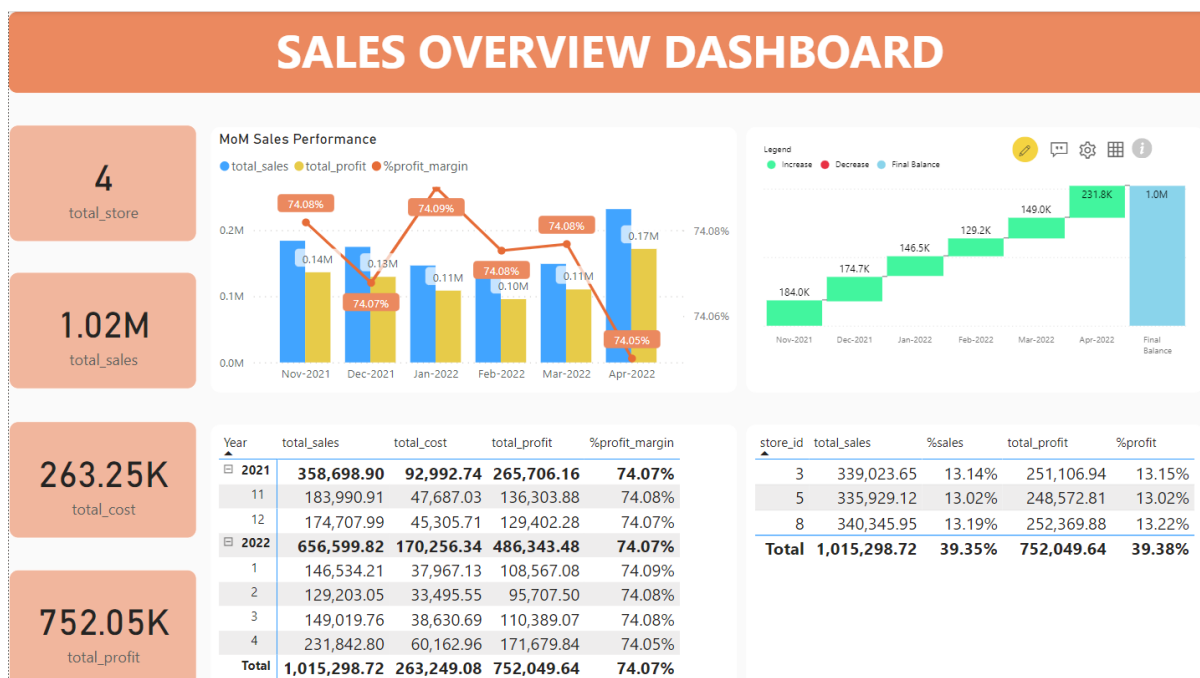
Hình 12: Biểu đồ thể hiện số năm trở thành khách hàng

Tất cả khách hàng đều có số năm gia nhập là từ 4-6 năm (tính tại thời điểm năm 2023). Số lượng khách hàng gia nhập 6 năm trước là cao nhất, với 991 khách hàng, tiếp đó là khách hàng 5 năm là 988 người và nhóm cuối cùng - 4 năm, là 272 khách hàng.

Từ năm 2020, bộ dữ liệu không ghi nhận thêm khách hàng mới.

Câu 3:

Tình hình hoạt động kinh doanh của Meki Group trong vòng 6 tháng gần nhất được biểu diễn dưới dạng dashboard như hình dưới:



Hình 13: Sales overview dashboard

Với hình trên, nhóm có một vài quan sát về hoạt động kinh doanh của Meki Group trong 6 tháng gần nhất (tháng 11,12 năm 2021 và tháng 1,2,3,4 năm 2022) như sau:

- Công ty có **4 store**: 3 store thực hiện việc bán hàng, 1 store dùng để làm warehouse dự trữ hàng hóa
- Doanh thu, chi phí và lợi nhuận đạt được trong 6 tháng gần nhất lần lượt là **1.02M**, **263.25K**, **705.05K**.
- Biên lợi nhuận dao động trong khoảng **74.05-74.09%** qua hầu hết 6 tháng gần nhất
- Tỷ lệ phần trăm đóng góp vào doanh thu được biểu diễn ở biểu đồ Waterfall với tháng 4 năm 2022 đem lại doanh thu vượt trội hơn 5 tháng trước với **231.8K**
- Lợi nhuận của mỗi cửa hàng đều duy trì ổn định khoảng **250,000** với tỉ lệ phần trăm tương đương nhau.

