



BI46_Final Project_Group 1

Brazilian E-commerce





Olist là một cửa hàng bách hóa lớn nhất ở các khu chợ tại Brazil giúp kết nối các doanh nghiệp nhỏ từ khắp Brazil với các kênh phân phối mà không gặp rắc rối.

Ngoài ra, những người bán có thể bán và phân phối sản phẩm của mình thông qua cửa hàng Olist và vận chuyển trực tiếp đến khách hàng qua các đối tác hậu cần của Olist.



Bối cảnh phân tích

Quản lý của công ty đang yêu cầu nhóm phân tích đánh giá hiệu quả kinh doanh của doanh nghiệp thông qua bộ dữ liệu để cung cấp một số phát hiện/quan điểm tiếp thị dưới dạng báo cáo về tình hình kinh doanh của công ty trong năm 2017 & 8 tháng đầu năm 2018.

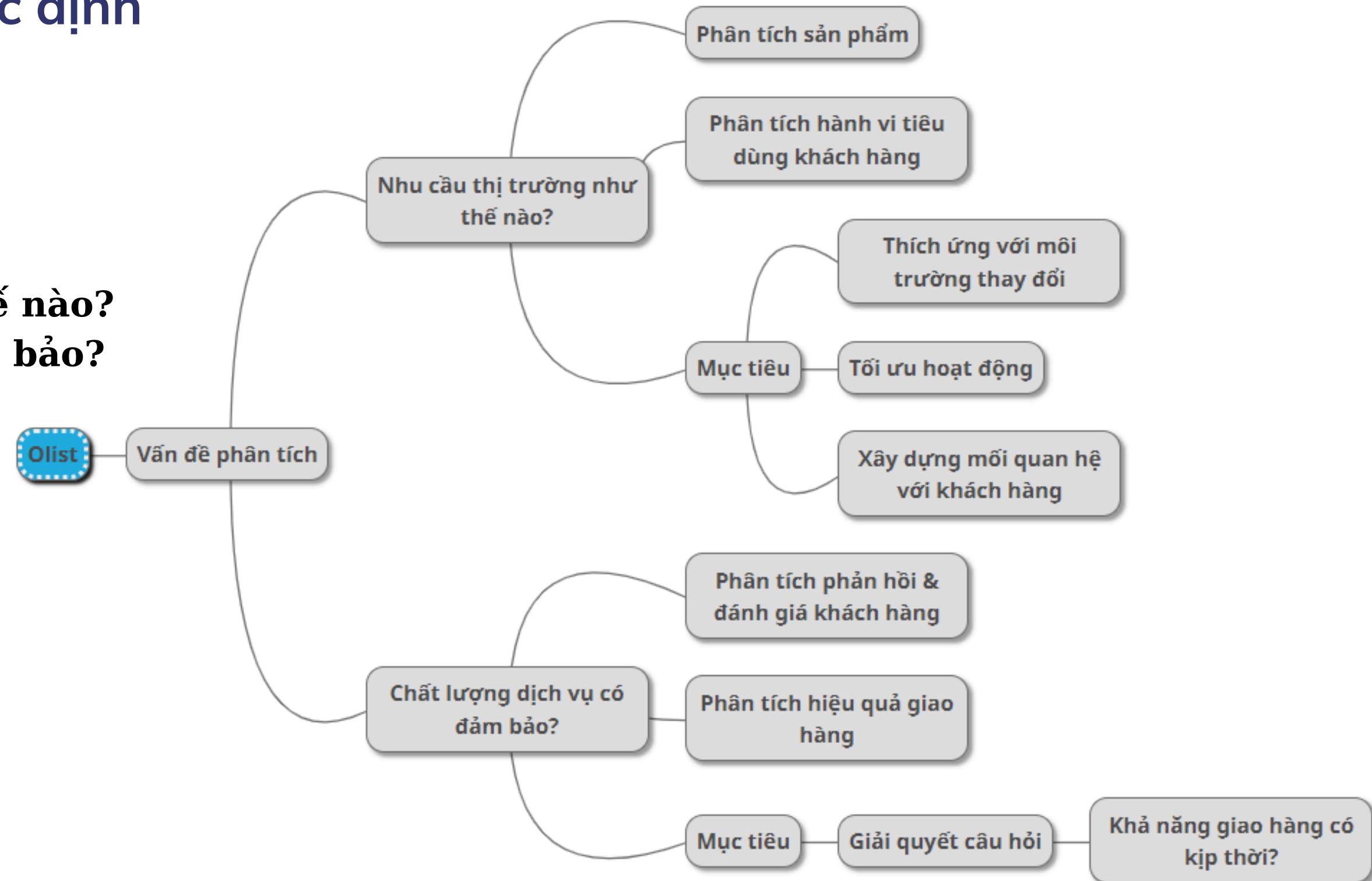
Thành viên Nhóm 1:
Hà Thúc Minh
Trần Đình Duy
Trần Nhật Khanh
Nguyễn Thị Thu Trang

Xác định vấn đề và xác định mục tiêu phân tích

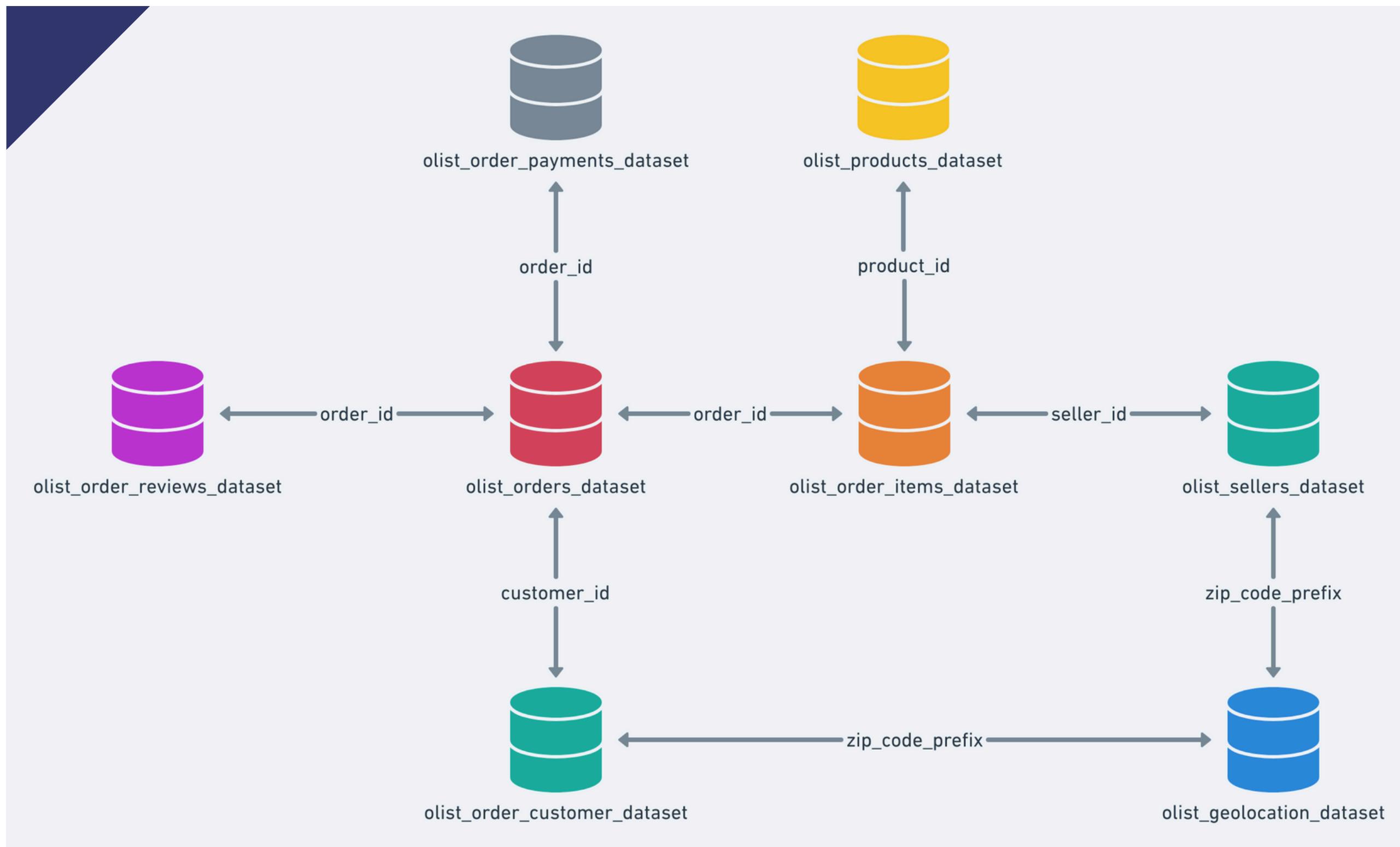
1. Xác định vấn đề

- **Nhu cầu thị trường như thế nào?**
- **Chất lượng dịch vụ có đảm bảo?**

2. Xác định định mục tiêu phân tích



Data Model/Relation



Mô tả dữ liệu (Data dictionary):

olist_customer_dataset	
customer_id	mã khách hàng. Mỗi đơn hàng có một customer_id duy nhất.
customer_unique_id	mã định danh duy nhất của một khách hàng.
customer_zip_code_prefix	năm chữ số đầu tiên của mã zip của khách hàng
customer_city	tên thành phố
customer_state	mã khu vực

olist_geolocation_dataset	
geolocation_zip_code_prefix	năm chữ số đầu tiên của mã zip của vùng
geolocation_lat	vĩ độ
geolocation_lng	kinh độ
geolocation_city	tên thành phố
geolocation_state	mã khu vực

olist_order_payments_dataset	
order_id	mã đơn hàng
payment_sequential	số phương thức thanh toán sử dụng
payment_type	thức thanh toán
payment_installments	thời gian thanh toán
payment_value	khoản thanh toán

Mô tả dữ liệu (Data dictionary):

olist_order_items_dataset	
order_id	mã đơn hàng
order_item_id	mã thứ tự của đơn hàng
product_id	mã sản phẩm
seller_id	mã nhân viên bán hàng
shipping_limit_date	giới hạn thời gian vận chuyển của người bán để xử lý đơn hàng cho đối tác hậu cần
price	giá sản phẩm
freight_value	giá trị vận chuyển

olist_order_reviews_dataset	
review_id	mã phản hồi
order_id	mã đơn hàng
review_score	điểm đánh giá (từ 1 - 5 sao)
review_comment_title	tiêu đề đánh giá
review_comment_message	phản hồi đánh giá
review_creation_date	thời gian gửi bảng khảo sát
review_answer_timestamp	thời gian trả lời khảo sát

olist_orders_dataset	
order_id	mã đơn hàng
customer_id	mã khách hàng
order_status	trạng thái đơn hàng (đã giao, đã vận chuyển,...)
order_purchase_timestamp	thời gian mua hàng
order_approved_at	thời gian phê duyệt thanh toán
order_delivered_carrier_date	thời gian đơn hàng được tiếp nhận xử lý bên đối tác phân phối
order_delivered_customer_date	thời gian giao hàng thực tế cho khách hàng
order_estimated_delivery_date	thời gian giao hàng ước tính đã được thông báo cho khách hàng tại thời điểm mua hàng

