

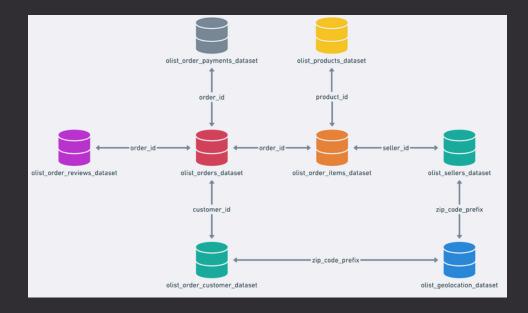
## Giới thiệu



#### FINAL TEST LEVEL 2 - KHOÁ HỌC X-DATA

Sử dụng Brazilian E-Commerce được cung cấp bởi Olist. 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.

Chi tiết về Dataset có thể tham khảo tại đây: <u>Link</u>







Sử dụng bất kỳ công cụ nào mà bạn quen thuộc để EDA bộ dữ liệu trên và thực hiện phân tích, xây dựng các mô hình về Customer Segmentation sau đó trả lời các câu hỏi:

- 1. Nhóm khách hàng Best Customers có đặc điểm như thế nào? Số lượng?
- 2. Đâu là các khách hàng rời bỏ? Số lượng?
- 3. Nhóm khách hàng có khả năng Churned có đặc điểm như thế nào?
- **4.** Trả lời các **Business Questions** và thực hiện **Exploratory Data Analysis** để giúp Olist hiểu rõ hơn về nền tảng thương mại của mình và tối ưu hóa các cơ hội phát triển sẵn có, hãy trả lời các câu hỏi bên dưới:
- a. Tổng doanh thu mà Olist thu được là bao nhiêu và thay đổi như thế nào theo thời gian? (1,5đ)
- **b.** Có bao nhiều đơn đặt hàng được đặt trên Olist và số lượng các đơn hàng thay đổi như thế nào theo tháng hoặc mùa? (1,5đ)
- c. Các danh mục sản phẩm phổ biến nhất trên Olist là gì và doanh số bán hàng của chúng khác nhau như thế nào? (1,5đ)
- **d.** Giá trị Average Order Value (AOV) trên Olist là bao nhiêu và giá trị này thay đổi như thế nào theo Product Category hoặc Payment Method?

Phân tích RFM để phân nhóm khách hàng. Nhóm khách hàng sẽ được phân loại thành 5 nhóm: **Best Customer**, **At Risk**, **New Customer**, **Churn** và **Other**. Như hình bên dưới.

#### Customer segmentation based on RFM Score

Nhóm	Mô tả	Điều kiện RFM		
Best Customers	Mua thường xuyên, gần đây, chi tiêu cao	R=4, F=4, M=4		
At Risk	Trước kia mua nhiều nhưng đã lâu không quay lại	R=1, F=4		
New Customers	Mới mua lần đầu gần đây	R=4, F=1		
Churn	Lâu rồi không quay lại, chi tiêu thấp	R=1, F=1, M=1		
Other	Nhóm khách hàng còn lại	Khác với những điều kiện trên		

Kết quả trả về sau khi phân tích RFM. Dựa vào đây chúng ta có thể trả lời các câu hỏi 1, 2 và 3.

_		RFM Summary by Segment											
Recency		Frequency		Monetary		у	customer_unique_id						
ean	min	max	mean	min	max	mean	min	max	count				
64	1	119	1	1	15	39194	15998	459800	1735				
157	353	699	1	1	6	16620	550	799800	5663				
40	5	729	1	1	9	14554	85	1344000	78641				
63	5	119	1	1	1	14244	440	601986	5868				
61	353	700	1	1	1	2944	490	4890	1496				
6	57 40	14 1 157 353 140 5 13 5	14 1 119 157 353 699 140 5 729 13 5 119	14 1 119 1 157 353 699 1 140 5 729 1 133 5 119 1	14 1 119 1 1 157 353 699 1 1 140 5 729 1 1 133 5 119 1 1	14 1 119 1 1 15 157 353 699 1 1 6 140 5 729 1 1 9 153 5 119 1 1 1	14	14	14 1 119 1 1 15 39194 15998 459800   157 353 699 1 1 6 16620 550 799800   40 5 729 1 1 9 14554 85 1344000   13 5 119 1 1 14244 440 601986				

#### Phân tích từng nhóm

#### **Best Customers**

- \* Recency trung bình = 63.8 → Họ mua hàng \*\*gần đây\*\*.
- \* Tần suất mua = 1.3, cao nhất toàn bảng.
- \* Chi tiêu trung bình = 15,998, rất cao.
- \* Đây là nhóm khách VIP, trung thành. Có 1.735 khách.

