**Final Project Report**

**on**

**Order Management of E-Commerce Business**

**Towards partial fulfilment for course**

**CST2205-Data Modeling**

**Submitted by:**

**Kathan Patel** (041081063)

**Rohan Mistry** (041081134)

**Rishabh Panchal** (040906654)

**Aman Bhatu** (041055122)

**Tanay Shah** (041055663)

Table of Contents

[Table of Figures 3](#_Toc163501754)

[Executive Summary 4](#_Toc163501755)

[Data Models & Analysis 5](#_Toc163501756)

[Order Placement 5](#_Toc163501757)

[Order Fulfilment and Shipping 7](#_Toc163501758)

[Invoicing 10](#_Toc163501759)

[Returns and Exchanges 13](#_Toc163501760)

[Customer Reviews and Feedback 16](#_Toc163501761)

[Business Matrix and Final Model 20](#_Toc163501762)

[Business Matrix 20](#_Toc163501763)

[Final Model 21](#_Toc163501764)

[Conclusion 22](#_Toc163501765)

[Recommendations & Suggestions 23](#_Toc163501766)

[References 24](#_Toc163501767)

Table of Figures

[Figure 1 4](#_Toc163501494)

[Figure 2 5](#_Toc163501495)

[Figure 3 6](#_Toc163501496)

[Figure 4 7](#_Toc163501497)

[Figure 5 8](#_Toc163501498)

[Figure 6 9](#_Toc163501499)

[Figure 7 10](#_Toc163501500)

[Figure 8 11](#_Toc163501501)

[Figure 9 11](#_Toc163501502)

[Figure 10 11](#_Toc163501503)

[Figure 11 12](#_Toc163501504)

[Figure 12 13](#_Toc163501505)

[Figure 13 14](#_Toc163501506)

[Figure 14 15](#_Toc163501507)

[Figure 15 16](#_Toc163501508)

[Figure 16 17](#_Toc163501509)

[Figure 17 18](#_Toc163501510)

[Figure 18 20](#_Toc163501511)

# Executive Summary

This project report delves into the comprehensive study and improvement strategies of an e-commerce business's order management processes, utilizing the Brazilian E-Commerce Public Dataset by Olist[1]. The project meticulously analyses five critical business processes: Order Placement, Order Fulfillment and Shipment, Invoicing, Exchange & Returns, and Customer Review/Feedback.

Order Placement: Analysis of sales order data revealed important seasonal trends and geographical factors, informing sales optimization decisions. Orders showed a significant increase from 2016 to 2018, indicating increased customer engagement.

Order Fulfillment and Shipment: Order management opened areas for improvement in delivery timelines and delivery. Despite the decrease in average processing time, there has been an alarming increase in cases of delays and unavailability of orders.

Invoicing: Managing the invoicing process demonstrated operational efficiencies along with a great sales focus. In the second quarter, demand increased significantly, with toys, phones and games ranking as the top selling items.

Exchanges and Returns: The study focused on the 16% rate of returns, breeding strategies to reduce return rates and facilitate the exchange process.

Customer Reviews and Feedback: Customer reviews and feedback were essential to improve service delivery. A comparison of historical and current feedback highlighted growing positive consumer sentiment. However, the points of dissatisfaction pointed to the need to actively address factors affecting customer satisfaction.

The data model was created using Microsoft Power BI tool.

Throughout our project, we maintained best practices such as consistent naming of surrogate keys and optimizing model performance. The common calendar table served as a foundational element, ensuring that all date-related information was consistent. Our findings and subsequent Power BI data models provided actionable insights. For example, improving product descriptions can reduce returns, and improving inventory management can streamline operations. By understanding the drivers of customer issues, we can adjust our business processes to ensure a more satisfying shopping experience, maintain strong initial sales and reduce post-sale processing activities.

# Data Models & Analysis

## Order Placement

Order placement is the first stage of the ordering process, denoting the moment when a consumer indicates that they would like to buy a good or service. During this stage, the consumer makes item selections, specifies quantity, including the shipping address and estimated delivery date.

The particular business objectives this data model deals with are:

1. Utilizing order placement data enables us to analyze the seasonal trends in customer purchasing behavior, we can devise strategies to optimize sales performance across all seasons.
2. By leveraging order placement data, we can identify opportunities to enhance sales in specific geographic locations through the introduction of related products tailored to the preferences and needs of local customers.

