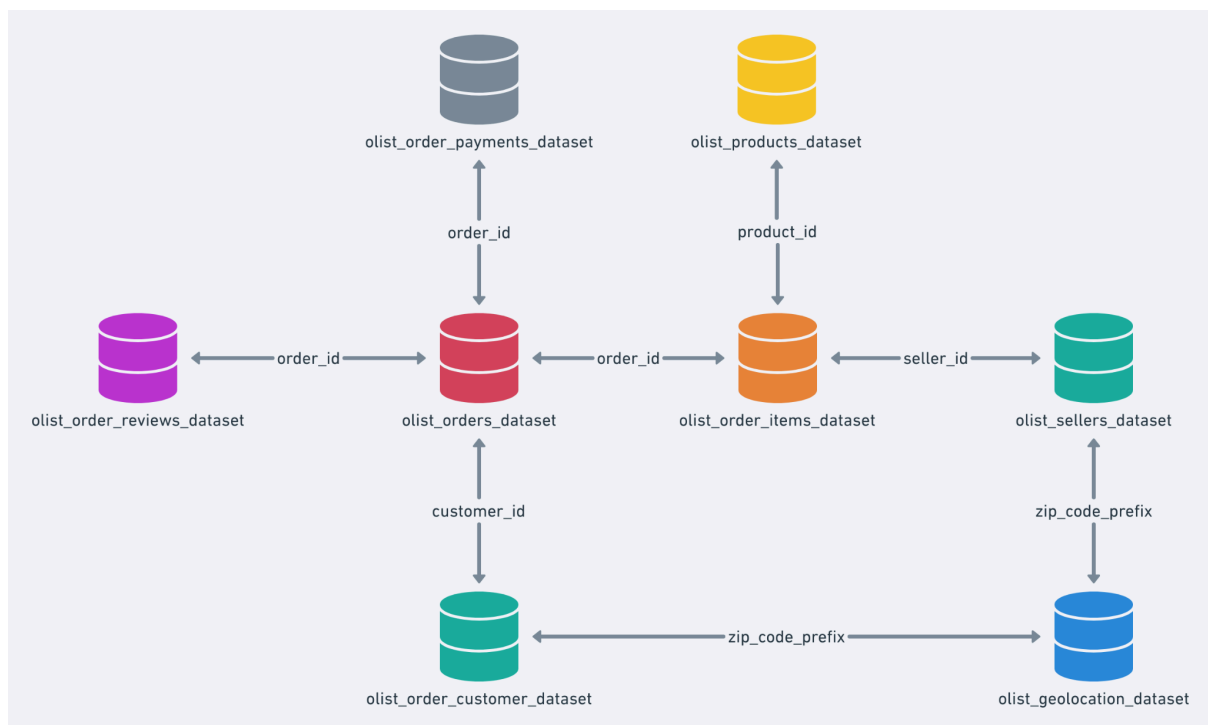


PROJECT: PHÂN TÍCH BRAZILIAN E-COMMERCE PUBLIC DATASET - OLIST

1. Business Understanding

Bộ dữ liệu được cung cấp bởi Olist, một tập đoàn siêu thị lớn nhất ở Brazil. Olist kết nối các doanh nghiệp nhỏ từ khắp Brazil với các kênh bán hàng một cách dễ dàng thông qua một hợp đồng duy nhất. Những người bán hàng này có thể bán sản phẩm của mình thông qua cửa hàng của Olist và vận chuyển trực tiếp đến khách hàng bằng cách sử dụng các đối tác hậu cần của Olist. Sau khi khách hàng mua sản phẩm từ Olist Store, người bán sẽ nhận được thông báo thực hiện đơn hàng đó. Sau khi khách hàng nhận được sản phẩm hoặc đến hạn giao hàng dự kiến, khách hàng sẽ nhận được một bản khảo sát về mức độ hài lòng qua email, nơi họ có thể đưa ra ghi chú về trải nghiệm mua hàng và viết ra một số nhận xét.