Câu 4:

Sau khi loại bỏ dữ liệu duplicate. Tiến hành phân khúc khách hàng dựa trên phương pháp RFM kết hợp tứ phân vị.

a) Lý thuyết

RFM (Recency – Frequency – M (Monetary Value): là một phần của Marketing Analysis và được sử dụng để phân tích giá trị khách hàng (Customer Value), từ đó giúp doanh

nghiệp có thể phân tích ra từng nhóm khách hàng mà mình đang có, từ đó có những chiến dịch marketing hoặc chăm sóc đặc biệt.

Cấu trúc mô hình RFM

RFM được tạo nên bởi ba yếu tố (điểm cho mỗi chỉ số từ 1 đến 5 tượng trưng từ tệ đến tốt), điểm của mỗi khách hàng sẽ được kết hợp từ điểm 3 yếu tố R,F,M ví dụ 112, 334.

Các khách hàng được phân nhóm như thế nào ?

Đầu tiên các giá trị R,F,M của khách hàng được tính như sau:

R : Khoảng thời gian mua hàng gần nhất đến ngày cuối cùng trên toàn bộ dữ liệu

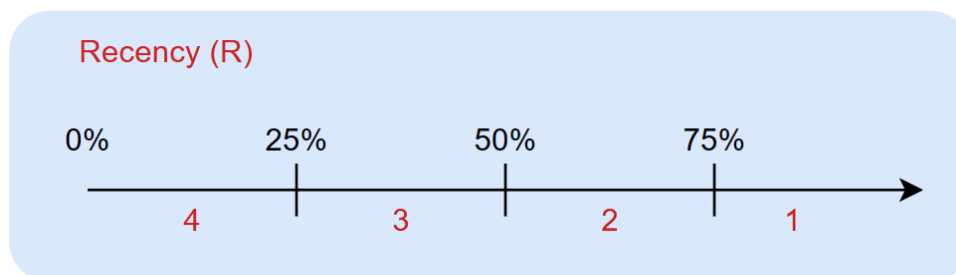
F: Tổng số lần mua của khách hàng đó

M: Tổng số tiền khách hàng đã mua sản phẩm

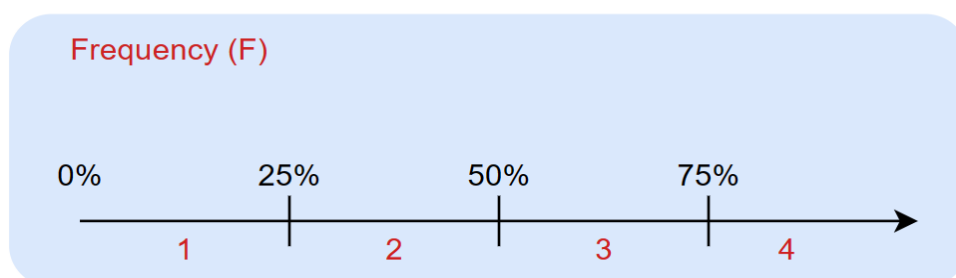
Tiếp theo, tiến hành đánh giá điểm của khách hàng sử dụng phương pháp RFM và tứ phân vị. Dữ liệu các cột được sắp xếp từ bé đến lớn. Lần lượt giá trị R,F,M của mỗi khách hàng được phân chia thành 4 đoạn như sau

- $0\% \leq X < 25\%$
- $25\% \leq X < 50\%$
- $50\% \leq X < 75\%$
- $75\% \leq X < 100\%$

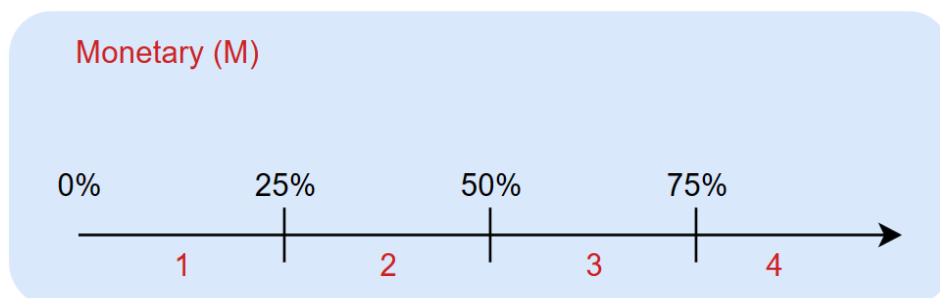
Vì giá trị Recency (R) càng thấp là càng tốt nên cách đánh giá sẽ ngược lại F, M. Cụ thể như sau



