

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
ĐẠI HỌC KINH TẾ TP HỒ CHÍ MINH  
TRƯỜNG CÔNG NGHỆ VÀ THIẾT KẾ**



## **ĐỒ ÁN MÔN HỌC HỆ HỖ TRỢ QUẢN TRỊ THÔNG MINH**

**ĐỀ TÀI: PHÂN TÍCH TÌNH HÌNH KINH DOANH CỦA  
OLIST STORE GIAI ĐOẠN 2016 – 2018 VÀ DỰ ĐOÁN XU  
HƯỚNG PHÁT TRIỂN NGÀNH THƯƠNG MẠI ĐIỆN TỬ  
TẠI BRAZIL**

**Mã Lớp Học phần: 24D1INF50908501**

**Nhóm Sinh Viên:**

1. Phan Thị Bích Diệu
2. Nguyễn Nữ Hoàng Khánh
3. Phùng Trần Văn Quang
4. Nguyễn Trương Thanh Việt
5. Trần Quang Vinh
6. Nguyễn Thanh Vy

**Khóa: K47**

**Giảng Viên Hướng Dẫn: ThS. Phạm Thị Thanh Tâm**

**TP. Hồ Chí Minh, Tháng 05 năm 2024**

# LỜI CẢM ƠN

***Kính gửi cô Phạm Thị Thanh Tâm,***

Nhóm chúng em xin gửi lời cảm ơn đến Cô trong suốt quá trình giảng dạy và truyền đạt nguồn kiến thức quý báu trong môn học Hệ hỗ trợ quản trị thông minh. Những kiến thức chúng em tích lũy được cùng với sự hướng dẫn, chỉ bảo nhiệt tình từ Cô mà chúng em có thể hoàn thành một cách tốt nhất bài báo cáo kết thúc học phần này.

Tuy nhiên, bài làm của chúng em không thể tránh khỏi những sai sót nên nhóm chúng em rất mong nhận được sự góp ý và chỉ bảo từ cô để rút ra những kinh nghiệm quý giá, bổ sung nguồn kiến thức và phục vụ cho việc làm các bài báo cáo sau này.

Nhóm kính chúc cô luôn vui vẻ, hạnh phúc và đạt được nhiều thành tựu hơn nữa trong cuộc sống cũng như trên con đường truyền đạt những kiến thức bổ ích cho sinh viên như chúng em.

***Xin chân thành cảm ơn cô!***

# MỤC LỤC

LỜI CẢM ƠN .....	2
DANH MỤC HÌNH ẢNH .....	5
DANH MỤC BẢNG.....	6
CHƯƠNG 1: TỔNG QUAN.....	7
1.1. Lý do chọn đề tài .....	7
1.2. Mục đích, mục tiêu của đề tài.....	7
1.3. Đối tượng và phạm vi thực hiện .....	8
1.4. Phương pháp thực hiện .....	8
1.5. Bố cục của đề tài.....	9
1.6. Phân công công việc .....	9
CHƯƠNG 2: MÔ TẢ DOANH NGHIỆP .....	11
2.1. Giới thiệu doanh nghiệp .....	11
2.2. Thực trạng của doanh nghiệp .....	11
2.3. Bài toán của doanh nghiệp và mục tiêu cần giải quyết .....	13
CHƯƠNG 3: QUÁ TRÌNH ETL .....	14
3.1. Giai đoạn 1: Trích xuất dữ liệu (Extract) .....	14
3.2. Giai đoạn 2: Chuyển đổi dữ liệu (Transform) .....	25
3.3. Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load).....	32
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU .....	40
4.1. Phân tích tổng quan tình hình doanh nghiệp (Dashboard Overview) .....	40
4.2. Phân tích nhánh Product.....	42
4.3. Phân tích nhánh Review .....	46

4.4. Phân tích nhánh Location .....	48
4.5. Phân tích nhánh Date .....	51
4.6. Phân tích nhánh Payment .....	54
CHƯƠNG 5: ĐỀ XUẤT GIẢI PHÁP .....	56
CHƯƠNG 6: KẾT LUẬN .....	59
6.1. Kết luận.....	59
6.2. Hướng phát triển.....	60
PHỤ LỤC.....	61

## DANH MỤC HÌNH ẢNH

Hình 3. 1 Kiểm tra chất lượng bảng olist_customers_dataset .....	20
Hình 3. 2 Kiểm tra chất lượng bảng olist_order_reviews_dataset.....	20
Hình 3. 3 Kiểm tra chất lượng bảng olist_order_payments_dataset.....	21
Hình 3. 4 Kiểm tra chất lượng bảng olist_order_items_dataset .....	21
Hình 3. 5 Kiểm tra chất lượng bảng product_category_name_translation .....	21
Hình 3. 6 Kiểm tra chất lượng bảng olist_geolocation_dataset.....	22
Hình 3. 7 Kiểm tra chất lượng bảng olist_sellers_dataset .....	22
Hình 3. 8 Kiểm tra chất lượng bảng olist_products_dataset.....	22
Hình 3. 9 Kiểm tra chất lượng bảng olist_orders_dataset .....	23
Hình 3. 10 Kết quả thống kê bảng olist_customers_dataset.....	23
Hình 3. 11 Kết quả thống kê bảng olist_order_reviews_dataset.....	23
Hình 3. 12 Kết quả thống kê bảng olist_order_payments_dataset.....	24
Hình 3. 13 Kết quả thống kê bảng olist_order_items_dataset.....	24
Hình 3. 14 Kết quả thống kê bảng product_category_name_translation.....	24
Hình 3. 15 Kết quả thống kê bảng olist_geolocation_dataset .....	24
Hình 3. 16 Kết quả thống kê bảng olist_sellers_dataset.....	24
Hình 3. 17 Kết quả thống kê bảng olist_products_dataset .....	25
Hình 3. 18 Kết quả thống kê bảng olist_orders_dataset.....	25
Hình 3. 19 Mô hình dữ liệu .....	39
Hình 4. 1 Dashboard phân tích tổng quan doanh nghiệp.....	41
Hình 4. 2 Dashboard phân tích nhánh Product.....	44
Hình 4. 3 Dashboard phân tích nhánh Review .....	47
Hình 4. 4 Dashboard phân tích nhánh Location .....	49
Hình 4. 5 Dashboard phân tích nhánh Date .....	53
Hình 4. 6 Dashboard phân tích nhánh Payment.....	55

## DANH MỤC BẢNG

<i>Bảng 3. 1 Bảng olist_customers_dataset</i> .....	15
<i>Bảng 3. 2 Bảng olist_sellers_dataset</i> .....	16
<i>Bảng 3. 3 Bảng olist_orders_dataset</i> .....	16
<i>Bảng 3. 4 Bảng olist_order_items_dataset</i> .....	17
<i>Bảng 3. 5 Bảng olist_order_payments_dataset</i> .....	18
<i>Bảng 3. 6 Bảng olist_order_reviews_dataset</i> .....	18
<i>Bảng 3. 7 Bảng olist_geolocation_dataset</i> .....	19
<i>Bảng 3. 8 Bảng olist_products_dataset</i> .....	19
<i>Bảng 3. 9 Bảng product_category_name_translation</i> .....	20
<i>Bảng 3. 10 Các cột dữ liệu cần xoá</i> .....	26
<i>Bảng 3. 11 Bảng dim_customers</i> .....	32
<i>Bảng 3. 12 Bảng dim_date</i> .....	33
<i>Bảng 3. 13 Bảng dim_geolocation</i> .....	34
<i>Bảng 3. 14 Bảng dim_order_payments</i> .....	34
<i>Bảng 3. 15 Bảng dim_order_reviews</i> .....	35
<i>Bảng 3. 16 Bảng dim_products</i> .....	35
<i>Bảng 3. 17 Bảng fact_order_items</i> .....	36

# CHƯƠNG 1: TỔNG QUAN

## 1.1. Lý do chọn đề tài

Thương mại điện tử đã xuất hiện và tăng trưởng một cách nhanh chóng tại Brazil. Chúng ta có thể biết đến một số công ty, sàn thương mại điện tử nổi tiếng trên thế giới như Amazon, Walmart, Home Depot, Alibaba, Reliance Industries,... Và Olist store là một trong những doanh nghiệp thương mại điện tử hàng đầu tại Brazil. Phân tích tình hình kinh doanh của Olist store sẽ cung cấp cái nhìn sâu rộng về thị trường thương mại điện tử Brazil trong giai đoạn năm 2015-2018. Trong giai đoạn này, thương mại điện tử tại Brazil đang trải qua sự tăng trưởng nhanh chóng, và Olist store là một phần của cơn sóng này. Hiểu rõ về kinh nghiệm của Olist store có thể cung cấp thông tin quý giá về cách thức vận hành và chiến lược thành công trong ngành này. Không chỉ vậy, bộ dữ liệu kinh doanh từ năm 2016-2018 của Olist Store đã được công bố và đăng tải trên Kaggle cho nên việc chọn Olist Store làm đối tượng nghiên cứu đem lại sự thuận tiện cho việc phân tích và nghiên cứu. Dữ liệu này có thể được sử dụng để phân tích xu hướng, đánh giá hiệu suất và dự đoán phát triển trong tương lai. Sau đại dịch Covid-19, suy thoái kinh tế và những hậu quả mà đại dịch để lại, việc dự đoán xu hướng phát triển thương mại điện tử tại Brazil từ kết quả của Olist store có thể giúp các doanh nghiệp khác trong ngành hiểu rõ hơn về những thách thức và cơ hội mà họ có thể đối mặt trong tương lai khi gia nhập vào ngành thương mại điện tử trong giai đoạn hiện nay. Bên cạnh đó, nghiên cứu này có thể đóng góp vào việc thúc đẩy nghiên cứu trong lĩnh vực thương mại điện tử và đưa ra những phân tích chi tiết về thị trường Brazil, giúp cải thiện chiến lược kinh doanh và quản lý rủi ro cho các doanh nghiệp.

## 1.2. Mục đích, mục tiêu của đề tài

Mục tiêu của đề tài là hiểu rõ về hoạt động kinh doanh của Olist store thông qua việc phân tích các chỉ số kinh doanh như doanh số bán hàng, lợi nhuận, tỷ lệ chuyển đổi, và thị phần của Olist store từ năm 2016 đến 2018 nhằm hiểu rõ hơn về hiệu suất hoạt động của công ty trong giai đoạn 2016-2018. Ngoài ra, nghiên cứu này còn nhằm xác

định các yếu tố ảnh hưởng đến sự phát triển của Olist store thông qua phân tích các yếu tố nội và ngoại lực như chiến lược kinh doanh, cạnh tranh, yêu cầu của khách hàng, và điều kiện thị trường để hiểu rõ những yếu tố nào đã ảnh hưởng đến sự phát triển của Olist store trong thời gian đã qua. Từ đây, các doanh nghiệp mới gia nhập ngành có thể phân tích được ưu nhược điểm, những bài học cần thiết cũng như xác định được chìa khóa thành công của một doanh nghiệp khi mới bắt đầu gia nhập ngành thương mại điện tử tại Brazil. Không chỉ vậy, việc dự đoán xu hướng phát triển thông qua các dữ liệu về tình hình kinh doanh của Olist store cùng với các yếu tố thị trường lớn hơn để dự đoán xu hướng phát triển của ngành thương mại điện tử tại Brazil trong tương lai là vô cùng cần thiết khi các doanh nghiệp có thể dựa vào những phân tích này để có những chiến lược cạnh tranh hợp lý và chiến lược phát triển bền vững, đột phá khi phần lớn thị phần của thương mại điện tử trên thị trường hiện nay đang được chiếm đóng bởi các ông lớn như Amazon, eBay, Alibaba,...Điều này có thể bao gồm dự đoán về tốc độ tăng trưởng, mô hình kinh doanh mới xuất hiện, và thách thức tiềm ẩn. Và, dựa trên những kết quả phân tích và dự đoán, nghiên cứu này đề xuất các chiến lược và hướng đi mới để Olist store có thể tối ưu hóa hiệu suất kinh doanh và cũng như để ngành thương mại điện tử tại Brazil tiếp tục phát triển bền vững trong tương lai.

### **1.3. Đối tượng và phạm vi thực hiện**

Đối tượng nghiên cứu trong nghiên cứu này có thể xác định là tình hình kinh doanh của Olist store từ năm 2016 đến năm 2018 và xu hướng phát triển ngành thương mại điện tử tại Brazil.

Phạm vi nghiên cứu: Dữ liệu về tình hình kinh doanh từ năm 2016-2018 của Olist Store tại Brazil.

### **1.4. Phương pháp thực hiện**

Nhóm thực hiện thống kê mô tả bằng R studio; chuyển đổi dữ liệu, nạp dữ liệu và trực quan hóa dữ liệu thông qua phần mềm Power BI.



### 1.5. Bố cục của đề tài

Bố cục của đề tài được chia thành 6 chương. Cụ thể bao gồm:

Chương 1: Tổng quan

Chương 2: Mô tả doanh nghiệp

Chương 3: Quá trình ETL

Chương 4: Phân tích dữ liệu

Chương 5: Đề xuất giải pháp

Chương 6: Kết luận

### 1.6. Phân công công việc

Họ và tên	MSSV	Nhiệm vụ	Đánh giá mức độ hoàn thành
Phan Thị Bích Diệu	31211022667	<ul style="list-style-type: none"><li>- Tìm kiếm dữ liệu, tải dữ liệu, tổng hợp dữ liệu.</li><li>- Thực hiện Giai đoạn 3 (Load) trong quá trình ETL.</li><li>- Tạo và phân tích dashboard Location.</li><li>- Tổng hợp và chỉnh sửa hoàn thiện các file Power BI</li></ul>	100%
Nguyễn Nữ Hoàng Khánh	31211027957	<ul style="list-style-type: none"><li>- Thực hiện các mục: Lý do chọn đề tài, mục đích, mục tiêu của đề tài, đối tượng và phạm vi thực hiện và phương pháp thực hiện đề tài.</li><li>- Tìm kiếm và mô tả doanh nghiệp.</li><li>- Tạo và phân tích dashboard Overview.</li><li>- Thực hiện đưa ra đề xuất và giải pháp cho đề tài.</li></ul>	100%
Phùng Trần Văn Quang	31211022629	<ul style="list-style-type: none"><li>- Tạo và phân tích dashboard Date và Payment</li></ul>	100%

Nguyễn Trương Thanh Việt	31211027356	<ul style="list-style-type: none"> <li>- Thực hiện Giai đoạn 1 (Extract) trong quá trình ETL.</li> <li>- Đánh giá chất lượng bộ dữ liệu.</li> <li>- Tạo và phân tích dashboard Review.</li> <li>- Kết luận</li> </ul>	100%
Trần Quang Vinh	31211022720	<ul style="list-style-type: none"> <li>- Thực hiện các mục: Thống kê mô tả, Phân tích nhánh Product, Hướng phát triển của đề tài</li> <li>- Viết code R phân tích thống kê mô tả bộ dữ liệu</li> <li>- Xử lý trên Excel chuyển dữ liệu gốc thành các sheet Dim &amp; Fact, phục vụ quá trình ETL</li> <li>- Tạo và phân tích dashboard Product.</li> </ul>	100%
Nguyễn Thanh Vy	31211027689	<ul style="list-style-type: none"> <li>- Thực hiện Giai đoạn 2 (Transform) trong quá trình ETL.</li> <li>- Thực hiện đưa ra đề xuất và giải pháp cho doanh nghiệp.</li> <li>- Tổng hợp bài báo cáo (file word)</li> </ul>	100%

## **CHƯƠNG 2: MÔ TẢ DOANH NGHIỆP**

### **2.1. Giới thiệu doanh nghiệp**

Olist là một nền tảng thương mại điện tử tiên tiến được thành lập tại Brazil vào năm 2015. Olist cung cấp giải pháp toàn diện cho các doanh nghiệp nhỏ và vừa (SMEs) để bắt đầu và phát triển kinh doanh trực tuyến của họ một cách hiệu quả. Một phần quan trọng của mô hình kinh doanh của Olist là việc họ tập trung vào việc hợp tác với các doanh nghiệp nhỏ, giúp chúng tối ưu hóa hoạt động thương mại điện tử của mình thông qua việc cung cấp các công cụ và dịch vụ đa dạng.

Olist cung cấp một nền tảng thương mại điện tử tích hợp, cho phép các doanh nghiệp quản lý đồng thời các hoạt động như quản lý sản phẩm, đặt hàng, quản lý kho, giao hàng và dịch vụ khách hàng. Điều này giúp giảm bớt gánh nặng cho các doanh nghiệp nhỏ và vừa, cho phép họ tập trung hơn vào việc phát triển sản phẩm và tăng cường trải nghiệm khách hàng.