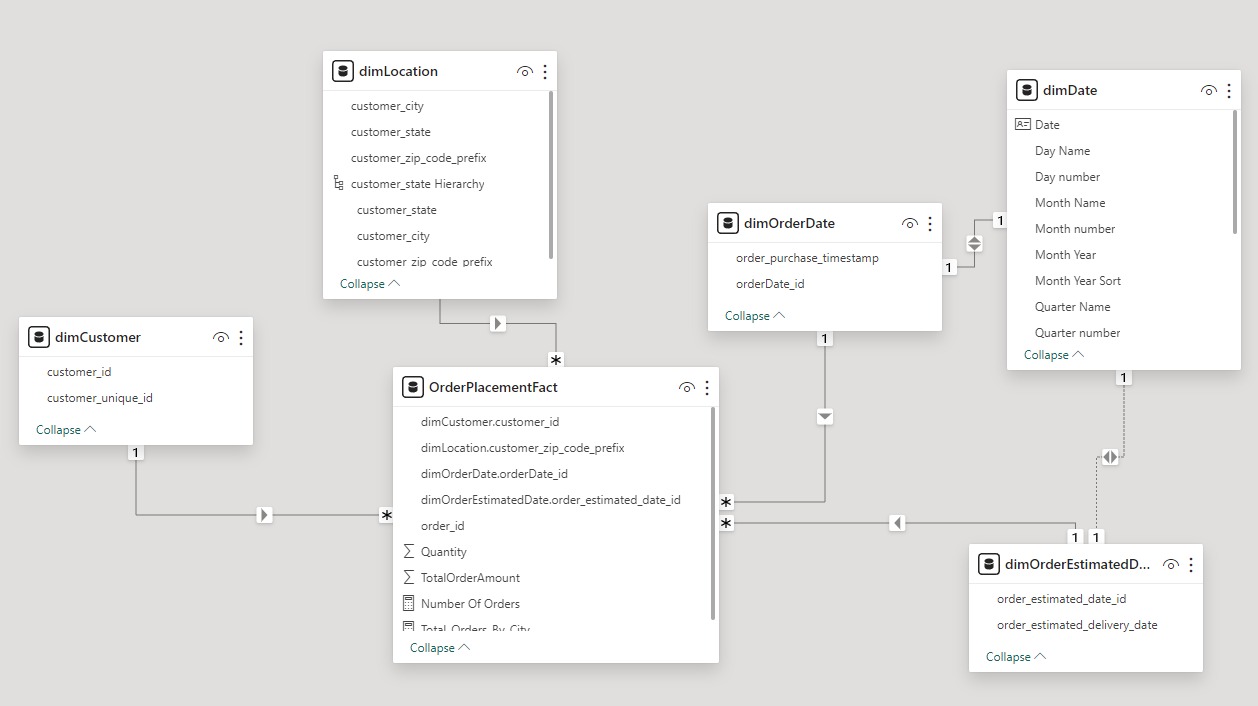


Figure 1

**Data Model:**

The fact table named “OrderPlacementFact” is a transactional type of fact table. The grain of each row in the fact table is at order level. “order\_id” is the degenerate dimension. Dimensions included in the model are

(1) dimLocation - Conformed Dimension

(2) dimCustomer - Conformed Dimension

(3) dimOrderDate - Roleplaying Dimension

(4) dimDate - Roleplaying Dimension

(5) dimOrderEstimatedDeliveryDate - Roleplaying Dimension

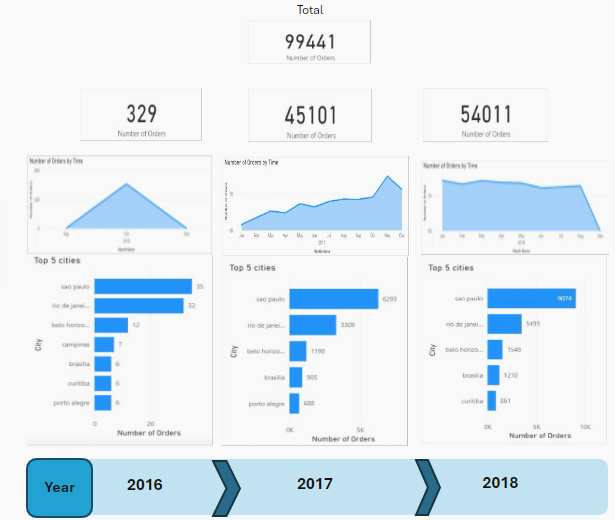


Figure 2

Analysis:

In analysis, throughout 3-year period, 99441 orders were placed using this ecommerce website. First year, in 2016, 329 orders were places which increased to 45101 in 2017 and that hiked to 54011 orders in 2018. Around average of 20% increase in number of orders placed throughout the years. Furthermore, for monthly breakdown, for the first-year highest sales were seen in mid of last quarter of the year, which stayed consistent in following year, and atlast, last year had same order placed increased to average of 7000 per month, which was consistent till year end and there was deep decline in last month in number of order placed. Sao Paulo was the city with highest order placed throughout three years of period. Using this analysis, one can improve our sales throughout the year by incorporating marketing strategies and different approaches, as well as, based on customer’s usage and needs, one can also scale this business geographically.