KH order → Olist connect Business → Business sell



2. Preparation data

Nguồn data: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data>

Thời gian thu thập: 04/09/2016 - 17/10/2018

3. Cleaning data

❖ Change Type

- Date/Time → Date
- Text → Number: product_length_cm

```
types(dbo_olist_products_data$
```

1.2	product_length_cm	1.2
100%	Valid	100%

- Number → Text : zip_code

❖ Thêm cột :

- Thêm cột R,F,M, R_score, R_score, M_score để phân loại khách hàng

```

1 RFM = SUMMARIZE(
2     olist_orders_dataset, olist_orders_dataset[customer_id]
3     , "R"
4     , DATEDIFF(MAX(olist_orders_dataset[order_purchase_timestamp]), DATE(2019,01,01), DAY)
5     , "F"
6     , DISTINCTCOUNT(olist_orders_dataset[order_id])
7     , "M"
8     , SUM(olist_order_items_dataset[price])
9 )

```

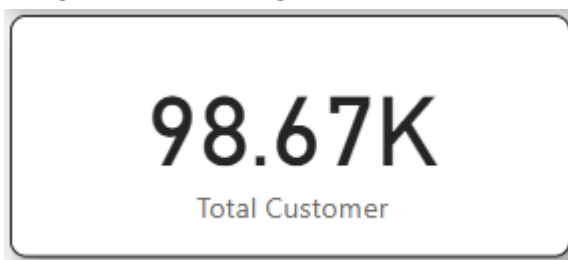
Dax cột R_score	Dax cột F_score	Dax cột M_score
<pre> 1 R_score = 2 var percent_80 = PERCENTILE.EXC(RFM[R], 0.8) 3 var percent_60 = PERCENTILE.EXC(RFM[R], 0.6) 4 var percent_40 = PERCENTILE.EXC(RFM[R], 0.4) 5 var percent_20 = PERCENTILE.EXC(RFM[R], 0.2) 6 RETURN 7 SWITCH(TRUE() 8 , RFM[R] >= percent_80, 1 9 , RFM[R] >= percent_60, 2 10 , RFM[R] >= percent_40, 3 11 , RFM[R] >= percent_20, 4 12 , 5) </pre>	<pre> 1 F_score = 2 var percent_80 = PERCENTILE.EXC(RFM[F], 0.8) 3 var percent_60 = PERCENTILE.EXC(RFM[F], 0.6) 4 var percent_40 = PERCENTILE.EXC(RFM[F], 0.4) 5 var percent_20 = PERCENTILE.EXC(RFM[F], 0.2) 6 RETURN 7 SWITCH(TRUE() 8 , RFM[F] >= percent_80, 5 9 , RFM[F] >= percent_60, 4 10 , RFM[F] >= percent_40, 3 11 , RFM[F] >= percent_20, 2 12 , 1) </pre>	<pre> 1 M_score = 2 var percent_80 = PERCENTILE.EXC(RFM[M], 0.8) 3 var percent_60 = PERCENTILE.EXC(RFM[M], 0.6) 4 var percent_40 = PERCENTILE.EXC(RFM[M], 0.4) 5 var percent_20 = PERCENTILE.EXC(RFM[M], 0.2) 6 RETURN 7 SWITCH(TRUE() 8 , RFM[M] >= percent_80, 5 9 , RFM[M] >= percent_60, 4 10 , RFM[M] >= percent_40, 3 11 , RFM[M] >= percent_20, 2 12 , 1) </pre>

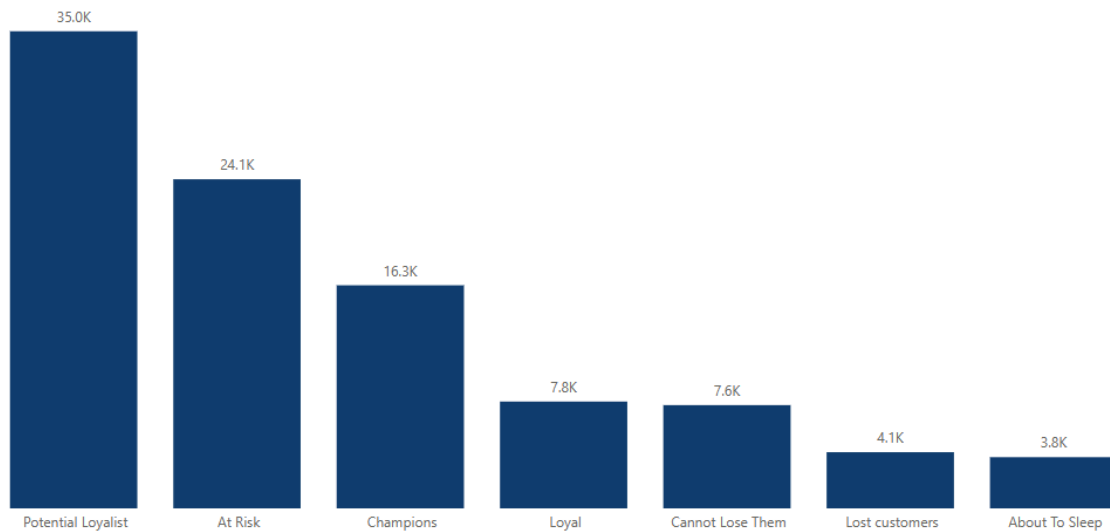
- Thêm cột Year

4. EDA

4.1 Phân tích Khách hàng

Tổng số Khách hàng: 98.67K





❖ **Nhóm Best customers: Potential Loyalist, Loyal, Champion (Chiếm khoảng 60% tổng số khách hàng)**

Champions (16.3K) – Nhóm khách hàng tốt nhất

Đặc điểm

- Mua hàng thường xuyên với giá trị cao.
- Trung thành và có thể là người giới thiệu sản phẩm/dịch vụ.
- Khả năng tương tác cao với thương hiệu.