Một điểm đặc biệt khác của Olist là họ cung cấp thị trường trực tuyến, giúp các doanh nghiệp nhỏ và vừa tiếp cận một cơ hội lớn hơn để tiếp cận khách hàng. Bằng cách kết nối với Olist, các doanh nghiệp có thể tiếp cận một lượng lớn người mua hàng trực tuyến đã có sẵn trên nền tảng, giúp tăng doanh số bán hàng và mở rộng quy mô kinh doanh.

### **2.2. Thực trạng của doanh nghiệp**

#### *2.2.1. Tăng trưởng doanh số bán hàng*

Từ năm 2016 đến 2018, Olist Store đã ghi nhận sự tăng trưởng đáng kể trong doanh số bán hàng. Sự tăng trưởng này phản ánh sự thành công của các chiến lược kinh doanh như mở rộng thị trường, tối ưu hóa trải nghiệm mua sắm, và tăng cường quảng cáo và tiếp thị.

#### *2.2.2. Mở rộng danh mục sản phẩm và dịch vụ*

Trong thời gian này, Olist Store đã không ngừng mở rộng danh mục sản phẩm và dịch vụ của mình để đáp ứng nhu cầu ngày càng đa dạng của khách hàng. Họ đã không chỉ mở rộng về độ phong phú của sản phẩm mà còn mở rộng về các dịch vụ hỗ trợ khác như dịch vụ vận chuyển, bảo hành và hỗ trợ khách hàng.

Việc mở rộng danh mục sản phẩm và dịch vụ giúp Olist Store thu hút một đối tượng khách hàng đa dạng hơn và tạo ra nhiều cơ hội kinh doanh mới.

#### *2.2.3. Tăng cường mối quan hệ với các đối tác*

Olist Store đặc biệt chú trọng vào việc xây dựng và tăng cường mối quan hệ với các đối tác cung cấp sản phẩm và dịch vụ. Họ đã thực hiện các hợp đồng đối tác chiến lược và cam kết cung cấp sản phẩm chất lượng và dịch vụ đáng tin cậy. Việc tăng cường mối quan hệ với các đối tác giúp Olist Store đảm bảo nguồn cung ổn định và chất lượng, đồng thời cũng mở ra cơ hội hợp tác chiến lược và tối ưu hóa chi phí.

#### *2.2.4. Chiến lược tiếp thị và quảng cáo*

Olist Store đã triển khai một loạt các chiến lược tiếp thị và quảng cáo để tăng cường nhận thức thương hiệu và thu hút khách hàng mới. Các chiến dịch tiếp thị bao gồm quảng cáo trên các nền tảng truyền thông xã hội, tìm kiếm trực tuyến, và email marketing.

Bằng cách tập trung vào việc xây dựng một thương hiệu mạnh mẽ và thu hút khách hàng mới, Olist Store đã có thể tăng cường vị thế của mình trên thị trường và tạo ra một lượng khách hàng trung thành.

#### *2.2.5. Phát triển công nghệ*

Olist Store đã đầu tư mạnh mẽ vào phát triển công nghệ để cải thiện trải nghiệm mua sắm của khách hàng và tối ưu hóa hoạt động kinh doanh. Olist đã phát triển và cải thiện các ứng dụng di động, giao diện người dùng trên trang web, và hệ thống quản lý đơn hàng và giao hàng.

Việc sử dụng công nghệ tiên tiến giúp Olist Store tối ưu hóa quy trình kinh doanh, tăng cường hiệu suất và cải thiện trải nghiệm mua sắm của khách hàng.

#### *2.2.6. Thị phần và vị thế trong ngành*

Nhờ vào các chiến lược kinh doanh hiệu quả và sự tăng trưởng ấn tượng, Olist Store đã có thể tăng cường vị thế của mình trong ngành thương mại điện tử tại Brazil. Olist đã tăng thị phần của mình và trở thành một trong những người dẫn đầu trong lĩnh vực này tại Brazil.

Nêu chiến lược kinh doanh của Olist Store, cách họ tăng thị phần và vị thế trong ngành.

Sự tăng cường vị thế của Olist Store không chỉ là kết quả của hiệu suất kinh doanh tích cực mà còn là kết quả của các nỗ lực trong việc xây dựng một thương hiệu mạnh mẽ và mối quan hệ đối tác chiến lược.

### **2.3. Bài toán của doanh nghiệp và mục tiêu cần giải quyết**

Là một công ty mới thành lập tại Brazil từ năm 2015, Olist Store phải đối diện với những thách thức khó khăn để bắt kịp xu hướng phát triển của ngành thương mại điện tử. Các bài toán đó có thể được xác định như sau:

- Tăng doanh số bán hàng: Một trong những mục tiêu chính của Olist có thể là tăng doanh số bán hàng bằng cách thu hút nhiều khách hàng mới hơn và tăng tần suất mua hàng của khách hàng hiện tại. Đồng thời, mở rộng thị phần và phát triển các chất lượng sản phẩm nhằm thúc đẩy sự phát triển của doanh nghiệp một cách bền vững là một trong những mục tiêu chính của Olist Store.
- Phân tích và thấu hiểu khách hàng: Để triển khai các chiến dịch marketing một cách hiệu quả, Olist Store cần thấu hiểu nhu cầu của khách hàng, điểm thỏa mãn khách hàng là điều gì để từ đó có thể xác định đúng và đưa ra các chiến lược cụ thể cho từng tệp khách hàng với những hành vi tiêu dùng khác nhau.
- Xây dựng mối quan hệ với khách hàng: Nhằm giữ chân khách hàng cũ và thu hút khách hàng mới một cách hiệu quả, xây dựng mối quan hệ với khách hàng

là một trong những yếu tố cần thiết trong quá trình phát triển và mở rộng thị trường.

- Tối ưu hóa chi phí vận hành: Tìm cách giảm chi phí vận hành mà không ảnh hưởng đến chất lượng dịch vụ và sản phẩm.

## CHƯƠNG 3: QUÁ TRÌNH ETL

### 3.1. Giai đoạn 1: Trích xuất dữ liệu (Extract)

#### 3.1.1. Trích xuất dữ liệu

Bộ dữ liệu trên được lấy trực tiếp trên trang web của kaggle, một nền tảng trực tuyến cho cộng đồng Machine Learning (ML) và Data Science. Kaggle cho phép người dùng chia sẻ, tìm kiếm các bộ dữ liệu; tìm hiểu và xây dựng models. Ngoài ra, đó còn là nơi để những người nghiên cứu về data science có thể chia sẻ kiến thức, dữ liệu và những người mới bắt đầu học về machine learning có thể áp dụng những kiến thức của mình vào những project thực tế.

#### 3.1.2. Mô tả bộ dữ liệu

Đây là tập dữ liệu công khai về thương mại điện tử của Brazil về các đơn đặt hàng được thực hiện tại Olist Store. Tập dữ liệu có thông tin về 100 nghìn đơn đặt hàng từ năm 2016 đến năm 2018 được thực hiện tại nhiều thị trường ở Brazil. Các tính năng của nó cho phép xem đơn đặt hàng từ nhiều khía cạnh: từ trạng thái đơn hàng, giá cả, thanh toán và hiệu suất vận chuyển đến vị trí của khách hàng, thuộc tính sản phẩm và cuối cùng là đánh giá do khách hàng viết.

Tập dữ liệu này được cung cấp rộng rãi bởi Olist, một trong những cửa hàng bách hóa lớn nhất ở các khu chợ ở Brazil. Olist kết nối các doanh nghiệp nhỏ từ khắp Brazil với các kênh mà không gặp rắc rối và chỉ bằng một hợp đồng duy nhất. Những người bán này có thể bán sản phẩm của mình thông qua Cửa hàng Olist và vận chuyển trực tiếp đến khách hàng thông qua 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 ghi chú về trải nghiệm mua hàng và viết ra một số nhận xét.

Bộ dữ liệu của Olist bao gồm 9 bảng:

- olist\_customers\_dataset
- olist\_sellers\_dataset
- olist\_orders\_dataset
- olist\_order\_items\_dataset
- olist\_order\_payments\_dataset
- olist\_order\_reviews\_dataset
- olist\_geolocation\_dataset
- olist\_products\_dataset
- product\_category\_name\_translation

Các bảng được trình bày chi tiết dưới đây:

*Bảng 3. 1 Bảng olist\_customers\_dataset*

STT	Tên thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	customer_id	Character	Khoá của tập dữ liệu order, mỗi đơn hàng sẽ có một customer_id duy nhất	
2	customer_unique_id	Character	Mã số khách hàng	Unique
3	customer_zip_code_prefix	Double	5 chữ số đầu tiên của mã zip của khách hàng	
4	customer_city	Character	Thành phố của khách hàng	

5	customer_state	Character	Bang của khách hàng	
---	----------------	-----------	---------------------	--

*Bảng 3. 2 Bảng olist\_sellers\_dataset*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	seller_id	Character	Mã số nhân viên bán hàng	Unique
2	seller_zip_code_prefix	Double	5 chữ số đầu tiên của mã zip của nhân viên	
3	seller_city	Character	Thành phố của nhân viên	
4	seller_state	Character	Bang của nhân viên	

*Bảng 3. 3 Bảng olist\_orders\_dataset*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	order_id	Character	Mã số đơn hàng	Unique
2	customer_id	Character	Khóa của tập dữ liệu customers. Mỗi đơn hàng có một customer_id duy nhất.	
3	order_status	Character	Tình trạng đơn hàng	
4	order_purchase_timestamp	Double	Thời gian mua hàng	
5	order_approved_at	Double	Thời gian xác nhận thanh toán	
6	order_delivered_carrier_date	Double	Thời gian xác nhận đơn hàng đã được xử lý	



			bởi đơn vị vận chuyển	
7	order_delivered_customer_date	Double	Thời gian đơn hàng đã giao đến khách hàng	
8	order_estimated_delivery_date	Double	Thời gian dự kiến giao hàng (được thông báo đến khách hàng khi thực hiện thanh toán)	

*Bảng 3. 4 Bảng olist\_order\_items\_dataset*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	order_id	Character	Mã số đơn hàng	Unique
2	order_item_id	Double	Số lượng mặt hàng có trong đơn hàng	
3	product_id	Character	Mã số sản phẩm	
4	seller_id	Character	Mã số nhân viên bán hàng	
5	shipping_limit_date	Double	Ngày giới hạn hoàn tất đơn hàng của người bán hàng để bàn giao cho đơn vị vận chuyển	
6	price	Double	Giá mặt hàng	
7	freight_value	Double	Giá trị vận chuyển mặt hàng (nếu một đơn hàng có nhiều hơn một mặt hàng, giá trị vận chuyển sẽ được chia cho các mặt hàng)	

*Bảng 3. 5 Bảng olist\_order\_payments\_dataset*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	order_id	Character	Mã số đơn hàng	Unique
2	payment_sequential	Double	Nếu khách hàng có thể thanh toán đơn hàng bằng nhiều phương thức, một chuỗi sẽ được tạo ra để đáp ứng tất cả các khoản thanh toán.	
3	payment_type	Character	Hình thức thanh toán	
4	payment_installments	Double	Số lần thanh toán trả góp	
5	payment_value	Double	Giá trị giao dịch	

*Bảng 3. 6 Bảng olist\_order\_reviews\_dataset*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	review_id	Character	Mã số đánh giá	Unique
2	order_id	Character	Mã số đơn hàng	
3	review_score	Double	Điểm đánh giá mức độ hài lòng của khách hàng cho đơn hàng (từ 1 đến 5)	
4	review_comment_title	Character	Tiêu đề nhận xét từ bài đánh giá của khách hàng	
5	review_comment_message	Character	Bình luận từ bài đánh giá của khách hàng	
6	review_creation_date	Double	Thời gian tạo bài đánh giá	
7	review_answer_timestamp	Double	Thời gian phản hồi bài đánh giá	

Bảng 3. 7 Bảng olist\_geolocation\_dataset

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	geolocation_zip_code_prefix	Double	5 chữ số đầu tiên của mã zip	
2	geolocation_lat	Double	Vĩ độ	
3	geolocation_lng	Double	Kinh độ	788
4	geolocation_city	Character	Tên thành phố	
5	geolocation_state	Character	Tên bang	


Bảng 3. 8 Bảng olist\_products\_dataset

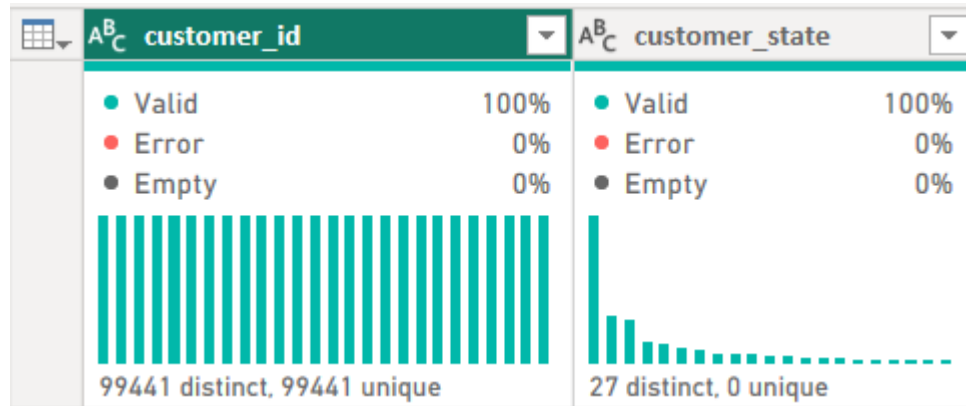
STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	product_id	Character	Mã số sản phẩm	Unique
2	product_category_name	Character	Tên gốc của sản phẩm (tiếng Bồ Đào Nha)	
3	product_name_lenght	Double	Số lượng ký tự của tên sản phẩm	
4	product_description_lenght	Double	Số lượng ký tự của mô tả sản phẩm	
5	product_photos_qty	Double	Số lượng hình ảnh của sản phẩm	
6	product_weight_g	Double	Cân nặng của sản phẩm (gam)	
7	product_length_cm	Double	Chiều dài sản phẩm (cm)	
8	product_height_cm	Double	Chiều cao sản phẩm (cm)	
9	product_width_cm	Double	Chiều rộng sản phẩm (cm)	

Bảng 3. 9 Bảng *product\_category\_name\_translation*


STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	product_category_name	Character	Tên gốc sản phẩm bằng Tiếng Bồ Đào Nha	
2	product_category_name_english	Character	Tên sản phẩm bằng Tiếng Anh	

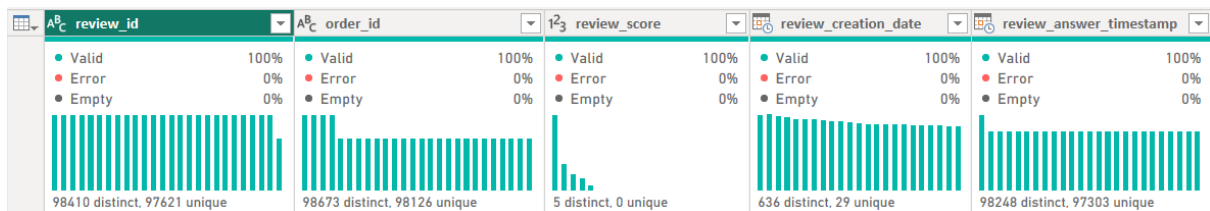
### 3.1.3. Đánh giá chất lượng của dữ liệu

 *olist\_customers\_dataset*




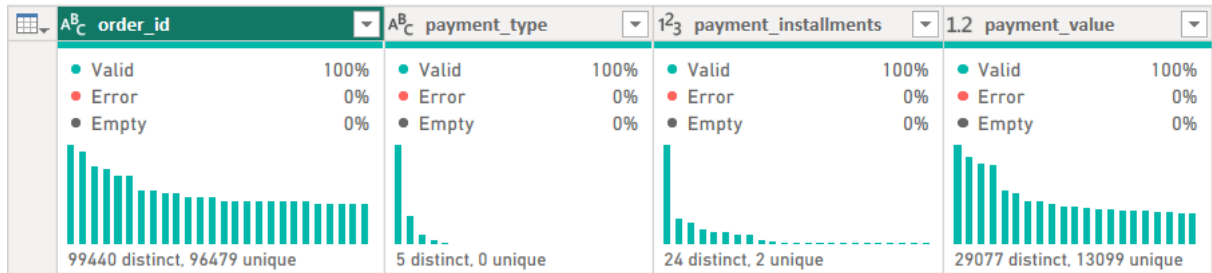
Hình 3. 1 Kiểm tra chất lượng bảng *olist\_customers\_dataset*

