

Olist Data Analysis: Exploratory data analysis and customer segmentation

HV. Nguyễn Hồng Nhung

Overviews

Tổng số đơn hàng: 99k với doanh thu 13,59M.

2016 → giữa 2017:

Doanh thu **rất thấp và tăng chậm**, gần như giai đoạn khởi động.

Giữa 2017 → đầu 2018:

Doanh thu **tăng đều và rõ rệt**, chứng tỏ business bắt đầu scale.

Đầu 2018 → giữa 2018:

Doanh thu **dao động mạnh hơn nhưng xu hướng vẫn tăng**, mức trung bình cao hơn hẳn 2017.

=> DN đang **tăng trưởng tốt theo thời gian**

Revenue **bắt đầu scale mạnh từ 2017**

Có **biến động lớn** → phụ thuộc **campaign/seasonality** (cuối 2017, đầu 2018)

13.59M

Revenue

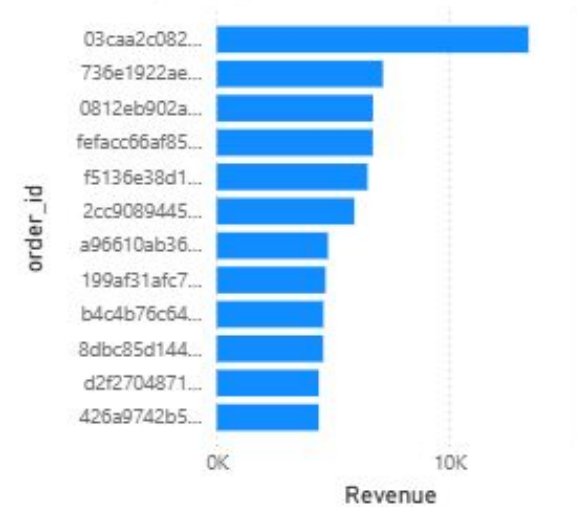
99K

Total Orders

99K

Customers

Revenue by order_id



Revenue by Year, Quarter, Month and Day



1. Nhóm khách hàng Best Customers có đặc điểm như thế nào? Số lượng?

Nhóm Best Customers (10324 chiếm khoảng 10.7% tổng số khách hàng).

Tuy nhiên, phần lớn khách trong nhóm này chỉ mua hàng một lần (frequency \approx 1), nhưng có giá trị chi tiêu cao (monetary trung bình \sim 334).

Điều này cho thấy nhóm Best Customers chủ yếu là những khách hàng có giá trị đơn hàng lớn, thay vì mua hàng thường xuyên.

Ngoài ra, recency trung bình khoảng 5–6 tháng cho thấy một phần khách hàng trong nhóm này chưa quay lại gần đây, tiềm ẩn nguy cơ churn.

Như vậy, Best Customers không phải là khách trung thành, mà là khách có giá trị đơn hàng cao.

Số lượng: 10324
Tỷ lệ: 0.10743423243423243

```
#Đặc điểm của best_customers  
best_customers[['recency', 'frequency', 'monetary']].describe()
```

	recency	frequency	monetary
count	10324.00000	10324.00000	10324.00000
mean	164.62563	1.204572	334.621589
std	82.15620	0.513496	333.422210
min	0.00000	1.000000	86.730000
25%	95.00000	1.000000	166.347500
50%	151.00000	1.000000	235.795000
75%	221.00000	1.000000	363.702500
max	397.00000	17.000000	7274.880000

2. Đây là các khách hàng rời bỏ? Số lượng?

Churn Customers có các đặc điểm:

- Lâu rồi chưa mua
- Mua ít
- Chi tiêu thấp

Số lượng Churn Customers: 11.474 khách hàng, chiếm khoảng 11,94% tổng số KH theo data thu thập được

```
#Churn rate
#Theo RFM: Churn customers = Recency cao (lâu rồi chưa mua) + Frequency thấp (mua ít) + Monetary thấp (chi tiêu thấp)
churn_customers = rfm[
    (rfm['R_score'] <= 2) &
    (rfm['F_score'] <= 2) &
    (rfm['M_score'] <= 2)
]
```

```
#Số lượng khách rời bỏ
num_churn = churn_customers.shape[0]
total = rfm.shape[0]

print("Số lượng churn:", num_churn)
print("Tỷ lệ:", num_churn / total)
```

```
Số lượng churn: 11474
Tỷ lệ: 0.1194014319014319
```

3. Nhóm khách hàng có khả năng Churned có đặc điểm như thế nào ?

Nhóm khách hàng rời bỏ chiếm khoảng 11.94% tổng số khách hàng. Đáng chú ý, 100% khách hàng trong nhóm này chỉ thực hiện một lần mua duy nhất (frequency = 1), cho thấy họ không quay lại sau lần mua đầu tiên.