Hình 14: Đánh giá điểm M dựa vào tứ phân vị

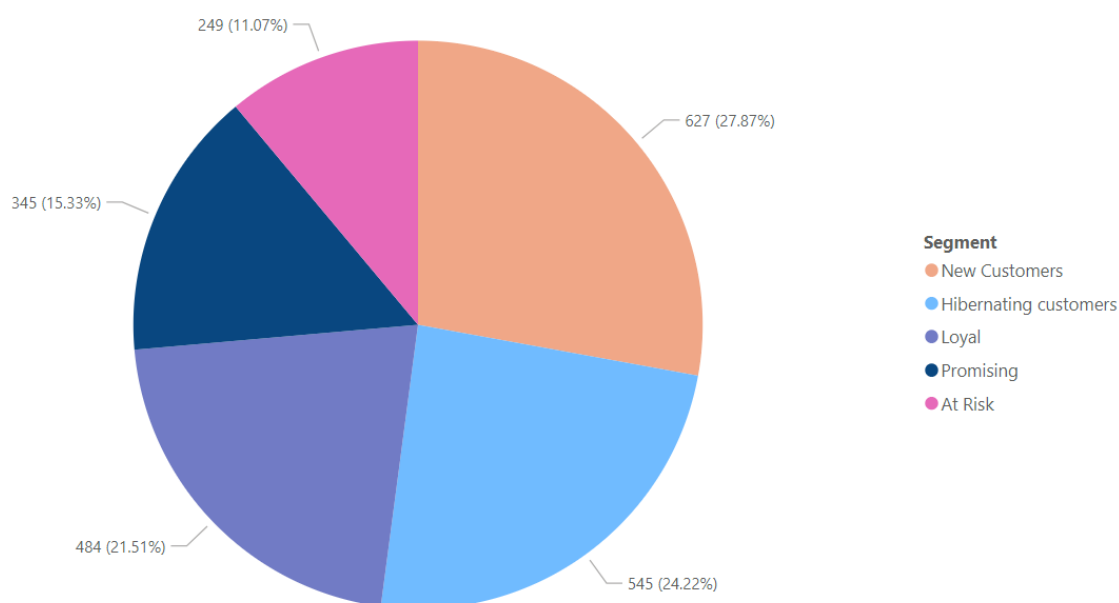


Hình 15: Đánh giá điểm M dựa vào tứ phân vị



Hình 16: Đánh giá điểm M dựa vào tứ phân vị

b) Kết quả phân khúc khách hàng



Hình 17: Biểu đồ tròn số lượng các phân khúc khách hàng

Segment	Avg.R	Avg.F	Avg.M
Hibernating customers	174.16	5.67	310.95
At Risk	135.67	10.10	601.09
Promising	53.66	9.95	627.92
New Customers	26.71	6.62	350.44
Loyal	13.17	29.15	3,197.74

Hình 18: Bảng kết quả chi tiết phân tích RFM các phân khúc khách hàng

Bảng 2: Mô tả phân khúc khách hàng

Phân khúc	Điểm RFM	Đặc tính
Loyal	344, 433, 434, 443, 444	Vừa mới mua, thường xuyên đặt hàng và chi tiêu nhiều nhất.
New Customers	311, 312, 321, 322, 323, 331, 332, 341, 342, 411, 412, 421, 422, 423, 431, 432	Mua lần đầu và gần đây.
Promising	233, 234, 241, 243, 313, 314, 324, 333, 334, 343, 413, 414, 424, 441, 442	Có tiềm năng trở thành khách hàng trung thành vài tháng trước. Thường xuyên chi tiêu một số tiền lớn. Nhưng lần mua hàng cuối cùng là vài tuần trước đây.
Hibernating customers	111, 112, 113, 114, 121, 122, 123, 131, 132, 211, 222, 231	Đã từng mua hàng của doanh nghiệp nhưng đã lâu chưa quay lại
At Risk	124, 133, 134, 141, 142, 143, 144, 212, 213, 214, 221, 223, 224, 232, 242, 244.	Không thể bỏ lỡ họ nhưng đang mất dần dần với lần mua gần nhất đã lâu và giá trị đơn hàng, tần suất thấp hơn