Potential Loyalist (35.0K) – Nhóm có tiềm năng trung thành và chiếm tỷ lệ cao nhất

Đặc điểm:

- Đã mua hàng nhiều lần nhưng chưa hoàn toàn trung thành.
- Có thể đang xem xét sản phẩm/dịch vụ khác.
- Nhóm khách hàng lớn nhất trong biểu đồ.

Loyal (7.8K) – Nhóm khách hàng trung thành

Đặc điểm:

- Mua hàng thường xuyên trong thời gian dài.
- Không có dấu hiệu rời bỏ nhưng cần duy trì sự quan tâm.
- Có thể chuyển thành "Champions" nếu chăm sóc tốt.

❖ **Nhóm khách hàng rời bỏ: Lost Customer 4.1K (khoảng 4.17%)**

Đặc điểm:

- Đã không mua hàng trong thời gian dài.
- Có thể đã chuyển sang đối thủ.

❖ **Nhóm khách hàng có nguy cơ rời bỏ: At Risk, Cannot Lose Them, About To Sleep (chiếm khoảng 36%)**

At Risk (24.1K) – Nhóm có nguy cơ rời đi

Đặc điểm:

- Đã từng mua hàng nhưng gần đây ít tương tác.

Cannot Lose Them (7.6K) – Nhóm khách hàng cực kỳ quan trọng

Đặc điểm:

- Đã từng là khách hàng trung thành nhưng đang có dấu hiệu rời đi.
- Có giá trị cao, mất họ sẽ ảnh hưởng doanh thu.

About To Sleep (3.8K) – Nhóm sắp không còn mua hàng

Đặc điểm:

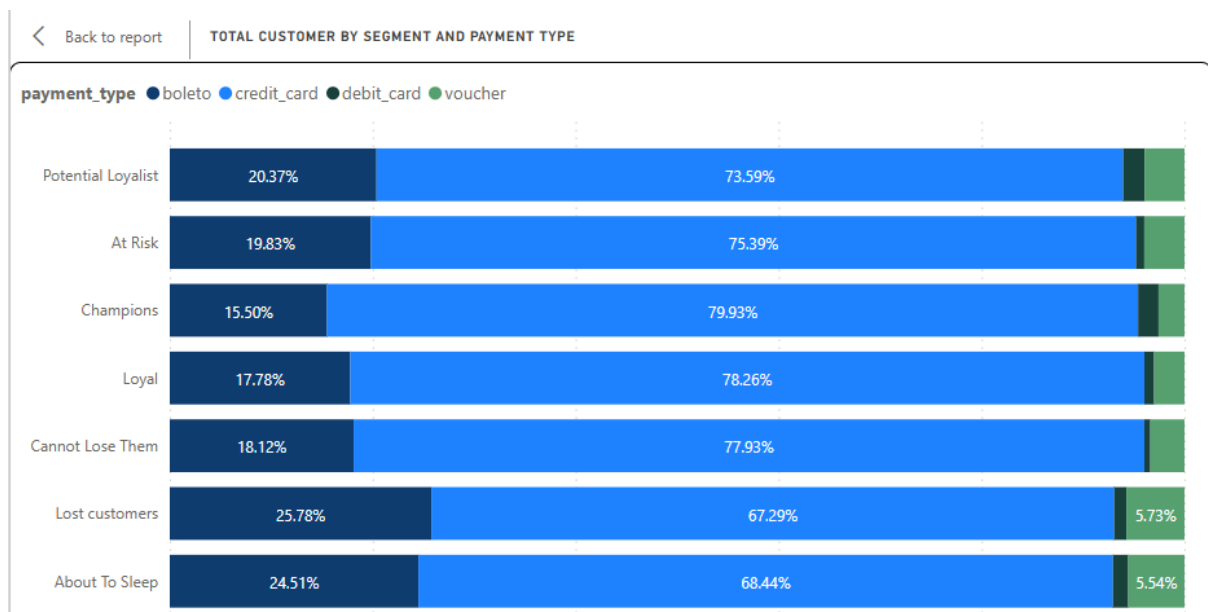
- Đã có dấu hiệu ngừng mua hàng nhưng chưa hẳn rời đi.
- Có khả năng chuyển thành nhóm Lost Customers nếu không chăm sóc.