Ngoài ra, số tiền chi tiêu của nhóm này tương đối thấp (monetary trung bình ~54), và thời gian kể từ lần mua gần nhất khá dài (recency trung bình ~7 tháng).

Điều này cho thấy nhóm churn customers chủ yếu là những khách hàng mới, có giá trị thấp và không được chuyển đổi thành khách hàng trung thành.

Như vậy, vấn đề chính của doanh nghiệp không phải là mất khách hàng trung thành, mà là không giữ chân được khách hàng sau lần mua đầu tiên cao.

```
#Đặc điểm của churn customers  
churn_customers[['recency', 'frequency', 'monetary']].describe()
```

	recency	frequency	monetary
count	11474.000000	11474.0	11474.000000
mean	218.244553	1.0	53.882634
std	96.475125	0.0	18.766410
min	36.000000	1.0	0.000000
25%	137.000000	1.0	38.935000
50%	220.000000	1.0	54.735000
75%	304.000000	1.0	69.040000
max	397.000000	1.0	86.720000

4. Business Questions và thực hiện Exploratory Data

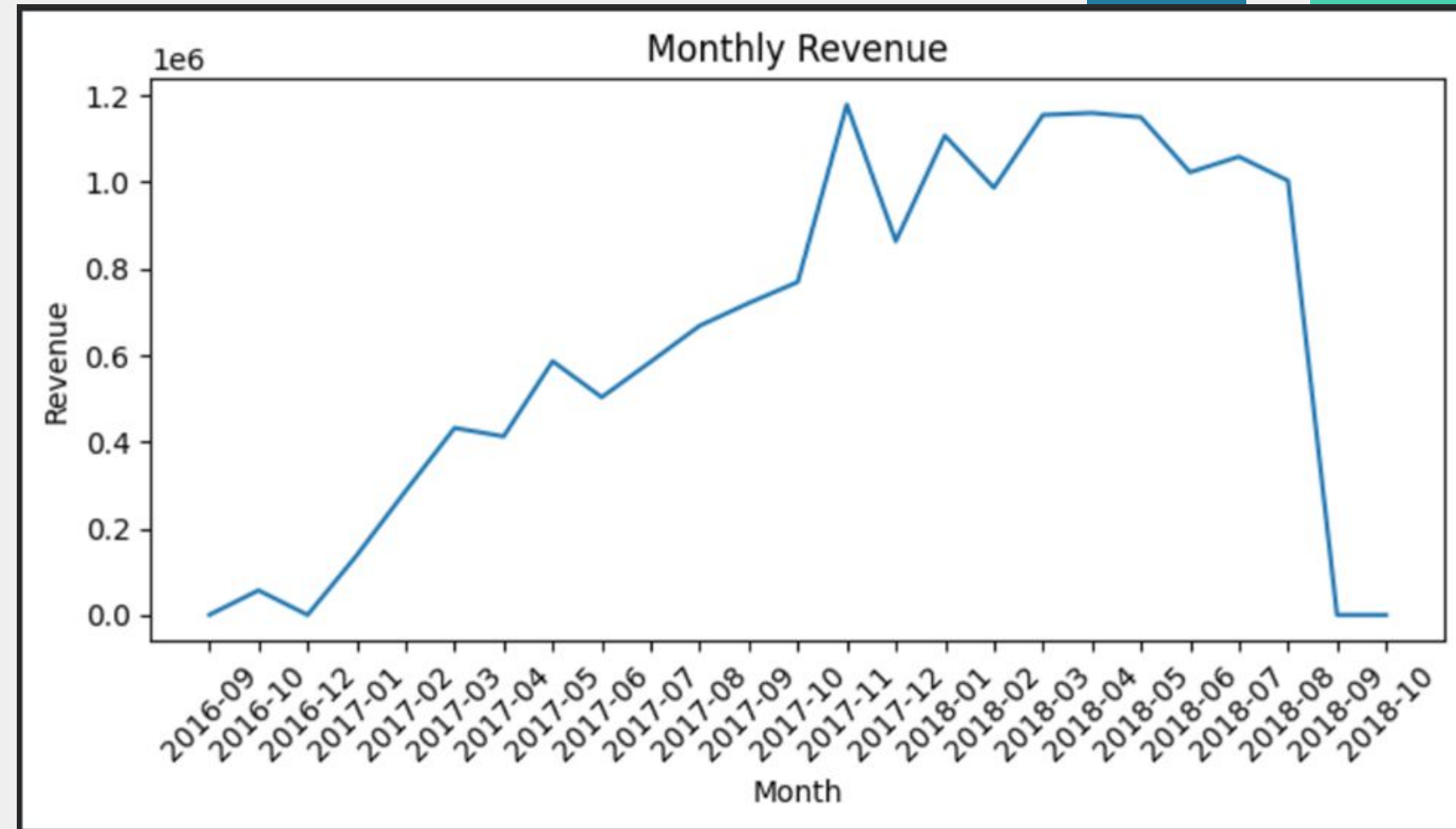
a. Tổng doanh thu và xu hướng theo thời gian