c) Phân tích kết quả

Ta có thể thấy lượng khách hàng mới chiếm tỉ lệ khá nhiều do đó doanh nghiệp đang làm tốt công việc tiếp thị và quảng cáo đến các khách hàng mới. Tuy nhiên nhóm khách hàng “Hibernating customers” cũng chiếm tỉ lệ không nhỏ. Điều này nói lên rằng doanh nghiệp đang tập trung quá nhiều vào tệp khách hàng mới, vì vậy trong tương lai nên chú tâm hơn với các nhóm khách hàng cũ nhằm duy trì lòng trung thành của khách hàng đối với doanh nghiệp

Câu 5:

a) Quý I năm 2021

Bảng 3: Doanh thu, hoa hồng, phần trăm đóng góp vào doanh thu của mỗi nhân viên trong Quý I - 2021

staff_id	first_name	last_name	sales	commission	%sales
12	Britanni	Jorden	49026.80	24.513.400	12.02%
42	Kylie	Candace	32542.14	16.271.070	7.98%

26	Joelle	Christen	30545.80	15.272.900	7.49%
45	Pandora	Neville	27895.14	13.947.570	6.84%
16	Darren	Xu	26020.36	13.010.180	6.38%
17	Quail	Octavia	25026.73	12.513.365	6.14%
20	Ronan	Magee	24443.56	12.221.780	5.99%
14	Damon	Sasha	22090.56	11.045.280	5.42%
30	Amela	Chadwick	20076.88	10.038.440	4.92%
44	Tamekah	Maya	19306.42	9.653.210	4.73%
29	Orson	Benedict	17928.77	8.964.385	4.4%
27	Ainsley	Evelyn	17432.83	8.716.415	4.27%
43	Tatum	Laurel	17139.31	8.569.655	4.2%
28	Joseph	Byron	15822.16	7.911.080	3.88%
15	Remedios	Mari	15519.58	7.759.790	3.81%
6	Karen	Cupps	13043.91	6.521.955	3.2%
41	Adrian	Macon	9720.03	4.860.015	2.38%
7	Kelsey	Cameron	5519.21	2.759.605	1.35%
19	Peter	Paloma	4783.56	2.391.780	1.17%
18	Ezekiel	Rashad	3899.95	1.949.975	0.96%
9	Caldwell	Veda	3057.10	1.528.550	0.75%
8	Hamilton	Emi	2450.25	1.225.125	0.6%
25	Aline	Melanie	2190.39	1.095.195	0.54%
10	Ima	Winifred	1742.60	871.300	0.43%
13	Berk	Derek	646.15	323.075	0.16%

- Bảng trên thể hiện các chỉ số về tổng doanh thu, hoa hồng và tỉ lệ đóng góp vào doanh thu của mỗi nhân viên trong quý I năm 2021, được sắp xếp theo doanh thu giảm dần.
- Britanni Jorden là nhân viên có doanh thu cao nhất, với tỉ lệ đóng góp là **12.02%**.
- Doanh thu trung bình của mỗi nhân viên là: **16314.81**

b) Quý I năm 2022

Bảng 4: Doanh thu, hoa hồng, phần trăm đóng góp vào doanh thu của mỗi nhân viên trong Quý I - 2022

