

Table of Contents

Reporting of Analysis:	2
Introduction:	2
Business Problem:	2
Stakeholders:	3
Data Source:	3
Storyline and Visualization:	5
Objective of Analysis:	5
Analysis:	6
Data Model:	6
Dashboard:	7
Executive Summary Dashboard	7
Exploratory Analysis Dashboard	8
Customer Behavior Analytics	9
Seller Analysis Dashboard	10
Delivery Analysis Dashboard	11
Forecasting Dashboard	12
Insights:	12
Recommendations:	13

Reporting of Analysis:

The Brazilian E-Commerce Public Dataset is a comprehensive collection of retail data encompassing 100,000 orders from Olist, placed between October 2016 and September 2018 across diverse states. This dynamic dataset evolves over time as new information is progressively added. It provides a wealth of metrics such as pricing, order numbers and statuses, payment methods, freight costs, and customer reviews, offering a well-rounded understanding of e-commerce patterns and trends.

Introduction:

Businesses have consistently strived to retain their customer base's engagement and satisfaction with the services they offer. In order to stay competitive in today's market, it is crucial for them to incorporate the latest technological advancements into their service delivery mechanisms.

Over a decade ago, when the Internet was still an emerging technology, numerous industries sought to leverage its potential as a communication conduit between various enterprises. This was a formative step in harnessing the power of digital transformation.

In the current decade, industries have evolved to provide services that are intricately tailored to meet the unique requirements of each customer. This level of personalization and precision necessitates the use of artificial intelligence, which is now integral to offering such tailored services.

Business Problem:

Olist, an online store with its corporate headquarters in Sao Paulo, Brazil, serves as a central hub for numerous small businesses and customers seeking to purchase their products. The company has shared a dataset on Kaggle containing details of 100,000 orders placed across various marketplaces between 2016 and 2018.

Customer reviews on e-commerce websites significantly influence our purchasing decisions. As such, Olist can utilize this feedback to streamline their product offerings, potentially eliminating products that consistently receive poor reviews. Conversely, they could choose to promote products that have been well-received by consumers.

A Power BI-generated dashboard provides a comprehensive view of overall customer satisfaction with purchased products. This platform employs both descriptive and predictive analysis to give a detailed account of past consumer behaviors and forecast future trends.

Stakeholders:

Olist's organizational structure is clearly defined and hierarchical, ensuring efficient administration and smooth operations. Clear communication channels are made possible by the organization's hierarchical structure, ensuring that every decision made at the top is effectively trickled down.

The CEO, who acts as the main decision-maker, is at the top of this structure. The CEO's choices, from evaluating and approving budgets to assuring the effectiveness of subordinate teams, are essential to the direction of the business. As a shareholder, the CEO, who is frequently the chairman or owner, has a large financial stake in the business.

In the hierarchy, the Marketing Team is in second place. The promotion of Olist's goods and services is greatly aided by this team, which uses cutting-edge technology to remain competitive. As stakeholders, the marketing division makes sure the business is visible and appealing to potential customers.

The Finance Team comes after the Marketing Team. The management of Olist's accounts and maintenance of the company's profitability depend on the work of this team. They manage cash flows, debts, loans, depreciation, income, taxes, and other financial matters while staying within set budgetary parameters. Additionally, this team offers monetary options like conditional coupons to boost sales or control the marketing budget necessary for new projects.

The Information Technology (IT) Team is the next group, which is essential to any e-commerce firm. Furthermore, the Olist IT infrastructure is managed by this team, which also provides assistance to other teams inside the company. They are essential to maintaining e-commerce operations efficiently and without hiccups. The organization's E-commerce Team, another essential part, manages the online sales procedure. They engage with consumers, make sure transactions go well, and help to increase customer loyalty and retention.

The CEO, Marketing, Finance, IT, and E-commerce departments, however, are the main parties participating in and impacted by this initiative.

It is easy to pinpoint the stakeholders who would be most impacted and invested in the project's outcomes by comprehending this organizational structure. For the project to be a success and for Olist to continue to develop and prosper, it is essential to address their particular needs and concerns.