< Back to report

customer_state	About To Sleep	At Risk	Cannot Lose Them	Champions	Lost customers	Loyal	Potential Loyalist	Total
SP	1578	8962	2625	6505	1619	2845	15824	39958
RJ	423	3337	1028	1964	467	1103	3968	12290
MG	415	2816	851	1896	485	944	3763	11170

- Bảng SP có số lượng khách hàng cao nhất (39,958 khách hàng), chiếm tỷ lệ áp đảo so với các bang khác.
- Các bang có lượng khách hàng cao tiếp theo là: RJ - 12,290 MG - 11,170

→ SP, RJ và MG là những bang có dân số đông và nền kinh tế phát triển mạnh, do đó chỉ 3 bang đã chiếm tổng lượng khách hàng khoảng 64% tổng khách hàng.

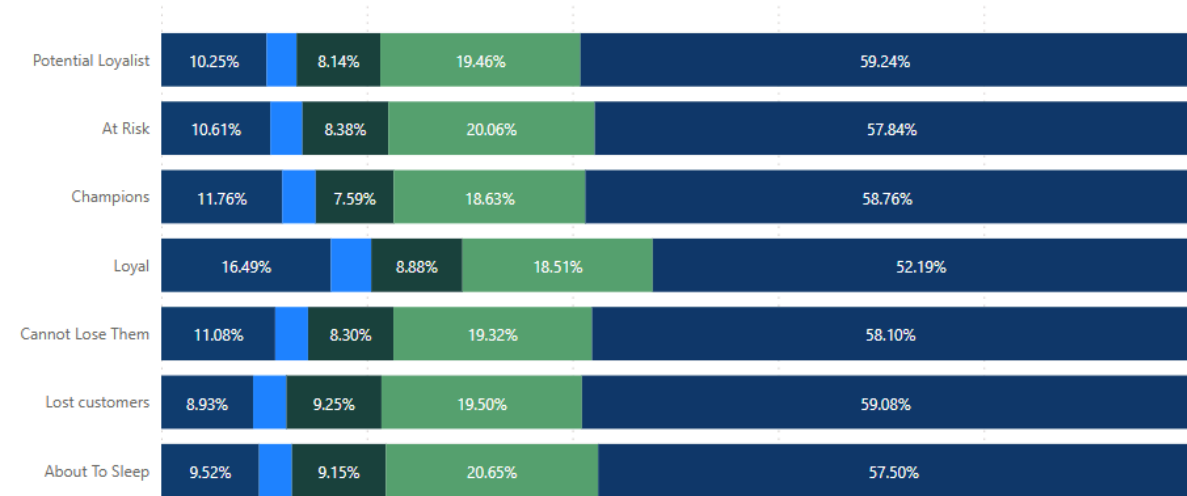


Dựa vào biểu đồ, ta thấy rằng **Lost Customers** và **About To Sleep** có tỷ lệ sử dụng phương thức thanh toán **boleto** (hơn 5%) và **voucher** (hơn 3%) cao hơn so với các nhóm khác. Cho thấy

- Nhóm KH này mua hàng không thường xuyên, bị thu hút bởi khuyến mãi.
- Thích thanh toán một lần, tránh sử dụng thẻ tín dụng.
- Ít tương tác với thương hiệu và dễ rời bỏ.

→ **Đề xuất** : Cá nhân hóa ưu đãi lớn để khuyến khích mua hàng quay lại và mua thường xuyên.

review_score ● 1 ● 2 ● 3 ● 4 ● 5



- Hầu hết các nhóm khách hàng đều có tỷ lệ **review 5 sao cao nhất** (~52-59%), cho thấy phần lớn khách hàng có trải nghiệm tốt.
- Tuy nhiên nhóm Loyal (Khách hàng trung thành) có tỷ lệ review 1 sao cao nhất (16.49%), trong khi các nhóm có nguy cơ rời bỏ hoặc đã rời bỏ như Lost Customers (~8.93%) và About to Sleep (~9.52%) lại có tỷ lệ thấp hơn.

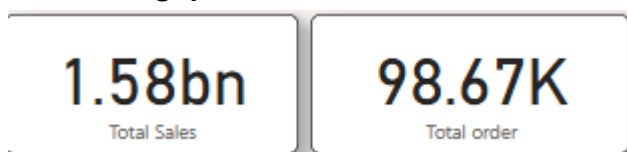
→ Khách hàng trung thành (Loyal, Champions, Cannot Lose Them) đã mua hàng nhiều lần, họ có kỳ vọng cao hơn về chất lượng sản phẩm, dịch vụ, và hậu mãi.