staff_id	first_name	last_name	sales	commission	%sales
12	Britanni	Jorden	50137.29	25.068.645	11.8%
42	Kylie	Candace	34270.84	17.135.420	8.07%
26	Joelle	Christen	30728.14	15.364.070	7.23%
45	Pandora	Neville	29164.18	14.582.090	6.87%
16	Darren	Xu	26754.85	13.377.425	6.3%
17	Quail	Octavia	26702.52	13.351.260	6.29%
20	Ronan	Magee	25544.31	12.772.155	6.01%
14	Damon	Sasha	23053.27	11.526.635	5.43%
30	Amela	Chadwick	20643.07	10.321.535	4.86%
44	Tamekah	Maya	20072.75	10.036.375	4.73%
29	Orson	Benedict	18839.62	9.419.810	4.44%
27	Ainsley	Evelyn	18544.72	9.272.360	4.37%
43	Tatum	Laurel	17890.46	8.945.230	4.21%
15	Remedios	Mari	16260.01	8.130.005	3.83%
28	Joseph	Byron	16146.51	8.073.255	3.8%
6	Karen	Cupps	14689.53	7.344.765	3.46%
41	Adrian	Macon	10099.31	5.049.655	2.38%
7	Kelsey	Cameron	6051.75	3.025.875	1.42%
19	Peter	Paloma	4917.71	2.458.855	1.16%
18	Ezekiel	Rashad	4237.45	2.118.725	1.0%
9	Caldwell	Veda	3080.80	1.540.400	0.73%
8	Hamilton	Emi	2587.60	1.293.800	0.61%
25	Aline	Melanie	1850.18	925.090	0.44%
10	Ima	Winifred	1794.60	897.300	0.42%
13	Berk	Derek	695.55	347.775	0.16%

- Bảng trên thể hiện các chỉ số về tổng doanh thu, hoa hồng và tỉ lệ đóng góp vào doanh thu của mỗi nhân viên trong quý I năm 2021, được sắp xếp theo doanh thu giảm dần.
- Britanni Jorden là nhân viên có doanh thu cao nhất, với tỉ lệ đóng góp là **11.8%**.
- Doanh thu trung bình của mỗi nhân viên là: **16990.28**

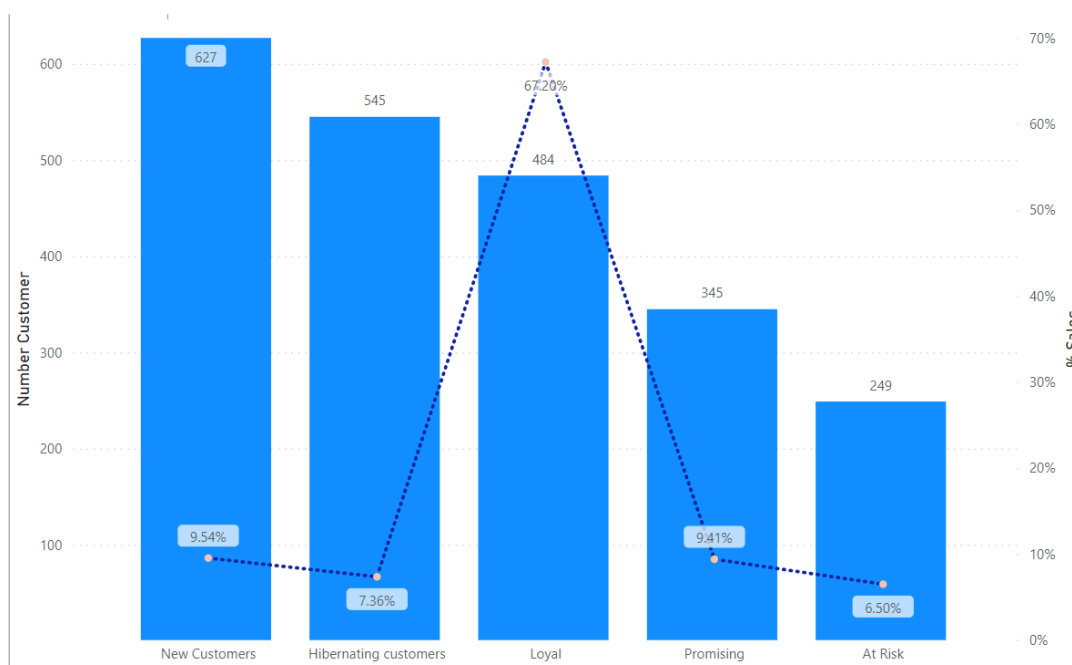
c) So sánh

- Tổng doanh thu của công ty trong quý I năm 2022 là **424757.01** tăng **4.14%** so với cùng kỳ năm 2021 (với doanh thu là **407870.19**)
- Số lượng nhân viên không đổi trong 2 năm 2021 và 2022. Do đó, doanh thu trung bình của mỗi nhân viên trong quý I năm 2022 tăng so với cùng kỳ năm trước.
- Về xếp hạng doanh thu từng nhân viên, hầu như không có sự thay đổi vị trí giữ các nhân viên trong 2 thời kì được xem xét (trừ Remedios Mari và Joseph Byron)