#### At Risk (Có khả năng churned)

- \* Recency = 457 ngày, tức là đã rất lâu không mua lại.
- \* Nhưng: Frequency = 1.11, Monetary = 16,620 → Trước đây từng mua nhiều và chi tiêu cao.
- \* Nhóm này rất có giá trị, cần có những chiến lược để kéo họ lại, có tới 5.663 khách.

Kết quả trả về sau khi phân tích RFM. Dựa vào đây chúng ta có thể trả lời các câu hỏi 1, 2 và 3.

#### Other

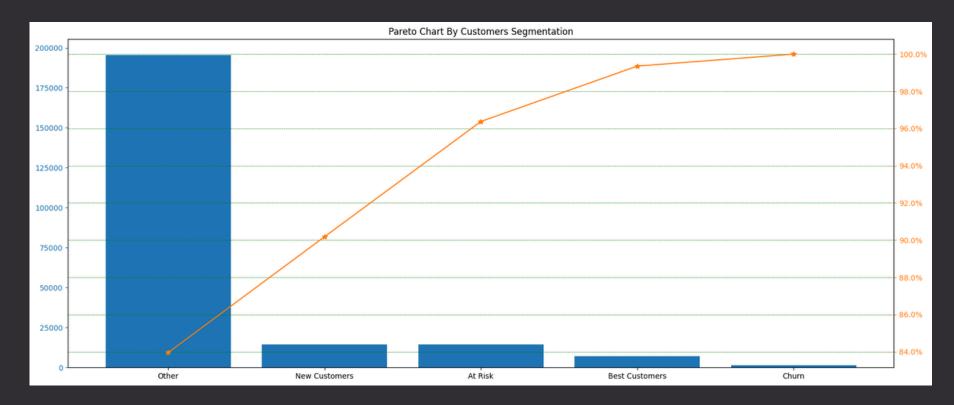
- \* Recency = 240 ngày (trung bình), Frequency = 1.02, Monetary = 14,554.
- \* Nhóm này là trung lập chỉ mua vài lần.
- \* Lớn nhất với 78.641 khách hàng → chiếm phần lớn doanh thu.

#### **New Customers**

- \* Recency thấp = 62.7 ngày → vừa mới mua.
- \* Nhưng Frequency = 1 (mới mua lần đầu), Monetary = 14,244.
- \* Nhóm tiềm năng, cần quan tâm nhiều hơn để biến thành khách trung thành. Có 5.868 khách.

#### Churn (Mất khách)

- \* Recency = 461 ngày → rất lâu không quay lại.
- \* Frequency = 1, Monetary = 2.944 → rất ít giá trị.
- \* Nhóm này có thể đã rời bỏ, khó tái kích hoạt. Có 1.496 khách.



#### Ý nghĩa các thành phần trên biểu đồ

#### 1. Nhóm Other

- Chiếm phần lớn tổng điểm Monetary (M) tức là nhóm này đóng góp doanh thu nhiều nhất cho Olis (84%).
- Đây là nhóm khách hàng trung lập, vãn lai, không thuộc các nhóm đặc biệt như Best Customer, At Risk, ...

#### 2. Các nhóm còn lại (New Customers, At Risk, Best Customers, Churn):

- Đóng góp Monetary nhỏ hơn rất nhiều so với nhóm "Other".
- Tuy nhiên, các nhóm như "Best Customers" và "At Risk" vẫn có vai trò quan trọng. Đặc biệt là nhóm "At Risk" khi mà doanh thu của nhóm này cao hơn nhóm "Best Customer" và nhóm "New Customer".