→ Khách hàng sắp rời bỏ có thể không quan tâm hoặc không đánh giá nữa, thay vào đó họ rời đi mà không phản hồi.

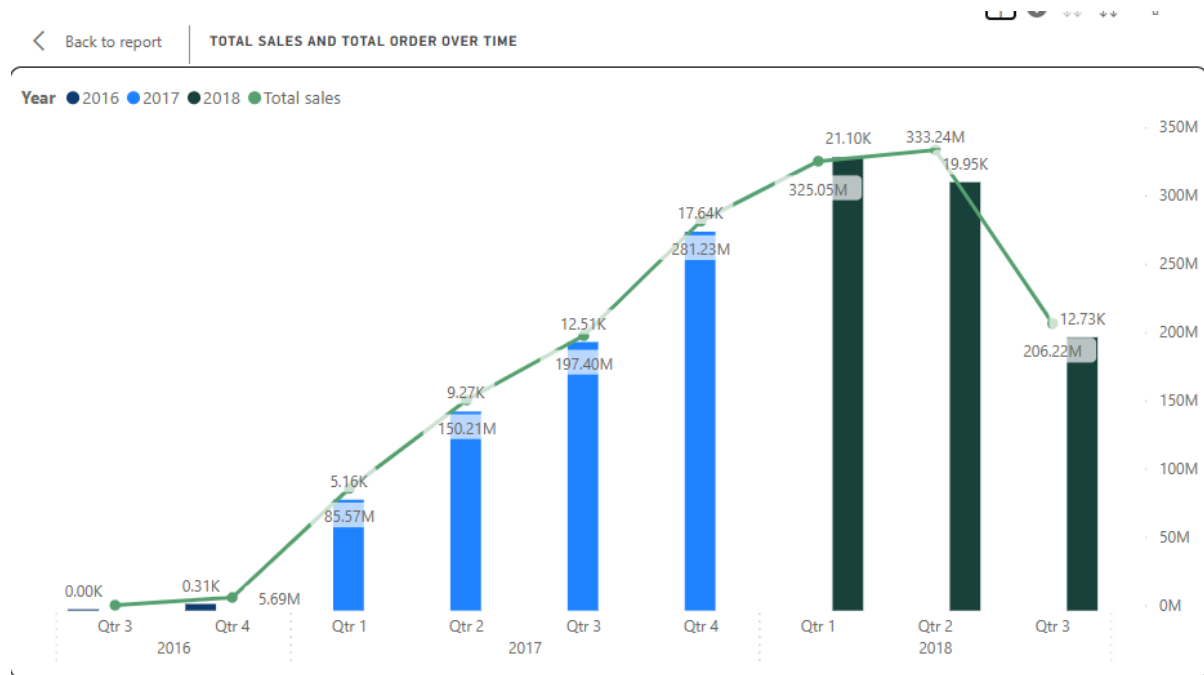
→ **Đề xuất:** Chủ động theo dõi đánh giá thấp của từng nhóm đặc biệt là các nhóm khách hàng trung thành và cung cấp hỗ trợ nhanh chóng để tối ưu hóa trải nghiệm.