## Order Fulfilment and Shipping

The order fulfilment and shipping business process tracks an order throughout its life cycle from the moment it is placed to the final delivery of the order. The journey of an order is marked by a series of status updates, which provide valuable data points for analysis. The various statuses involved in the journey of order are: created, approved, shipped, delivered, invoiced, cancelled or unavailable.

The particular business questions this data model deals with are:

(1) What number/percentage of orders are completed/delivered to customer?

(2) How much average time does it take for an order to reach the customer after it is placed?

(3) What is the number of orders that were delivered later than the estimated date?

Figure 3

**A screenshot of a computer

Description automatically generated**Data Model:

The fact table named “orderFact” is a transactional type of fact table. The grain of each row in the fact table is at order level. “order\_id” is the degenerate dimension. The conformed dimensions included in the model are (1) dimLocation (2) dimCustomer. The other dimension included is the order status dimension is named “dimOrderStatus”. Finally, the date dimension named “dimDate” is the role-playing dimension. It is connected to various fact dates such as “order\_purchase\_timestamp”, “order\_approved\_date”, “order\_delivered\_carrier\_date”, and “order\_delivered\_date”.

A screenshot of a cell phone

Description automatically generated

Figure 4

Analysis:

In this visual, the first metric is “Number of orders fulfilled”. This metric gives the total percentage of orders with the status of “delivered” during the year 2016. As it is evident from the visual that the percentage of orders successfully delivered to the customer is low which shoots up to nearly 96.5% in 2017 and increases to approximately 97.7%. This shows that the delivery process has improved throughout the years.

The next metric is average lead time which is the number of days it takes on an average for an order to delivered to the customer after the order has been placed. It is observed that the average lead time in 2016 was 20 days which is not consistent with the industry standards with other e-commerce businesses like Amazon providing delivery times as low as 2 days. The current lead time for this business is almost 10 times more than the industry leaders. This number has significantly reduced in 2017, to 13 days which indicates that the process improvements made by the company has been successful. However, this improvement is not significant in 2018.

The last metric for the analysis is the number of late deliveries which indicates the number of orders which were delivered later than the estimated delivery date given to the customer. 3 late deliveries in 2016 may indicate the number is low, however when we consider number of orders placed in 2016 was around 625 orders, this may not be a good indication. In 2017, the number of late deliveries shot up to approximately 2500 which is a bad indication and may lead to customer dissatisfaction. In 2018, this number almost doubled which indicates that the company has not been paying attention to it or the process improvement is place is harming the time it takes an order to be delivered. This is a serious concern and the company should definitely look into it.

A screenshot of a computer

Description automatically generated

Figure 5

Another analysis is the orders with “cancelled” or “unavailable” status. This is another concern for the business. This indicates the orders which are either cancelled by the customers or the customer ordered but the order was not available. Both of these statuses indicate customer dissatisfaction. The cancelled and unavailable orders were relatively low in 2016. However, both of these numbers shot up in 2017, even though the percentage seems to be decreasing. Lastly in 2018, the numbers seems to be decreasing, which indicates that the measures implemented are having positive effect and the company needs to keep on going.

In conclusion, the increase in late deliveries and presence of unavailable or cancelled ordered may cause customer dissatisfaction which may lead to customer choosing other e-commerce platforms.

## Invoicing

The objective of this business process is to optimize the E-Commerce invoicing process for improved financial accuracy, record-keeping, and deeper sales analysis. This involves meticulously recording invoices issued to customers for purchased products and tracking sales data derived from these invoices. By enhancing these aspects, the aim is to streamline operations, enabling better understanding of financial transactions, and facilitating informed decision-making for efficient business management.

Data Model:

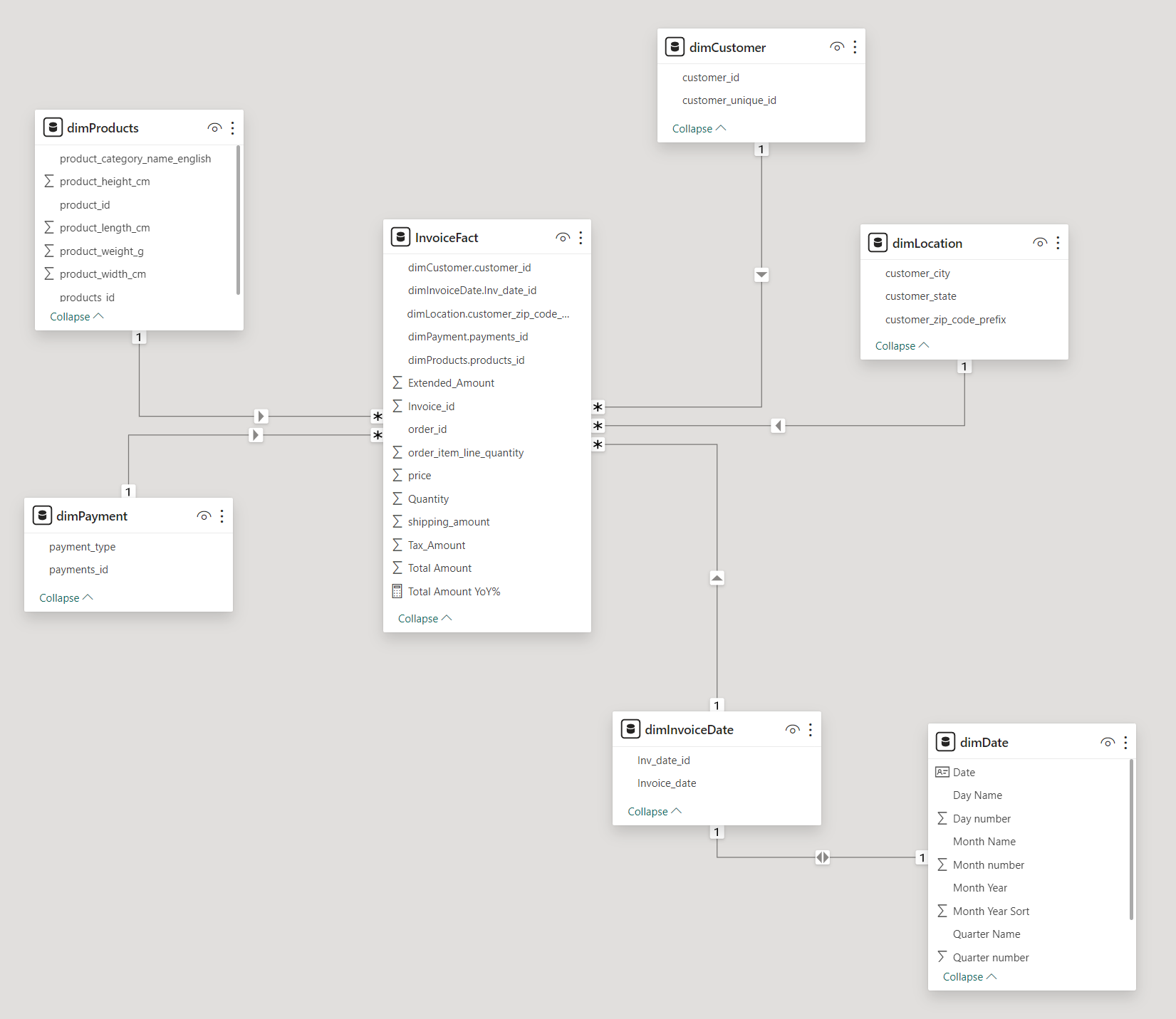
****

Figure 6

Fact Table: Transactional, Contains quantitative facts like Quantity, Price, Shipping Amount, Extended Amount, Tax Amount & Total Amount.

Dimensions:

* + Products (Conformed)
  + Customer (Conformed)
  + Location (Conformed)
  + Payment
  + Invoice Date – Date
  + Invoice\_id (Degenerate Dim)

Granularity: Invoice Data can be Drilled downed by Date & Location.

Analysis:

A screenshot of a computer

Description automatically generated

Figure 7

The figure 7 above shows the dashboard made with the Invoices data model, it can be filtered through Years & Quarters from the left, with significant information of Revenue over Time, Top Products Sold, Top Sales Location, Invoices by Payment Type, Products Sold, Returns & Exchanges with some key metrics at the top.

Overall Insights based on the invoices:

* + Over the span of 2 Years, Total Revenue was $**13.66M** from over **113K** Products Sold.
  + Most sold products were **Sports**, **Watches, Telephony and Toys**
  + Most orders are placed from customers residing in **Sao Paulo, Rio de Janeiro**
  + **Credit card** is the most common used method of payment

To understand the sales better, the dashboard helps to drill down the data by Date (Year, Quarters, Months, Date). Let’s look at the seasonality of the sales to better prepare the business and our fulfillment services and dive deeper into some analysis to help the business achieve higher sales.

A graph with a line

Description automatically generated

Figure 8

Figure 8 shows that on average from the Years 2017 & 2018, Sales trend seasonality starts strong with Q1 with **Q2** being the highest in sales then it dies down by Q4. This can give insight into when to have business ready with exceptional fulfillment & shipping with full inventory to meet the demands and grow.