Mô tả dữ liệu (Data dictionary):

olist_products_dataset	
product_id	mã sản phẩm
product_category_name	danh mục sản phẩm (tiếng Bồ Đào Nha)
product_name_length	độ dài tên sản phẩm
product_description_length	độ dài mô tả sản phẩm
product_photos_qty	số lượng ảnh sản phẩm được công bố
product_weight_g	trọng lượng sản phẩm, đo bằng grams
product_length_cm	chiều dài sản phẩm, đo bằng cm
product_height_cm	chiều cao sản phẩm, đo bằng cm
product_width_cm	chiều rộng sản phẩm, đo bằng cm

olist_sellers_dataset	
seller_id	mã nhân viên bán hàng
seller_zip_code_prefix	năm chữ số đầu tiên của mã zip của nhân viên
seller_city	tên thành phố của nhân viên
seller_state	khu vực của nhân viên

olist_product_category_name_transaction	
product_category_name	danh mục sản phẩm (tiếng Bồ Đào Nha)
product_category_name_english	danh mục sản phẩm (tiếng anh)

Tiền xử lý

Công cụ xử lý



Tiên xử lý

Clean Data

```
df_customers.info()
```

df_customers.info() output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99441 entries, 0 to 99440
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   customer_id      99441 non-null   object  
 1   customer_unique_id 99441 non-null   object  
 2   customer_zip_code_prefix 99441 non-null   int64  
 3   customer_city      99441 non-null   object  
 4   customer_state     99441 non-null   object  
dtypes: int64(1), object(4)
memory usage: 3.8+ MB
```

Bảng Customer

```
df_geolocation.info()
```

df_geolocation.info() output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27912 entries, 0 to 27911
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   geolocation_zip_code_prefix 27912 non-null   int64  
 1   geolocation_city          27912 non-null   object  
 2   geolocation_state         27912 non-null   object  
dtypes: int64(1), object(2)
memory usage: 654.3+ KB
```

Bảng Geolocation

```
[ ] df_payments.info()
```

df_payments.info() output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103886 entries, 0 to 103885
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   order_id        103886 non-null   object  
 1   payment_sequential 103886 non-null   int64  
 2   payment_type    103886 non-null   object  
 3   payment_installments 103886 non-null   int64  
 4   payment_value   103886 non-null   int64  
dtypes: int64(3), object(2)
memory usage: 4.0+ MB
```

Bảng Payment

Tiên xử lý

Clean Data

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 96189 entries, 0 to 99440
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   order_id         96189 non-null   object  
 1   customer_id     96189 non-null   object  
 2   order_status     96189 non-null   object  
 3   order_purchase_timestamp  96189 non-null   datetime64[ns]
 4   order_approved_at 96189 non-null   datetime64[ns]
 5   order_delivered_carrier_date 96189 non-null   datetime64[ns]
 6   order_delivered_customer_date 96189 non-null   datetime64[ns]
 7   order_estimated_delivery_date 96189 non-null   datetime64[ns]
dtypes: datetime64[ns](5), object(3)
memory usage: 6.6+ MB

# chuyển sang kiểu dữ liệu datetime
df_orders['order_purchase_timestamp'] = pd.to_datetime(df_orders['order_purchase_timestamp'])
df_orders['order_approved_at'] = pd.to_datetime(df_orders['order_approved_at'])
df_orders['order_delivered_carrier_date'] = pd.to_datetime(df_orders['order_delivered_carrier_date'])
df_orders['order_delivered_customer_date'] = pd.to_datetime(df_orders['order_delivered_customer_date'])
df_orders['order_estimated_delivery_date'] = pd.to_datetime(df_orders['order_estimated_delivery_date'])
```

Chuyển định dạng các cột sang kiểu dữ liệu datetime

```
[ ] df_orders.min()

order_id          00010242fe8c5a6d1ba2dd792cb16214
customer_id       00012a2ce6f8dcda20d059ce98491703
order_status      approved
order_purchase_timestamp  2016-09-04 21:15:19
order_approved_at 1900-01-01 00:00:00
order_delivered_carrier_date 1900-01-01 00:00:00
order_delivered_customer_date 1900-01-01 00:00:00
order_estimated_delivery_date 2016-09-30 00:00:00
dtype: object
```

Phát hiện dữ liệu lỗi năm 1900

```
[ ] #Xóa dữ liệu của dữ liệu năm 1900
df_orders = df_orders.drop(df_orders[df_orders['order_approved_at'].dt.year == 1900].index)
df_orders = df_orders.drop(df_orders[df_orders['order_delivered_carrier_date'].dt.year == 1900].index)
df_orders = df_orders.drop(df_orders[df_orders['order_delivered_customer_date'].dt.year == 1900].index)

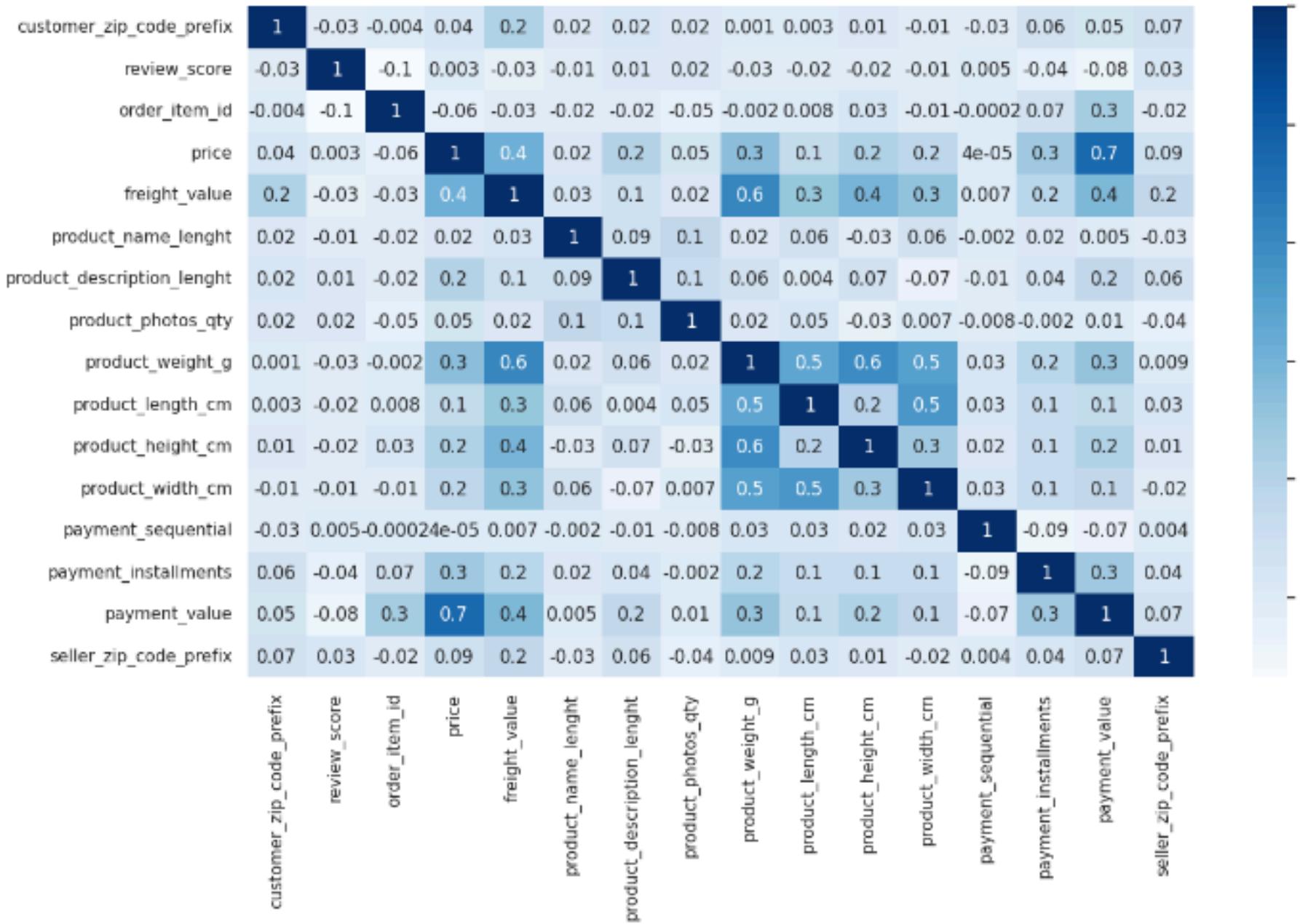
[ ] #Xóa dữ liệu của dữ liệu năm 2016
df_orders = df_orders.drop(df_orders[df_orders['order_approved_at'].dt.year == 2016].index)
df_orders = df_orders.drop(df_orders[df_orders['order_delivered_carrier_date'].dt.year == 2016].index)
df_orders = df_orders.drop(df_orders[df_orders['order_delivered_customer_date'].dt.year == 2016].index)
df_orders = df_orders.drop(df_orders[df_orders['order_purchase_timestamp'].dt.year == 2016].index)
df_orders = df_orders.drop(df_orders[df_orders['order_estimated_delivery_date'].dt.year == 2016].index)
```

Xóa dữ liệu lỗi năm 1900 và năm 2016

Phân tích dữ liệu



Phân tích khám phá data



Price: có tác động khá mạnh đến payment_value (0.7) trong khi có tác động tương đối đến freight_value (0.4), product_weight (0.3), payment_installments (0.3).

Freight_value: có tác động khá mạnh đến product_weight (0.6) và tác động tương đối đến payment_value (0.4), product_height (0.4) và price (0.4).

Product_weight: có tác động khá mạnh đến freight_value (0.6), product_length (0.5), product_height (0.6) và product_width (0.5).

Product_height: có tác động khá mạnh đến product_weight (0.6) và tác động tương đối đến payment_value (0.4) và product_width (0.3).

Product_width: có tác động tương đối đến product_weight (0.5) và product_length (0.5).

Payment_value: có tác động khá mạnh đến price (0.7) trong khi có tác động tương đối đến freight_value (0.4), order_item (0.3) và product_weight (0.3).

Phân tích tổng quan tình hình kinh doanh của công ty

- Tổng doanh thu của Olist \$ 1,512M.