Câu 6: 3 insights tâm đắc nhất từ câu 3 và câu 4

a) Khách hàng trung thành mang về doanh thu rất cao dù không nhiều khách hàng

Lượng khách hàng trung thành chiếm khoảng 20% nhưng lượng doanh thu được chiếm khoảng 70%. Doanh nghiệp nên áp dụng phương pháp Pareto để tiếp tục duy trì phong độ đối với nhóm khách hàng này.

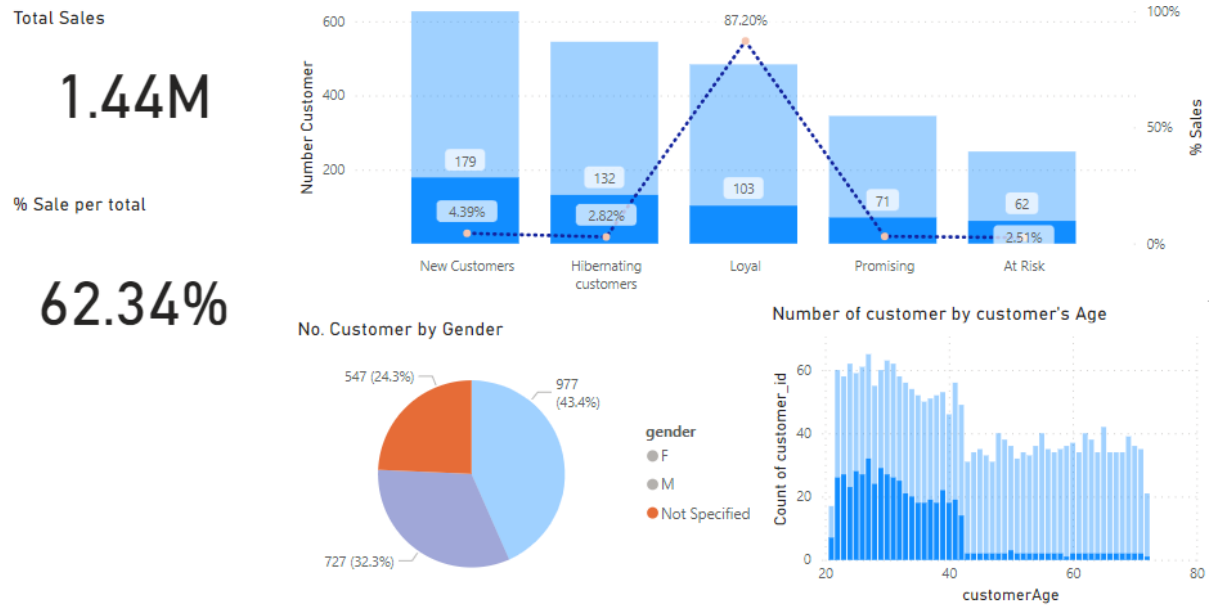


Hình 19: Biểu đồ cột và đường biểu thị số lượng khách hàng và phần trăm doanh thu

b) Những khách hàng chưa xác định được giới tính có chung nhóm độ tuổi và mang lại doanh thu rất cao

Nhìn vào hình dưới đây ta có thể thấy tập hợp các khách hàng chưa xác định được giới tính chiếm khoảng 90% nằm trong độ tuổi từ 20 đến 40. Hơn nữa, tập khách hàng này lại đóng góp vào doanh thu một lượng tương đối lớn ~ 63%. Có thể nói rằng những khách hàng này là lớp người trẻ, bận rộn nhưng vẫn thường xuyên mua sắm. Vậy lý do nào họ không thực hiện điền thông tin giới tính của mình ? Lý do khách quan có