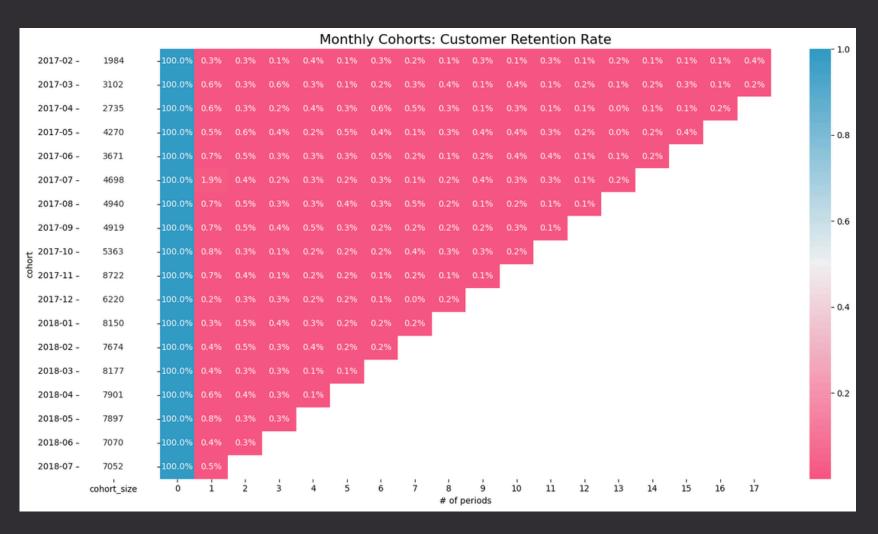
#### 3. Đường cộng dồn phần trăm (cumpercentage):

- Đường này cho thấy chỉ cần nhóm "Other" đã chiếm hơn 80% tổng điểm Monetary.
- Khi cộng thêm các nhóm khác, tỷ lệ này nhanh chóng đạt gần 100%.

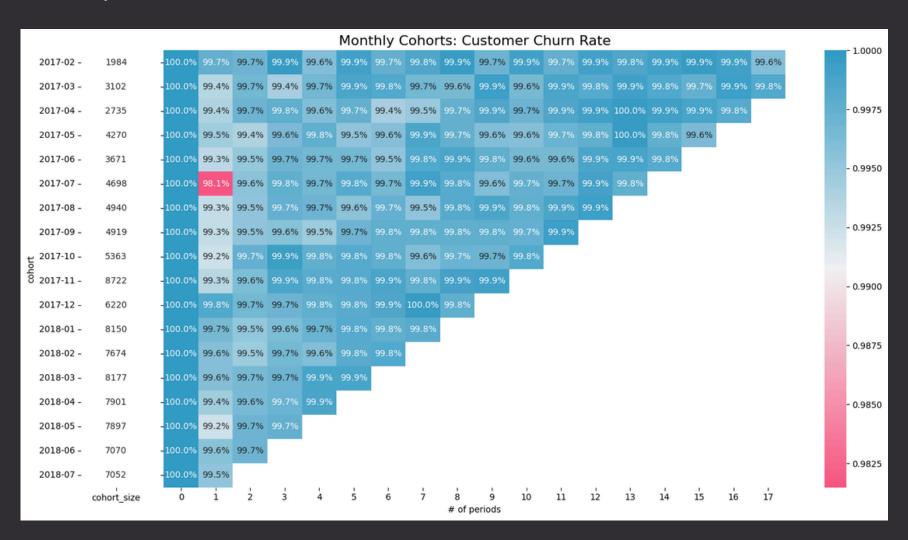
#### Tóm lại:

Biểu đồ cho thấy phần lớn doanh thu đến từ nhóm khách hàng "Other", còn các nhóm đặc biệt (Best, At Risk, New, Churn) chiếm tỷ trọng nhỏ hơn nhưng lại có giá trị chiến lược trong việc chăm sóc khách hàng.

Monthly Cohorts: Customer Retention Rate



Monthly Cohorts: Customer Churn Rate



#### Phân tích và nhận xét Cohort Retention & Churn

#### 1. Customer Retention (Giữ chân khách hàng)

- **Tỷ lệ giữ chân khách hàng cực thấp:** Ở tất cả các cohort, sau tháng đầu tiên, tỷ lệ khách hàng quay lại mua hàng chỉ còn khoảng 0.1% – 0.8%.

Ví dụ: Cohort tháng 2017-07 có 4.698 khách, nhưng chỉ còn 87 khách quay lại ở tháng tiếp theo (~1.9%), các tháng sau giảm mạnh chỉ còn vài chục khách.