Revenue

1512M

Total Products Sold
(items)

110K

Total Customers

96.19K

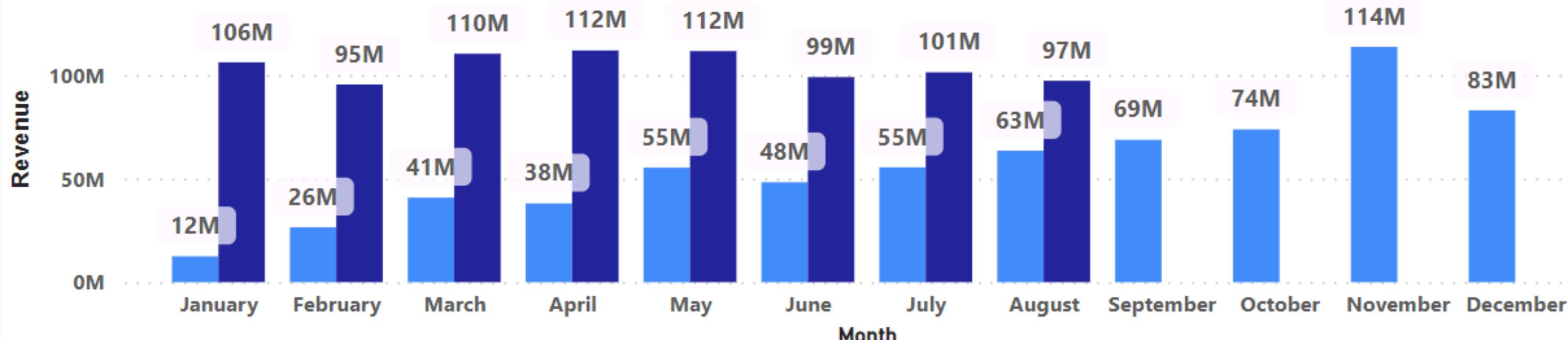
Total Sellers

3068

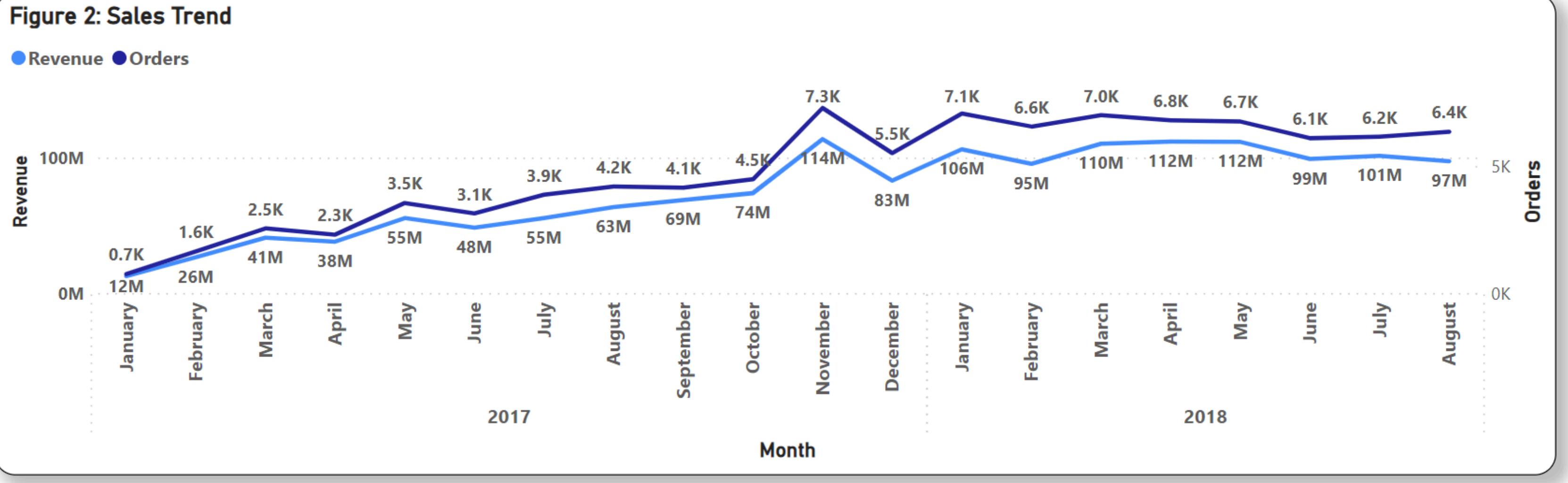
- Doanh thu có xu hướng tăng qua các khoảng thời gian, doanh thu 8 tháng đầu năm 2018 tăng trưởng mạnh hơn 2017
 - Doanh thu lớn nhất ghi nhận được thuộc về **114M** trong tháng 11 năm 2017
- > Thời điểm này có chương trình Black Friday

Figure 1: Revenue by month and year

Year ● 2017 ● 2018



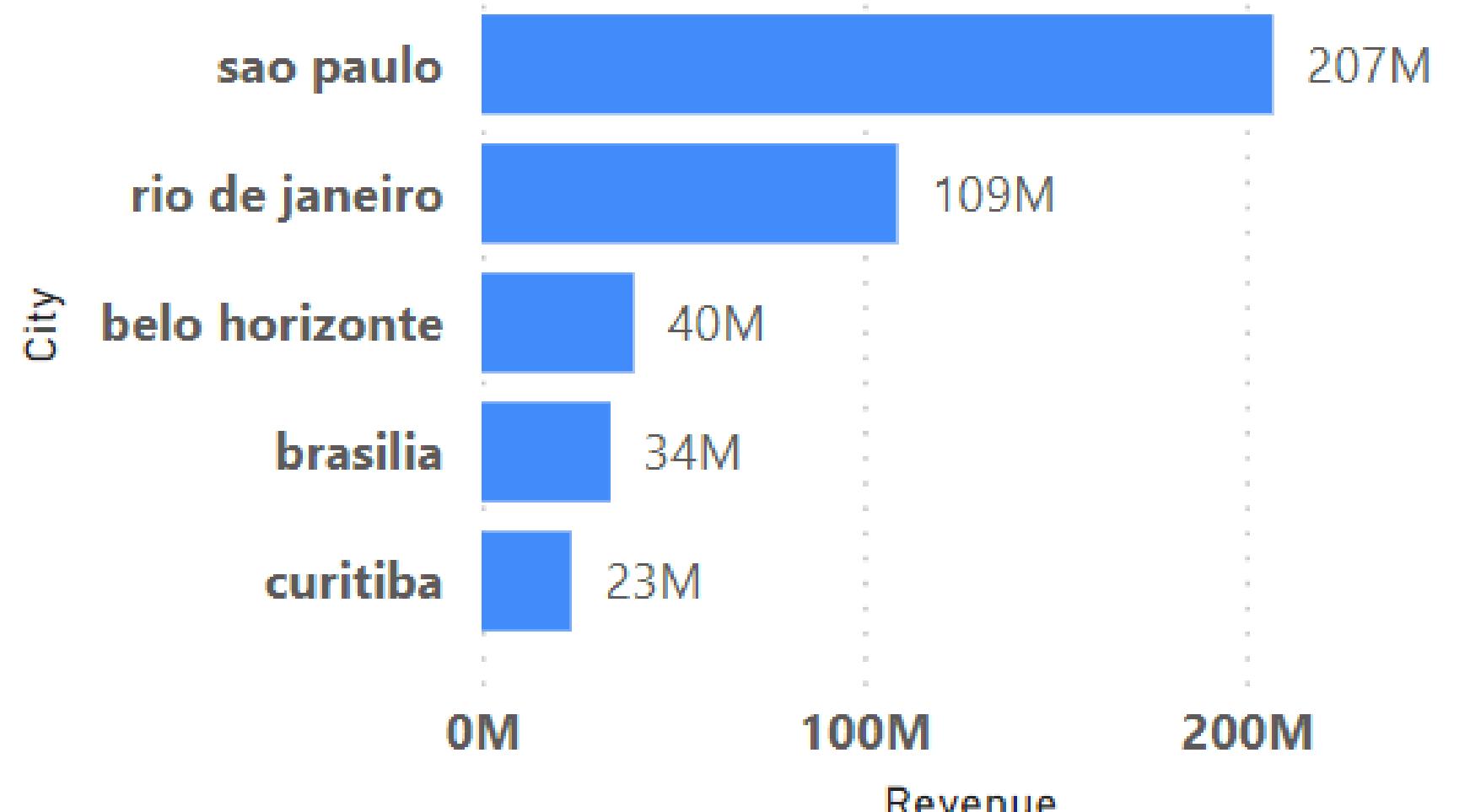
Phân tích tổng quan tình hình kinh doanh của công ty



- Sự phân bổ giữa doanh thu và số lượng đơn hàng tương đối tăng đồng đều nhau trong cùng 1 thời điểm
 - Olist có khả năng xử lý đơn hàng tương ứng với doanh thu tạo ra

Phân tích tổng quan tình hình kinh doanh của công ty

Figure 3: Top 5 Revenue by City



- São Paulo mang lại doanh thu lớn nhất với **207M** doanh số thứ 2 là Rio de Janeiro với **109M**
 - Là 2 thành phố lớn phát triển với dân số đông (São Paulo **19 triệu người**) & (Rio de Janeiro **6.7 triệu người**)
 - São Paulo và Rio de Janeiro là hai trung tâm kinh tế quan trọng tại Brazil

Phân tích về sản phẩm

- Tổng số sản phẩm đã bán trong năm 2017 & 8 tháng đầu năm 2018 là **110,280** sản phẩm
- Danh mục sản phẩm gồm **71 chủng loại** và **32,787** sản phẩm

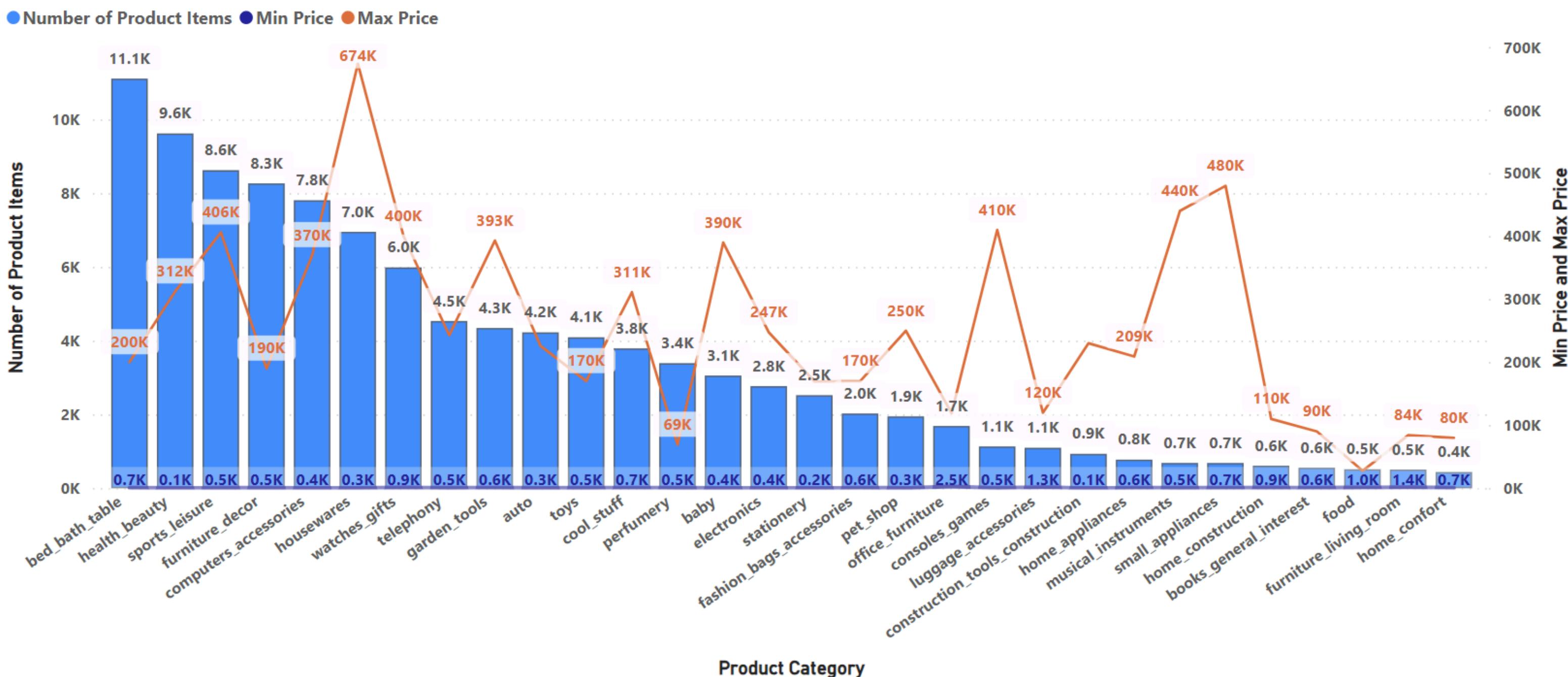
Total Products Sold
(items)