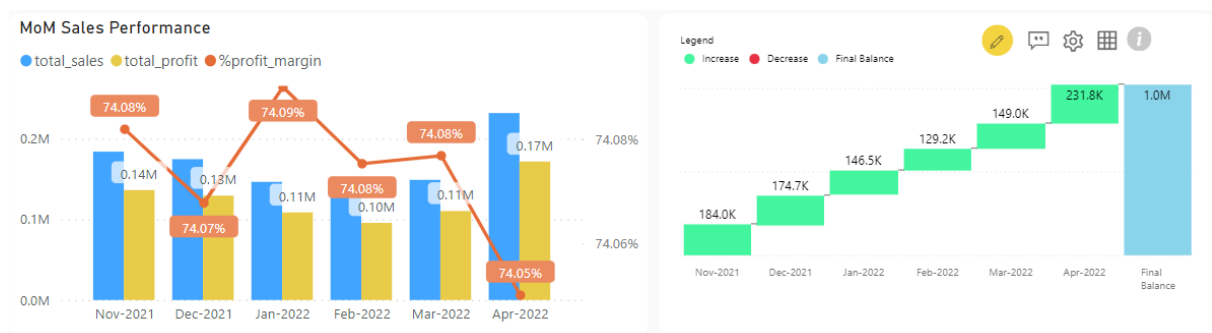
thể là do đặc tính của nhóm này không có nhiều thời gian hoặc không muốn khai báo thông tin cá nhân vì lý do nhạy cảm. Lý do chủ quan là do bản thân doanh nghiệp chưa có các chương trình đủ hấp dẫn để khuyến khích khách hàng điền thông tin khảo sát. Có thể áp dụng các chương trình như tặng mã giảm giá khi điền đầy đủ thông tin cá nhân,...



Hình 20: Các biểu đồ thể hiện doanh số, phần trăm doanh số, số lượng khách hàng theo độ tuổi & giới tính và biểu đồ Hình 13.

c) Doanh thu tháng 4 năm 2022 tăng vượt bậc

Doanh thu và lợi nhuận của tháng 4 năm 2022 mặc dù vượt trội hơn 5 tháng trước đó, tuy nhiên biên lợi nhuận của tháng này có xu hướng giảm nhiều hơn.



Hình 21: Doanh thu và biên lợi nhuận

Nguyên nhân là do vào tháng 4 năm 2022, Meki Group đã thực hiện nhiều chiến dịch giảm giá tại hai cửa hàng 5 và 8 dẫn đến số lượng sản phẩm bán ra tăng vọt đến 22.89%, tuy nhiên phải chịu lượng lớn chi phí. Điều này có nghĩa chiến dịch giảm giá này có thể làm tăng nhu cầu mua hàng của khách hàng nhưng lại đem lại tổn thất cho công ty.

MonthYear	5	8	Total
Nov-2021	6.97%	12.15%	19.12%
Dec-2021	6.34%	10.26%	16.60%
Jan-2022	4.70%	9.87%	14.57%
Feb-2022	4.31%	7.94%	12.25%
Mar-2022	5.23%	9.34%	14.57%
Apr-2022	8.57%	14.33%	22.89%
Total	36.11%	63.89%	100.00%

Hình 22: Matrix biểu hiện phần trăm số lượng hàng hóa giảm giá ở mỗi cửa hàng

Tuy nhiên, ở cửa hàng 3 vào tháng 4 năm 2022 không thực hiện chiến dịch giảm giá sản phẩm nhưng doanh thu và lợi nhuận đem lại vẫn ở mức ổn định so với 2 cửa hàng 3 và 5 khi thực hiện chiến dịch này.

store_id	total_sales	%sales	total_profit	%profit
3	77,277.53	2.99%	57,231.10	3.00%
5	77,058.07	2.99%	57,007.75	2.99%
8	77,507.20	3.00%	57,441.00	3.01%
Total	231,842.80	8.99%	171,679.84	8.99%

Hình 23: Matrix biểu hiện doanh thu và lợi nhuận ở mỗi cửa hàng

Nhận xét trên đưa ra 2 hướng:

Hướng 1: nếu mục đích của chiến dịch nhằm quảng bá sản phẩm mới và tăng nhu cầu sản phẩm cũ thì doanh nghiệp đã thành công với chiến dịch này với mức biên lợi nhuận giảm không quá cao.

Hướng 2: nếu mục đích của chiến dịch nhằm mở rộng thị trường hàng hóa thì đây là một chiến dịch thất bại, doanh nghiệp nên áp dụng chiến dịch khác phù hợp hơn để giúp tăng doanh thu nhưng vẫn đảm bảo mức chi phí ổn định.