4.2 Phân tích doanh thu

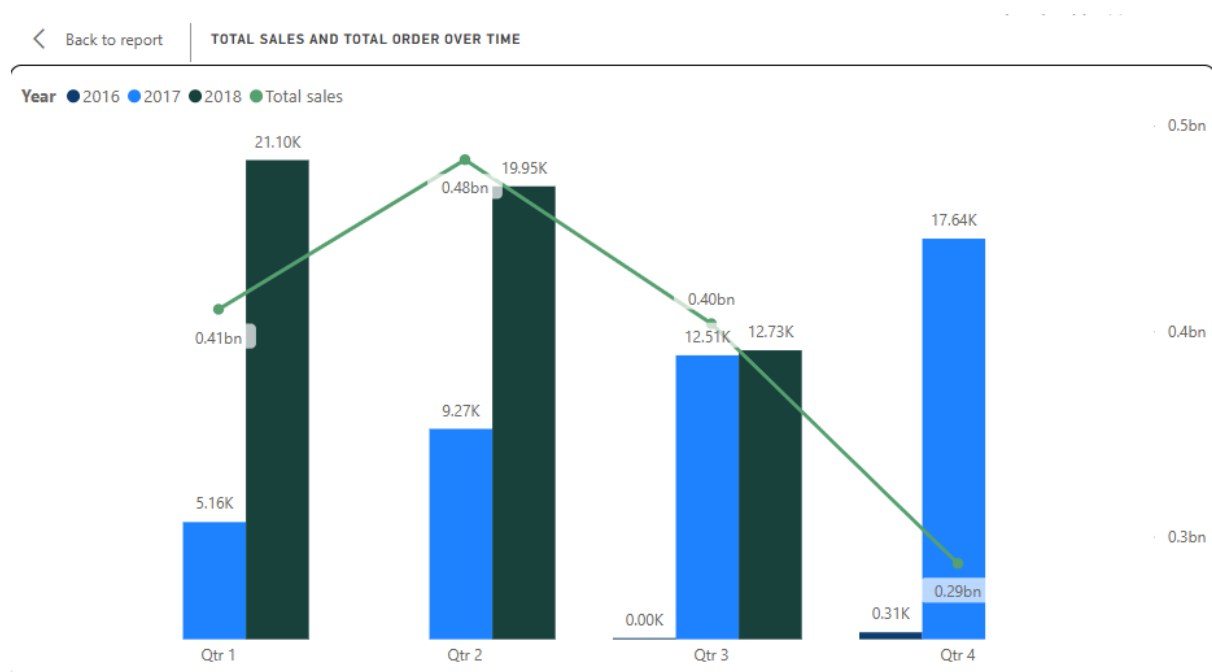
❖ Tổng quan



- Từ ngày 04/09/2016 - 17/10/2018, tổng doanh thu của Olist là 1,5 tỷ và tổng số lượng đơn hàng là 98,67K đơn



- Từ Q4/2016 - Q2/2018, doanh thu của Olist có xu hướng tăng mạnh, từ 5.6M lên 333M nhưng từ Q3/2018 lại có xu hướng giảm còn 206.22M (giảm khoảng 38%)



Doanh thu có xu hướng như vậy là do tổng số đơn hàng giữa các năm có sự đối lập nhau trong cùng quý

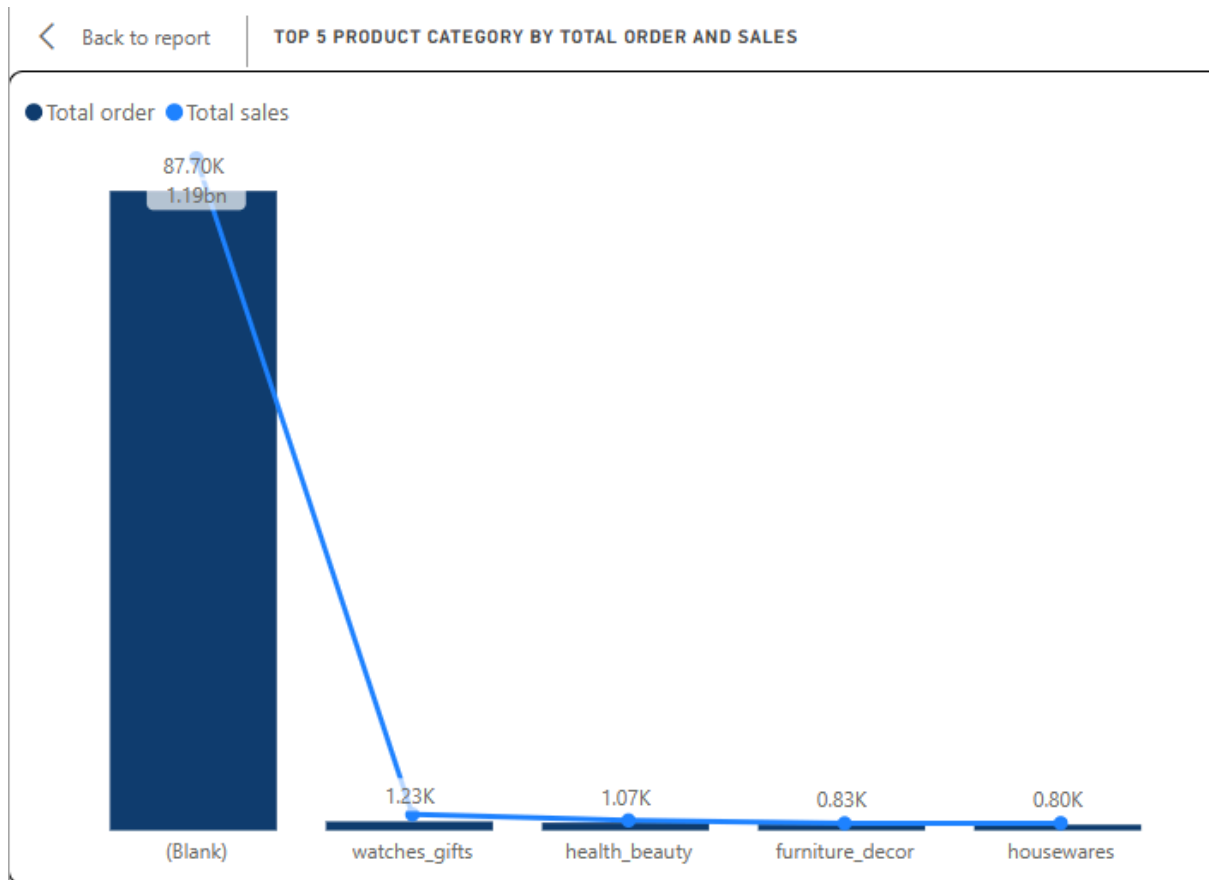
- Năm 2017: Tổng số đơn hàng ít hơn khoảng $\frac{1}{4}$ (Q1) và $\frac{1}{2}$ (Q2) so với năm 2018 nhưng lại có xu hướng tăng từ Q1 - Q4
- Năm 2018: Tổng số đơn hàng có xu hướng giảm từ Q1 - Q3