Data Source:

The plan is to leverage the dataset generously provided by Olist, the largest department store in Brazilian marketplaces. Olist functions as a crucial link, connecting small businesses from all over Brazil to various distribution channels, thereby streamlining administrative tasks and consolidating legal

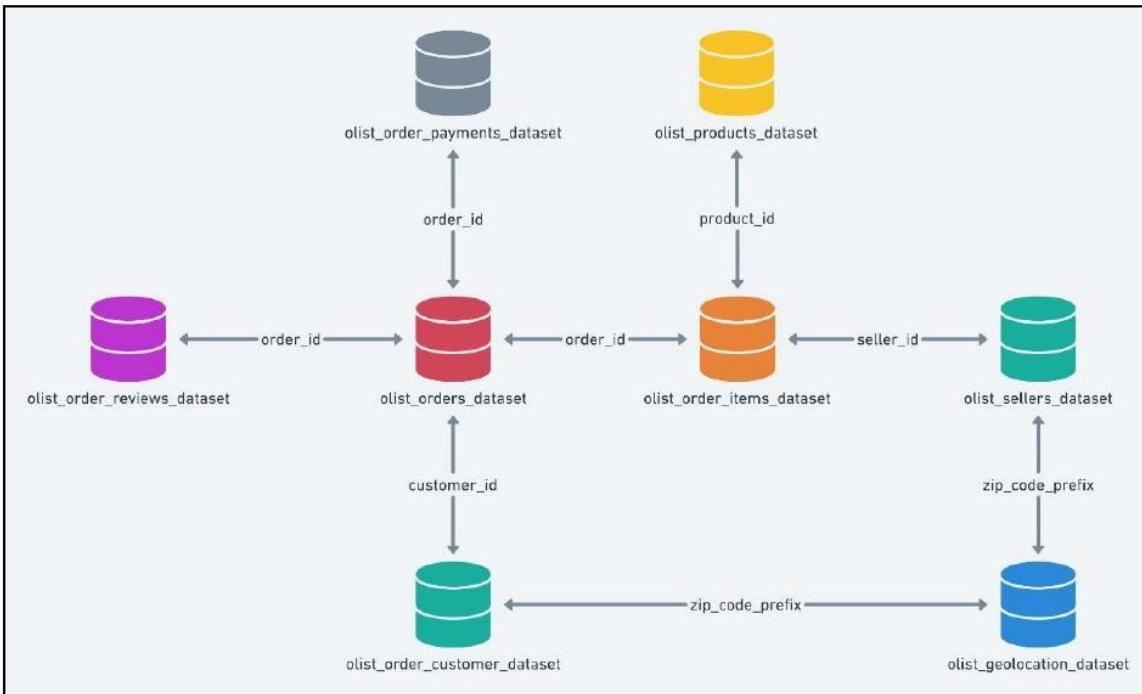
obligations. This setup allows merchants to sell their products through the Olist Store. Following a sale, Olist's logistics partners are responsible for the direct delivery of the items to the purchasers.

The data encompasses over 100,000 unique customer and order identifiers. It is comprehensively structured to include detailed information on orders, customers, products, sellers, payments, geolocations, and reviews, offering a holistic view of the entire commercial process.

Figure 2 Description of tables

Table	Description
olist_orders_dataset	It is linked to 4 table and connects all the information pertaining to an order.
olist_order_items_dataset	It includes information on a purchased item, including the price, shipping date, and other facts.
olist_order_reviews_dataset	It includes information about any reviews that the customer may have written about a certain item that he had purchased. It includes information about a product, including its ID, category name, and measurements.
olist_products_dataset	It includes information about a product, including its ID, category name, and measurements.
olist_order_payments_dataset	This table's data relates to the payment-related information connected to a certain order. Describes this company's customer base in detail.
olist_customers_dataset	Describes this company's customer base in detail.
olist_sellers_dataset	All of the sellers who have registered with this company are listed in this table along with their connected information.
olist_geolocation_dataset	Both the geographical location of the buyers and the sellers is included.

Figure 1 Relationship Diagram (Entity Relationship)



Storyline and Visualization:

The line manager has requested the use of business intelligence tools to thoroughly analyze the provided datasets and present the findings and recommendations in a comprehensive report format. This dataset, rich with information on 100,000 orders placed across various Brazilian marketplaces, facilitates a multifaceted perspective of each transaction. Such insights include customer location, product attributes, order status, pricing, payment processes, and freight performance. Moreover, the dataset incorporates a geolocation component that associates Brazilian zip codes with corresponding latitude and longitude coordinates.