A black arrow pointing to a calendar

Description automatically generated

Figure 9

Figure 9 shows the Top products being sold by our business, 2017 had strong sales which helps us narrow down highest sales products like **sports leisure, toys, watches gifts, telephony**, etc. Some of the products remain consistent for the following year, but we can see a drop in sales of toys and a growth in sales in watches gifts and telephony meanwhile sports leisure being our strong selling product.

**A blue and purple squares

Description automatically generated**

Figure 10

Figure 10 shows the aggregated Top Locations for our sales over 2016-2018, The ones that stands out is **Sao Paulo** and **Rio de Janeiro.** This gives us insight into which location our customers reside and which locations to focus on for any promotions, offers to better our sales and to improve our fulfillment and shipping.

Returns and Exchanges

Introduction:

Our data example had two main fact tables - Return Fact and Exchange Fact, both of Transactional type. The combined concepts were products, customers, and locations, which were made to match other data. Degraded dimensions such as Return Invoice ID and Exchange Invoice ID were used to maintain the granularity of transaction data without the need for separate dimension tables Primary keys include unique identifiers for customers, products, and orders, which ensured that each record was unique.

A screenshot of a computer

Description automatically generated

Figure 11

1. **dimLocation**: Contains location-related details for customers such as **customer\_city**, **customer\_state**, and **customer\_zip\_code\_prefix**.
2. **dimCustomer**: Holds customer information with fields like **customer\_id** and **customer\_unique\_id**.
3. **dimProducts**: Details about products including **product\_category\_name\_english**, **product\_id**, dimentions, and **products\_id**.
4. **ExchangeFact**: This fact table seems to record data about product exchanges. It references **dimCustomer**, **dimLocation**, and **dimProducts** and includes fields like **Exchange Difference Amount**, **Exchange ID**, **Exchange Invoice Id**, **Number of Exchanges**, and **Order ID.**
5. **ReturnFact**: It includes references to **dimCustomer**, **dimLocation**, **dimProducts**, and **dimReturn**. It has fields such as **Number of Returns**, **Order ID**, **Price**, **Return Invoice ID**, and **Return Shipping Count**.

Each of these entities is connected, suggesting how facts interrelate with each other , such as customer details link to product returns or exchanges. The arrows among tables generally defines relationships, like foreign keys, which may be one-to-one (1:1) or one-to-many (1:N), displaying how each access in a single desk can relate to 1 or multiple entries in some other.

Deep Dive into the Problem:

Our problem statement revolves around understanding the causes of relapse and change and finding ways to reduce these events. The key performance indicators that determined this were the number of items returned, return processing time and associated costs including freight costs.

Insights and Solutions through Power BI Visualization:

A graph of sales

Description automatically generated

Figure 12

Analysis showed that while sports leisure sold 9K units, its rivals also faced 2K exchanges. This increased customer engagement still indicates that the conversion process can be greatly improved. Watches gifts and the telephone teams followed with a significant return, highlighting the need to reassure customers and improve product satisfaction. In contrast, toys and stationery showed proportionately lower or adjusted returns, despite declining sales.

A blue circle with a red and orange center

Description automatically generated with medium confidence

Figure 13

Through Power BI visualization, we found that the sales success rate was 82% of transactions, with a profitability of 16% and a conversion rate of 2%. This distinction between sales and post-sales activities highlighted the robustness of our pre-sales process. However, important gains required a review of strategies to identify and address the causes of dissatisfaction.

Strategic Recommendations:

In response to the question, we suggest providing targeted surveys of customer satisfaction, especially in the sport leisure category, to identify specific reasons for switching. For watches gifts and phones, it is recommended to increase online product descriptions and images to reduce returns. In addition, having a strategy for toys and stationery could fine-tune the supply chain and better align with customer expectations.

In summary, the survey highlights the sales strengths and highlights opportunities to improve post-sale customer experiences. By implementing the proposed strategies, The aim is to provide a better customer journey, thereby reducing profitability and variability and improving overall satisfaction.

## Customer Reviews and Feedback

Overview

The Customer Review and Feedback Analysis Data Model gathers customer feedback at various points in time about our products. It collects their detailed opinions, ratings, and thoughts, and ties this information to their backgrounds and specific products. Through this process, we can better understand how our customers feel, see what's trending in their opinions, and make products and services that meet their needs.

By thoroughly analyzing customer feedback, one of our primary objectives is to elevate the quality of our services. This process will allow us to identify specific areas where our services can be improved in order to meet and exceed the expectations of our customers. We also seek to understand prevailing trends through our customers' feedback, which is invaluable in adapting our business strategies. By taking into account our customers' needs and preferences, we are able to evolve our service offerings dynamically and stay ahead of the competition.