Chuẩn đoán

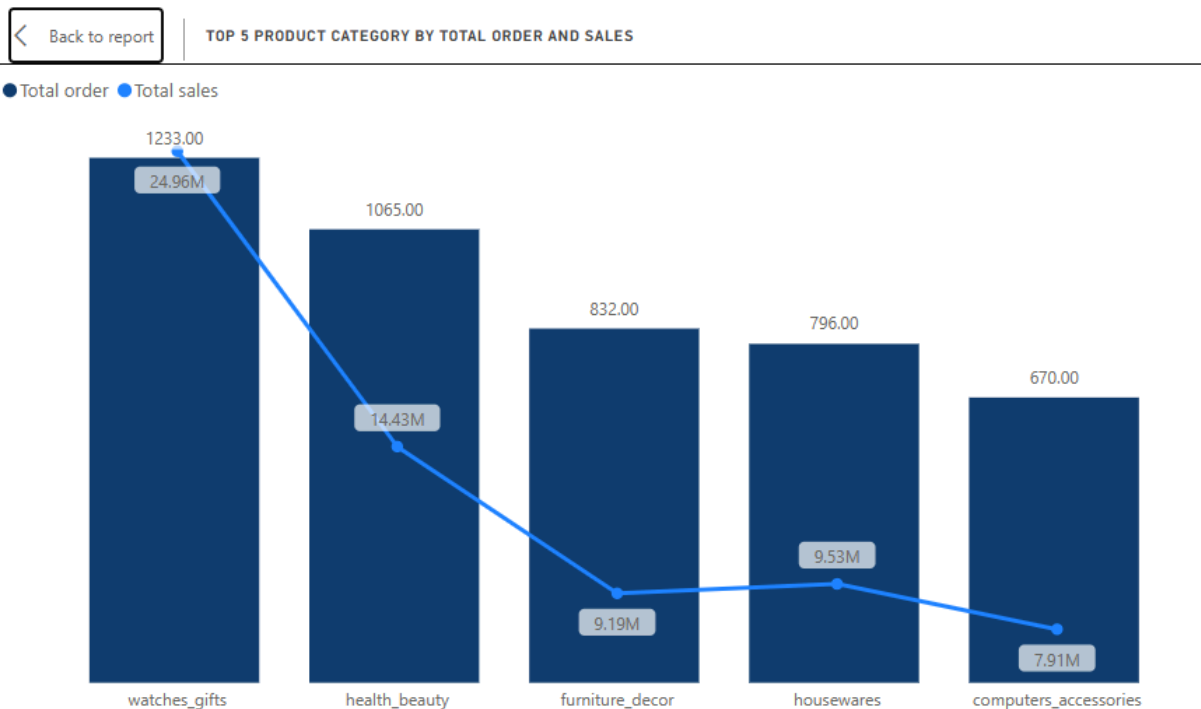
→ Từ Q4/2016 - Q2/2018: Olist có sự tăng trưởng ấn tượng nhờ mở rộng thị trường, chiến lược marketing tốt và sự tin tưởng từ khách hàng.

→ Từ Q3/2018: Doanh thu giảm 38%, có thể do bão hòa thị trường, hiệu quả marketing giảm, yếu tố mùa vụ hoặc cạnh tranh khốc liệt.