Once an order is placed through the Olist Store, the relevant vendor is notified to complete the order. After the product has been delivered or the projected delivery date has passed, the customer is invited to fill out a satisfaction survey via email. This process allows them to rate their shopping experience and offer constructive feedback.

For this analysis, Microsoft Power BI has been chosen as the tool for data visualization. This platform stands out due to its interactive dashboards and compatibility with a diverse array of data sources.

Objective of Analysis:

The main objective of the analysis would be able to answer below questions:

- What is the company's total number of customers, orders, and orders per customer?
- Which states have the most customers?
- How many orders were there each month?
- What are the top five product categories and the lowest five?
- Display the consumer demographics, sales trends, order categories, and order changes over time for the company.

Analysis:

Data Model:

The dataset, sourced from Kaggle, has been imported into Power BI. As it currently contains only tables and corresponding keys for each file of the dataset, pre-processing is required to ensure precise analytics. We commence by constructing a data model, the structure of which can be viewed in the accompanying Figure 3.

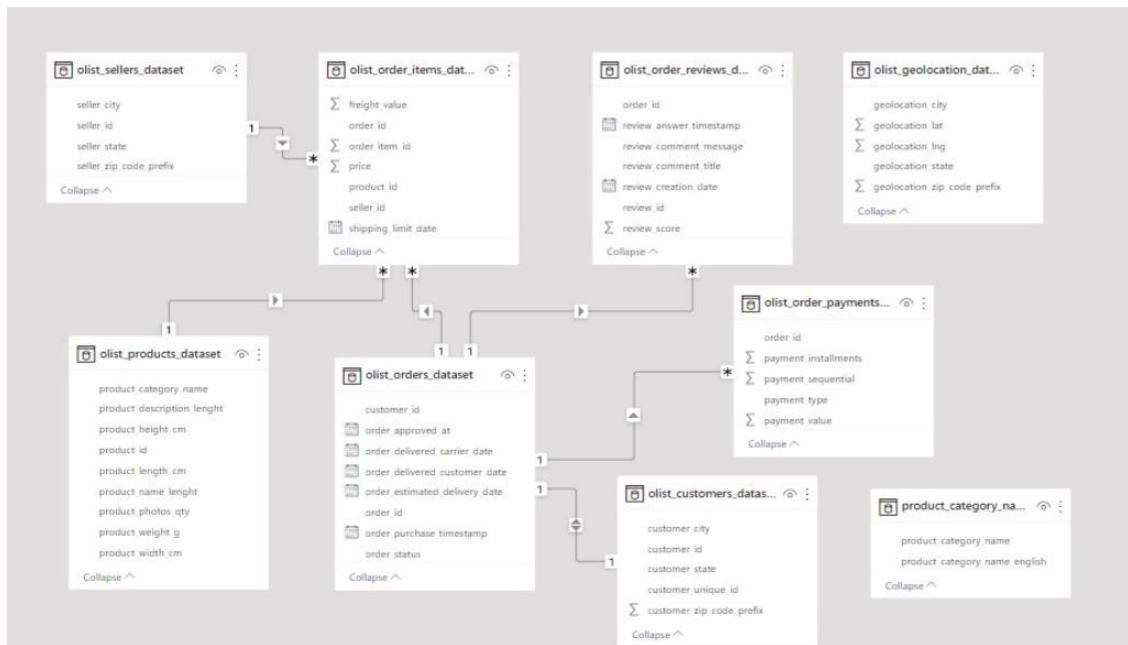


Figure 3 Initial schema

Consequently, depending on the interactions between the tables, connections or linkages between them are essential. The 'olist_geolocation' and 'product_category' tables cannot be incorporated as they don't have any apparent relation to the existing data model. Temporal information is also not explicitly stated as efficient time management is crucial when applying Power BI to the remaining tables. Furthermore, the variables are categorized based on the nature of the data they contain, whether it be textual or numerical. It's evident that the dataset encompasses information regarding transportation logistics, customers, vendors, and a range of products sold, each with a unique identifier. The final schema of the refined data model can be viewed in the following Figure 4.