110K

Total Products Sold
(distinct items)

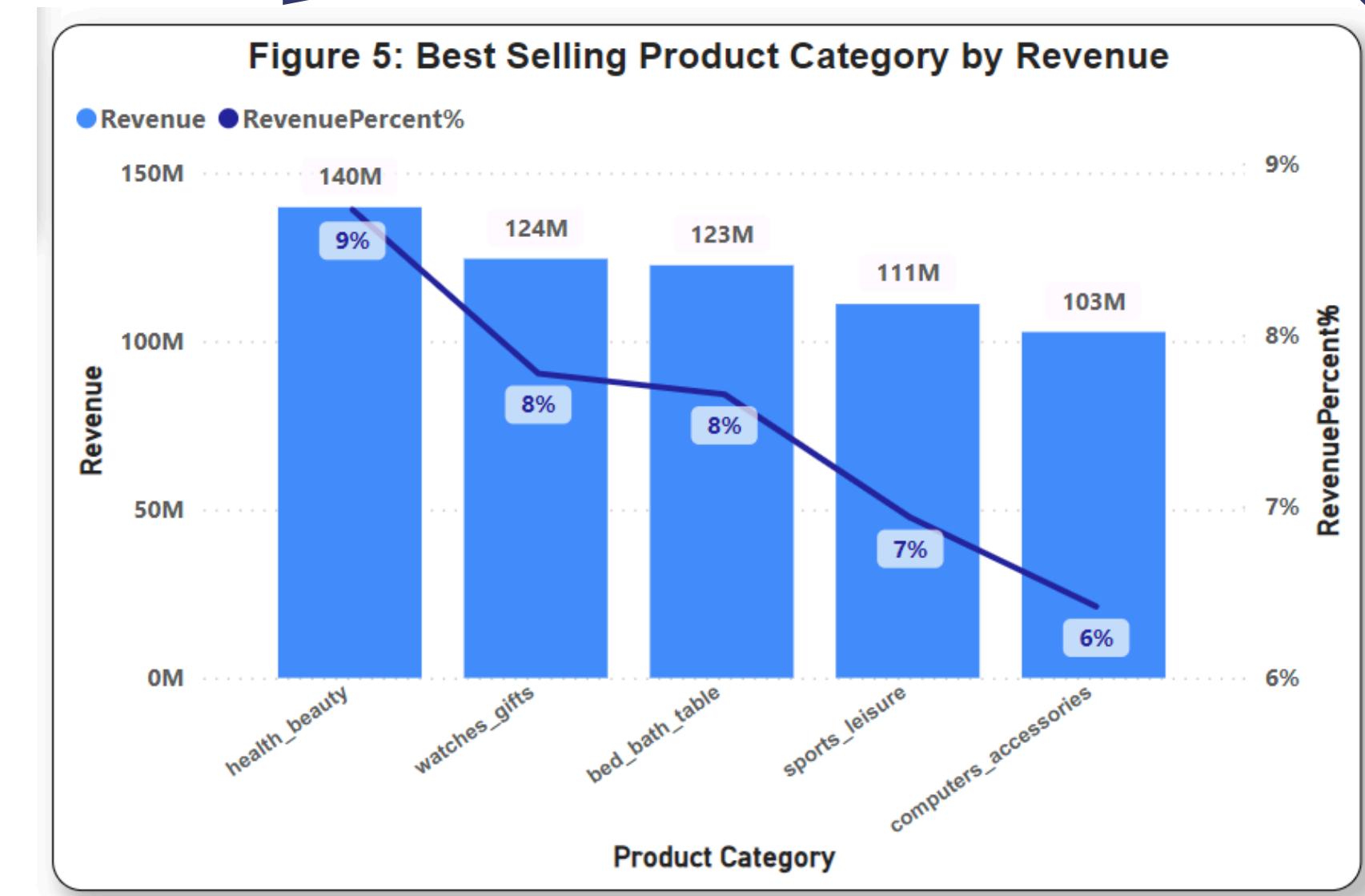
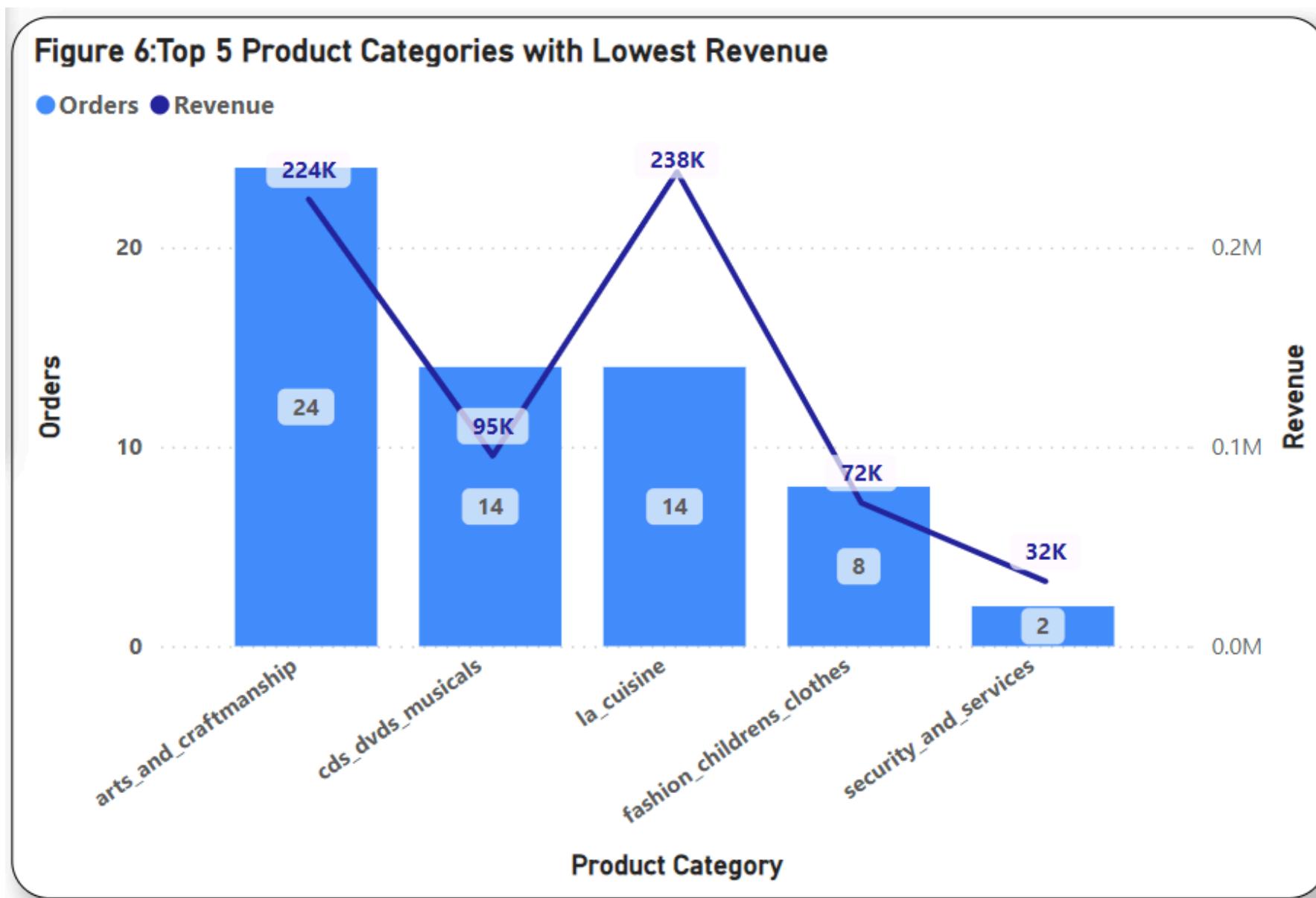
32K

Figure 4: Number of Product Items, Min Price and Max Price by Product Categories



Phân tích về sản phẩm

- Top danh mục sản phẩm bán chạy nhất:
 - Health_beauty là hạng mục dẫn đầu, chiếm khoảng **9%** theo sau là Watches_gifts và Bed_bath_table
 - Đề xuất doanh nghiệp lập kế hoạch chiến lược, quản lý hàng tồn kho và các chiến dịch tiếp thị, nhằm mục tiêu vào các danh mục sản phẩm phổ biến và sinh lợi nhất.

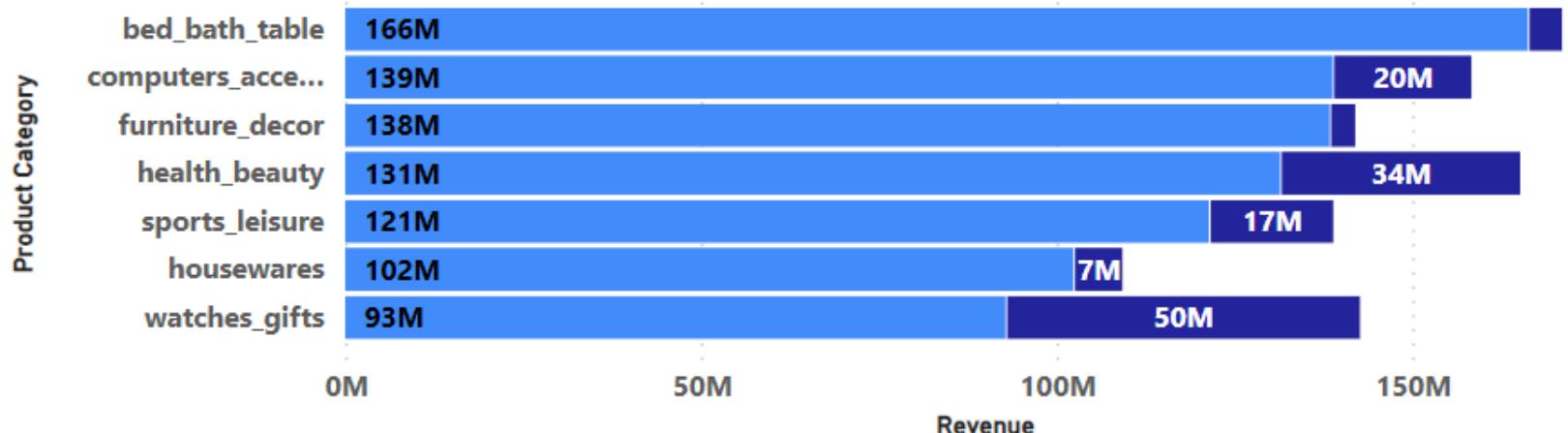


- Top chủng loại sản phẩm bán kém nhất:
 - Danh mục Security and services ít được đặt hàng nhất với tổng 2 đơn tương ứng khoảng \$32K

Phân tích về sản phẩm

Figure 7: Payment_product_price by Product_category_name

● Payment_product_price_less_50000 ● Payment_product_price_more_50000



Các sản phẩm có giá dưới 50,000 USD được mua sắm nhiều nhất

Khách hàng ưa thích các sản phẩm ở mức giá dưới 50000 USD ở hầu hết các chủng loại sản phẩm