- Tổng doanh thu: \$ 15,8k
- Doanh thu của Olist có xu hướng tăng trưởng mạnh từ năm 2016 đến 2018, đặc biệt tăng nhanh trong năm 2017.

Doanh thu đạt đỉnh vào cuối năm 2017 và đầu năm 2018, cho thấy khả năng có ảnh hưởng từ các mùa mua sắm cao điểm.

Tuy nhiên, doanh thu có sự biến động giữa các tháng, phản ánh yếu tố mùa vụ trong hành vi tiêu dùng.

Ngoài ra, dữ liệu cho thấy doanh thu giảm về 0 trong các tháng cuối năm 2018, nhiều khả năng do dữ liệu chưa đầy đủ.



4. Business Questions và thực hiện Exploratory Data

b. 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 đơn hàng: 99.441
- Số lượng đơn hàng trên Olist tăng trưởng mạnh từ năm 2016 đến 2017, đặc biệt bùng nổ trong năm 2017.

Đỉnh điểm số đơn hàng đạt vào tháng 11/2017, có thể liên quan đến các sự kiện mua sắm lớn như Black Friday.

Trong năm 2018, số lượng đơn hàng duy trì ổn định ở mức cao, cho thấy DN đã đạt được sự phát triển bền vững.

Ngoài ra, dữ liệu cho thấy sự sụt giảm về 0 trong các tháng cuối năm 2018, nhiều khả năng do dữ liệu chưa đầy đủ



4. Business Questions và thực hiện Exploratory Data

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?

- Top category theo doanh thu: Top 3:

1. beleza_saude (Beauty & Health) ~1.45M
2. relorios_presentes (Watches & Gifts) ~1.3M
3. cama_mesa_banho (Bed, Bath & Table) ~1.25M

=> Các danh mục liên quan đến chăm sóc cá nhân, quà tặng và gia dụng chiếm doanh thu cao nhất trên nền tảng.

- Nhóm category chính có thể chia thành 3 nhóm:

1. Nhóm Lifestyle / tiêu dùng cá nhân: beleza_saude; relorios_presentes; cama_mesa_banho

=> Đây là nhóm nhu cầu thiết yếu và tiêu dùng thường xuyên

2. Nhóm giải trí / thể thao: esporte_lazer

3. Nhóm công nghệ & nhà cửa: informatica_acessorios; moveis_decoracao; utilidades_domesticas.

- Sự chênh lệch doanh thu theo biểu đồ cho thấy

Doanh thu giữa các danh mục có sự chênh lệch đáng kể (~2.5 lần giữa cao nhất 1,45m, ở giữa 0,6m), cho thấy thị trường tập trung vào một số ngành hàng chủ lực