 *olist\_order\_reviews\_dataset*



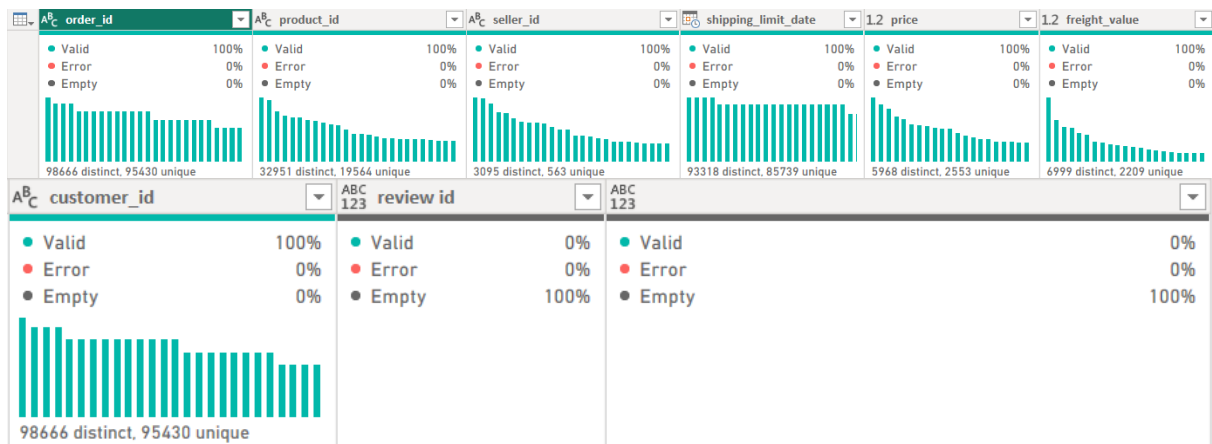
Hình 3. 2 Kiểm tra chất lượng bảng *olist\_order\_reviews\_dataset*

 *olist\_order\_payments\_dataset*



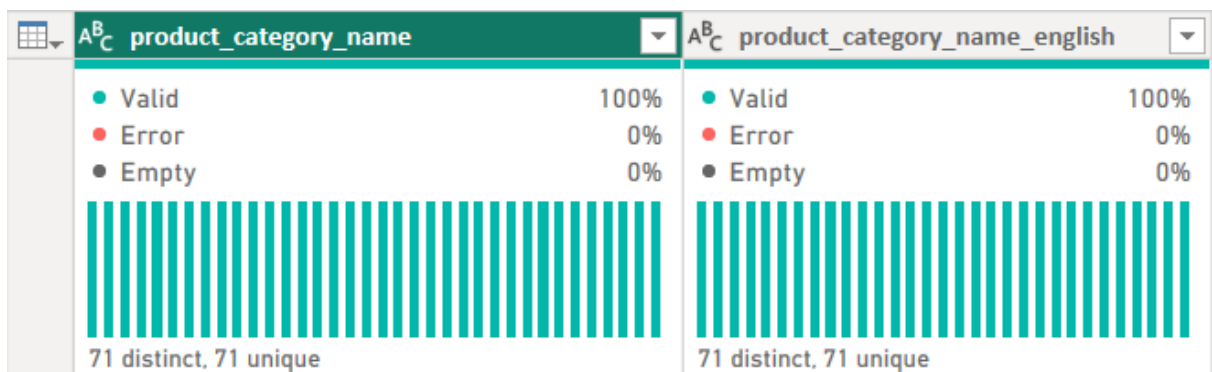
Hình 3. 3 Kiểm tra chất lượng bảng olist\_order\_payments\_dataset

+ olist\_order\_items\_dataset



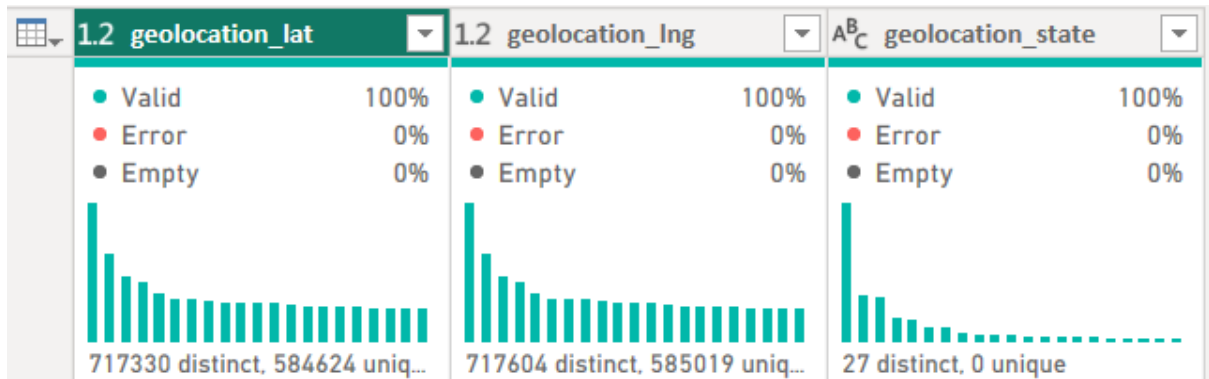
Hình 3. 4 Kiểm tra chất lượng bảng olist\_order\_items\_dataset

+ product\_category\_name\_translation



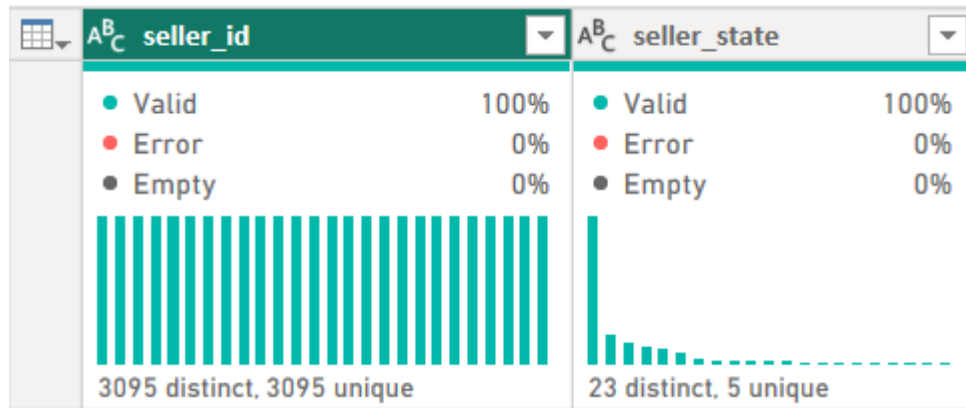
Hình 3. 5 Kiểm tra chất lượng bảng product\_category\_name\_translation

+ olist\_geolocation\_dataset



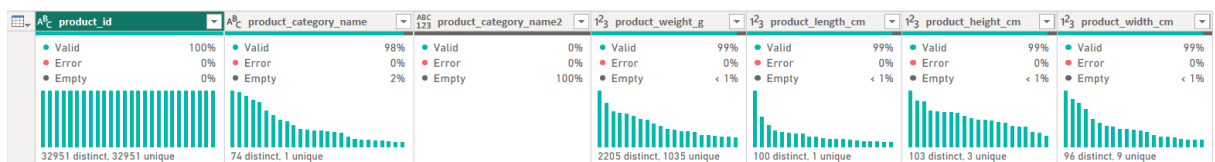
Hình 3. 6 Kiểm tra chất lượng bảng olist\_geolocation\_dataset

olist\_sellers\_dataset



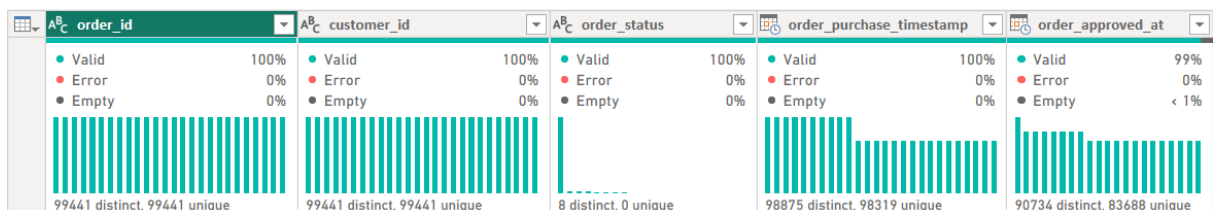
Hình 3. 7 Kiểm tra chất lượng bảng olist\_sellers\_dataset

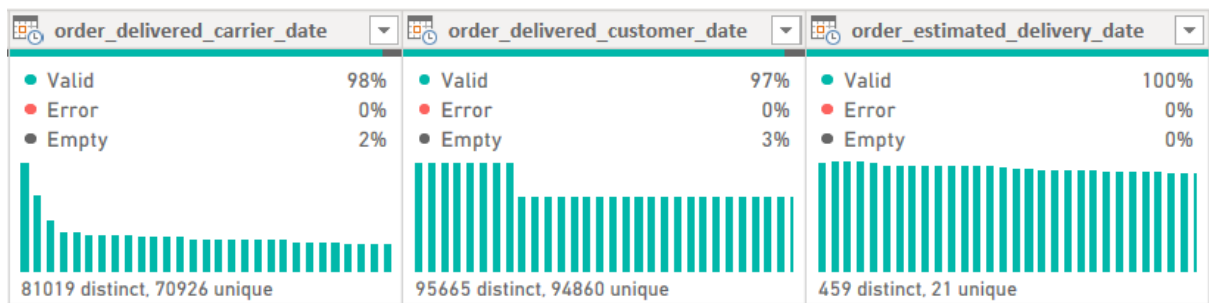
olist\_products\_dataset



Hình 3. 8 Kiểm tra chất lượng bảng olist\_products\_dataset

olist\_orders\_dataset





Hình 3. 9 Kiểm tra chất lượng bảng olist\_orders\_dataset

Nhìn vào các bảng trên ta có thể nhận thấy rõ ràng rằng các bảng dữ liệu đều mang đến kết quả hợp lệ khá cao trên 90%. Đồng thời thì phần trăm dữ liệu bị lỗi và rỗng chiếm tỷ lệ khá ít. Từ đó cho thấy đây là một bộ dữ liệu đáng tin cậy và có thể dùng nó để tiến hành phân tích và giúp mang lại các kết quả tốt trong quá trình thực hiện đồ án.

### 3.1.4. Thống kê mô tả

Nhóm đã thực hiện thống kê mô tả trên RStudio và thu lại được các kết quả sau:

#### 📊 Bảng olist\_customers\_dataset

```
> summary(customerdataset)
customer_id      customer_unique_id  customer_zip_code_prefix  customer_city      customer_state
Length:99441      Length:99441      Min.   : 1003              Length:99441      Length:99441
Class :character  Class :character  1st Qu.:11347             Class :character  Class :character
Mode  :character  Mode  :character  Median :24416             Mode  :character  Mode  :character
                        Mean   :35137
                        3rd Qu.:58900
                        Max.   :99990
```

Hình 3. 10 Kết quả thống kê mô tả bảng olist\_customers\_dataset

#### 📊 Bảng olist\_order\_reviews\_dataset

```
> summary(reviewsdataset)
review_id      order_id      review_score  review_comment_title  review_comment_message  review_creation_date  review_answer_timestamp
Length:99224      Length:99224      Min.   :1.000      Length:99224      Length:99224      Min.   :2016-10-02 00:00:00.00      Min.   :2016-10-07 18:32:28.00
Class :character  Class :character  1st Qu.:4.000      Class :character  Class :character  1st Qu.:2017-09-23 00:00:00.00      1st Qu.:2017-09-27 01:53:27.25
Mode  :character  Mode  :character  Median :5.000      Mode  :character  Mode  :character  Median :2018-02-02 00:00:00.00      Median :2018-02-04 22:41:47.50
                        Mean   :4.086
                        3rd Qu.:5.000
                        Max.   :5.000
                        Mean   :2018-01-12 20:49:23.95
                        3rd Qu.:2018-05-16 00:00:00.00
                        Max.   :2018-08-31 00:00:00.00
                        Mean   :2018-01-16 00:23:56.98
                        3rd Qu.:2018-05-20 12:11:21.50
                        Max.   :2018-10-29 12:27:35.00
```

Hình 3. 11 Kết quả thống kê mô tả bảng olist\_order\_reviews\_dataset

#### 📊 Bảng olist\_order\_payments\_dataset

```
> summary(paymentsdataset)
order_id      payment_sequential payment_type      payment_installments payment_value
Length:103886 Min.   : 1.000      Length:103886 Min.   : 0.000      Min.   : 0.00
Class :character 1st Qu.: 1.000      Class :character 1st Qu.: 1.000      1st Qu.: 56.79
Mode  :character Median : 1.000      Mode  :character Median : 1.000      Median : 100.00
                Mean  : 1.093      Mean  : 2.853      Mean  : 154.10
                3rd Qu.: 1.000      3rd Qu.: 4.000      3rd Qu.: 171.84
                Max.   :29.000      Max.   :24.000      Max.   :13664.08
```

Hình 3. 12 Kết quả thống kê bảng *olist\_order\_payments\_dataset*

#### Bảng *olist\_order\_items\_dataset*

```
> summary(itemsdataset)
order_id      order_item_id      product_id      seller_id      shipping_limit_date      price      freight_value
Length:112650 Min.   : 1.000      Length:112650 Length:112650 Min.   :2016-09-19 00:15:34.00 Min.   : 0.85 Min.   : 0.00
Class :character 1st Qu.: 1.000      Class :character Class :character 1st Qu.:2017-09-20 20:57:27.50 1st Qu.: 39.90 1st Qu.: 13.08
Mode  :character Median : 1.000      Mode  :character Mode  :character Median :2018-01-26 13:59:35.00 Median : 74.99 Median : 16.26
                Mean  : 1.198      Mean  : 1.198      Mean  :2018-01-07 15:36:52.19 Mean  :120.65 Mean  : 19.99
                3rd Qu.: 1.000      3rd Qu.: 1.000      3rd Qu.:2018-05-10 14:34:00.75 3rd Qu.: 134.90 3rd Qu.: 21.15
                Max.   :21.000      Max.   :21.000      Max.   :2020-04-09 22:35:08.00 Max.   :6735.00 Max.   :409.68
```

Hình 3. 13 Kết quả thống kê bảng *olist\_order\_items\_dataset*

#### Bảng *product\_category\_name\_translation*

```
> summary(translatedataset)
product_category_name      product_category_name_english
Length:71                  Length:71
Class :character           Class :character
Mode  :character           Mode  :character
```

Hình 3. 14 Kết quả thống kê bảng *product\_category\_name\_translation*

#### Bảng *olist\_geolocation\_dataset*

```
> summary(geolocationdataset)
geolocation_zip_code_prefix geolocation_lat geolocation_lng geolocation_city geolocation_state
Min.   : 1001      Min.   : -36.61      Min.   : -101.47      Length:1000163      Length:1000163
1st Qu.:11075      1st Qu.: -23.60      1st Qu.: -48.57      Class :character     Class :character
Median :26530      Median : -22.92      Median : -46.64      Mode  :character     Mode  :character
Mean   :36574      Mean   : -21.18      Mean   : -46.39
3rd Qu.:63504      3rd Qu.: -19.98      3rd Qu.: -43.77
Max.   :99990      Max.   : 45.07      Max.   : 121.11
```

Hình 3. 15 Kết quả thống kê bảng *olist\_geolocation\_dataset*

#### Bảng *olist\_sellers\_dataset*

```
> summary(sellersdataset)
seller_id      seller_zip_code_prefix seller_city      seller_state
Length:3095      Min.   : 1001      Length:3095      Length:3095
Class :character 1st Qu.: 7094      Class :character Class :character
Mode  :character Median :14940      Mode  :character Mode  :character
                Mean  :32291
                3rd Qu.:64553
                Max.   :99730
```

Hình 3. 16 Kết quả thống kê bảng *olist\_sellers\_dataset*



### Bảng olist\_products\_dataset

```
> summary(productsdataset)
```

product_id	product_category_name	product_name_length	product_description_length	product_photos_qty	product_weight_g	product_length_cm	product_height_cm	product_width_cm
Length:32951	Length:32951	Min. : 5.00	Min. : 4.0	Min. : 1.000	Min. : 0	Min. : 7.00	Min. : 2.00	Min. : 6.0
Class :character	Class :character	1st Qu.:42.00	1st Qu.: 339.0	1st Qu.: 1.000	1st Qu.: 300	1st Qu.: 18.00	1st Qu.: 8.00	1st Qu.: 15.0
Mode :character	Mode :character	Median :51.00	Median :395.0	Median : 1.000	Median : 700	Median : 25.00	Median : 13.00	Median : 20.0
		Mean :48.48	Mean : 771.5	Mean : 2.189	Mean : 2276	Mean : 30.82	Mean : 16.94	Mean : 23.2
		3rd Qu.:57.00	3rd Qu.: 972.0	3rd Qu.: 3.000	3rd Qu.:1900	3rd Qu.: 38.00	3rd Qu.: 21.00	3rd Qu.: 30.0
		Max. :76.00	Max. :3992.0	Max. :20.000	Max. :40425	Max. :105.00	Max. :105.00	Max. :118.0
		NA's :610	NA's :610	NA's :610	NA's :2	NA's :2	NA's :2	NA's :2