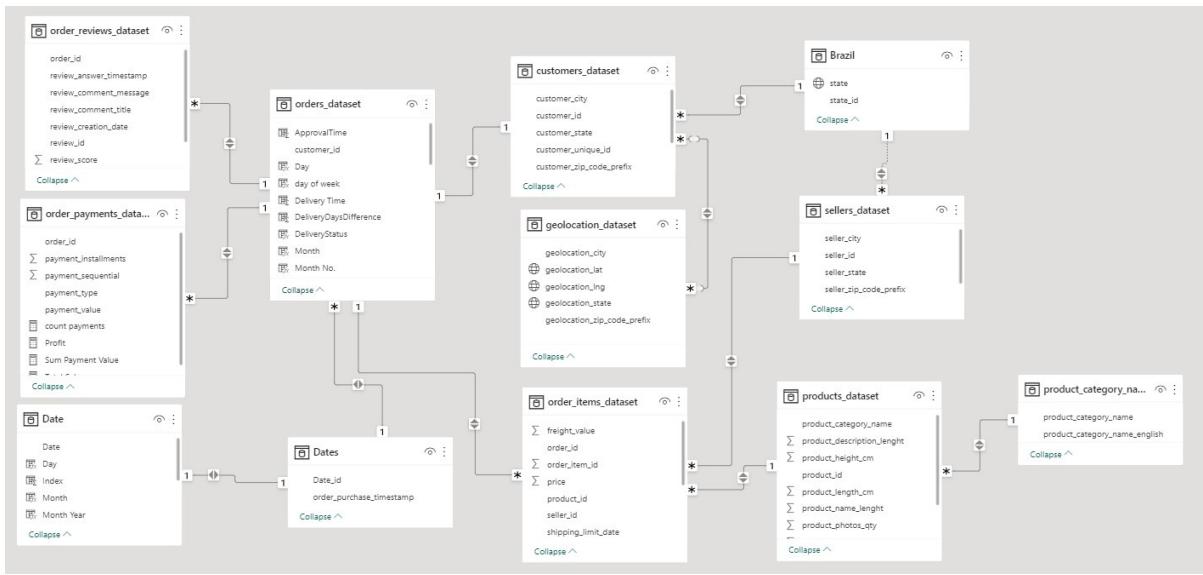
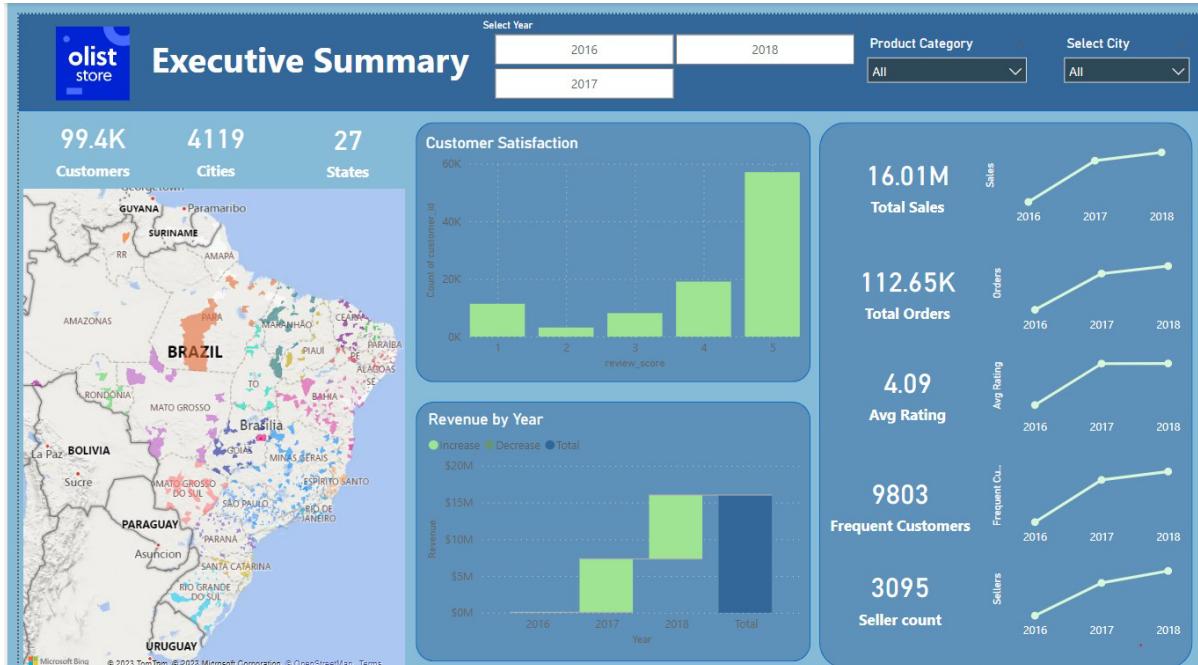