- Số lượng khách quay lại giảm rất nhanh:
- Đa số khách hàng chỉ mua 1 lần, rất ít khách hàng trung thành.
- Các cohort mới (2018) cũng có xu hướng tương tự, không có dấu hiệu cải thiện retention.

#### 2. Customer Churn (Mất khách)

- **Tỷ lệ mất khách cực cao**: Sau tháng đầu tiên, gần như toàn bộ khách hàng của mỗi cohort đều rời bỏ, không quay lại mua hàng.

Ví dụ: Cohort 2017-11 có 8.722 khách, nhưng chỉ còn 57 khách quay lại ở tháng tiếp theo, tức là hơn 99% khách đã churn.

- Churn diễn ra ngay sau lần mua đầu: Đa số khách hàng chỉ phát sinh giao dịch 1 lần rồi rời bỏ nền tảng.

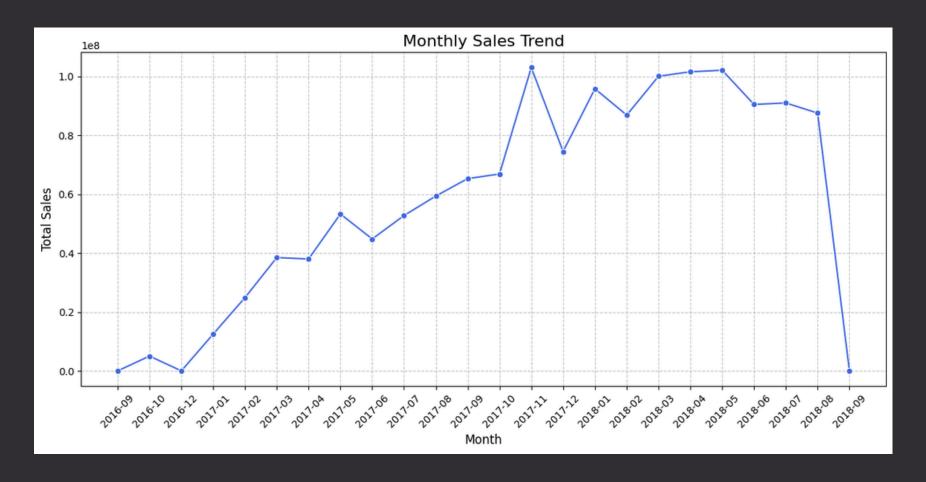
#### 3. Nhận xét tổng quan

- Đặc trưng của Olist: Khách hàng chủ yếu mua 1 lần, rất khó giữ chân.
- Dữ liệu không bị lỗi, mà phản ánh thực tế hành vi khách hàng trên Olist.

#### Tóm lại:

- Tỷ lệ giữ chân khách hàng của Olist rất thấp, tỷ lệ mất khách cực cao.
- Cần có chiến lược chăm sóc khách hàng sau bán, ưu đãi cho khách cũ, và cải thiện trải nghiệm để tăng retention.

Tổng doanh thu mà Olist thu được là bao nhiêu và thay đổi như thế nào theo thời gian () Tổng doanh thu Olist: 1394683022.0



#### Phân tích biểu đồ



#### 1. Xu hướng tăng trưởng:

- Doanh thu tăng mạnh từ cuối năm 2016 đến cuối năm 2017.
- Giai đoạn đầu (2016-09 đến 2017-03) doanh thu còn thấp mặc dù có tăng và sau đó tăng trưởng rõ rệt.
- Từ giữa năm 2017 đến giữa năm 2018, doanh thu duy trì ở mức cao, có nhiều tháng đạt đỉnh.

#### 2. Các đỉnh và đáy:

- Đỉnh doanh thu cao nhất rơi vào khoảng tháng 2017-11 và 2018-03/04/05.
- Một số tháng có biến động mạnh (ví dụ: tháng 2017-11 tăng vọt, sau đó tháng 2017-12 giảm rồi lại tăng trở lại).

#### 3. Giai đoạn ổn định:

- Từ đầu 2018 đến giữa 2018, doanh thu duy trì ổn định ở mức cao (Tuy có giảm ở tháng 2 nhưng sau đó đã tăng trở lại).