Hình 3. 17 Kết quả thống kê bảng olist\_products\_dataset

### Bảng olist\_orders\_dataset

```
> summary(ordersdataset)
```

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date
Length:99441	Length:99441	Length:99441	Min. :2016-09-04 21:15:19.00	Min. :2016-09-15 12:16:38.0	Min. :2016-10-08 10:34:01.00
Class :character	Class :character	Class :character	1st Qu.:2017-09-12 14:46:19.00	1st Qu.:2017-09-12 23:24:16.0	1st Qu.:2017-09-15 22:28:50.25
Mode :character	Mode :character	Mode :character	Median :2018-01-18 23:04:36.00	Median :2018-01-19 11:36:13.0	Median :2018-01-24 16:10:58.00
			Mean :2017-12-31 08:43:12.77	Mean :2017-12-31 18:35:24.0	Mean :2018-01-04 21:49:48.14
			3rd Qu.:2018-05-04 15:42:16.00	3rd Qu.:2018-05-04 20:35:10.0	3rd Qu.:2018-05-08 13:37:45.00
			Max. :2018-10-17 17:30:18.00	Max. :2018-09-03 17:40:06.0	Max. :2018-09-11 19:48:28.00
				NA's :160	NA's :1783

order_delivered_customer_date	order_estimated_delivery_date
Min. :2016-10-11 13:46:32.00	Min. :2016-09-30 00:00:00.00
1st Qu.:2017-09-25 22:07:22.25	1st Qu.:2017-10-03 00:00:00.00
Median :2018-02-02 19:28:10.50	Median :2018-02-15 00:00:00.00
Mean :2018-01-14 12:09:19.03	Mean :2018-01-24 03:08:37.73
3rd Qu.:2018-05-15 22:48:52.25	3rd Qu.:2018-05-25 00:00:00.00
Max. :2018-10-17 13:22:46.00	Max. :2018-11-12 00:00:00.00
NA's :2965	

Hình 3. 18 Kết quả thống kê bảng olist\_orders\_dataset

## 3.2. Giai đoạn 2: Chuyển đổi dữ liệu (Transform)

### 3.2.1. Gộp dữ liệu

Với mục đích tạo cho các bảng dữ liệu trở nên gọn hơn nhưng vẫn chứa đầy đủ nội dung để phục vụ cho việc phân tích, nhóm tiến hành thao tác gộp dữ liệu trên Microsoft Excel theo các phần được liệt kê trong phần tiếp theo.

### Bảng olist\_products\_dataset

Như đã đề cập ở phần mô tả dữ liệu, bảng product\_category\_name chứa các dữ liệu bao gồm tên sản phẩm (tiếng Tây Ban Nha) và tên sản phẩm phiên âm sang tiếng Anh.

- Sử dụng hàm VLOOKUP để chuyển dữ liệu của cột product\_category\_name\_english từ bảng product\_category\_name sang bảng olist\_products\_dataset dựa trên giá trị cột product\_category\_name.
- Xoá bảng product\_category\_name.
- Xoá cột product\_category\_name (tiếng Tây Ban Nha) của bảng olist\_products\_dataset.

- Đổi tên cột `product_category_name_english` thành `product_category_name` của bảng `olist_products_dataset`.

#### *Bảng `fact_order`*

- Đổi tên bảng `olist_orders_dataset` thành `fact_order`.
- Sử dụng hàm `VLOOKUP` để chuyển dữ liệu từ 5 cột `product_id`, `seller_id`, `shipping_limit_date`, `price`, `freight_value` của bảng `olist_order_items_dataset` sang bảng `fact_order` dựa trên `order_id`.
- Xóa bảng `olist_order_items_dataset`.
- Ở bảng `fact_order`, tạo thêm cột `review id`, `payment type`, `state of seller`. Với giá trị của cột `review id` và `payment_type` được truy vấn dựa trên `order_id`; giá trị của cột `state of seller` được truy vấn dựa trên `seller_id`.

### 3.2.2. *Xóa dữ liệu*

Nhóm tiến hành thao tác xóa cột được liệt kê trong bảng bên dưới vì đây là những cột bị thiếu dữ liệu hoặc không chứa thông tin cần thiết để phân tích về tình hình bán hàng của doanh nghiệp Olist.

*Bảng 3. 10 Các cột dữ liệu cần xóa*

Tên bảng	Tên cột cần xóa
olist_customer_dataset	customer_unique_id
	customer_zip_code_prefix
	customer_city
olist_sellers_dataset	seller_zip_code_prefix
	seller_city
olist_order_payments_dataset	payment_sequential

olist_order_reviews_dataset	review_comment_title
	review_comment_message
olist_geolocation_dataset	geolocation_zip_code_prefix
	geolocation_zip_code_prefix
olist_products_dataset	product_name_lenght
	product_description_lenght
	product_photos_qty

### 3.2.3. Chuyển đổi dữ liệu

Các thao tác chuyển đổi dữ liệu sẽ được thực hiện trên phần mềm Power BI.

#### Bảng State code

Với mục đích nhằm lấy tên gọi của từng bang theo mã code đã có sẵn trong bộ dữ liệu ban đầu, nhóm sẽ thực hiện sao chép đường dẫn trang web chứa thông tin về mã code từng bang của Brazil trên Wikipedia.

The screenshot shows the Wikipedia page for 'Federative units of Brazil'. It includes three maps of Brazil at different historical points: 1889 (States at the start of the republic), 1943 (Border territories), and 1988 (Current states). Below the maps is a table listing the states and Federal District of Brazil with various statistics.

Flag and name	Code	Capital	Largest city	Area (km²) <sup>[20]</sup>	Population (May 2023) <sup>[21]</sup>	Density (per km², 2022)	GDP (R\$ millions, 2022) <sup>[22]</sup>	HDI (2022) <sup>[23]</sup>
Acre	AC	Rio Branco		164,123	830,018	6.34	21,000	0.719
Alagoas	AL	Maceió		27,779	3,127,683	125.52	73,000	0.683
Amapá	AP	Macapá		142,829	733,759	2.63	9000	0.740
Amazonas	AM	Manaus		1,559,159	3,941,613	2.58	103,000	0.733
Bahia	BA	Salvador		564,733	14,141,626	30.52	257,000	0.714
Ceará	CE	Fortaleza		148,921	8,794,957	60.33	157,000	0.735
Distrito Federal	DF	Brasília		5,780	2,817,381	493.00	249,000	0.850
Espírito Santo	ES	Vitória	Serra	46,096	3,833,712	80.63	125,000	0.772
Goiás	GO	Goiânia		340,112	7,056,495	18.46	202,000	0.769
Maranhão	MA	São Luís		331,937	6,776,699	19.03	102,000	0.687
Mato Grosso	MT	Cuiabá		903,366	3,658,649	4.01	142,000	0.774
Mato Grosso do Sul	MS	Campo Grande		357,146	2,880,308	7.83	107,000	0.766
Minas Gerais	MG	Belo Horizonte		586,522	21,279,353	31.72	583,000	0.787
Pará	PA	Belém		1,247,955	8,639,532	7.02	156,000	0.698
Paraíba	PB	João Pessoa		56,470	4,175,326	78.93	60,000	0.722
Paraná	PR	Curitiba		199,308	11,623,091	43.46	417,000	0.792

Sau đó, trở lại Power BI → Chọn “Get data from web”.

The screenshot shows the 'From Web' dialog box in Power BI. It has two tabs: 'Basic' (selected) and 'Advanced'. The 'URL' field contains the text 'https://en.wikipedia.org/wiki/Federative\_units\_of\_Brazil'. There are 'OK' and 'Cancel' buttons at the bottom right.

Sau khi thực hiện các bước trên, ta sẽ có được bảng State code với 2 cột được đưa vào sử dụng để phân tích là: *Flag and name* và *Code*, như hình bên dưới.

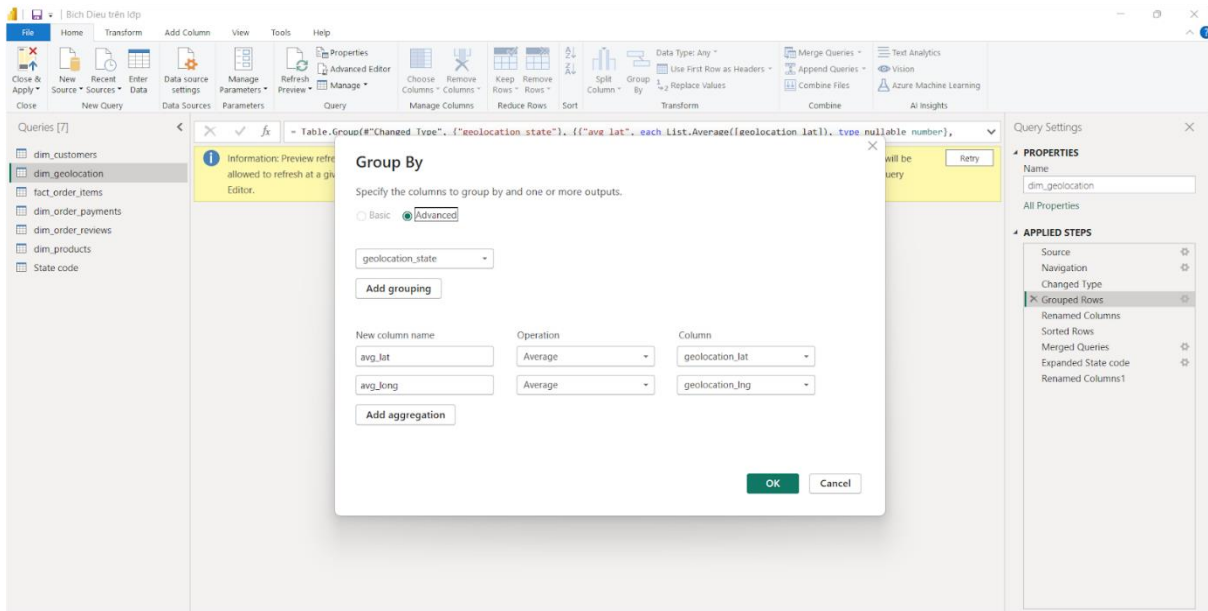
Flag and name		Code
Valid	100%	Valid 100%
Error	0%	Error 0%
Empty	0%	Empty 0%
1	Acre	AC
2	Alagoas	AL
3	Amapá	AP
4	Amazonas	AM
5	Bahia	BA
6	Ceará	CE
7	Distrito Federal	DF
8	Espírito Santo	ES
9	Goiás	GO
10	Maranhão	MA
11	Mato Grosso	MT
12	Mato Grosso do Sul	MS
13	Minas Gerais	MG
14	Pará	PA
15	Paraná	PR

📊 *Bảng dim\_geolocation*

**Changed Type:** Thay đổi kiểu dữ liệu của các cột sau:

- geolocation\_lat: decimal number
- geolocation\_lng: decimal number

**Grouped Rows:** Nhóm các giá trị theo State code để lấy ra chỉ số trung bình của vĩ độ (Latitude) và kinh độ (Longitude) đại diện cho GPS theo từng bang trong bản đồ.



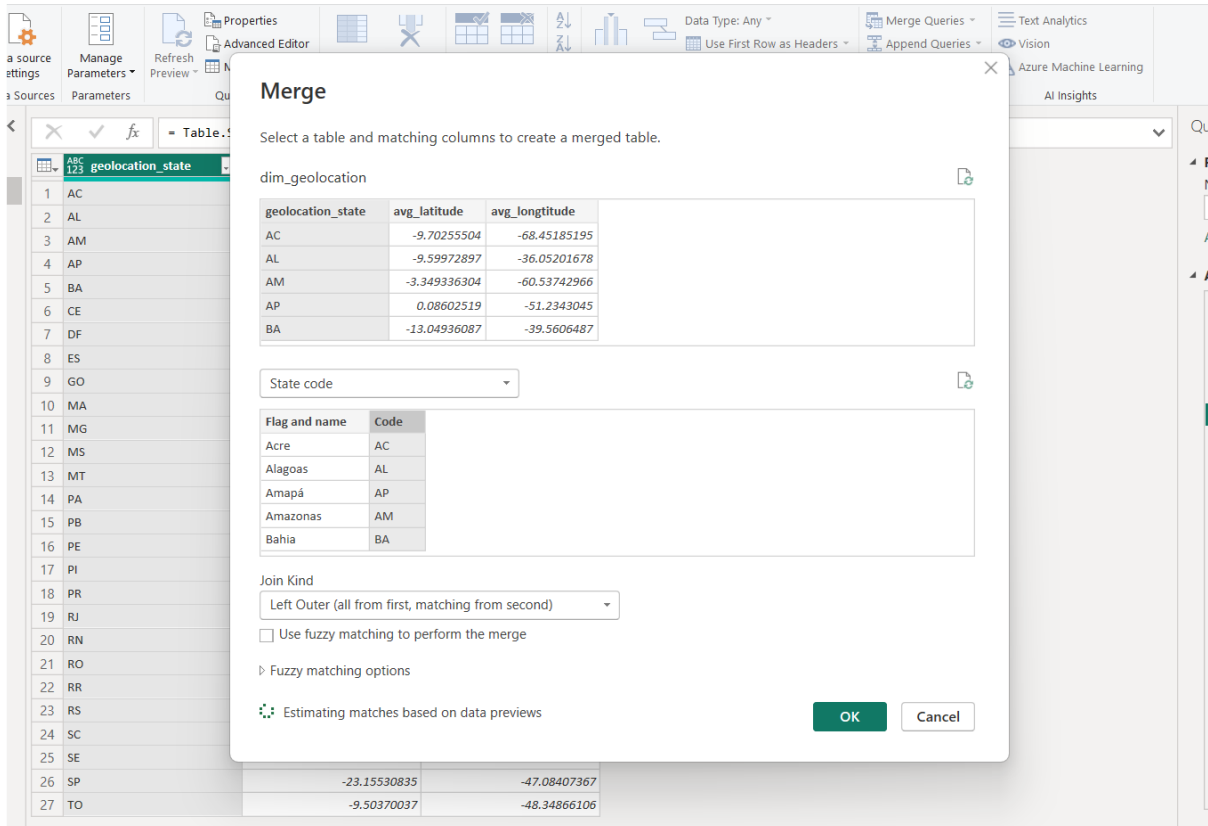
**Renamed Columns:** Tiến hành đổi tên các cột sau:

- avg\_lat thành avg\_latitude
- avg\_long thành avg\_longitude

```
= Table.RenameColumns("#"Grouped Rows",{{"avg_lat", "avg_latitude"}, {"avg_long", "avg_longitude"}})
```

**Sorted Rows:** Lọc giá trị cột geolocation\_state theo thứ tự tăng dần (từ A đến Z).

**Merged Queries:** Gộp bảng dim location và bảng State code để lấy tên gọi của từng bang theo mã và đổi tên cột State code.Flag and name thành State name.