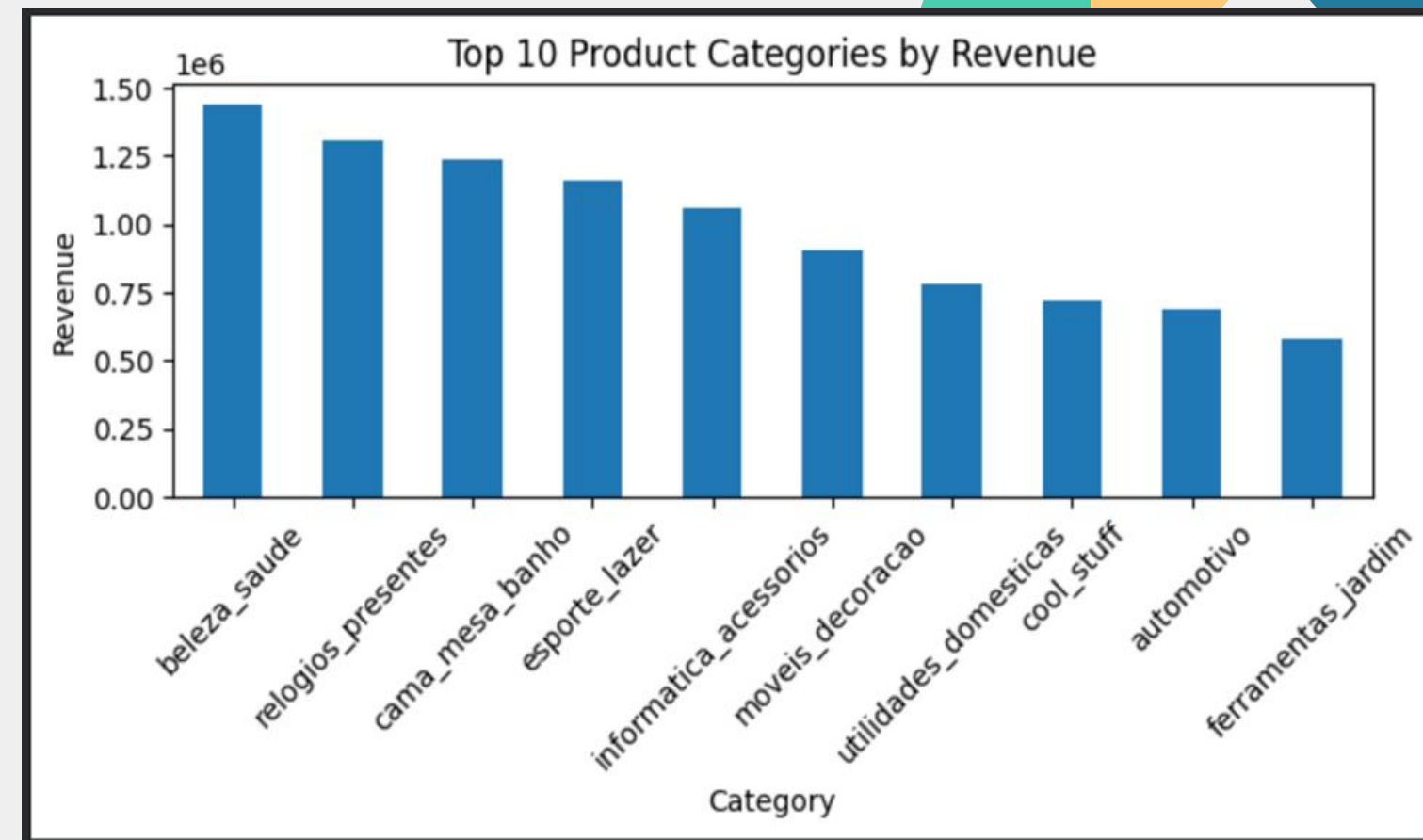
- Top 10 Category đã chiếm phần lớn doanh thu.

- Các danh mục có doanh thu cao chủ yếu là sản phẩm tiêu dùng phổ biến, cho thấy hành vi mua hàng của khách hàng tập trung vào các nhu cầu thiết yếu và quà tặng

==> Các danh mục sản phẩm phổ biến nhất trên Olist bao gồm Beauty & Health, Watches & Gifts, và Bed, Bath & Table.

Những danh mục này đóng góp doanh thu cao nhất, phản ánh nhu cầu lớn đối với các sản phẩm tiêu dùng cá nhân và gia dụng.