Data Model:

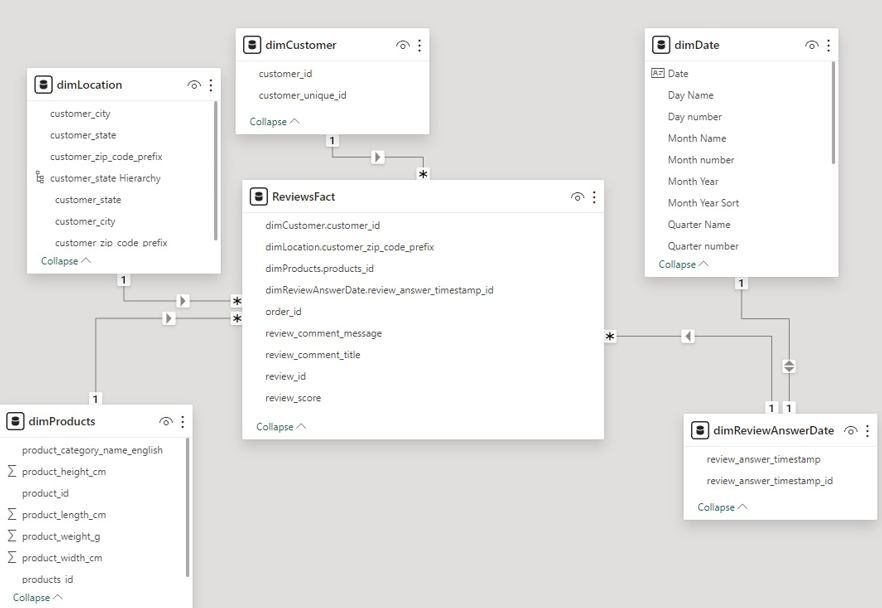


Figure 14

The data model presented here is a well-structured, transactional system designed to capture and analyze customer reviews and feedback comprehensively. Central to this model is the **‘ReviewsFact’** table which serves as the primary repository for storing detailed feedback, encompassing review messages, titles, and scores. This table is intricately linked to a series of dimensions that add depth and context: **‘dimProducts’** introduces product-related data, **‘dimCustomer’** infuses customer demographics, **‘dimLocation’** gives geographic insights, and **‘dimReviewAnswerDate’** captures the timing of each review. Notably, **‘dimReviewAnswerDate’** is connected to the **‘dimDate’** calendar table, providing a link to standardized date information and allowing for time-based analysis.