### *Bảng fact\_order\_items*


**Changed Type:** Thay đổi kiểu dữ liệu của các cột sau:

- order\_value: decimal number
- price: decimal number
- freight\_value: decimal number
- shipping\_limit\_date: date
- order\_purchase\_timestamp: date
- order\_approved\_at: date
- order\_delivered\_carrier\_date: date
- order\_delivered\_customer\_date: date
- order\_estimated\_delivery\_date: date

### *Bảng dim\_order\_reviews*


**Changed Type:** Thay đổi kiểu dữ liệu của các cột sau:

- review\_creation\_date: date
- review\_answer\_timestamp: date
- review\_score: whole number

 *Bảng dim\_order\_payments*

**Changed Type:** Thay đổi kiểu dữ liệu của các cột sau:

- payment\_value: decimal number
- payment\_installments: whole number
- payment\_type: text

 *Bảng dim\_products*

**Changed Type:** Thay đổi kiểu dữ liệu của các cột sau:

- product\_weight\_g: decimal number
- product\_length\_cm: decimal number
- product\_height\_cm: decimal number
- product\_width\_cm: decimal number

### 3.3. Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load)

#### 3.3.1. Giới thiệu các bảng Dim → n Dim

 *Bảng dim\_customers*

*Bảng 3. 11 Bảng dim\_customers*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả
1	customer_id	Text	5 chữ số đầu tiên của mã zip
2	customer_state	Text	Bang của khách hàng



### *Bảng dim\_date*

Nhóm tiến hành tạo bảng dim\_date bằng cách sử dụng hàm DAX trong Power BI, bao gồm năm, quý, tháng, tuần, ngày. Bảng được tạo ra để có cái nhìn toàn diện về dữ liệu ngày, cho phép phân tích dựa trên thời gian một cách linh hoạt và tạo mối quan hệ phân cấp ở các mức độ chi tiết khác nhau (chẳng hạn như ngày, tuần, tháng, v.v.), tạo điều kiện thuận lợi cho khả năng phân tích trong các dashboard.

```
dim_date = ADDCOLUMNS(
    CALENDAR(DATE(2016,1,1), DATE(2018,12,31)),
    "Year", YEAR([Date]),
    "QuarterNum", INT(FORMAT ([Date], "Q" )),
    "Quarter", "Q" & FORMAT ([Date], "Q" ),
    "QuarterYear", "Q" & FORMAT ([Date], "Q" ) & "/" & FORMAT ([Date], "YYYY" ),
    "QuarterYearNum", FORMAT ([Date], "YYYY" ) & FORMAT ([Date], "Q" ),
    "Month", MONTH([Date]),
    "Month Short", FORMAT ([Date], "mmm" ),
    "Week", FORMAT ([Date], "ww" ),
    "WeekDay", WEEKDAY([Date]),
    "Day", DAY([Date]),
    "Day Short", FORMAT ([Date], "ddd" ),
    "Fiscal Year", "FY"& IF(MONTH([Date]) <=6, Year([Date]), YEAR([Date])+1),
    "Month Fiscal sort", IF(MONTH([Date]) <=6, MONTH([Date])+6,MONTH([Date])-6),
    "Working Date", IF(WEEKDAY([Date])=1, "No", "Yes"),
    "MonthYear", FORMAT ([Date], "mm" ) & "/" & FORMAT ([Date], "YYYY" )
)
```

*Bảng 3. 12 Bảng dim\_date*

STT	Thuộc tính	Kiểu dữ liệu
1	Date	Date/time
2	Year	Whole number
3	QuarterNum	Whole number
4	Quarter	Text
5	QuarterYear	Text
6	QuarterYearNum	Text

7	Month	Whole number
8	Month Short	Text
9	Week	Text
10	WeekDay	Whole number
11	Day	Whole number
12	Day Short	Text
13	Fiscal Year	Text
14	Month Fiscal sort	Whole number
15	Working Date	Text
16	MonthYear	Text

🚦 *Bảng dim\_geolocation*

*Bảng 3. 13 Bảng dim\_geolocation*


STT	Thuộc tính	Kiểu dữ liệu	Mô tả
1	geolocation_state	Text	Mã bang
2	avg_latitude	Decimal number	Vĩ độ
3	avg_longitude	Decimal number	Kinh độ
4	State name	Text	Tên bang

🚦 *Bảng dim\_order\_payments*

*Bảng 3. 14 Bảng dim\_order\_payments*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả
1	order_id	Text	Mã số đơn hàng
2	payment_type	Text	Hình thức thanh toán

3	payment_installments	Whole number	Số lần thanh toán trả góp
4	payment_value	Whole number	Giá trị giao dịch

 *Bảng dim\_order\_reviews*

*Bảng 3. 15 Bảng dim\_order\_reviews*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	review_id	Text	Mã số đánh giá	
2	order_id	Text	Mã số đơn hàng	
3	review_score	Whole number	Điểm đánh giá mức độ hài lòng của khách hàng cho đơn hàng (từ 1 đến 5)	
4	review_creation_date	Date	Thời gian tạo bài đánh giá	
5	review_answer_timestamp	Date	Thời gian phản hồi bài đánh giá	
6	product_id	Text	Mã số sản phẩm	
7	Review_Performance	Text	Hiệu suất đánh giá	<code>Review_Performance = SWITCH (dim_order_reviews[review_score], 1, "Very Bad", 2, "Bad", 3, "Average", 4, "Good", "Very Good")</code>


 *Bảng dim\_products*

*Bảng 3. 16 Bảng dim\_products*

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
-----	------------	--------------	-------	---------

1	product_id	Text	Mã số sản phẩm	
2	product_weight_g	Whole number	Cân nặng của sản phẩm (gam)	
3	product_length_cm	Whole number	Chiều dài sản phẩm (cm)	
4	product_height_cm	Whole number	Chiều cao sản phẩm (cm)	
5	product_width_cm	Whole number	Chiều rộng sản phẩm (cm)	
6	product_category_name	Text	Tên sản phẩm (English)	
7	product_name	Text	Tên sản phẩm (in hoa)	product_name = UPPER ( (SUBSTITUTE(dim_products[product_category_name], "_", " ")) )

### 3.3.2. Giới thiệu bảng Fact

 Bảng fact\_order\_items

Bảng 3. 17 Bảng fact\_order\_items

STT	Thuộc tính	Kiểu dữ liệu	Mô tả	Ghi chú
1	order_id	Text	Mã số đơn hàng	
2	product_id	Text	Mã số sản phẩm	
3	seller_id	Text	Mã số nhân viên bán hàng	
4	shipping_limit_date	Date	Ngày giới hạn hoàn tất đơn hàng của người bán hàng để bàn giao cho đơn vị vận chuyển	

5	price	Decimal number	Giá mặt hàng	
6	freight_value	Decimal number	Giá trị vận chuyển mặt hàng (nếu một đơn hàng có nhiều hơn một mặt hàng, giá trị vận chuyển sẽ được chia cho các mặt hàng)	
7	order_value	Decimal number		<pre>order_value = Table.AddColumn( olist_order_items_dataset__2_Table, "order_value", each [price] + [freight_value])</pre>
8	customer_id	Text		
9	review id	Text		
10	order_status	Text		
11	order_purchase_timestamp	Date		
12	order_approved_at	Date		
13	order_delivered_carrier_date	Date		
14	order_delivered_customer_date	Date		
15	order_estimated_delivery_date	Date		
16	payment type	Text		

17	state of seller_ID	Text		
18	Lead Time	Whole number		Lead Time = DATEDIFF(fact_order_items[order_purchase_timestamp], fact_order_items[order_delivered_customer_date], DAY)
19	On time delivery	Text		On time delivery = IF(fact_order_items[order_delivered_customer_date] < fact_order_items[order_estimated_delivery_date], "Ontime", "Late")
20	Processing Time	Whole number		Processing Time = DATEDIFF(fact_order_items[order_purchase_timestamp], fact_order_items[order_delivered_carrier_date], HOUR)

### 3.3.3. Mô hình dữ liệu (Data Model)

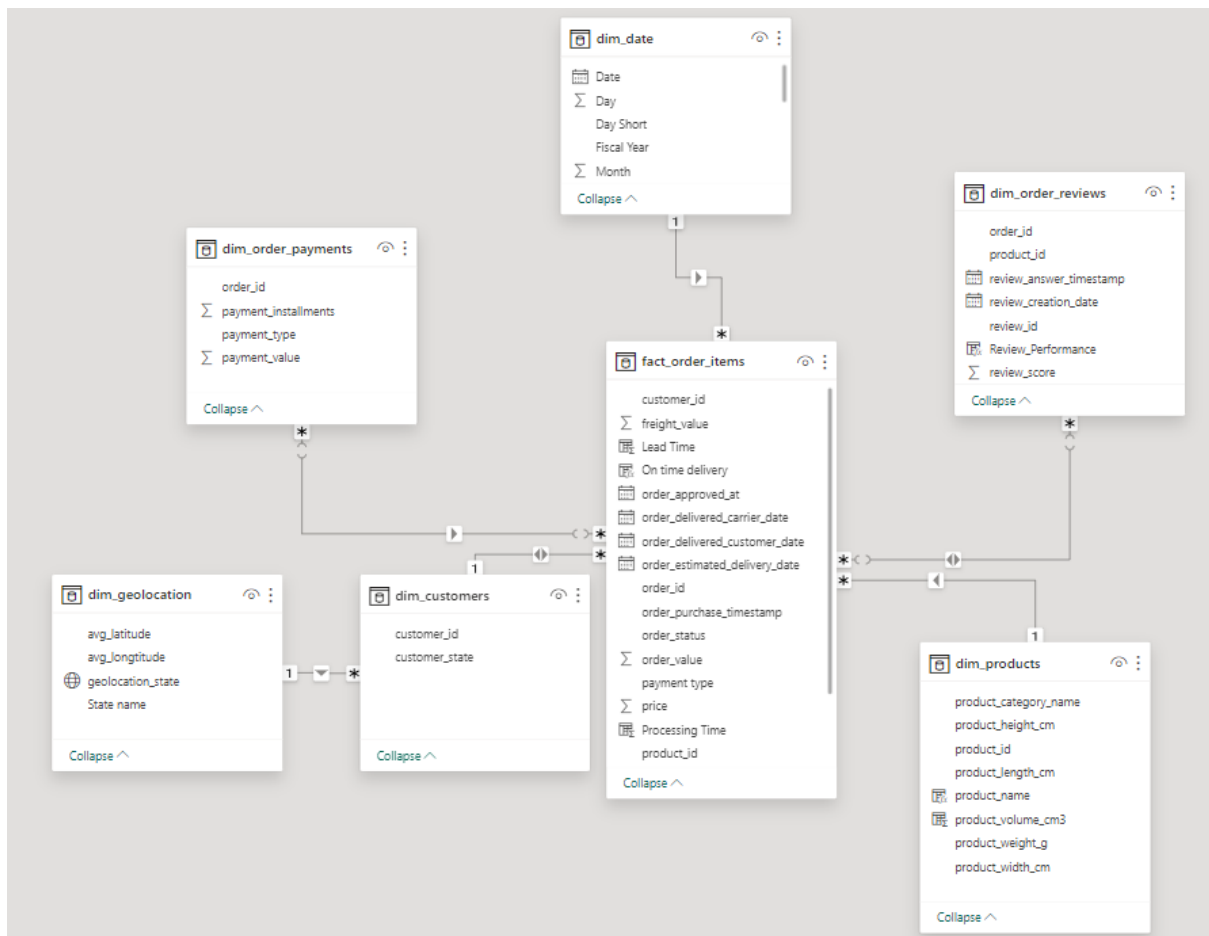
Các bảng Dimension được mô hình hóa với bảng Fact qua các cột chung.

Bảng fact\_order\_items được nối với:

- Bảng dim\_products qua cột product\_id

- Bảng dim\_customers qua cột customer\_id
- Bảng dim\_products qua cột product\_id
- Bảng dim\_order\_payments qua cột order\_id
- Bảng dim\_order\_reviews qua cột review\_id

Bảng dim\_geolocation được nối với bảng dim\_customers thông qua geolocation\_state)



Hình 3. 19 Mô hình dữ liệu

## CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU

### 4.1. Phân tích tổng quan tình hình doanh nghiệp (Dashboard Overview)

#### 4.1.1. Chart

- *Line and Stacked Column Chart*: So sánh doanh thu và số lượng đơn hàng qua từng năm của Olist Store
- *Doughnut Charts*: Thể hiện tỉ trọng đơn hàng được giao đúng giờ/ trễ giờ và so sánh tỷ trọng hình thức thanh toán giữa các đơn hàng
- *Bar chart*: So sánh khối lượng tiêu thụ của từng sản phẩm và so sánh mức độ hài lòng của khách hàng theo đơn hàng
- *Map visualization*: Thể hiện phân bố khách hàng theo doanh thu của Olist Store.

#### 4.1.2. Measure

*Total Sales (Revenue)*: Tổng doanh thu trong 3 năm của Olist Store

```
1 Total sales = CALCULATE(SUM(fact_order_items[order_value]))
```

*Total Customer*: Tổng số khách hàng trong 3 năm của Olist Store

```
1 Total Customers = DISTINCTCOUNT(fact_order_items[customer_id])
```

*Total Orders*: Tổng số đơn hàng trong 3 năm

```
1 Total Orders = DISTINCTCOUNT(fact_order_items[order_id])
```

*Total State*: Tổng số bang

```
1 Total State = DISTINCTCOUNT(dim_geolocation[State name])
```



*Total Seller:* Tổng số nhà bán hàng đang hoạt động tại Olist Store

```
1 Total Seller = DISTINCTCOUNT(fact_order_items[seller_id])
```

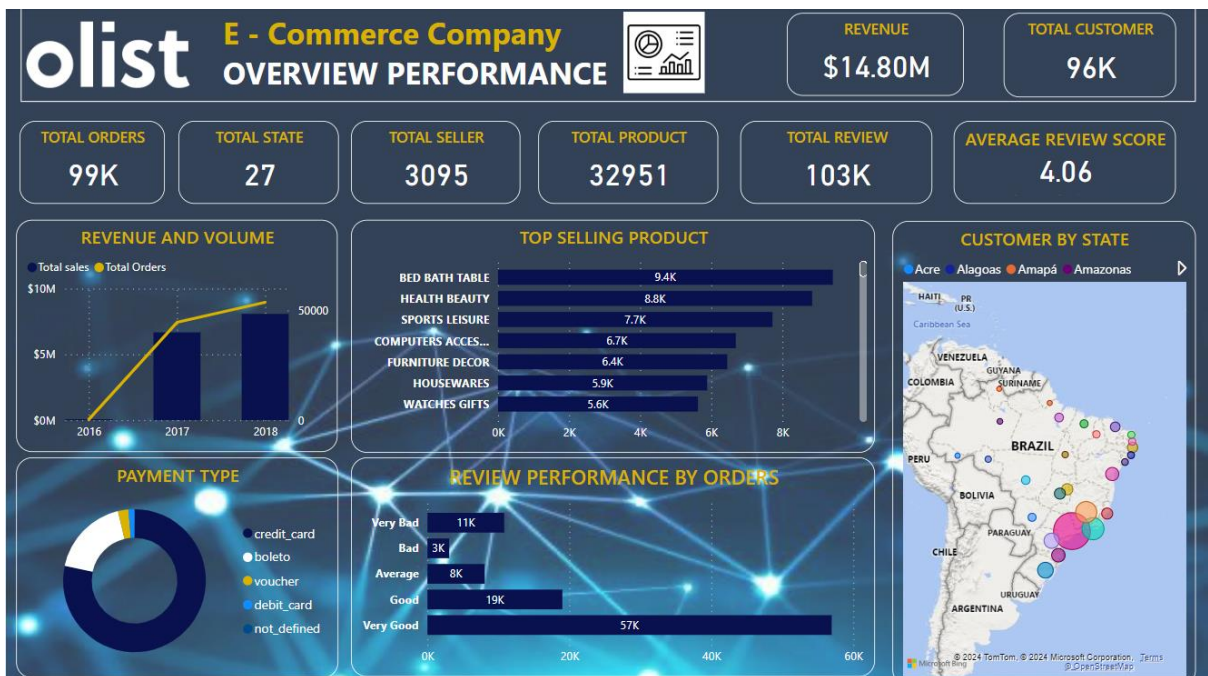
*Total Product:* Tổng số sản phẩm kinh doanh tại Olist Store

```
1 Total Product = CALCULATE(  
2 |   COUNTROWS(dim_products)  
3 )
```

*Total Review:* Tổng số đánh giá của khách hàng về đơn hàng của Olist Store

```
1 Total Review = COUNT(dim_order_reviews[review_id])
```

#### 4.1.3. Data Storytelling



Hình 4. 1 Dashboard phân tích tổng quan doanh nghiệp

Tổng quan về tình hình kinh doanh của doanh nghiệp trong 3 năm 2016, 2017, 2018 được trực quan hóa bằng hệ hỗ trợ quản trị thông minh - Power BI. Trong đó, ta có thể thấy, doanh thu đạt được 14,8 triệu USD, phục vụ cho hơn 96 nghìn khách hàng, tổng số đơn hàng đạt mức 99 nghìn đơn hàng, cung cấp dịch vụ cho 27 bang tại Nam

Mỹ, Bắc Mỹ và Ấn Độ, số lượng nhà bán hàng gia nhập đạt con số 3095 nhà bán hàng, có hơn 103 nghìn lượt đánh giá, điểm đánh giá trung bình của các đơn hàng là 4.06/5.

Cụ thể, doanh thu trong ba năm qua liên tục tăng, nhất là trong giai đoạn 2016-2017, số lượng đơn hàng tăng mạnh từ mức hơn 300 đơn hàng lên đến 44 nghìn đơn hàng. Với sự tăng nhanh về đơn hàng, ta có thể thấy doanh thu trong hai năm này cũng tăng rõ rệt khi ở năm 2016, doanh thu đạt khoảng hơn 53 nghìn USD, thì đến năm 2017, doanh thu đạt được hơn 6,6 triệu USD. Năm 2018, số lượng đơn hàng và doanh thu tiếp tục tăng tuy nhiên mức độ tăng trưởng chậm lại khi doanh thu đạt mức 8 triệu USD và số lượng đơn hàng là 53 nghìn - tăng 54,6% so với năm trước đó. Các sản phẩm được bán tại Olist Store vô cùng đa dạng. Trong đó, sản phẩm bán chạy nhất là bed bath table với hơn 9,4 nghìn lượt bán.