Figure 4 Final schema

Dashboard:

Executive Summary Dashboard



This dashboard is designed to provide a comprehensive view of the company's performance over the years by highlighting key metrics such as total revenues, total orders, reviews, regular customers, and the number of registered sellers. As these aspects form the core of any business operation, the main focus lies on revenue generation and customer satisfaction. From 2016 to 2018, the company has demonstrated significant growth, reflected in the data with an average of 16M in total sales and an order

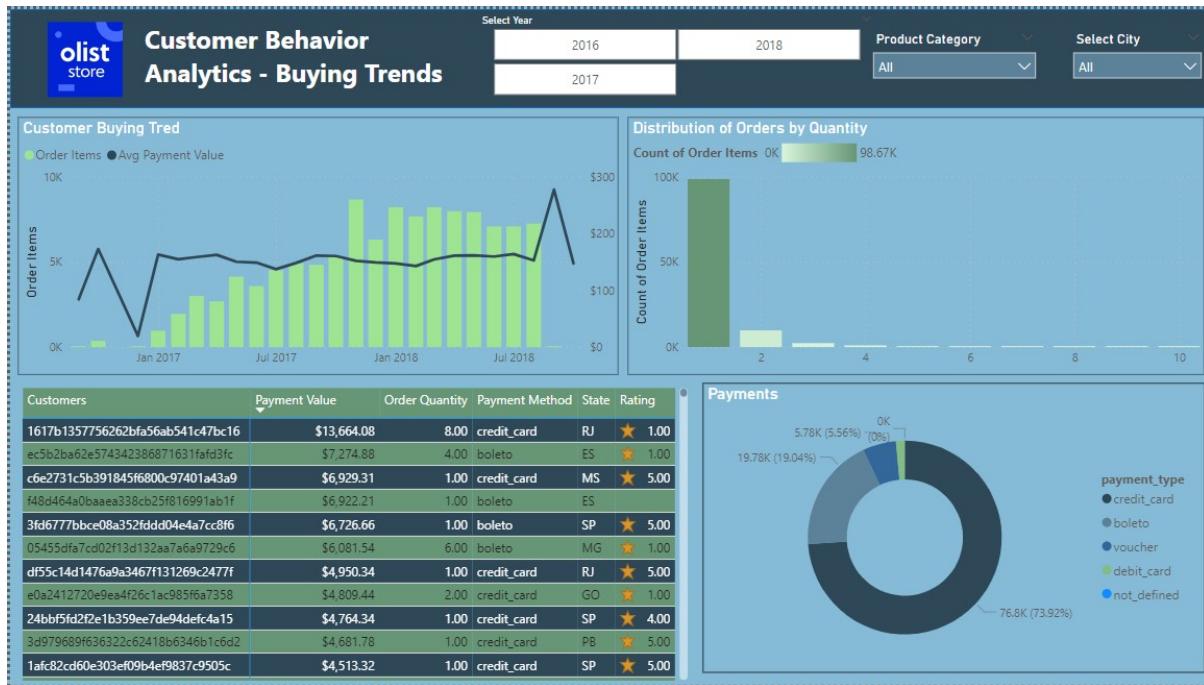
quantity of 113K. This sustained expansion signifies the company's successful strategies and effective operations.