Ngoài ra, doanh thu giữa các danh mục có sự chênh lệch đáng kể, cho thấy thị trường tập trung vào một số ngành hàng chủ lực, trong khi các danh mục khác đóng góp ít hơn.



4. Business Questions và thực hiện Exploratory Data

d. Giá trị AOV trên Olist là bao nhiêu và giá trị này thay đổi như thế nào theo Product Category hoặc Payment Methods?

- AOV: 159,3

1. AOV theo category có sự chênh lệch lớn:

- computers ~1300+ (cao vượt trội)

- Nhóm tiếp theo: ~600 → ~300

- Nhóm thấp: ~250–300

-> AOV khác biệt rất lớn giữa các danh mục sản phẩm, đặc biệt nhóm công nghệ có giá trị đơn hàng cao vượt trội so với các nhóm còn lại.

2. Nhóm giá trị cao (High AOV). Top:

- computers (~1300+)

- small appliances (oven, coffee...)

- home appliances

=> Các sản phẩm điện tử và gia dụng lớn có giá trị cao → kéo AOV lên mạnh.

3. Nhóm trung bình

- musical instruments

- small appliances

- telephony

=> Nhóm này có giá trị đơn hàng trung bình, ổn định

4. Nhóm thấp

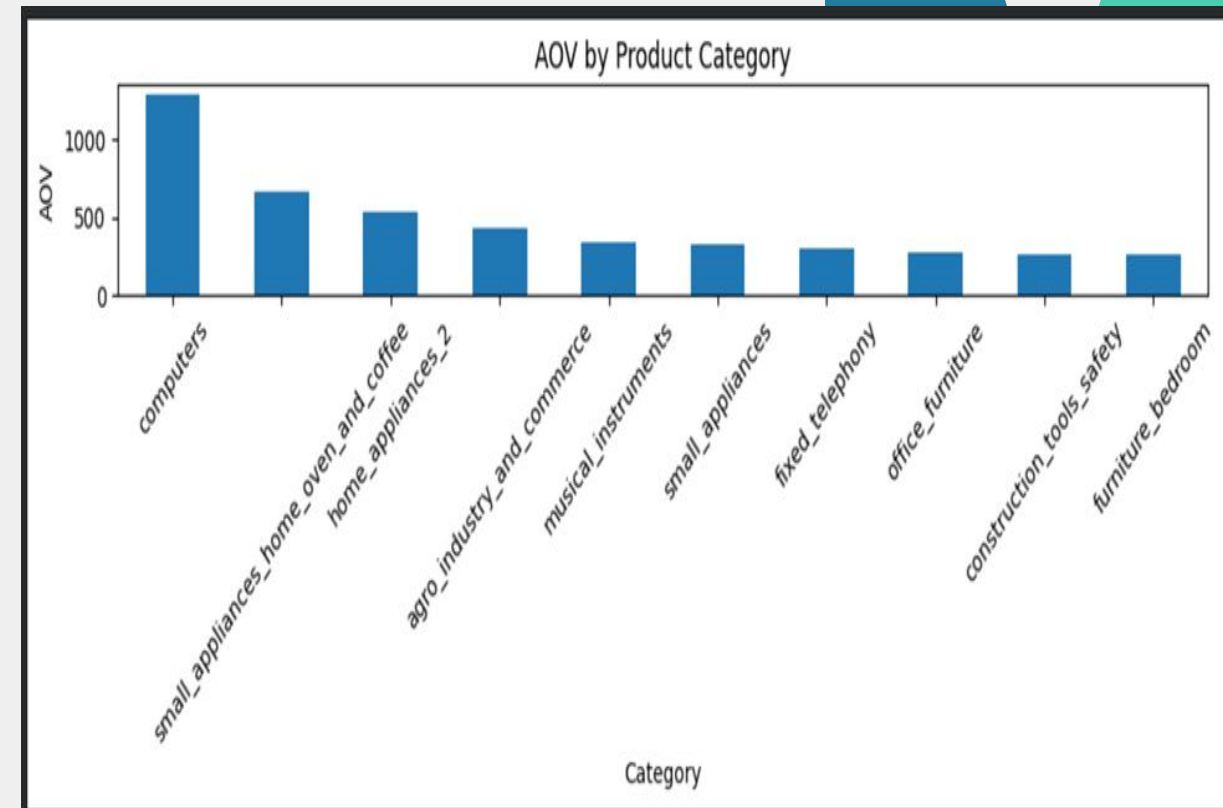
- furniture

- tools

- office

=> Đây là các sản phẩm giá thấp hoặc mua lẻ → AOV thấp hơn.