#### 4. Sut giảm cuối chuỗi:

- Tháng cuối cùng (2018-09) doanh thu gần như bằng 0, do dữ liệu chưa được cập nhật cho tháng này.

#### Nhận xét tổng quan



- Olist có tốc độ tăng trưởng doanh thu rất tốt trong năm đầu tiên.
- Sau khi đạt đỉnh, doanh thu duy trì ổn định ở mức cao, cho thấy hoạt động kinh doanh đã vào giai đoạn phát triển bền vững.

#### Tóm lại:

Biểu đồ cho thấy Olist phát triển mạnh mẽ, doanh thu tăng nhanh và duy trì ổn định sau khi đạt đỉnh. Đây là dấu hiệu tích cực về tăng trưởng.

Có bao nhiêu đơn đặt hàng được đặt trên Olist và số lượng các đơn hàng thay đổi như thế nào theo tháng hoặc mùa?

Tổng số lượng đơn hàng: 96523



#### Phân tích biểu đồ



#### 1. Xu hướng tăng trưởng:

- Số lượng đơn hàng tăng mạnh từ cuối năm 2016 đến cuối năm 2017.
- Giai đoạn đầu (2016-09 đến 2017-03) số lượng đơn hàng còn thấp mặc dù có tăng nhưng sau đó tăng trưởng rõ rệt.
- Từ giữa năm 2017 đến giữa năm 2018, số đơn hàng duy trì ở mức cao, có nhiều tháng đạt đỉnh.

#### 2. Các đỉnh và đáy:

- Đỉnh cao nhất rơi vào khoảng tháng 2017-11 và 2018-03/04/05 (Điều này cũng giải thích tại sao doanh thu ở các tháng này đạt đỉnh)
- Một số tháng có biến động mạnh (ví dụ: tháng 2017-11 tăng vọt, sau đó tháng 2017-12 giảm rồi lại tăng trở lại).

#### 3. Giai đoan ổn đinh:

- Từ đầu 2018 đến giữa 2018, số đơn hàngduy trì ổn định ở mức cao (Tuy có giảm ở tháng 2 nhưng sau đó đã tăng trở lại).

#### 4. Sụt giảm cuối chuỗi:

- Tháng cuối cùng (2018-09) số đơn hàng gần như bằng 0, do dữ liệu chưa được cập nhật cho tháng này.

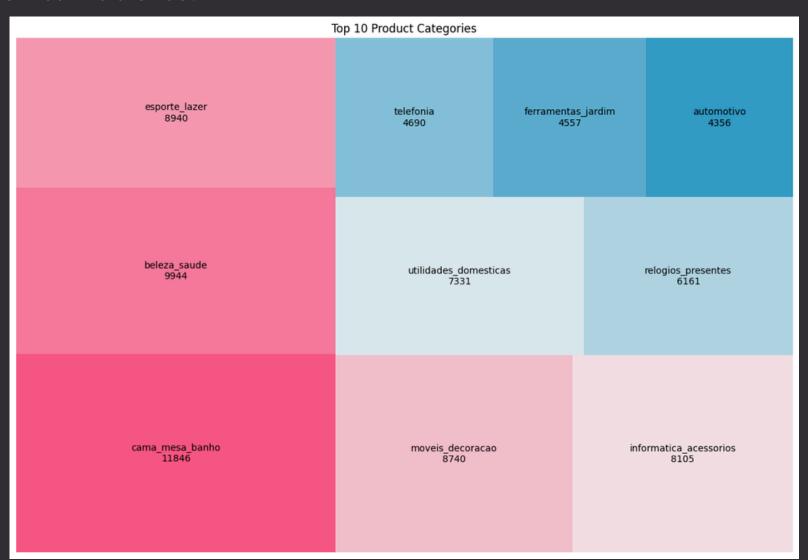
#### Nhận xét tổng quan:

- Olist có tốc độ tăng trưởng đơn hàng rất tốt.
- Sau khi đạt đỉnh, số đơn hàng duy trì ổn định ở mức cao.