Figure 8: Number of Orders by Product Price

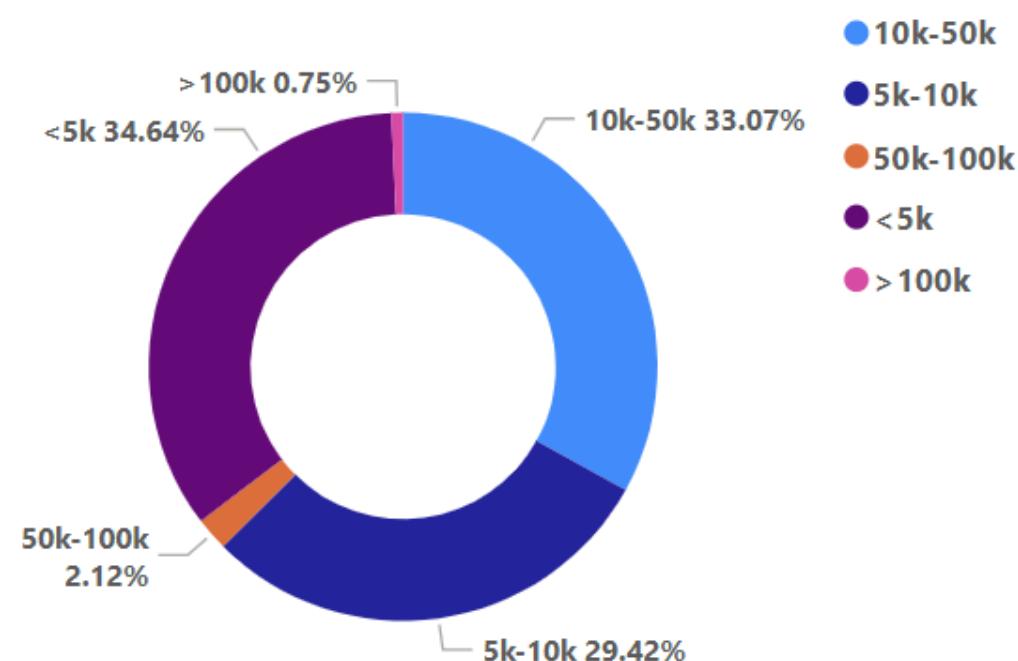
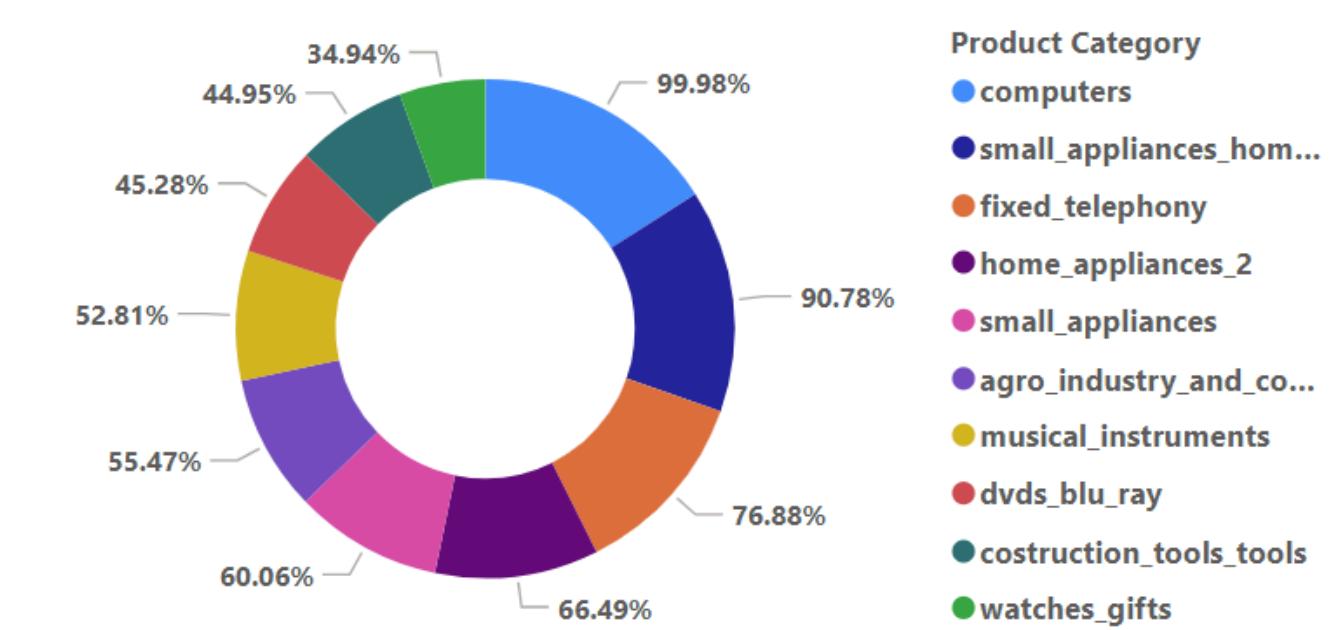


Figure 9: Top 10 of Percent_payment_product_price_more_50000 by Product_categories



Phân tích về hành vi khách hàng

Total Customers

96.19K

Orders

96.19K

Active User

100%

Mỗi một khách hàng chỉ thực hiện duy nhất 1 lần đặt hàng

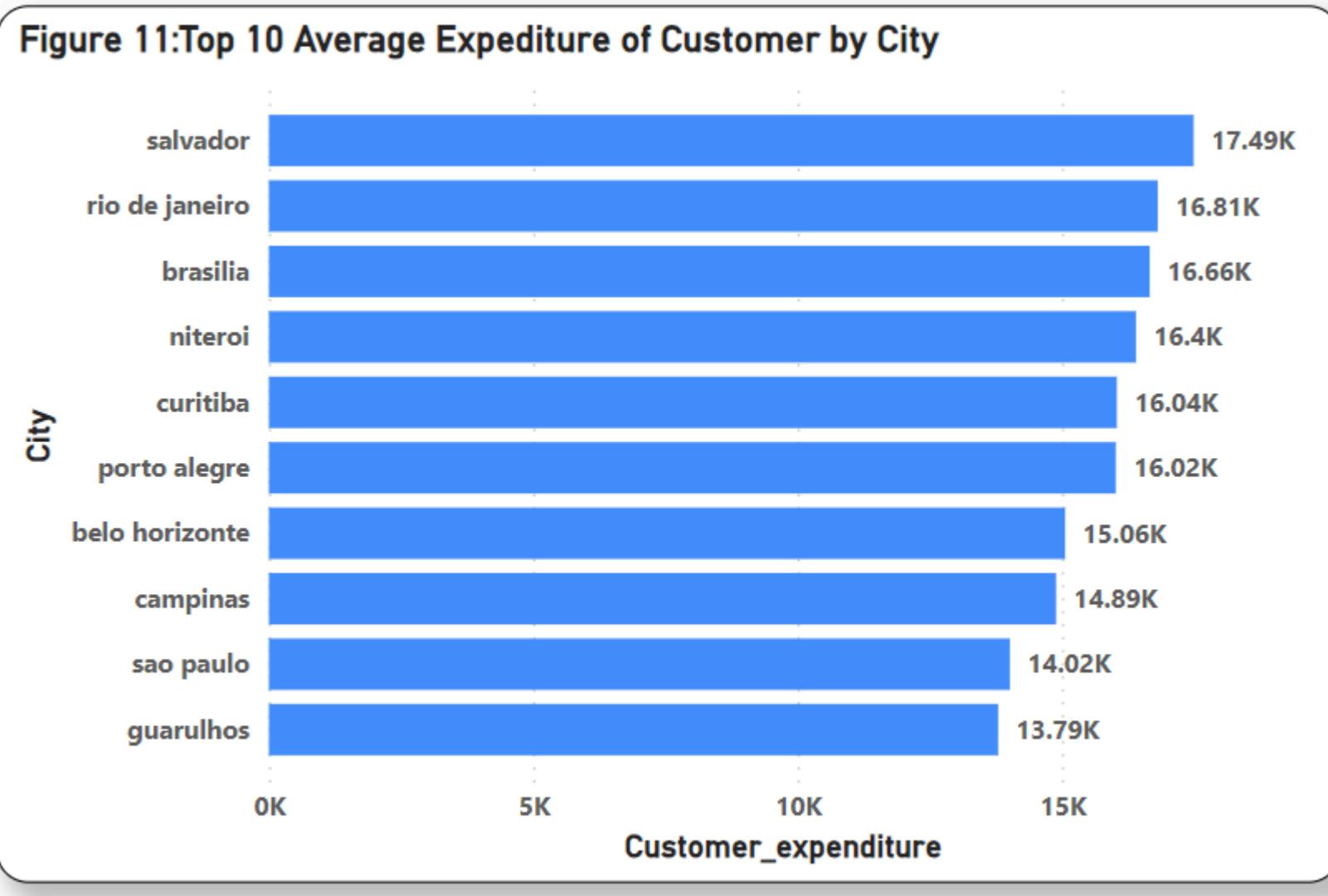
Figure 10:Top 10 Number of customer by city



Thị trường tiêu thụ chính của Olist là São Paulo (**15,66K đơn**) và Rio de Janeiro (**6,88K đơn**)

Phân tích về hành vi khách hàng

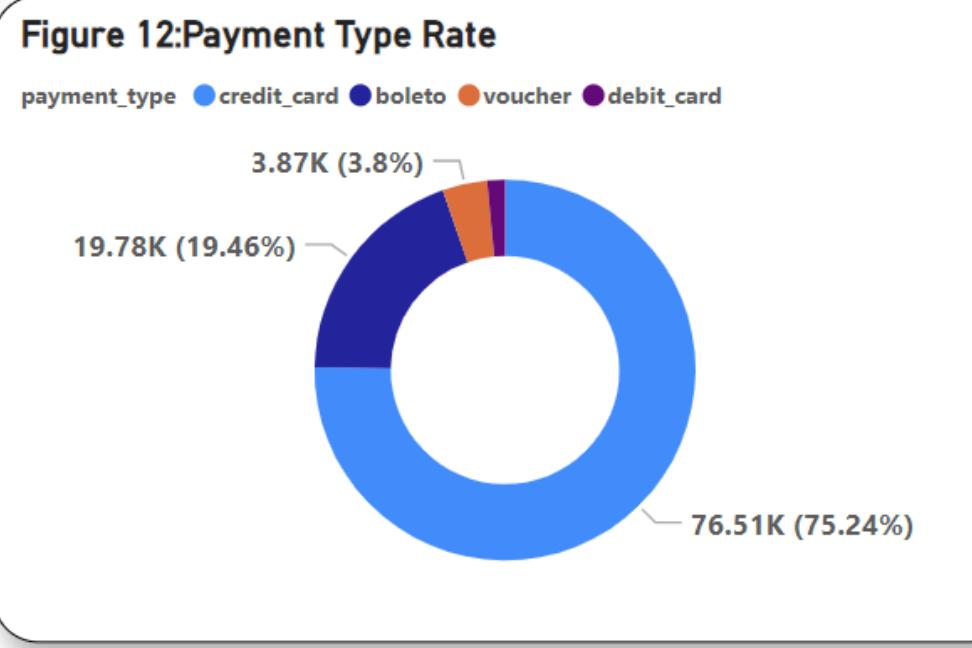
Figure 11: Top 10 Average Expediture of Customer by City