**Analysis:**

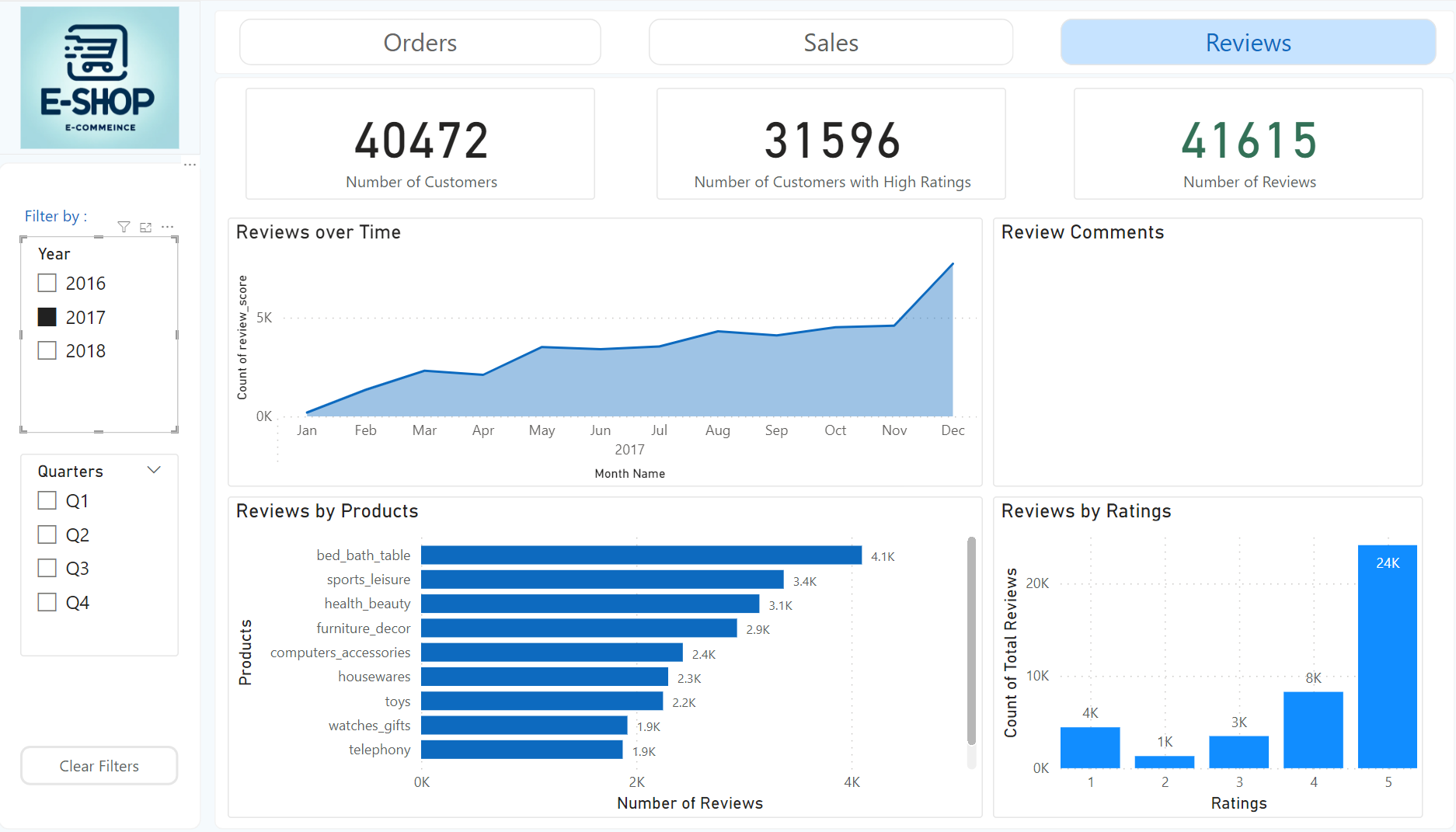


Figure 15

The dashboard presents a positive narrative of customer interactions and satisfaction. In total, 55,288 customers placed orders, and 42,510 of these customers rated their experience highly. As a result of this validation, we can be confident that our services and products are of high quality.

Based on the review frequency over the months, we can gain insight into patterns of customer engagement. As part of the analysis, the fluctuations observed will be examined in greater detail in order to determine the underlying factors influencing the shoppers' behavior, such as seasonal or promotional factors.

Dashboard filters, known as slicers, allow users to examine data in depth across different years and quarters. With these slicers, we are able to isolate specific time periods in order to discern trends and isolate events or changes that may have influenced customer behavior.

Based on customer feedback, a word cloud can be used to visualize the most prominent themes. Qualitative data is essential for identifying what resonates with our customers and pinpointing areas for improvement.

According to the analysis of product categories, health, beauty, and bed, bath, and table products receive the most reviews, indicating potential high engagement sectors. Based on this information, we will conduct a detailed review of these categories in order to ensure that they continue to meet and exceed customer expectations.

According to the 'Reviews by Ratings' distribution, the majority of our feedback is overwhelmingly positive, with 5-star ratings predominating. Our commitment to quality is reflected in this positive customer feedback. Lower ratings, however, are a valuable indicator of areas that need to be addressed, and strategies will be developed to address these issues.

As a result, the dashboard serves as a testament to our successful customer engagement strategy and provides a wealth of information to guide our ongoing efforts to enhance our offerings.



Figure 16

We have observed a positive trend in customer satisfaction over the last three years when examining the customer ratings. Our 5-star ratings increased significantly in 2016, indicating that our products and services were well received by our customers.

In 2017, we continued to see high levels of satisfaction, as evidenced by an even higher rate of 5-star ratings. As a result of customer-centric initiatives and operational improvements implemented in response to feedback from the previous year, this peak is a reflection of the effectiveness of our customer-centric initiatives.

In 2018, although our 5-star ratings remained above the 2016 benchmark, we did observe a slight decline from last year. This presents us with an opportunity to examine closely the customer experience strategies and operational changes that have been implemented between 2017 and 2018. In order to maximize customer satisfaction, we will need to understand the nuances behind this slight decline.

Furthermore, the reduction in 1-star and 2-star ratings from 2016 to 2018 is indicative of the successful mitigation of customer complaints. It is evident from this decrease that we are committed to resolving issues efficiently and improving the overall quality of our service. In spite of this progress, the slight decline in top ratings in 2018 has prompted us to reevaluate our current approaches and adapt to changing market dynamics and customer expectations.

Based on these ratings trends, we will be able to develop strategies that will be effective in the future. In addition to restoring our upward trajectory of customer satisfaction, our objective is to preemptively address factors that may negatively affect our ratings. As long as we keep track of customer feedback and are agile in our response, we will be able to enhance customer relationships and foster loyalty among our customers.