Về phân bố khách hàng, khách hàng của Olist Store đến từ Nam Mỹ, khu vực Trung Đông, và một số bang tại Mỹ và Canada. Dù vậy, khách hàng của Olist Store phần lớn đến từ khu vực Nam Mỹ. Trong đó, có hơn 41 nghìn (~41,7%) khách hàng đến từ bang Sao Paulo - Brazil.

Phần lớn các đơn hàng của Olist Store được thanh toán theo hình thức credit card với 78,34% đơn hàng lựa chọn phương thức thanh toán này. Thấp nhất là hình thức thanh toán bằng debit card với 1,36%.

Trong 3 năm hoạt động với hơn 103 nghìn lượt đánh giá đơn hàng, Olist Store nhận được đa số các lượt đánh giá tích cực. Trong đó có hơn 57 nghìn lượt đánh giá đơn hàng ở mức cực kỳ tốt, 19 nghìn lượt đánh giá đơn hàng ở mức tốt. Đạt mức 78,35% đơn hàng nhận được các đánh giá tích cực. Tuy nhiên số lượng đơn hàng được đánh giá ở mức cực kỳ tệ khá cao khi có hơn 11 nghìn lượt đánh giá ở mức này. Như vậy, Olist Store cần giám sát và có các chính sách cải thiện mới để giảm mức đánh giá tiêu cực này, nhằm đem lại hiệu quả trải nghiệm của khách hàng ở mức cao hơn.

#### **4.2. Phân tích nhánh Product**

#### 4.2.1. Chart

- *Chart Scatter plot*: Phân tích mối tương quan giữa chi phí vận chuyển và doanh thu bán hàng của từng loại sản phẩm
- *Chart Treemaps*: Phân tích doanh thu hoặc số lượng đơn hàng cho từng loại ngành hàng
- *Table*: Thể hiện chi tiết thông tin đi sâu đến từng loại sản phẩm

#### 4.2.2. Measure

*Sale by Category*: Tính % doanh thu của từng loại ngành hàng so với tổng doanh thu

```
% Sales by Category =  
VAR Overall_Sales= CALCULATE(SUM(fact_order_items[order_value]), ALLSELECTED(dim_products[product_category_name]))  
RETURN  
    DIVIDE(SUM(fact_order_items[order_value]),Overall_Sales)
```

*Order by Category*: Tính % lượng đơn hàng của từng loại ngành hàng so với tổng doanh thu

```
% Order By Category =  
VAR Overall_Order= CALCULATE(DISTINCTCOUNT(fact_order_items[order_id]), ALLSELECTED(dim_products[product_category_name]))  
RETURN  
    DIVIDE(DISTINCTCOUNT(fact_order_items[order_id]),Overall_Order)
```

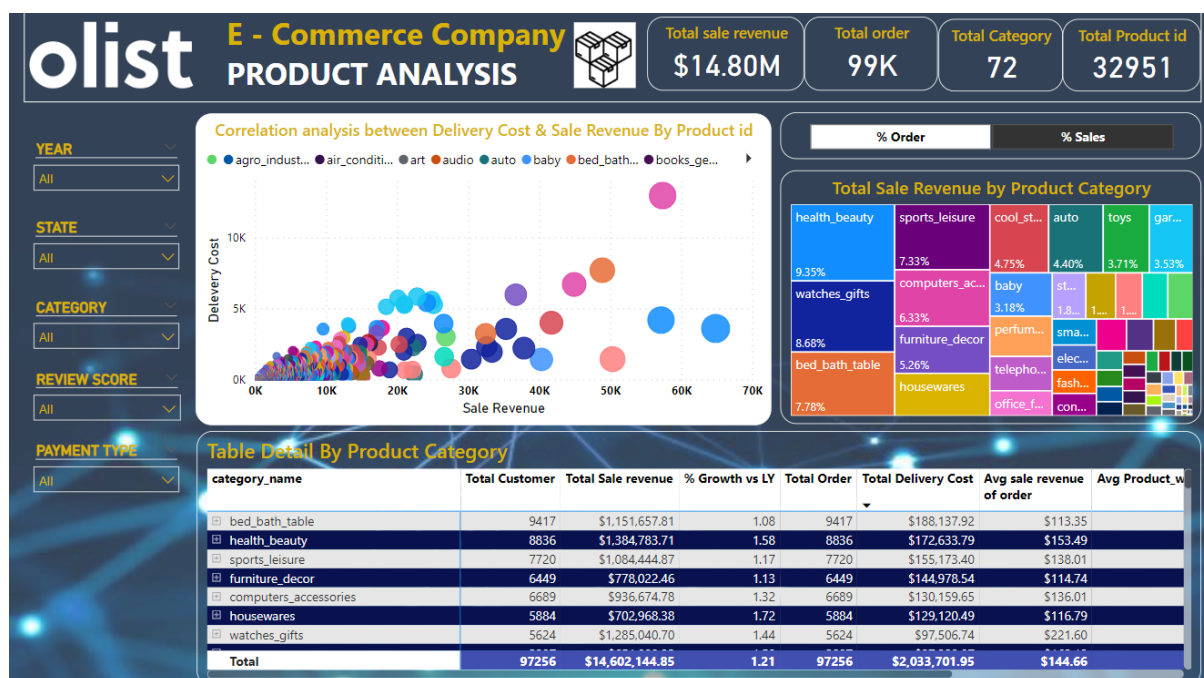
*Total KPI*: Tạo button chuyển đổi 2 giá trị % Sale by Category và % Order by Category

```
Total KPI = {  
    ("% Orders", NAMEOF('dim_products'[% Order By Category]), 0),  
    ("% Sale Revenue", NAMEOF('dim_products'[% Sales by Category]), 1)  
}
```

*product\_volume-cm3*: Thể hiện thể tích của sản phẩm

```
1 product_volume_cm3 = dim_products[product_height_cm]*dim_products[product_length_cm]*dim_products[product_width_cm]
```

#### 4.2.3. Data Storytelling



Hình 4. 2 Dashboard phân tích nhánh Product

Về tổng quan, Olist Store đang kinh doanh tổng cộng 72 ngành hàng và 32,951 loại sản phẩm, trong đó top 3 ngành hàng chủ lực mang về doanh thu cao nhất là heath\_beauty (9.3%), watches\_gifts (8.6%), bed\_bath\_table (7.7%). Song song, 3 ngành hàng có số lượng đơn đặt hàng nhiều nhất lần lượt là bed\_bath\_table, heath\_beauty, sports\_leisure với gần 26,000 đơn hàng chiếm 26% tổng số đơn hàng của Olist Store trong 3 năm (2016-2018). Trái ngược lại với top 3 doanh thu/lượng đơn cao nhất, ta có ngành hàng cds\_dvds\_musicals, fashion\_childrens\_clothes, security\_and\_services là 3 ngành hàng có Doanh thu thấp nhất, chỉ chiếm 0.02% tổng số đơn hàng và 0.01% tổng doanh số bán hàng.

Phân tích sâu hơn vào từng loại sản phẩm (Product id), ta sử dụng chart scatter plot để phân tích mối tương quan giữa chi phí vận chuyển và doanh thu bán hàng, từ đó đưa ra kế hoạch chiến lược tối đa hóa doanh thu hoặc tối ưu hóa chi phí cho từng ngành hàng. Ở đây xét đến top 3 ngành hàng có lượng đơn hàng cao nhất.

Ở ngành hàng bed\_bath\_table, biểu đồ thể hiện rằng có mối tương quan tích cực cao, nghĩa là chi phí vận chuyển càng cao thì doanh thu bán hàng càng cao và ngược lại.

Điều này cho ta thấy được rằng, ngành hàng đã được tối ưu hóa doanh thu khá tốt, bằng chứng là ngành hàng này cũng có doanh thu lớn thứ 3 trong tất cả các ngành hàng.

Ở ngành hàng `heath_beauty`, biểu đồ thể hiện rằng 2 biến chi phí vận chuyển và doanh thu bán hàng có mối tương quan tích cực thấp, tương tự như ở ngành hàng `bed_bath_table`, chi phí vận chuyển càng cao thì doanh thu bán hàng càng cao. Tuy nhiên có vì có mối tương quan thấp, doanh nghiệp cần tập trung vào các loại sản phẩm (Product id) có chi phí vận chuyển cao nhưng doanh thu còn thấp, những sản phẩm ấy cần được tối ưu hóa chi phí, tiêu biểu là product id “19c91...” có tổng chi phí vận chuyển cao nhất là 5,4 nghìn USD tuy nhiên lợi nhuận chỉ đạt khoảng 24 nghìn USD, đứng top 4 doanh thu. Từ đó Olist store có thể phân tích sâu hơn các yếu tố về thời gian giao hàng, địa điểm bán hàng, trọng lượng của sản phẩm.. nhằm đưa ra hành động hiệu quả như: tạm dừng kinh doanh sản phẩm này, điều chỉnh các tuyến bán hàng nhằm tối ưu chi phí vận chuyển sản phẩm.

Ở ngành hàng `sports_leisure`, ta thấy các sản phẩm tập chung chủ yếu trong khoảng doanh thu từ 0-5 nghìn USD và chi phí vận chuyển từ 0-1 nghìn USD. Và Product id: “cec09...” là một điểm bất thường (outlier), cần doanh nghiệp tập chung phân tích khi nó có chi phí vận chuyển cao bất thường. Olist store cần có các chiến lược giúp cải thiện chi phí và đẩy mạnh doanh thu của dòng sản phẩm này.

Bên cạnh dữ liệu về chi phí vận chuyển và doanh thu bán hàng, nhóm cũng phân tích sâu hơn về data của từng loại sản phẩm như Cân nặng (Avg Product\_weight), Thể tích (Avg Product\_volume). Cụ thể `furniture_mattress_and_upholstery` là ngành hàng có khối lượng trung bình cao nhất đạt khoảng 13kg và thể tích đạt 77 dm<sup>3</sup>, Việc sản phẩm sở hữu trọng lượng và thể tích cao có thể ảnh hưởng chi phí vận chuyển đồng thời là lượng mua của khách hàng giảm xuống. Bằng chứng là ngành hàng `furniture_mattress_and_upholstery` chỉ có tổng cộng 38 đơn hàng (0.04% tổng lượng đơn hàng), kéo theo đó là doanh thu 5,9 nghìn USD, đứng thứ 13 sản phẩm có doanh thu thấp nhất.

### 4.3. Phân tích nhánh Review

#### 4.3.1. Chart

- *Gauge*: Dùng để đo lường độ hài lòng của khách hàng và cho biết điểm trung bình được rate bởi khách hàng.
- *Donut Chart* : Thể hiện tỷ trọng các đánh giá của khách hàng.
- *Area Chart*: Thể hiện sự thay đổi trong việc đánh giá theo thời gian.
- *Stacked Bar Chart*: Thể hiện số lượng đánh giá thông qua các Product Category khác nhau.
- *Clustered Bar Chart*: Thể hiện điểm trung bình được đánh giá của các Product Category khác nhau.

#### 4.3.2. Measure

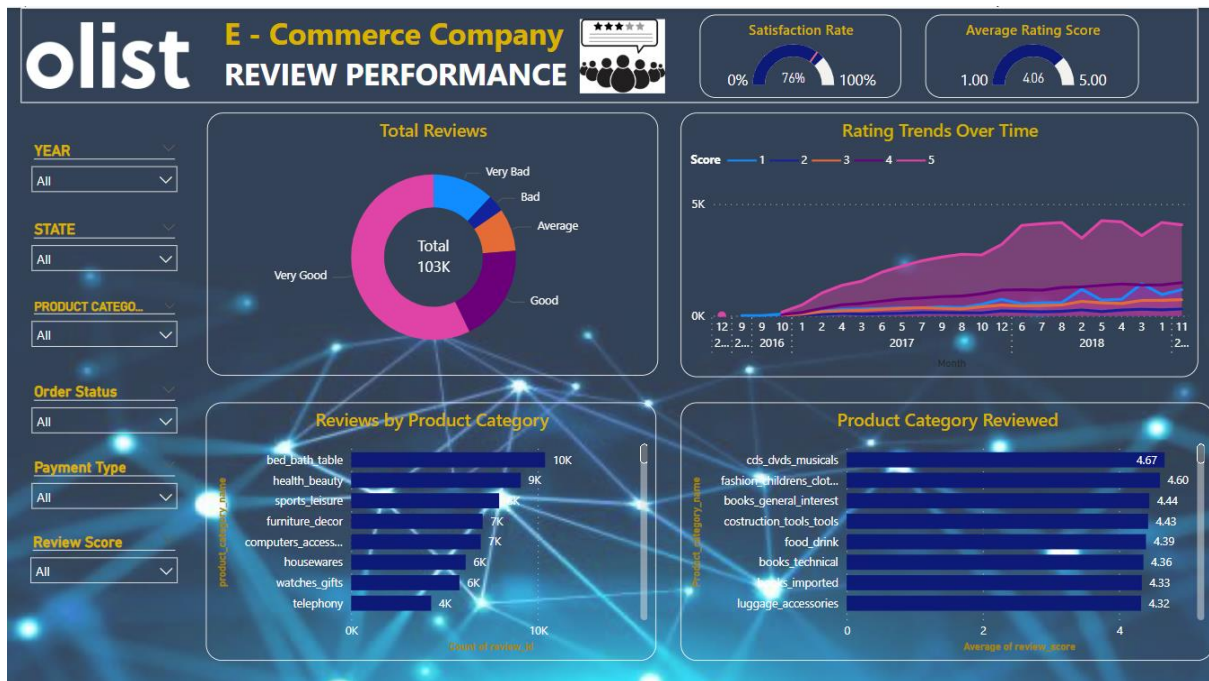
*Review Performance*: Thể hiện các mức độ đánh giá của khách hàng.

```
1 Review_Performance = SWITCH(  
2     dim_order_reviews[review_score],  
3     1,"Very Bad",  
4     2,"Bad",  
5     3,"Average",  
6     4,"Good",  
7     "Very Good")
```

*dim\_date*: Tính thời gian theo từng thời điểm như: năm, quý, tháng, tuần, ngày,...

```
1 dim_date = ADDCOLUMNS(  
2     CALENDAR(DATE(2016,1,1), DATE(2018,12,31)),  
3     "Year", YEAR([Date]),  
4     "QuarterNum", INT(FORMAT ([Date], "Q" )),  
5     "Quarter", "Q" & FORMAT ([Date], "Q" ),  
6     "QuarterYear", "Q" & FORMAT ([Date], "Q" ) & "/" & FORMAT ([Date], "YYYY" ),  
7     "QuarterYearNum", FORMAT ([Date], "YYYY" ) & FORMAT ([Date], "Q" ),  
8     "Month", MONTH([Date]),  
9     "Month Short", FORMAT ([Date], "mmm" ),  
10    "Week", FORMAT ([Date], "ww" ),  
11    "WeekDay", WEEKDAY([Date]),  
12    "Day", DAY([Date]),  
13    "Day Short", FORMAT ([Date], "ddd" ),  
14    "Fiscal Year", "FY"& IF(MONTH([Date]) <=6, Year([Date]), YEAR([Date])+1),  
15    "Month Fiscal sort", IF(MONTH([Date]) <=6, MONTH([Date])+6,MONTH([Date])-6),  
16    "Working Date", IF(WEEKDAY([Date])=1, "No", "Yes"),  
17    "MonthYear", FORMAT ([Date], "mm" ) & "/" & FORMAT ([Date], "YYYY" )
```

#### 4.3.3. Data Storytelling



Hình 4. 3 Dashboard phân tích nhánh Review

Đầu tiên là mức độ hài lòng của khách hàng và số điểm trung bình được thể hiện qua biểu đồ Gauge. Olist nhận được 76% độ hài lòng từ khách hàng cùng với đó là điểm trung bình khá cao với 4.06 điểm. Đây là một điểm số khả quan và tích cực giúp Olist có thêm động lực để có thể nâng cao thêm chất lượng từ đó nhận thêm được nhiều hơn nữa các đánh giá tích cực từ khách hàng.

Về tổng quan, Olist Store hiện tại nhận được tổng cộng 103 nghìn lượt review từ khách hàng với 5 loại đánh giá từ Very Good, Good, Average, Bad, Very Bad, trong đó số lượt đánh giá Very Good của Olist chiếm phần lớn nhất với gần 58,9 nghìn lượt. Song song, lượt đánh giá Bad chiếm tỷ lệ ít nhất với số lượt review chưa đến 3,5 nghìn lượt. Qua đó có thể thấy được rằng Olist hiện tại vẫn đang làm rất tốt và nhận được nhiều đánh giá khả quan từ khách hàng.

Tiếp theo là xu hướng đánh giá của khách hàng theo thời gian thông qua Area Chart. Các đánh giá được nhận 5 điểm chiếm phần lớn nhất dù là 2016, 2017 hay 2018. Ngoài ra lượng đánh giá tích cực còn có xu hướng tăng dần theo thời gian. Từ đây có



thể thấy theo thời gian thì lượt đánh giá tốt nhận về càng nhiều chứng tỏ Olist đã thành công trong việc làm tăng sự hài lòng của khách hàng.