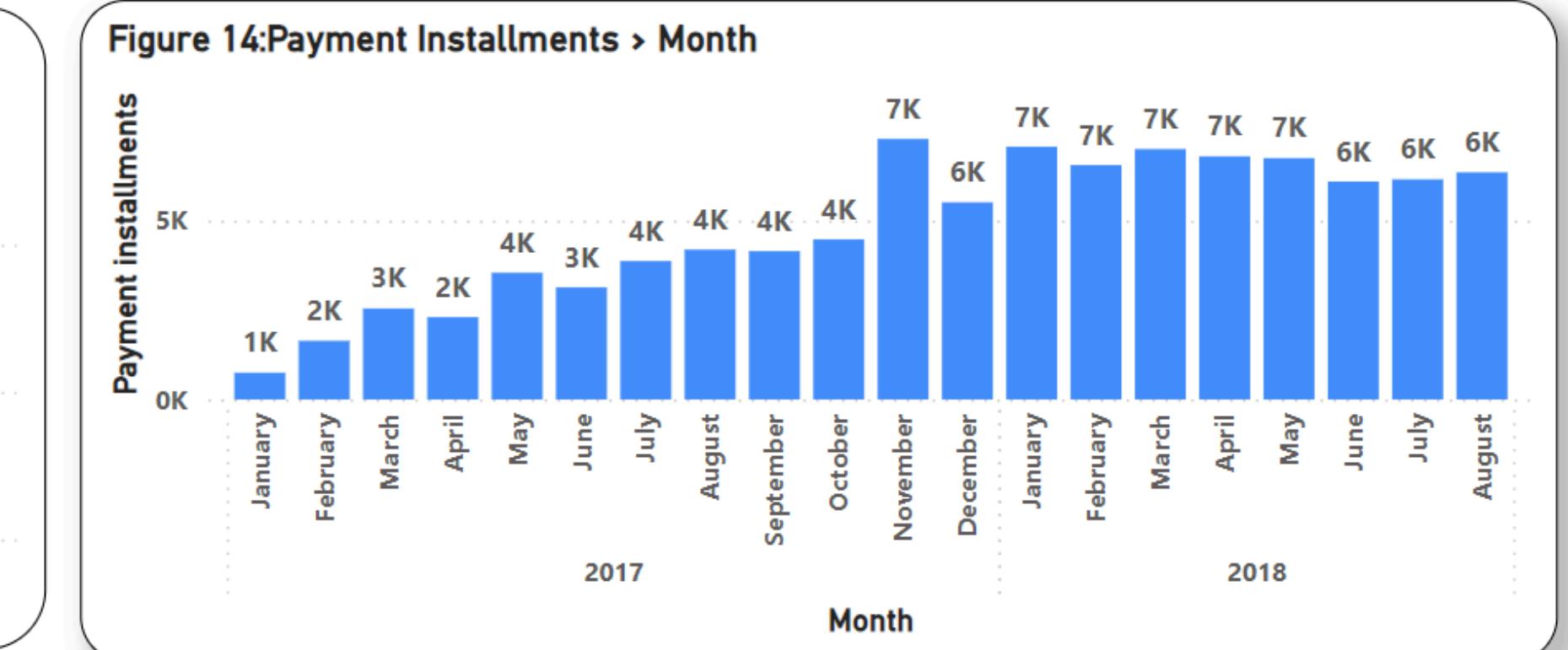
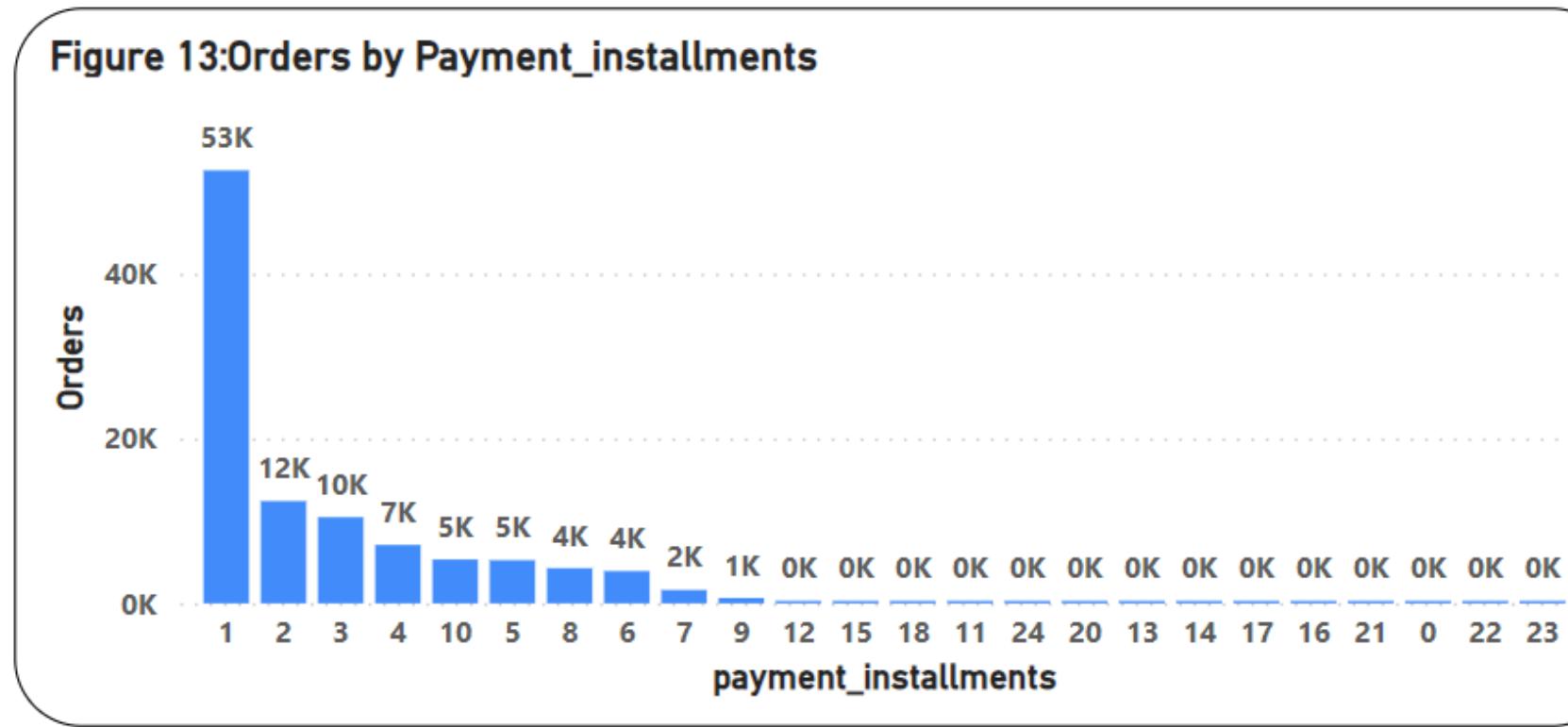
Khu vực **Salvador** ghi nhận mức chi tiêu trung bình một khách hàng cao nhất với **17,49K**

- **Rio de Janeiro:** mức chi tiêu và mức đóng góp vào doanh thu tốt -> cần tiếp tục phát huy và duy trì.
- **Sao Paulo:** doanh thu đóng góp cao nhưng mức chi tiêu của 1 khách hàng chưa tốt -> cần kích thích khả năng tiêu dùng của khách hàng lên.
- **Salvador:** doanh thu đóng góp thấp hơn nhưng mức chi tiêu 1 khách hàng cao nhất -> tập trung tiếp thị, thu hút và mở rộng tệp khách hàng để nâng cao số lượng người tiêu dùng.

Phân tích về hành vi khách hàng



- Khách hàng chuộng thanh toán qua **Credit card (73,92%)**
 - Debit card** (thẻ ghi nợ) ít được sử dụng để thanh toán nhất.
- => Nguyên nhân:
- Cơ sở hạ tầng thanh toán phát triển
 - Quảng bá từ các ngân hàng và tổ chức tín dụng, chính sách trả góp hấp dẫn trên thẻ tín dụng



- Có đến **hơn 50,000 thanh toán trả góp** (trong tổng **96,19K đơn hàng**) được thực hiện trong vòng 1 tháng
- Khách hàng bắt đầu có xu hướng chuộng thanh toán trả góp khi đặt hàng

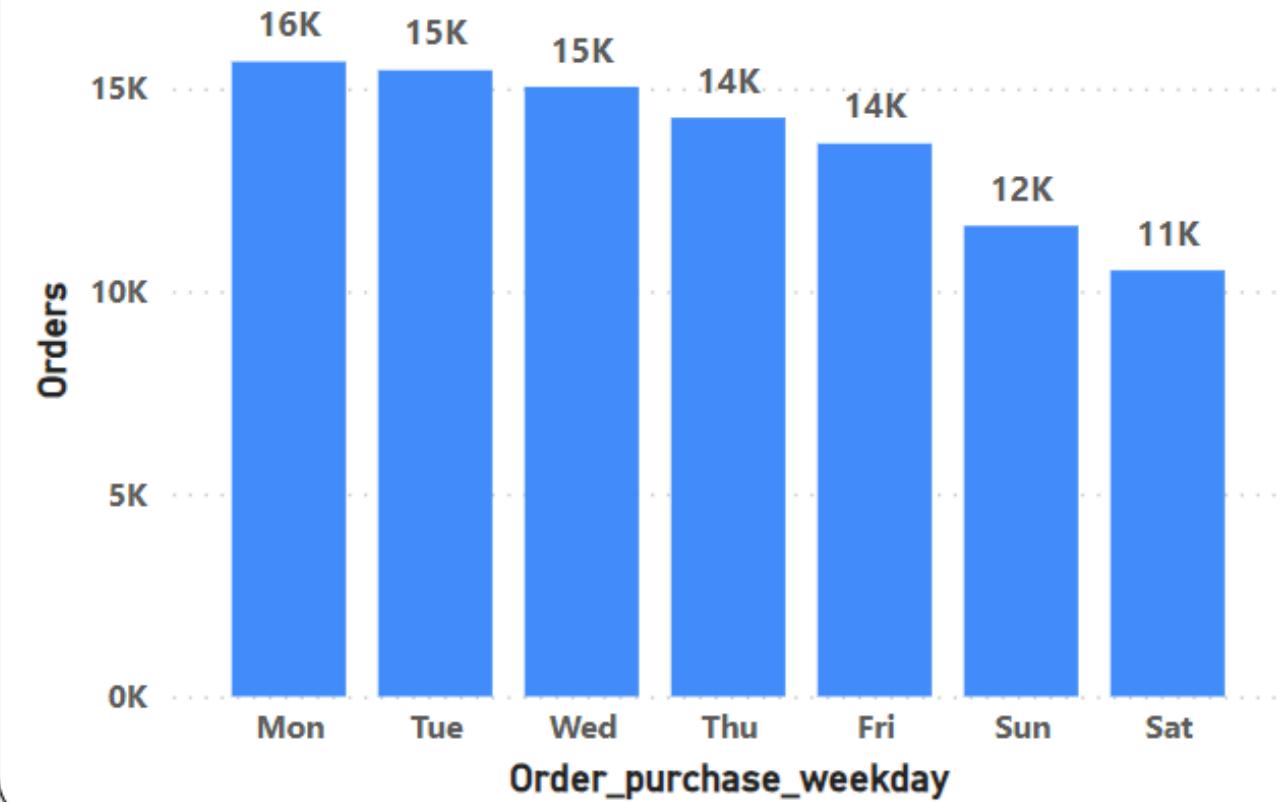
Phân tích về hành vi khách hàng

- Khoảng thời gian cuối tuần có khá ít lượng đặt hàng hơn so với các ngày trong tuần