==> AOV không chỉ phụ thuộc vào hành vi khách hàng mà còn phụ thuộc mạnh vào loại sản phẩm, trong đó các danh mục công nghệ và thiết bị điện tử có xu hướng tạo ra đơn hàng giá trị cao hơn đáng kể.



4. Business Questions và thực hiện Exploratory Data

AOV theo Payment Methods

- Credit card có AOV cao nhất

-> Khách thanh toán bằng credit card có xu hướng chi tiêu nhiều hơn. Có thể do trả góp không bị giới hạn tiền ngay lập tức

- Voucher có AOV thấp nhất: Khách dùng voucher thường có giá trị đơn hàng thấp hơn.

Có thể đây là kết quả săn deal, mua sản phẩm giá rẻ

- Các phương thức còn lại khá tương đồng: boleto & debit_card ~140–145 -> không chênh lệch nhiều

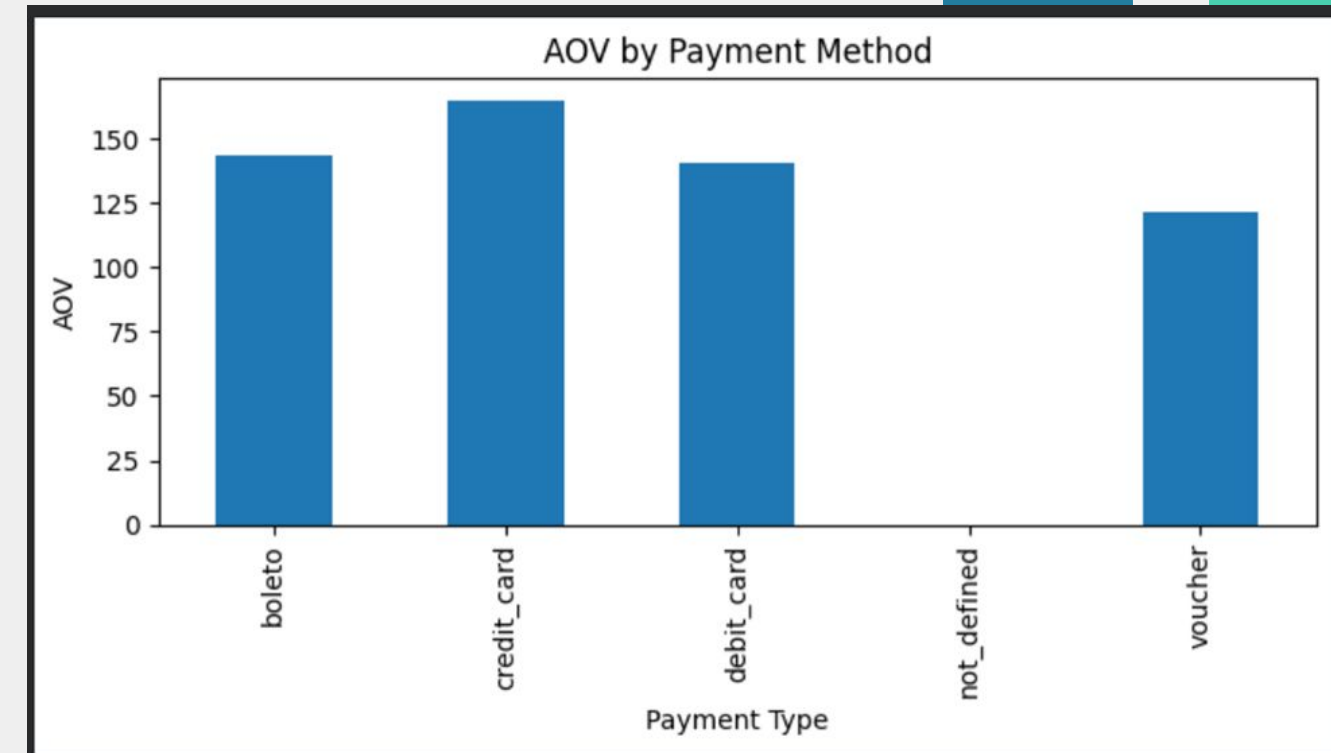
=> Các phương thức thanh toán truyền thống có mức chi tiêu tương đối ổn định.