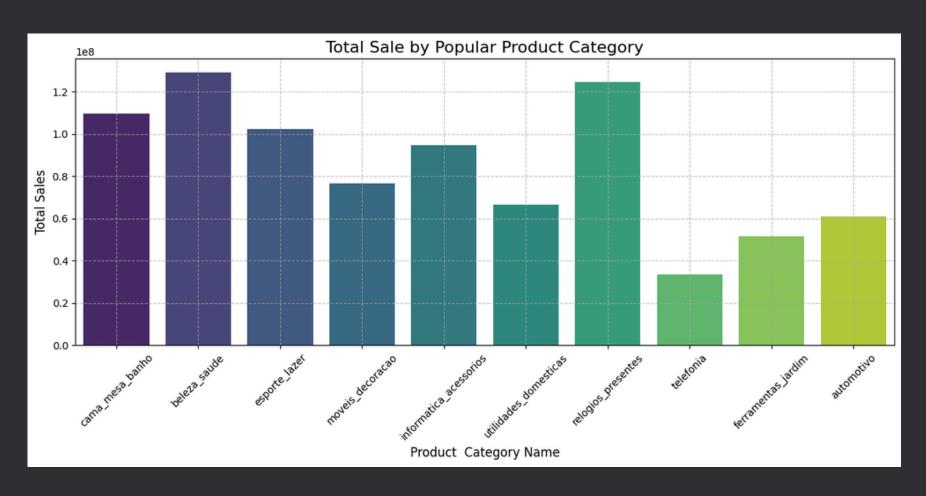
#### Tóm lại:

Biểu đồ cho thấy Olist phát triển mạnh mẽ về số lượng đơn hàng, tăng trưởng nhanh và duy trì ổn định sau khi đạt đỉnh.

Các danh mục sản phẩm phổ biến nhất trên Olist là gì và doanh số bán hàng của chúng L khác nhau như thế nào?



Các danh mục sản phẩm phổ biến nhất trên Olist là gì và doanh số bán hàng của chúng 👉 khác nhau như thế nào?



Các danh mục sản phẩm phổ biến nhất trên Olist là gì và doanh số bán hàng của chúng 🗡

#### Phân tích biểu đồ

- cama\_mesa\_banho có số lượng đơn hàng cao nhất nhưng doanh thu lại thấp hơn beleza\_saude và relogios\_presentes. Điều này cho thấy giá trị trung bình mỗi đơn hàng của danh mục cama\_mesa\_banho thấp hơn các danh mục còn lại.
- beleza\_saude tuy số đơn hàng ít hơn nhưng tổng doanh thu lại cao nhất, chứng tỏ giá trị đơn hàng trung bình của nhóm này rất cao. Đây là danh mục sản phẩm có thể mang lại lợi nhuận lớn dù không cần quá nhiều đơn hàng.
- automotivo cũng tương tự, số đơn hàng không nhiều nhưng doanh thu vẫn cao, cho thấy giá trị sản phẩm lớn.
- Các nhóm như **telefonia** và **ferramentas\_jardim** vừa có số đơn hàng thấp vừa doanh thu thấp.

#### Kết luận:

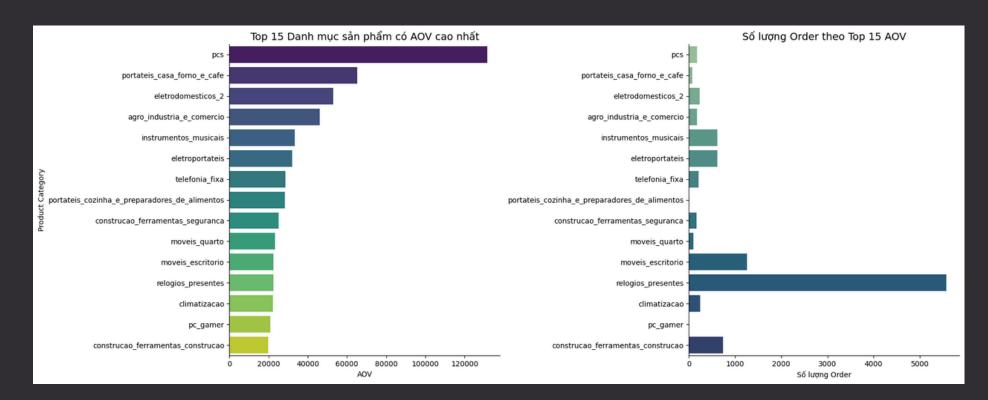
Không phải nhóm sản phẩm nào bán chạy nhất cũng mang lại doanh thu cao nhất. Doanh thu phụ thuộc vào cả số lượng đơn hàng và giá trị trung bình mỗi đơn hàng.

nào theo Product Category hoặc Payment Method?