Về số lượt đánh giá thông qua các ngành hàng thì ngành hàng nhận được lượt đánh giá nhiều nhất là bed\_bath\_table với 10 nghìn lượt đánh giá. Ngược lại có một số ngành lại không nhận được đánh giá nào như security\_and\_service, fashion\_childrens\_clothes,... Lượt đánh giá càng nhiều chứng tỏ ngành đó đang được khách hàng quan tâm. Từ đó, ta có thể biết được nhu cầu của khách hàng và tiến hành lên kế hoạch quảng bá sản phẩm hợp lý.

Điểm đánh giá trung bình của khách hàng thông qua các ngành hàng cho thấy được khách hàng đang hài lòng và không hài lòng đến sản phẩm nào để có thể cải thiện được chất lượng và trải nghiệm của người dùng khi tiến hành mua sản phẩm.

#### 4.4. Phân tích nhánh Location

##### 4.4.1. Chart

- *Stacked area chart*: thể hiện tổng số khách hàng và tổng số đánh giá theo bang
- *Line and stacked column chart*: thể hiện thông tin về đơn hàng đã giao và đơn hàng bị hủy theo bang
- *Matrix*: thể hiện thông tin chi tiết về tổng đơn hàng, tổng sản phẩm, doanh thu, trung bình giá trị đơn hàng và giá trị vận chuyển theo bang
- *Map*: tổng số người bán theo bang

##### 4.4.2. Measure

*Delivered orders*: tổng số đơn hàng đã giao

```
Delivered orders =  
| | CALCULATE(DISTINCTCOUNT(fact_order_items[order_id]),fact_order_items[order_status]="delivered")
```

*Canceled orders*: Tổng số đơn hàng bị hủy

```
Canceled orders =  
| | | CALCULATE(DISTINCTCOUNT(fact_order_items[order_id]), fact_order_items[order_status] = "canceled")
```

*Total Freight*: Tổng chi phí vận chuyển



```
Total Freight =
| | | SUM(fact_order_items[freight_value])
```

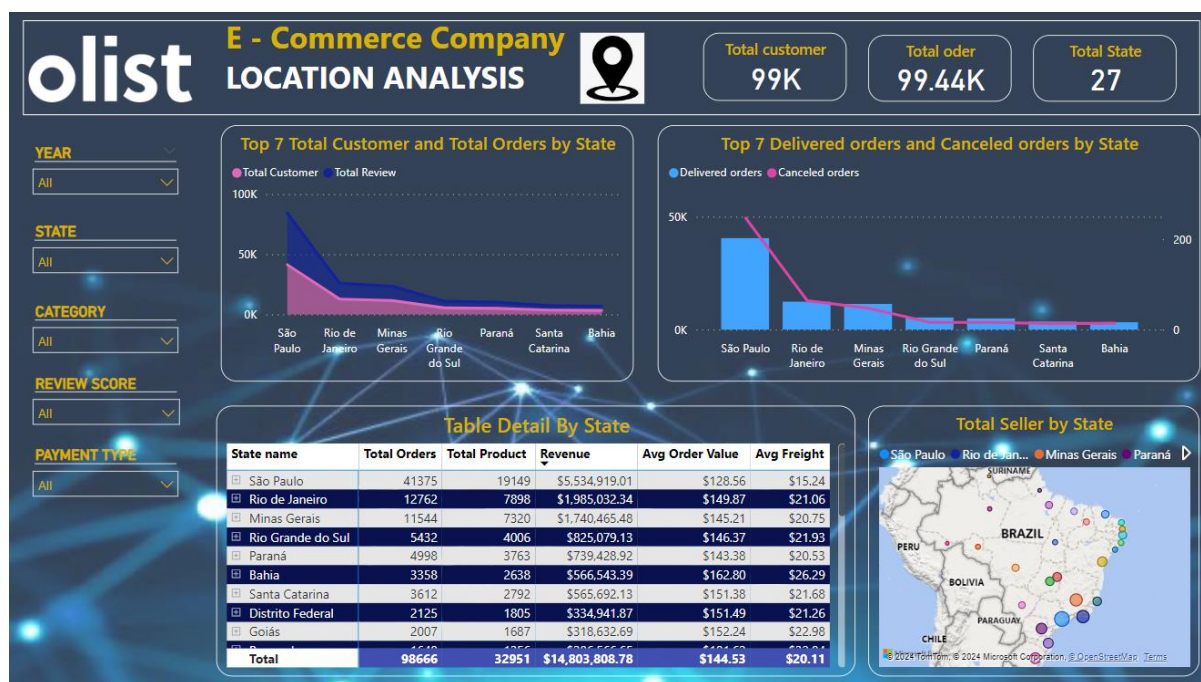
*Average Freight*: chi phí vận chuyển trung bình

```
Average Freight =
| | | | CALCULATE(AVERAGE(fact_order_items[freight_value]))
```

*Average Order Value*: Trung bình giá trị mỗi đơn hàng

```
Average Order Value =
| | | | DIVIDE ([Total sales],
| | | | CALCULATE(COUNTROWS(fact_order_items))
| | | )
```

#### 4.4.3. Data Storytelling



Hình 4. 4 Dashboard phân tích nhánh Location

Về tổng quan, Olist Store đang kinh doanh trên 27 bang và có tổng cộng 98,67 nghìn khách hàng. Trong đó bang có tổng số khách hàng cao nhất và tổng số đánh giá của khách hàng cao nhất là bang Sao Paulo, với số khách hàng là 41,375 và có 42,906

đánh giá. Đứng thứ hai là bang Rio De Janeiro, với 12,762 khách hàng và 13,120 lượt đánh giá. Ngược lại với hai bang này, Roraima là bang có số khách hàng và số đánh giá thấp nhất, lần lượt 46 khách hàng và 47 lượt đánh giá.

Tương tự, Sao Paulo cũng là bang có số đơn hàng đã giao cao nhất (40,501 đơn) và cũng có số đơn hàng bị hủy cao nhất (247 đơn). Đồng thời, bang Rio De Janeiro cũng đứng thứ hai về số đơn đã giao cũng như số đơn bị hủy. Qua hai phân tích trên, chúng ta có thể đưa ra các chiến lược kinh doanh cũng như chiến lược marketing nhằm tăng doanh số đơn hàng cũng như làm tăng số khách hàng cho các bang, đặc biệt là các bang có tình hình kinh doanh kém.

Bên cạnh những dữ liệu về khách hàng và tình trạng đơn hàng đơn hàng, nhóm cũng phân tích sâu hơn về tổng số sản phẩm, tổng đơn hàng, tổng doanh thu, trung bình giá trị đơn hàng, trung bình chi phí vận chuyển. Có thể thấy được Sao Paulo vẫn là bang đứng đầu về tổng số đơn hàng (41,375 đơn) và đứng đầu về tổng doanh thu (\$5,534,919). Nhưng về trung bình giá trị đơn hàng và trung bình phí vận chuyển thì Sao Paulo là bang có hai chỉ tiêu này đều ở mức thấp nhất. Chi phí vận chuyển thấp có thể được giải thích rằng vì lượng khách hàng cũng như đơn hàng ở Sao Paulo là lớn nhất nên Olist đã triển khai những chiến lược về vận chuyển thuận lợi với bang này nhằm tối thiểu hóa chi phí vận chuyển mang lại lợi ích cho cả doanh nghiệp và khách hàng. Song Song, Roraima cũng là bang có tổng đơn hàng và tổng doanh thu thấp nhất, lần lượt là 46 đơn và doanh thu là \$8,933. Tuy nhiên, chi phí vận chuyển của bang này là lớn nhất, điều này được giải thích rằng do tình hình kinh doanh không khả quan, doanh số còn thấp nên Olist sẽ không tập trung phát triển hệ thống vận chuyển ở đây, dẫn đến chi phí vận chuyển khá cao (khoảng 22% trên tổng giá trị đơn hàng).

Nhóm cũng phân tích thêm số lượng người bán tập trung theo từng bang, trong đó bang có nhiều người bán nhất là Sao Paulo (2549 người), thứ hai là Rio de Janeiro (1751 người), đứng thứ ba là Minas Gerais (1664 người). Ngược lại, Roraima là bang có ít người bán nhất (39 người), kế đến là Amapa (63 người).

Sau những phân tích trên, có thể thấy được tình hình kinh doanh của olist ở các bang Sao Paulo, Rio de Janeiro, Minas Gerais có nhiều điểm tích cực, doanh số luôn dẫn đầu trong ba năm, các chi phí phát sinh như chi phí vận chuyển luôn được tối thiểu hóa mang lại lợi ích cho doanh nghiệp và người mua hàng. Ngược lại, bang Roraima tình hình kinh doanh tổng cả ba năm đều ở mức thấp, có thể vì người bán không tập trung ở đây nhiều, lượng phân bố dân cư ít nên không có nhiều khách hàng tiềm năng, dẫn đến doanh số trong ba năm luôn ở mức thấp nhất. Vì vậy, Olist cần có những chiến lược quảng bá sản phẩm, tăng ưu đãi cho cả người bán và người mua ở khu vực này, giảm chi phí vận chuyển nhằm tạo động lực cho người bán và người mua tham gia vào thị trường nhiều hơn.

## 4.5. Phân tích nhánh Date

### 4.5.1. Chart

- *Ribbon chart*: Thể hiện tổng doanh thu theo các tháng trong năm
- *Stacked area chart*: Thể hiện tổng số khách hàng và người bán hàng trong suốt quá trình kinh doanh
- *Line and clustered column chart 1*: Phân tích các ngày trong tuần lượng số đơn hàng khoảng bao nhiêu và giá trị đơn hàng trung bình
- *Line and clustered column chart 2*: Thể hiện thời gian trung bình để xử lý đơn hàng và thực hiện giao hàng đến tay khách hàng

### 4.5.2. Measure

*Total Sales (Revenue)*: Tổng doanh thu trong 3 năm của Olist Store

```
1 Total sales = CALCULATE(SUM(fact_order_items[order_value]))
```

*Total Seller*: Tổng số nhà bán hàng đang hoạt động tại Olist Store

```
1 Total Seller = DISTINCTCOUNT(fact_order_items[seller_id])
```

*Total Customer*: Tổng số khách hàng trong 3 năm của Olist Store

```
1 Total Customers = DISTINCTCOUNT(fact_order_items[customer_id])
```

*Average Oder Value:* Trung bình giá trị mỗi đơn hàng

```
1 Average Order Value =  
2 | | | | | DIVIDE ([Total sales],  
3 | | | | | CALCULATE(COUNTROWS(fact_order_items))  
4 | | | | | )
```

*Average of Lead Time:* Trung bình thời gian thực hiện đơn hàng giao đến tay khách hàng

**Average of 'fact\_order\_items'[Lead Time]**

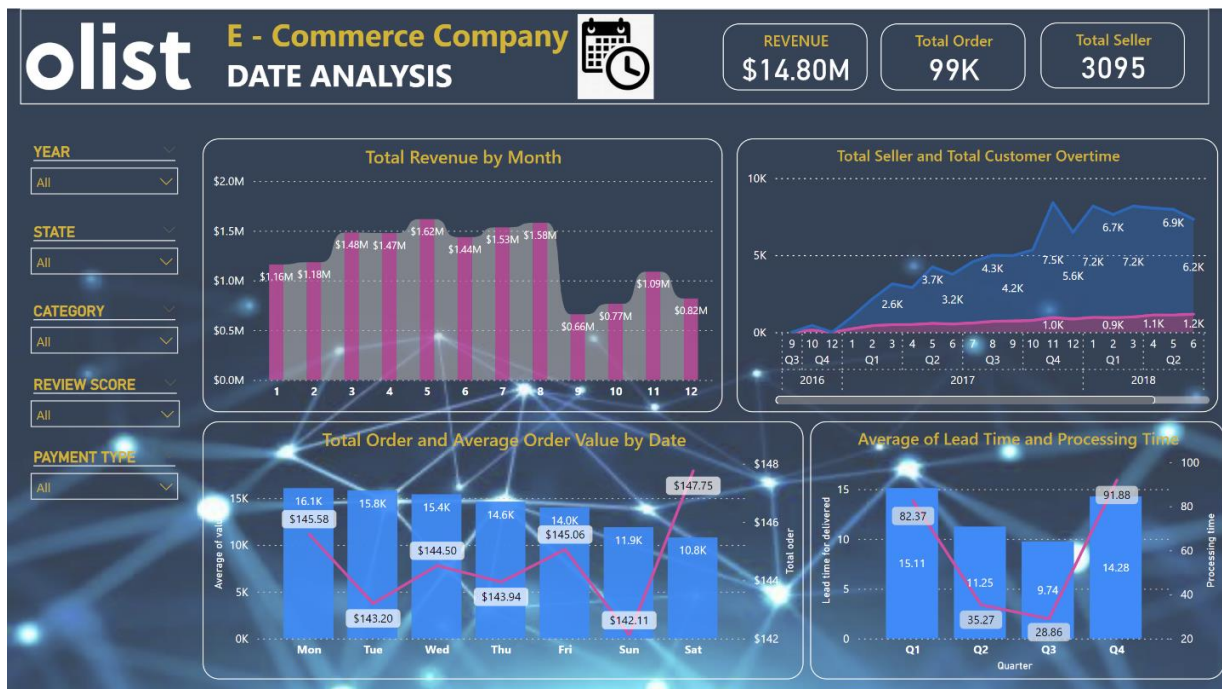
Average of Lead Time ▼ ✕

*Average of Processing Time:* Trung bình thời gian xử lý đơn hàng

**Average of 'fact\_order\_items'[Processing Time]**

Average of Processing... ▼ ✕

### 4.5.3. Data Storytelling



Hình 4. 5 Dashboard phân tích nhánh Date

Page Data Analysis phân tích tình hình kinh doanh của Olist Store theo thực thời gian từ tháng 9 năm 2016 đến tháng 9 năm 2018 tổng cộng là khoảng 24 tháng trong đó doanh thu khoảng giữa năm là quý II là cao nhất. Số lượng khách hàng cũng như nhân viên bán hàng có xu hướng tăng lên theo thời gian và đạt cao nhất tại tháng 8 năm 2018 với 6452 khách hàng và 1278 nhân viên bán hàng. Phân tích sâu hơn về thời gian thì mỗi ngày Olist Store có khoảng 145 đơn hàng và giá trị mỗi đơn hàng ước tính khoảng 13 đến 14 nghìn. Sự chênh lệch số lượng các ngày trong tuần không quá lớn cao nhất và thứ bảy khoảng 148 đơn và thấp nhất vào chủ nhật 142 đơn. Giá trị các đơn hàng có xu hướng giảm từ đầu tuần đến cuối tuần, cao nhất là vào thứ hai 16,1 nghìn giảm xuống thấp nhất 10,8 nghìn vào thứ bảy. Một góc nhìn khác cũng khá quan trọng là thời gian để xử lý và hoàn thành giao cho khách hàng, để thấy được sự cải tiến trong quy trình vận hành nên được phân tích theo quãng thời gian dài cụ thể là theo quý. Thời điểm ra báo cáo là khoảng tháng 9, quý III thì chúng ta thấy có sự cải thiện rõ rệt trong thời gian xử lý và giao hàng. Thời gian xử lý giảm từ khoảng 14 giờ xuống còn 10 giờ, thời gian giao hàng giảm từ 92 giờ còn khoảng 29 giờ.