=> Nguyên nhân:

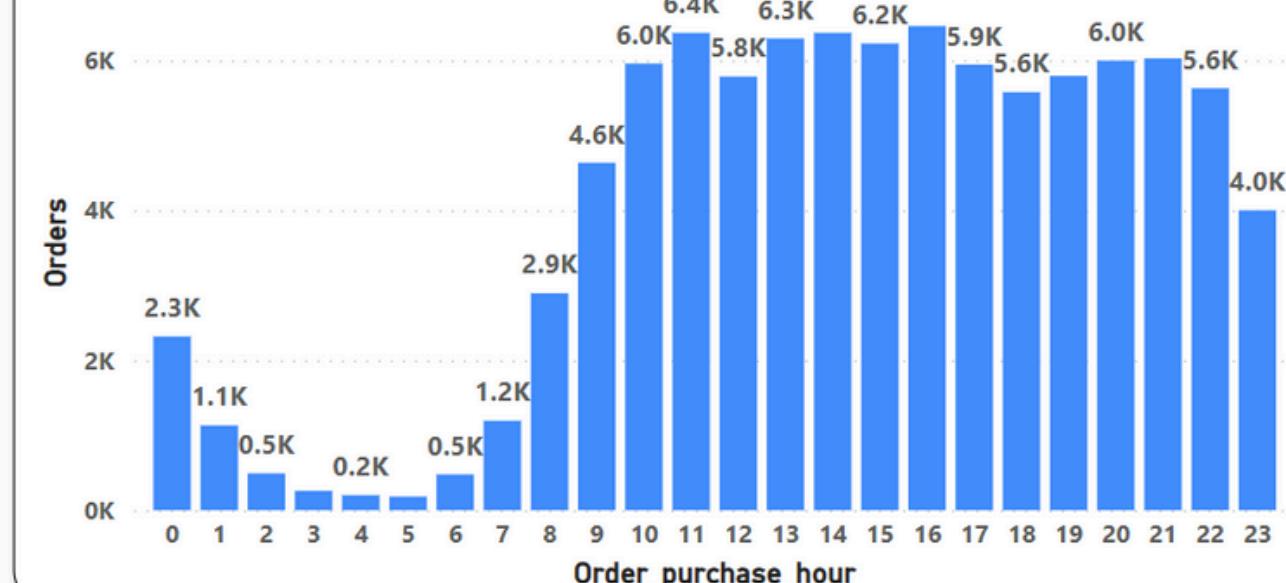
- Có khả năng đáp ứng đơn hàng nhanh hơn so với cuối tuần.
- Xu hướng áp dụng các chương trình khuyến mãi vào thường ngày

Figure 15:Order WeekDays



- Thời gian đặt hàng phổ biến của nhóm khách hàng thường kéo dài từ khoảng trưa đến tối

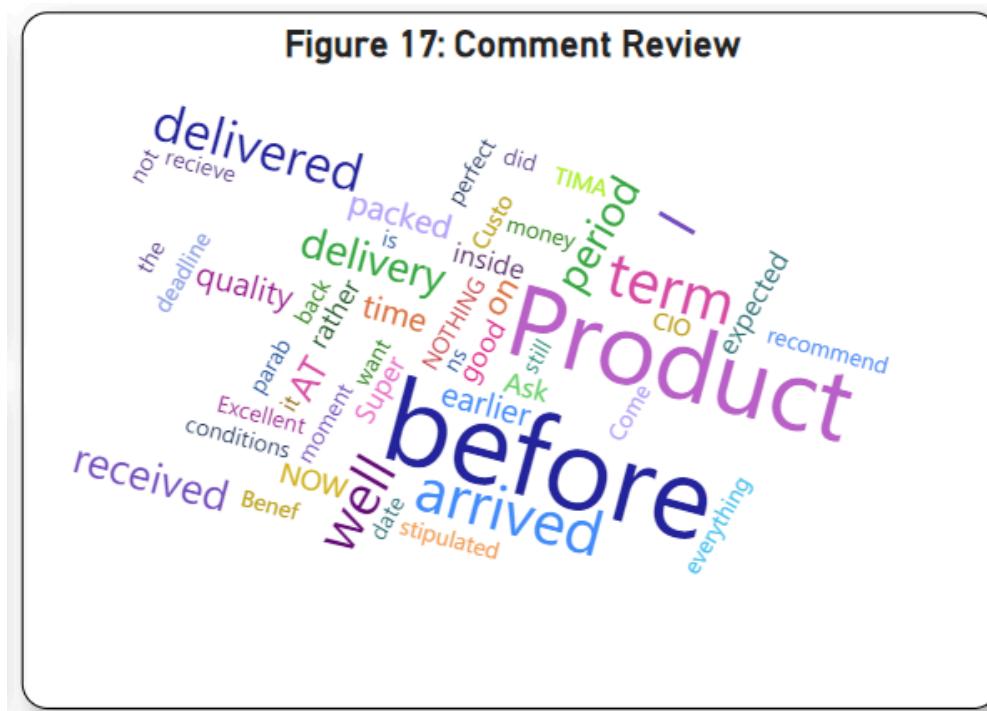
Figure 16:Order Hours



Phân tích về phản hồi và đánh giá

- Tổng số khách hàng đánh giá sản phẩm khoản **41 ngàn** với điểm đánh giá trung bình là **3.67**

Figure 17: Comment Review

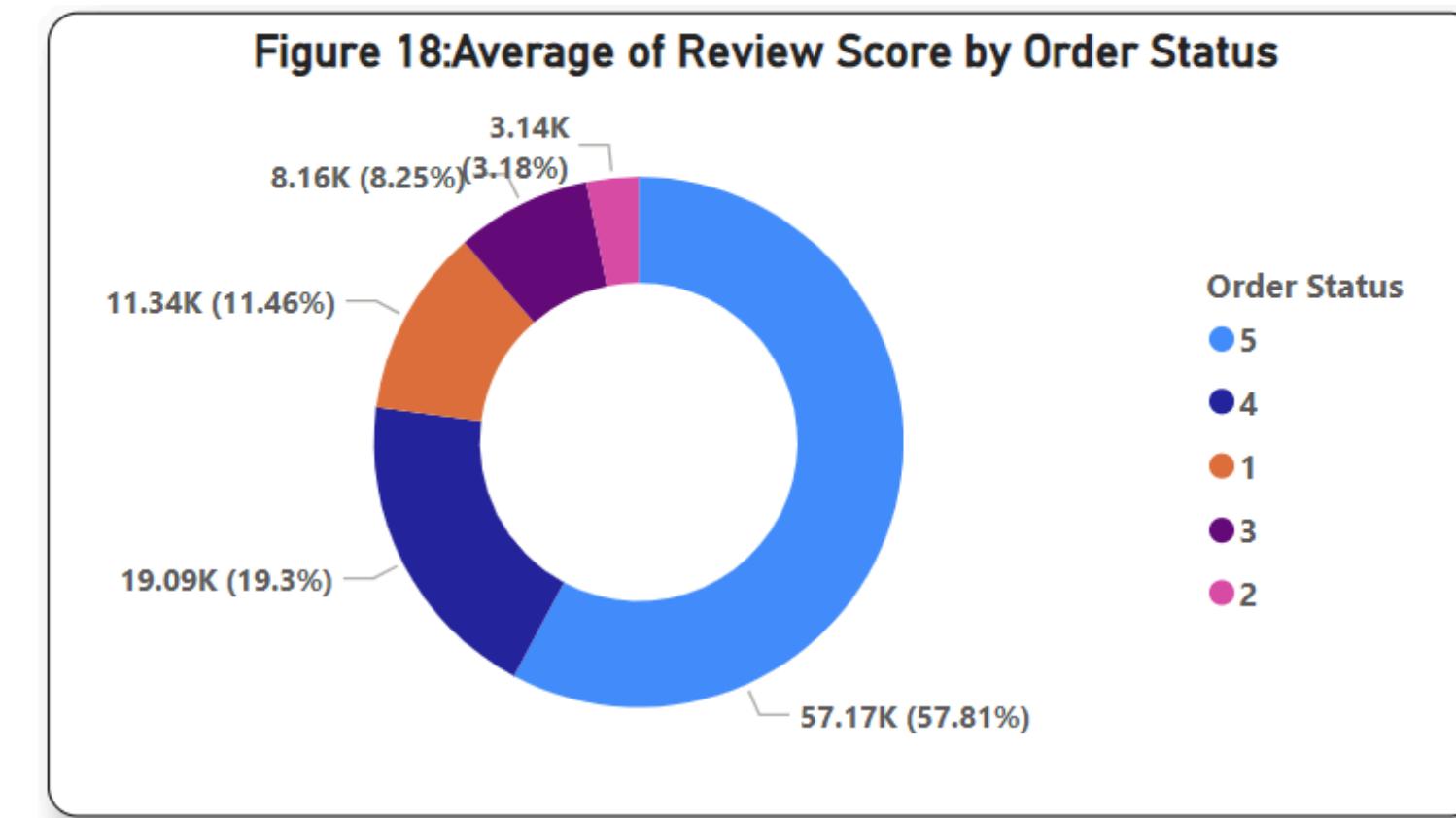


- Những comment phổ biến nhất là thường là **về thời gian giao hàng và chất lượng sản phẩm**

- Đơn hàng có review tích cực chiếm tỷ lệ cao nhất khoảng **65%**
 - Đơn hàng có review trung tính (3 sao) chiếm gần **9%**
 - Đơn hàng có review tiêu cực chiếm khoảng **26%**

=> Phần lớn khách hàng hài lòng với dịch vụ của Olist

Figure 18:Average of Review Score by Order Status



Phân tích về hiệu quả giao hàng

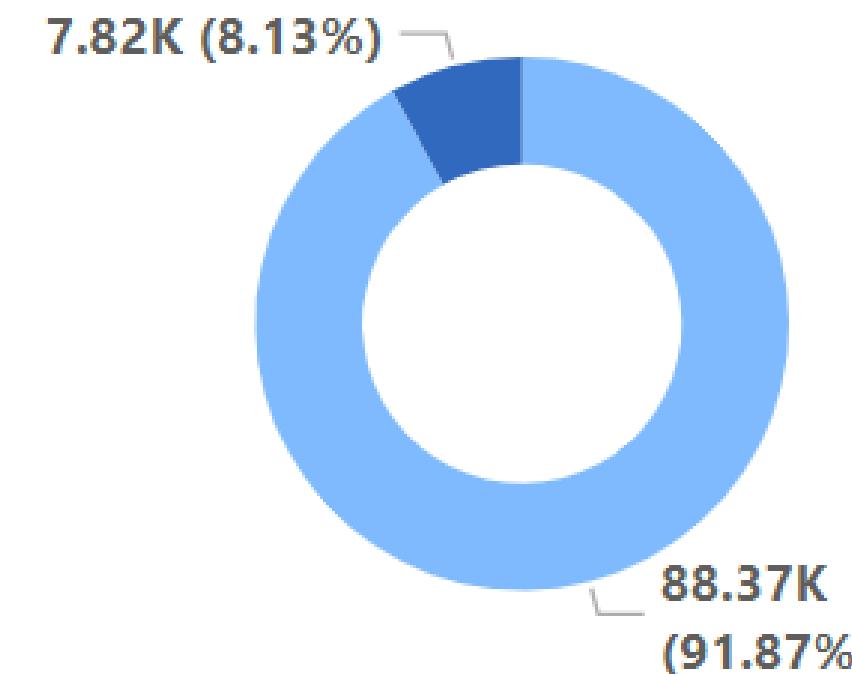


- Trung bình đơn hàng hoàn thành khoảng **12.7 ngày**, các đơn hàng giao đúng giờ là **10 ngày** và đơn hàng giao trễ tầm **13 ngày**