Giá trị Average Order Value (AOV) trên Olist là bao nhiều và giá trị này thay đổi như thế



Average Order Value: 14449.229945194409



Giá trị Average Order Value (AOV) trên Olist là bao nhiêu và giá trị này thay đổi như thế nào theo Product Category hoặc Payment Method?



#### Insight

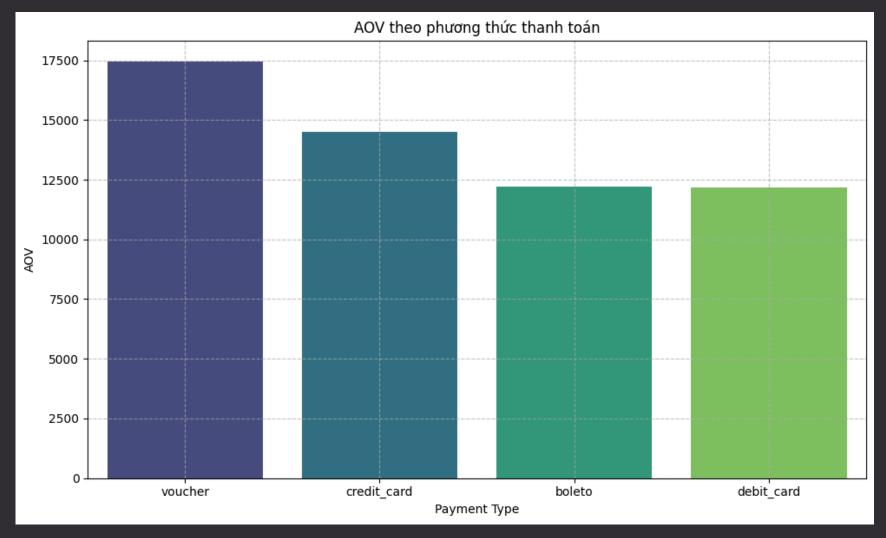
- Các danh mục như **pcs, portateis\_casa\_forno\_e\_cafe, eletrodomesticos\_2** có AOV rất cao (trên 50.000, thậm chí hơn 130.000), nhưng số lượng đơn hàng lại rất thấp. Điều này cho thấy đây là các sản phẩm giá trị lớn, không phổ biến, ít người mua.
- Các danh mục như **relogios\_presentes, moveis\_escritorio** có số lượng đơn hàng cao nhất trong top 15, nhưng AOV lại không cao bằng các nhóm trên. Đây là các sản phẩm vừa với sức mua của nhiều khách hàng.
- Phần lớn các danh mục có AOV cao đều là sản phẩm đặc thù, giá trị lớn, không phải là hàng tiêu dùng phổ biến.

#### Kết luận:

- Sản phẩm càng đặc thù, giá trị càng lớn thì AOV càng cao nhưng số lượng đơn hàng thường thấp.
- Các danh mục phổ biến hơn có AOV thấp hơn nhưng lại có số lượng đơn hàng lớn.



Giá trị Average Order Value (AOV) trên Olist là bao nhiêu và giá trị này thay đổi như thế nào theo Product Category hoặc Payment Method?



Giá trị Average Order Value (AOV) trên Olist là bao nhiêu và giá trị này thay đổi như thế nào theo Product Category hoặc Payment Method?



#### Insight

Giá trị AOV (Average Order Value - Giá trị đơn hàng trung bình) thay đổi rõ rệt theo từng phương thức thanh toán:

- **Voucher** có AOV cao nhất (~17,500), cho thấy các đơn hàng thanh toán bằng voucher thường có giá trị lớn hơn các phương thức khác.
- Credit card đứng thứ hai (~14,500), là phương thức phổ biến và cũng có giá trị đơn hàng trung bình cao.
- Boleto và debit card có AOV thấp hơn (~12,200), cho thấy các đơn hàng thanh toán bằng hai phương thức này thường có giá trị nhỏ hơn.

#### Kết luận:

Khách hàng sử dụng voucher hoặc thẻ tín dụng thường chi tiêu nhiều hơn cho mỗi đơn hàng so với các phương thức khác.