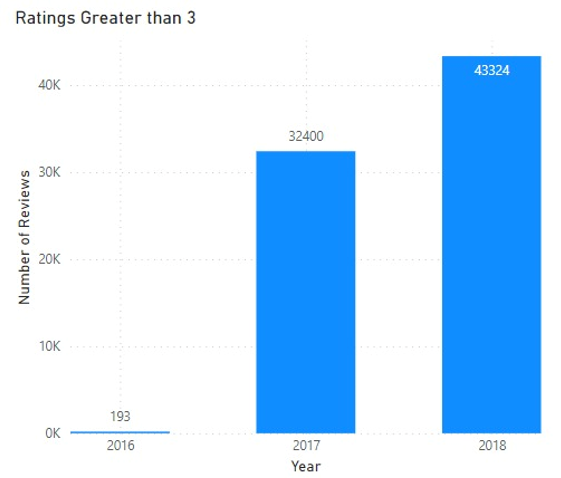


Figure 17

A significant increase in positive reviews was observed between 2016 and 2018, indicating an increase in customer satisfaction. The number of good reviews increased from 193 in 2016 to 43,324 by 2018. As a result of our continuous commitment to customer service excellence, we have seen substantial growth during this period, which may be attributed to enhanced product offerings and responsive customer service initiatives.

# Business Matrix and Final Model

## Business Matrix

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Products** | **Customers** | **Location** | **Date** | **Order Status** | **Payments** | **Returns** |
| Order Placement |  | X | X | X |  |  |  |
| Order Fulfilment |  | X | X | X | X |  |  |
| Invoicing | X | X | X | X |  | X |  |
| Exchange | X | X | X | X |  |  |  |
| Returns | X | X | X | X |  |  | X |
| Customer Review/Feedback | X | X | X | X |  |  |  |

## Final Model

The image below shows our final model which connects facts and dimensions from all the business processes.

**A screenshot of a computer

Description automatically generated**

Figure 18

# Conclusion

Our exhaustive study and subsequent improvements in the order management processes of an e-commerce business, using the rich Brazilian E-Commerce Public Dataset by Olist, have yielded significant insights across five key business processes: Order Placement, Order Fulfillment and Shipping, Invoicing, Exchanges and Returns, and Customer Reviews and Feedback.

Order Placement highlighted an upward trend in customer engagement, with a noteworthy rise in orders placed over a three-year period. This indicates that our strategic initiatives are resonating with customers and driving increased interaction with our platform.

Order Fulfillment and Shipping showed improvements in the delivery process yet flagged an increase in delayed and unavailable orders. This contrast points us towards specific operational areas needing targeted improvements to enhance the customer's end-to-end experience.

Invoicing presented us with insights into operational efficiencies, and sales trends that reflected a significant demand increase, particularly in specific product categories like toys, phones, and games. The ability to track and analyze financial transactions through invoicing has empowered us to make informed decisions for business management.

Exchanges and Returns provided an actionable focus with a 16% return rate, leading to strategies aiming to lower this figure and streamline the exchange process. This is crucial for maintaining customer trust and reducing logistical overheads.

Customer Reviews and Feedback were instrumental in capturing the voice of the customer, revealing a growing positive sentiment. The rich feedback, when analyzed, also pointed towards potential improvements that can elevate the customer experience further.

Each of the data models created using Microsoft Power BI has been integral in providing us with actionable insights. Our meticulous approach to best practices, such as consistent naming conventions and performance optimization, has ensured the reliability of our data insights. The common calendar table, in particular, has been pivotal in maintaining consistent date-related information across the models.

In conclusion, our in-depth data modeling not only unearthed pivotal insights for the e-commerce industry but also solidified a methodical path towards perpetual improvement. Our collective endeavors have not only resolved immediate operational challenges but have also laid a foundation for sustained growth and heightened customer satisfaction.

We have observed that each business process interconnects with another, underscoring the importance of a holistic view in operational strategy. By persistently aligning our business processes with customer feedback and market demands, we have been able to maintain strong sales momentum and address post-sale activities more effectively. Our commitment to refining the customer journey through targeted enhancements has positioned us to proactively adapt to customer expectations, reduce return and exchange rates, and bolster overall satisfaction.

# Recommendations & Suggestions

* To drive the number of orders higher, the company should increase the variety of products and increase the marketing efforts in the region where the sales are relatively low.
* The company should focus on late deliveries and should consider improving the delivery process to get the lead times under 5 days which may improve late deliveries as well.
* The products which have more number of review score less than 2 should be investigated and the quality of products should be improved or else discarded given their sales were not very high.
* Root Cause Analysis for orders with order status cancelled and unavailable along with orders that were exchanged or returned should be investigated and appropriate solutions.

# References

1. “Brazilian E-Commerce Public Dataset by OLIST,” Oct. 01, 2021. <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>