- Giai đoạn từ năm 2017 đến tháng 8 năm 2018, hoạt động tỷ lệ đơn hàng giao trễ chiếm **8%**
 - Hầu hết Olist đều đáp ứng kỳ vọng đặt hàng của khách hàng

Figure 18: Dispatch Rate

Status ● On-Time Deliveries ● Late Deliveries



Phân tích về hiệu quả giao hàng

Figure 19: Dispatch by Yr, > Qtr, > Month

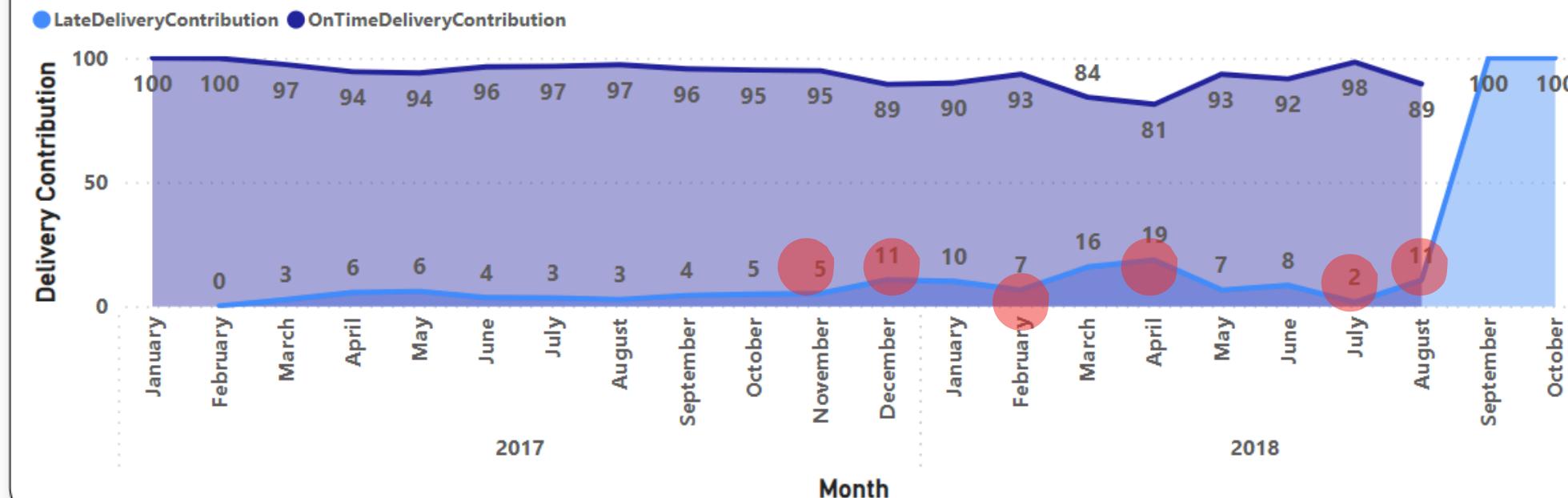
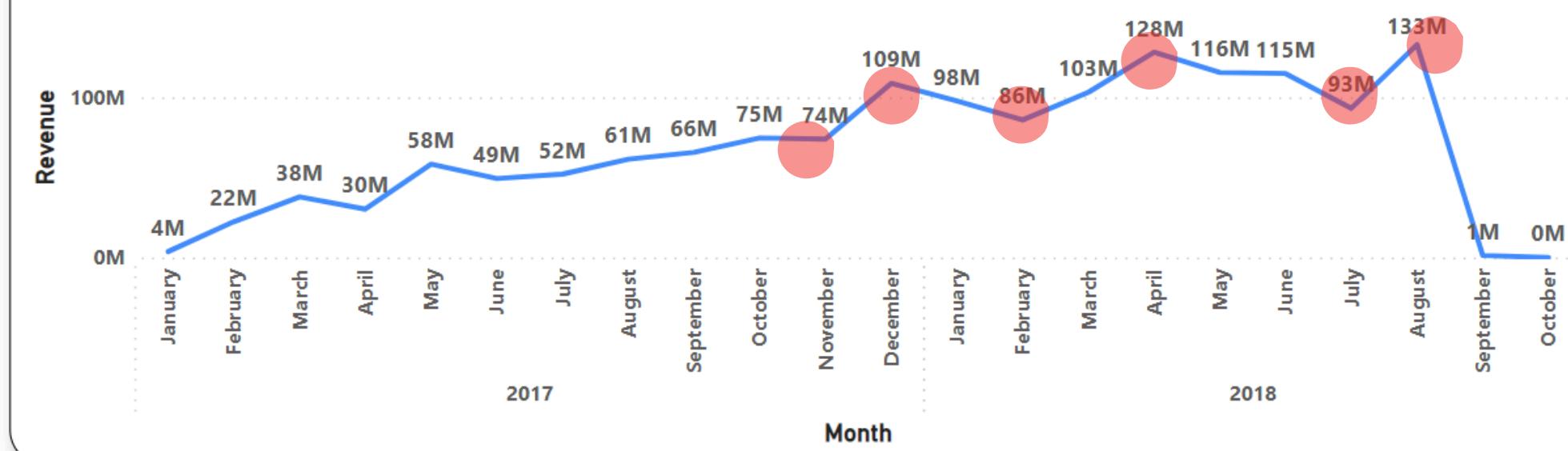


Figure 20: Delivery Performance by Product > Year > Quater > Month



- Tỷ lệ giao hàng trễ bắt đầu tăng từ khi doanh số tăng vào năm 2018 và tỷ lệ giao hàng trễ cao nhất vào tháng 4 (2018) -> Olist đang quá tải khi có đơn đặt hàng nhiều
 - Bổ sung nhân công và hàng hóa, phân bổ lịch trình để đáp ứng nhu cầu đặt hàng.

KẾT LUẬN & ĐỀ XUẤT CHIẾN LƯỢC



KẾT LUẬN & ĐỀ XUẤT CHIẾN LƯỢC

VẤN ĐỀ 1: NHU CẦU THỊ TRƯỜNG NHƯ THẾ NÀO?

- **Sản phẩm:**
 - Nhu cầu khách hàng ưa thích hầu hết các sản phẩm có giá bình dân (dưới 50KUSD)
 - Nhóm sản phẩm được khách hàng thường xuyên mua sắm là: Health_beauty, Watches_gifts, Bed_bath_table và ít nhất là Security_and_services, cds_dvds_musicals, home_comfort_2
 - Nhà phân phối sản phẩm:
 - **Tối ưu hóa chi phí lưu kho để tăng khả năng cạnh tranh về giá cả.**
 - **Dự đoán nhu cầu mua sắm của các mặt hàng một cách hiệu quả để tránh bị tồn đọng không cần thiết**
 - Khách hàng mua sản phẩm:
 - **Duy trì kho hàng để đảm bảo có đủ sản phẩm trong danh mục sản phẩm đóng góp doanh thu cao.**
 - **Quảng cáo tiếp thị, cung cấp gói ưu đãi mua 1 tặng 1 đối với các dòng sản phẩm ít ưa chuộng Security và DVD.**

KẾT LUẬN & ĐỀ XUẤT CHIẾN LƯỢC

VẤN ĐỀ 1: NHU CẦU THỊ TRƯỜNG NHƯ THẾ NÀO?

- **Hành vi tiêu dùng của khách hàng:**

- Thị trường tiêu thụ chính của Olist tập trung tại 2 thành phố lớn của Brazil là São Paulo và Rio de Janeiro. Ngoài ra, khách hàng tại Salvador có mức trung bình chi tiêu cao nhất.
- Hầu hết khách hàng đặt hàng vào các ngày trong tuần và ưa chuộng hình thức thanh toán credit card.
 - Nhà phân phối sản phẩm:
 - **Mở rộng các điểm bán và phân phối sản phẩm ở các khu vực có mức chi tiêu trung bình cao như Salvador.**
 - Khách hàng mua sản phẩm:
 - **Đề xuất các chương trình khuyến mãi, chăm sóc khách hàng nhằm thu hút sự quay lại sử dụng dịch vụ và kích thích khả năng tiêu dùng.**
 - **Cung cấp thẻ thành viên cho những khách hàng quay trở lại sử dụng dịch vụ từ 2 lần trở lên theo tuần/tháng/quý.**
 - **Hợp tác với ngân hàng để hỗ trợ thanh toán trả góp cho người tiêu dùng với mức lãi suất thấp.**

KẾT LUẬN & ĐỀ XUẤT CHIẾN LƯỢC

VẤN ĐỀ 2: CHẤT LƯỢNG DỊCH VỤ CÓ ĐẢM BẢO

- **Tình trạng giao hàng:**

- Đa số gần như Olist đều hoàn thành việc giao hàng trong thời gian dự kiến.
- Tỷ lệ giao hàng trễ tăng khi có sự đột biến về số lượng đơn hàng được đặt cao hơn so với dự tính.

- Nhà phân phối sản phẩm:

- **Lên kế hoạch trước và dự đoán được nhu cầu đặt hàng tại các dịp tặng quà để tránh việc giao hàng trễ.**
- **Điều động nhân lực vào các ngày trong tuần so với cuối tuần để đáp ứng nhu cầu đặt hàng.**

- Khách hàng mua sản phẩm:

- **Thiết lập lộ trình giao hàng tối ưu và dự đoán khung giờ giao hàng hiệu quả để rút ngắn thời gian giao hàng.**

KẾT LUẬN & ĐỀ XUẤT CHIẾN LƯỢC

VẤN ĐỀ 2: CHẤT LƯỢNG DỊCH VỤ CÓ ĐẢM BẢO

- **Phản hồi & đánh giá:**

- Các sản phẩm về quà tặng tăng đột biến trong các dịp lễ, làm ảnh hưởng đến việc giao hàng.
- Nhiều khách hàng có trải nghiệm tích cực nhưng vẫn có số lượng khá đáng kể thể hiện sự không hài lòng trong quá trình vận chuyển đơn hàng.
- Hầu hết các đánh giá đề cập đến thời gian giao hàng (kịp thời/không kịp thời) và chất lượng bảo quản sản phẩm (còn nguyên/bị hư hỏng).

- Nhà phân phối sản phẩm:

- **Thiết lập danh sách tiêu chí đánh giá phân phối để giám sát quy trình tiếp nhận hàng hóa và kiểm tra chất lượng hàng vận chuyển.**

- Khách hàng mua sản phẩm:

- **Xây dựng cộng đồng trực tuyến, các kênh online để tăng trải nghiệm sử dụng dịch vụ tạo cơ hội tương tác với khách hàng.**
 - **Đào tạo nhân viên chuyển đóng gói bài bản, chắc chắn và nhân viên chuyển giao về bảo tồn sản phẩm tránh bị hư hại.**

XIN CẢM ƠN