### **Word Cloud of Review Comments**



### **Word Cloud of Review Comments**

#### **Phân tích Word Cloud Review Comments:**

- Các từ xuất hiện lớn nhất là: produto (sản phẩm), não recebi (không nhận được), foi entregue (đã giao hàng), prazo (thời hạn/giao hàng), muito bom (rất tốt), chegou antes (đến sớm), bem embalado (đóng gói tốt).
- Nhiều từ liên quan đến trải nghiệm giao hàng: entregue, prazo, chegou, antes, entrega rápida (giao hàng nhanh).
- Một số từ thể hiện phản hồi tích cực: muito bom, ótimo produto (sản phẩm rất tốt), bem embalado.
- Tuy nhiên, cũng có nhiều từ thể hiện phản hồi tiêu cực: não recebi (không nhận được), não foi entregue (không được giao), ainda não (vẫn chưa).

#### Nhận xét:

- Chất lượng sản phẩm và dịch vụ giao hàng được khách hàng nhắc đến nhiều nhất.
- Đa số khách hàng hài lòng với sản phẩm (muito bom, ótimo produto), nhưng vẫn còn nhiều trường hợp khách hàng phàn nàn về việc không nhận được hàng hoặc giao hàng chậm (não recebi, não foi entregue).
- Các phản hồi tích cực về đóng gói và giao hàng nhanh cũng xuất hiện, cho thấy một phần khách hàng hài lòng với dịch vụ.
- Do đó, Olist nên tiếp tục duy trì chất lượng sản phẩm và cải thiện quy trình giao hàng để giảm các phản hồi tiêu cực về việc không nhận được hàng.



## Kết Luận

Dựa trên các phân tích dữ liệu vừa rồi có thể đề xuất một số hướng cải thiện cho Olist như sau:

#### 1. Cải thiện quy trình giao hàng

- Nhiều khách hàng phàn nàn về việc **không nhận được hàng** hoặc **giao hàng chậm** ("não recebi", "não foi entregue", "prazo").
- Olist nên hợp tác chặt chẽ hơn với các đối tác vận chuyển, tăng cường kiểm soát tiến độ giao hàng, cập nhật trạng thái đơn hàng minh bạch cho khách hàng.

#### 2. Tăng cường chăm sóc nhóm khách hàng giá trị cao

- Nhóm **Best Customers** và **At Risk** có giá trị chi tiêu cao, cần có chương trình chăm sóc riêng như ưu đãi, voucher, nhắc nhở quay lại mua hàng, khảo sát ý kiến để giữ chân họ.

#### 3. Chú trọng nhóm khách hàng mới

- Nhóm **New Customers** là tiềm năng, nên có các chương trình chào mừng, giảm giá lần đầu, hướng dẫn sử dụng dịch vụ để tăng trải nghiệm và tỷ lệ quay lại.

#### 4. Kiểm soát chất lượng sản phẩm

- Đảm bảo sản phẩm đúng mô tả, đóng gói tốt ("bem embalado"), tránh hàng lỗi, hàng không đúng cam kết để giảm phản hồi tiêu cực.

# Kết Luận

#### 5. Tối ưu danh mục sản phẩm

- Các danh mục có **AOV cao** nhưng ít đơn hàng nên được quảng bá đúng đối tượng, còn các danh mục phổ biến nên tập trung vào tăng số lượng đơn hàng và tối ưu giá trị trung bình.

#### 6. Khuyến khích thanh toán qua các phương thức có AOV cao

- Đẩy mạnh các chương trình ưu đãi cho khách hàng thanh toán bằng voucher hoặc credit card để tăng giá trị đơn hàng trung bình.

#### 7. Phản hồi nhanh các đánh giá tiêu cực

- Chủ động liên hệ, giải quyết các trường hợp khách hàng phản ánh không nhận được hàng hoặc không hài lòng để tăng sự tin tưởng và giữ chân khách.

#### Tóm lại:

Olist nên tập trung vào nâng cao trải nghiệm giao hàng, chăm sóc khách hàng giá trị, kiểm soát chất lượng sản phẩm, tối ưu hóa các danh mục để nhằm mục đích tăng trưởng bền vững và giữ chân được khách hàng. Việc tăng **retention** dù chỉ 1–2% cũng sẽ tạo ra sự khác biệt lớn về doanh thu lâu dài.