==> Giá trị đơn hàng trung bình (AOV) khác nhau đáng kể giữa các phương thức thanh toán.

Cụ thể, khách hàng sử dụng credit card có AOV cao nhất (khoảng 165), cho thấy họ có xu hướng chi tiêu nhiều hơn.

Ngược lại, các đơn hàng sử dụng voucher có AOV thấp nhất (khoảng 120), phản ánh hành vi săn khuyến mãi hoặc mua các sản phẩm giá thấp.

Các phương thức thanh toán khác như boleto và debit card có mức AOV trung bình và tương đối ổn định.



SUMMARY

Thông qua phân tích dữ liệu thăm dò và phân khúc khách hàng, một số thông tin chi tiết quan trọng về nền tảng thương mại điện tử của Olist đã được xác định.

- Nền tảng này đã chứng tỏ sự tăng trưởng mạnh mẽ cả về doanh thu và số lượng đơn đặt hàng theo thời gian, đặc biệt là trong năm 2017 và đầu năm 2018. Điều này cho thấy việc mở rộng thành công và tăng cường sự chấp nhận của khách hàng. Tuy nhiên, sự hiện diện của dữ liệu không đầy đủ trong những tháng sau đó cho thấy cần phải xác thực dữ liệu một cách cẩn thận khi diễn giải các xu hướng gần đây.
- Doanh thu tập trung ở một số danh mục sản phẩm chính như Sắc đẹp & Sức khỏe, Đồng hồ & Quà tặng và các sản phẩm liên quan đến Gia dụng. Những danh mục này phản ánh mức tiêu dùng thiết yếu và theo định hướng lối sống, cho thấy sức mạnh cốt lõi của Olist nằm ở nhu cầu tiêu dùng hàng ngày. Đồng thời, việc phân bổ doanh thu cho thấy một mô hình dài hạn rõ ràng, trong đó một số ít danh mục đóng góp không tương xứng vào tổng doanh số bán hàng.
- Phân tích hành vi khách hàng cho thấy một thách thức kinh doanh quan trọng. Trong khi một bộ phận nhỏ khách hàng có giá trị cao (Best Customer) đóng góp đáng kể vào doanh thu thì một bộ phận lớn khách hàng chỉ mua một lần và không quay lại. Điều này cho thấy vấn đề chính không phải là mất đi khách hàng trung thành mà là không chuyển đổi được người mua lần đầu thành khách hàng thường xuyên.
- Giá trị đặt hàng trung bình (AOV) thay đổi đáng kể trên cả phương thức thanh toán và danh mục sản phẩm. Khách hàng sử dụng thẻ tín dụng có xu hướng có AOV cao hơn, trong khi người dùng phiếu thưởng thường tạo ra các đơn hàng có giá trị thấp hơn. Ngoài ra, các danh mục sản phẩm như máy tính và đồ điện tử có AOV cao hơn nhiều so với các danh mục như đồ nội thất hoặc dụng cụ. Điều này nêu bật tác động kép của cả hành vi thanh toán và loại sản phẩm lên giá trị giao dịch.

Nhìn chung, phân tích cho thấy Olist là một nền tảng đang phát triển với nhu cầu mạnh mẽ ở các danh mục chính nhưng vẫn phải đối mặt với những thách thức trong việc giữ chân khách hàng và tối ưu hóa giá trị. Doanh nghiệp cần giải quyết những vấn đề quan trọng này để duy trì tăng trưởng dài hạn và tối đa hóa hiệu quả kinh doanh.

The image features a light gray background with decorative geometric patterns in the corners. The top-left corner has a series of parallel teal lines radiating from a point. The top-right corner contains several overlapping semi-circles in teal, orange, and red. The bottom-left corner features overlapping semi-circles in red, teal, and dark blue. The bottom-right corner has a teal arc with parallel teal lines radiating from its center.

THANK YOU