❖ Doanh thu theo Sản phẩm



- Dựa vào biểu đồ có thể thấy một lượng đơn hàng các sản phẩm chưa được phân loại chiếm tới 88% tổng lượng đơn hàng (87.7K đơn)



Top 5 doanh mục sản phẩm có số lượng đơn hàng lớn nhất là:

- watches_gifts: 1233 đơn hàng
- health_beauty: 1065 đơn hàng
- furniture_decor: 832 đơn hàng
- housewares: 796 đơn hàng
- computer_accessories: 670 đơn hàng

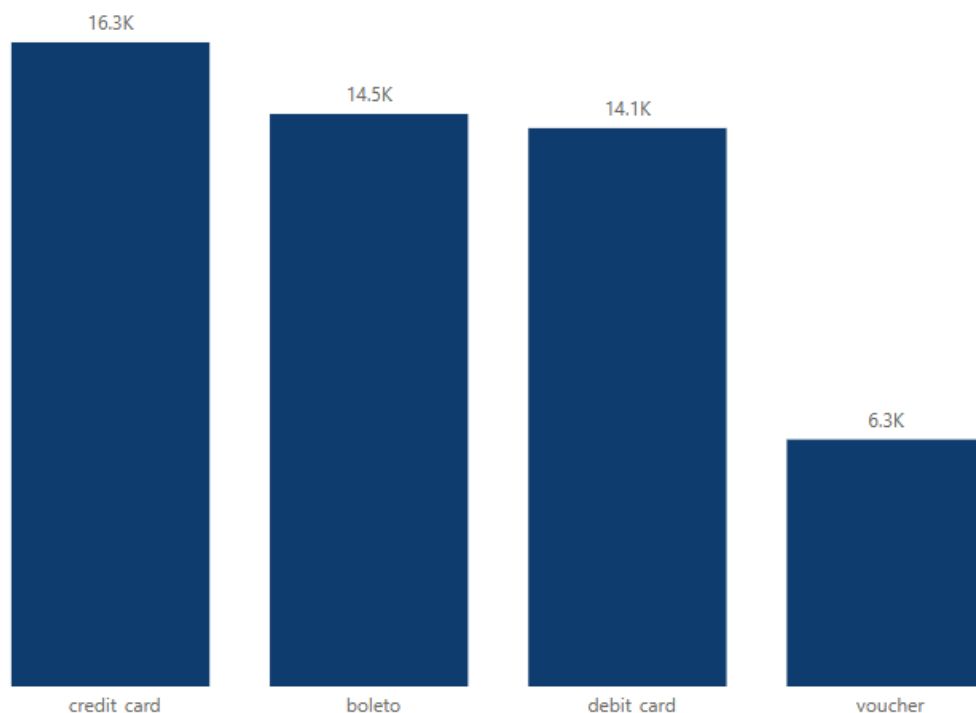
Số lượng đơn hàng không chênh lệch quá lớn giữa các doanh mục sản phẩm thuộc top 5.

Tuy nhiên doanh thu đến từ các doanh mục này có sự chênh lệch đáng kể:

- watches_gifts có doanh thu gấp từ 1.5 - 3 lần các danh mục còn lại
- housewares có số lượng đơn hàng thấp hơn furniture_decor nhưng lại có doanh thu cao hơn 0.34M

→ Do Average order value (AOV) của watches_gifts cao hơn các doanh mục sản phẩm khác và housewares có AOV cao hơn furniture_decor

product_category_name_english	AOV
watches_gifts	19,352.23
health_beauty	12,879.79
computers_accessories	11,016.05
housewares	10,410.78
furniture_decor	9,338.57



Dựa vào biểu đồ:

- Thẻ tín dụng (Credit Card) có AOV cao nhất: 16.3K
- Boleto và thẻ ghi nợ (Debit Card) có mức tương đương nhau: 14.5K và 14.1K
- Voucher có AOV thấp nhất: 6.3K (khoảng 38% so với thẻ tín dụng)

→ Người dùng thẻ tín dụng thường có sức mua cao hơn do có thể chi tiêu trước, trả sau. Bên cạnh đó thẻ tín dụng cũng có nhiều chương trình tích điểm

→ Người dùng thanh toán trực tiếp bằng tiền mặt hoặc tài khoản ngân hàng nên chi tiêu có thể bị giới hạn bởi số dư tài khoản.

→ AOV thấp nhất do các giao dịch bằng voucher thường có giá trị nhỏ, chủ yếu áp dụng cho các chương trình khuyến mãi.

→ **Đề xuất:** Khuyến khích sử dụng voucher cho các đơn hàng có giá trị cao hơn từ đó làm kích cầu và tăng giá trị đơn hàng