## 4.6. Phân tích nhánh Payment

### 4.6.1. Chart

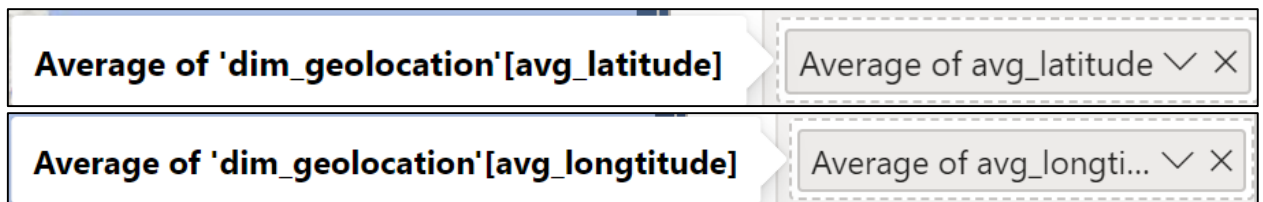
- Donut chart: Phân tích tỷ lệ tổng số đơn đặt hàng theo phương thức thanh toán
- Clustered Column chart: So sánh trung bình giá trị mỗi đơn hàng ứng với mỗi phương thức thanh toán
- Table: Thống kê chi tiết số lượt sử dụng phương thức thanh toán theo từng nhóm sản phẩm
- Map: Trực quan số khách hàng lựa chọn phương thức thanh toán theo từng khu vực

### 4.6.2. Measure

*Average of payment value: Giá trị trung bình mỗi loại thanh toán*

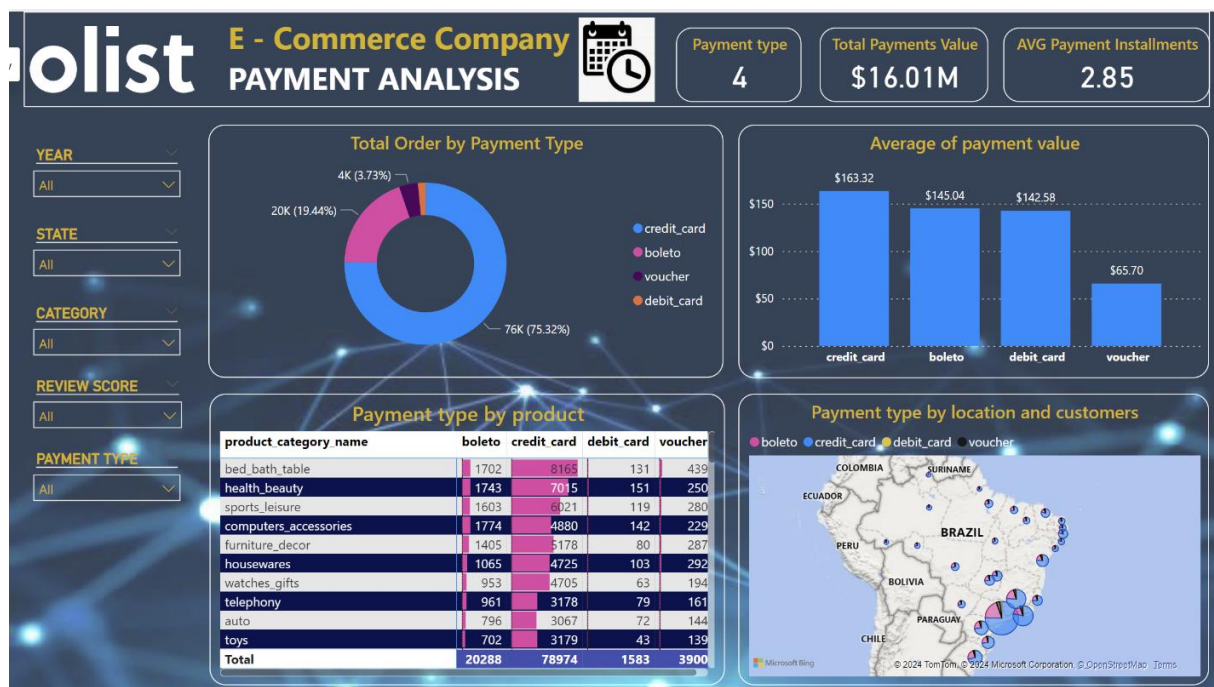


*Average of latitude and longitude: Kinh độ và vĩ độ trung bình*



### 4.6.3. Data Storytelling





Hình 4. 6 Dashboard phân tích nhánh Payment

Khách hàng có 4 lựa chọn để thanh toán đơn hàng credit, boleto, voucher và debit. Nhìn vào số lượt thanh toán ta thấy khách hàng ưa thích hình thức credit tiếp theo là boleto và chỉ ít là 2 phương thức còn lại.

Đi sâu hơn là giá trị thanh toán mỗi đơn hàng có chênh lệch như thế nào. Điểm chú ý được nhắc đến là phương thức voucher có giá trị thấp hơn đáng kể so với các phương thức còn lại. Giải thích cho hiện tượng này có lẽ chương trình marketing chưa đúng với nhu cầu thực của khách hàng, khách hàng chỉ muốn nhận được khoảng hời hơn là chú ý đến sản phẩm, đến doanh nghiệp. Tiếp tục phân tích sâu hơn là theo sản phẩm, theo xu hướng mua hàng, sản phẩm bán chạy nhất khách hàng rất ưa thích sử dụng phương thức credit chứng tỏ có sự tương quan.

Nhìn lại một lần nữa, theo các khu vực thì không có sự khác biệt quá nhiều giữa các khu vực với nhau, sự phổ biến của phương thức credit, thuận tiện của nó khiến nó được sử dụng rộng rãi khiến các phương thức còn lại bị lép vế.

## CHƯƠNG 5: ĐỀ XUẤT GIẢI PHÁP

Trong môi trường kinh doanh ngày càng cạnh tranh như hiện nay, việc duy trì và phát triển doanh số bán hàng không chỉ đòi hỏi sự sáng tạo mà còn yêu cầu sự đánh giá cẩn thận về các thách thức và cơ hội. Trong phần này, chúng tôi sẽ xem xét những giải pháp cụ thể để Olist có thể vượt qua những khó khăn hiện tại và đạt được mục tiêu kinh doanh. Từ việc tăng cường chất lượng dịch vụ đến việc tối ưu hóa chiến lược tiếp thị và phân phối, chúng tôi sẽ đề xuất các giải pháp thúc đẩy sự phát triển bền vững của doanh nghiệp và tạo ra một trải nghiệm tuyệt vời cho khách hàng.

*Thứ nhất*, về mặt sản phẩm, việc cải thiện chất lượng sản phẩm là ưu tiên hàng đầu. Bằng cách thiết lập các tiêu chuẩn và quy trình kiểm tra chất lượng nghiêm ngặt, Olist phải đảm bảo rằng chất lượng của tất cả các sản phẩm được bán trên nền tảng của họ đều đạt đến mức cao nhất. Đây là một yếu tố quan trọng để xây dựng lòng tin và lòng trung thành từ phía khách hàng. Ngoài ra, việc xác định và giải quyết các vấn đề về chất lượng thông qua những phản hồi từ khách hàng và cải tiến quy trình sản xuất cũng là một phần quan trọng trong việc phát triển sản phẩm. Từ việc phân tích dữ liệu bán hàng qua các năm, doanh nghiệp nên đề ra các biện pháp nâng cao chất lượng của các mặt hàng bán chạy, đồng thời cải thiện các sản phẩm chưa nhận được các phản hồi tích cực. Thêm vào đó, việc đánh giá và đổi mới thường xuyên cũng giúp doanh nghiệp duy trì tính cạnh tranh trên thị trường. Bằng cách theo dõi các xu hướng mới và so sánh với các đối thủ cạnh tranh, Olist có thể liên tục cải tiến sản phẩm của mình để đáp ứng được nhu cầu và sở thích ngày càng đa dạng của khách hàng.

*Thứ hai*, để thúc đẩy doanh số bán hàng, lượt tương tác của khách hàng và lượng mua sản phẩm, Olist Store có thể triển khai các sự kiện khuyến mãi theo mùa/ quý/ tháng, các chương trình như “Black Friday”, “11.11 - Quốc tế độc thân”,... Ngoài ra một số ưu đãi đặc biệt như giảm giá, quà tặng kèm theo, hoặc giao hàng miễn phí trong các đơn hàng đầu tiên sẽ kích thích sự quan tâm của khách hàng và thúc đẩy hành động mua sắm. Không chỉ vậy, các chương trình ưu đãi đặc biệt cho các khách hàng cũ nhằm củng cố lòng trung thành của họ đối với doanh nghiệp là vô cùng cần thiết. Một



số hoạt động như ngày hội thành viên, giá ưu đãi cho thành viên, miễn phí vận chuyển đối với các thành viên ở một số khu vực nhất định (VD: Sao Paulo, Acre,...) nhằm giữ chân khách hàng cũ, nâng cao tỷ lệ tái mua hàng và thu hút các khách hàng mới. Ngoài ra, dựa trên dữ liệu phân tích, ta có thể thấy xu hướng thanh toán chủ yếu của khách hàng hiện tại là bằng hình thức credit card, như vậy, nhằm tối ưu hóa chi phí vận hành, dễ dàng kiểm soát dòng tiền chi tiêu, Olist Store có thể kết hợp với các ngân hàng để cho ra các ưu đãi, khuyến mãi trực tuyến nhằm thu hút khách hàng mới thanh toán theo hình thức này. Điều này không chỉ đem lại lợi ích cho Olist Store trong tối ưu hóa chi phí vận hành mà còn giúp Olist Store dễ dàng hơn trong việc tiếp cận nhiều đối tượng khách hàng mới, đồng thời, giảm thiểu rủi ro về sai sót trong thanh toán.

*Thứ ba*, để phát triển mối quan hệ với khách hàng, doanh nghiệp cần tập trung vào việc cung cấp dịch vụ khách hàng tốt nhất và thúc đẩy trải nghiệm tích cực cho người mua hàng. Điều này có thể đạt được thông qua việc đảm bảo rằng mọi thắc mắc của khách hàng được giải quyết một cách nhanh chóng và hiệu quả, cùng với việc tạo ra một môi trường mua sắm trực tuyến thoải mái và tiện lợi. Bên cạnh đó, việc lắng nghe và phản hồi tích cực đối với ý kiến và đề xuất từ phía khách hàng cũng là một phần quan trọng trong việc xây dựng lòng tin và sự cam kết với sự hài lòng của họ. Đồng thời, phản hồi nhanh chóng và chuyên nghiệp đối với mọi đánh giá tiêu cực cũng giúp công ty thể hiện sự chăm sóc và quan tâm đến nhu cầu của khách hàng. Cuối cùng, doanh nghiệp cần cố gắng giảm thiểu thời gian giao hàng giữa thời gian dự kiến và thực tế, đồng thời khắc phục vấn đề trì hoãn giao hàng và đơn hàng không được giao đến. Doanh nghiệp nên áp dụng các biện pháp như phân tích vị trí địa lý, theo dõi hiệu suất cũng như cung cấp hệ thống theo dõi vận chuyển, cập nhật thông tin thường xuyên và tập trung vào phân phối người bán trong một khu vực cụ thể có thể giúp cải thiện hiệu suất giao hàng trong khu vực đó. Bằng cách thực hiện những chiến lược trên một cách có hiệu quả, Olist có thể xây dựng một mối quan hệ mạnh mẽ với khách hàng và tạo ra cơ hội tăng trưởng bền vững trong tương lai.

*Thứ tư*, nhằm tối ưu hóa chi phí vận hành, chúng tôi đề xuất Olist Store cần có những chiến lược về giá, chiến lược cắt giảm chi phí như giảm số lượng hàng trả lại sản phẩm bằng cách cung cấp hiển thị tốt hơn, đánh giá rõ ràng và mô tả chi tiết (Gautam & Kale, 2019), đàm phán với nhà cung cấp để chi phí mua sắm thấp, và nó sẽ giúp tăng lợi nhuận của doanh nghiệp (Gautam & Kale, 2019), đàm phán phí xử lý thẻ tín dụng thấp hơn và tìm kiếm hợp tác tích cực hơn với các ngân hàng thanh toán (Gautam & Kale, 2019). Ngoài ra, Olist Store cần lập kế hoạch kinh doanh nhằm xác định hướng đi tiếp theo cho doanh nghiệp, tiến hành nghiên cứu thị trường phù hợp với nhu cầu hiện tại của Olist Store giúp tiết kiệm chi phí nhưng vẫn đạt được hiệu quả vận hành.

*Cuối cùng*, để tối ưu hoá hiệu quả trên các phương tiện truyền thông, doanh nghiệp cần phải phát triển thêm các kênh quảng cáo đa dạng như mạng xã hội, trang web, email marketing và quảng cáo truyền thông. Bên cạnh đó, việc tăng cường chia sẻ trải nghiệm sản phẩm thông qua các phiên bản thử nghiệm miễn phí, demo sản phẩm trực tiếp trên mạng xã hội cũng góp phần quan trọng trong việc xây dựng lòng tin và quan hệ với khách hàng. Ngày nay, sức ảnh hưởng của các người nổi tiếng trên các phương tiện truyền thông đang ngày càng phát triển. Vì vậy, việc hợp tác với các đối tác chiến lược và những người có tầm ảnh hưởng trong ngành sẽ đóng vai trò quan trọng trong việc tăng cường nhận thức và uy tín của sản phẩm, từ đó thu hút và giữ chân khách hàng một cách hiệu quả hơn.

## CHƯƠNG 6: KẾT LUẬN

### 6.1. Kết luận

Trong thời đại mà công nghệ thông tin đang phát triển một cách mạnh mẽ như hiện nay thì việc ứng dụng các sản phẩm đã được nên từ công nghệ vào cuộc sống là một điều đương nhiên. Tuy nhiên, việc làm cách nào để có thể áp dụng nó và giúp cho công việc mang lại hiệu quả tốt nhất là điều mà không phải bất kỳ ai cũng làm được. Phải trải qua quá trình học hỏi, rèn luyện và đúc kết kinh nghiệm mới có thể giúp cho những dự án được thực hiện thông qua các hệ Hỗ trợ quản trị thông minh (Business Intelligence) đem đến hiệu quả cao.

Thông qua đồ án này, nhóm đã sử dụng rất nhiều công cụ để hỗ trợ cho việc phân tích, nghiên cứu, trình bày dữ liệu một cách hiệu quả và trực quan. Những công cụ đã được sử dụng có thể kể đến như Microsoft Power BI, Excel và RStudio. Việc sử dụng các công cụ trên đã giúp ích rất nhiều cho nhóm trong quá trình nghiên cứu về hoạt động kinh doanh của Olist Store. Từ đó cho thấy rằng tầm quan trọng và khả năng mà việc áp dụng các Hệ hỗ trợ quản trị thông minh có thể giúp cho người dùng khi thực hiện quá trình phân tích dữ liệu không chỉ về hoạt động kinh doanh mà còn trong nhiều lĩnh vực khác.

Sau khi thực hiện việc nghiên cứu về đề tài “Phân tích tình hình kinh doanh của Olist store và dự đoán xu hướng phát triển ngành thương mại điện tử tại Brazil”, nhóm đã đúc kết được những kết quả như sau: Đầu tiên là nhóm đã thực hiện được quá trình phân tích về tổng quan tình hình của doanh nghiệp. Thứ hai là phân tích các dữ liệu, xử lý những dữ liệu thừa và thực hiện thống kê mô tả. Cuối cùng là trực quan hóa dữ liệu, giúp cho việc phân tích và truyền đạt thông tin được thực hiện một cách rõ ràng và hiệu quả hơn.

Có thể nói rằng việc áp dụng các Hệ hỗ trợ quản trị thông minh đem lại rất nhiều lợi ích to lớn, trong đó điều quan trọng nhất phải kể đến đó là nó đã là một công cụ đắc lực giúp Olist Store nói riêng và các doanh nghiệp khác nói chung đưa ra những quyết

định chính xác và những hướng đi phù hợp với định hướng phát triển trong tương lai và làm gia tăng tính cạnh tranh trong quá trình kinh doanh của mình. Đồng thời, các kết quả thu thập được từ quá trình phân tích sẽ là tài liệu vô cùng quan trọng để Olist có thể hiểu rõ được tình hình của doanh nghiệp từ đó nghiên cứu ra giải pháp để cải thiện và giúp cho doanh nghiệp ngày càng phát triển hơn.

## **6.2. Hướng phát triển**

Trong tương lai, để Hệ hỗ trợ quản lý thông minh (business intelligence) đem lại hiệu quả phân tích tốt hơn cho Olist Store, nhóm sẽ ứng dụng thêm các tính năng mở rộng của phần mềm PBI như Microsoft Fabric. Giúp doanh nghiệp phản ứng tức thì với những thay đổi trong dữ liệu của họ bằng cách thiết lập một hệ thống phát hiện tự động cảnh báo cho nhóm về ngữ cảnh phù hợp để thực hiện hành động ( Data Factory, Synapse Data Engineering, Synapse Data Warehouse...). Bên cạnh đó, nhóm cũng có định hướng ứng dụng các mô hình học máy (machine learning) nhằm phân tích và dự báo các yếu tố ảnh hưởng đến doanh thu, lượng đơn hàng và khách hàng. Từ đó nắm rõ insights, nhằm đưa ra các quyết định đúng đắn hơn với sự trợ giúp của các mô hình dự báo.

## PHỤ LỤC

### Tài liệu tham khảo:

Lê T. (2022, March 21). 6 Lợi Ích Áp Dụng Thanh Toán Điện Tử Đối Với Doanh Nghiệp. *Abaha – Giải pháp xây dựng Mobile App TMDT chuyên nghiệp*.  
<https://abaha.vn/blog/loi-ich-cua-thanh-toan-dien-tu/>

*View of cost optimization in E-commerce industry: An analytical study.* (n.d.).

Thinkindiaquarterly.org. Retrieved May 5, 2024, from  
<https://thinkindiaquarterly.org/index.php/think-india/article/view/15468/10603>