Exploratory Analysis Dashboard



The geographic visualization indicates that Sao Paulo registers the highest number of orders, highlighting its key role in the market. An examination of order trends throughout the year reveals a consistent increase starting from January, with a marked surge in August, signifying the peak season. In contrast, September witnesses the least activity with the lowest number of orders. The total count of orders delivered 'In Advance' (88,649) considerably surpasses those classified as 'Late' (7,827). The data has been further dissected by product type to facilitate analysis of revenue, cost, profit, and order volume. According to the results, the 'bed_bath_table' category secures the largest number of orders, closely trailed by the 'health_beauty' category. These findings can shape business strategies and product emphasis to optimize sales.

Customer Behavior Analytics



This dashboard provides insights into customer purchasing behavior. From January to July 2018, the ordering of items seemed to maintain consistency, with an average payment value of \$160. Interestingly, the average payment value in September escalated to \$277.

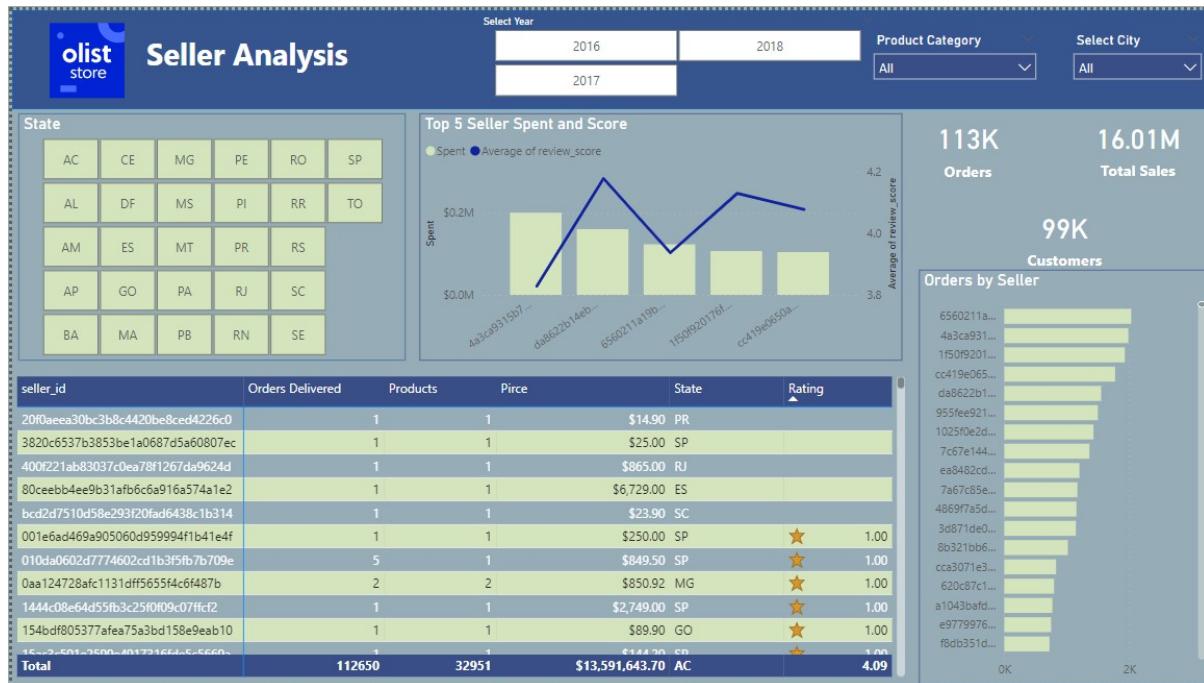
In terms of payment methods, credit cards lead the way, accounting for 74% of all payments. Boleto follows suit, making up approximately 20% of the payments. The three highest-spending customers each purchased over 21 items, with expenditure exceeding \$4,600.



From the observed data, it is evident that the majority of new customers register in the months of August and May. Throughout the year, the count of ordered items varies, ranging from 4,838 to 12,158, with August marking the peak season.

A noteworthy 9.86% of customers are classified as frequent buyers, with an average spending of \$154 on their purchases. The product category 'bed_bath_table' leads in terms of order volume. The top 5 categories combined account for approximately 46,000 orders, which represents 46% of the total orders, with the 'security' category registering the fewest orders.

Seller Analysis Dashboard



Spent and total Average of review_score are negatively correlated with each other. Across all 3,095 seller_id, Count of Order Items ranged from 1 to 2033.

4a3ca9315b744ce9f8e9374361493884 accounted for 28.83% of Spent.

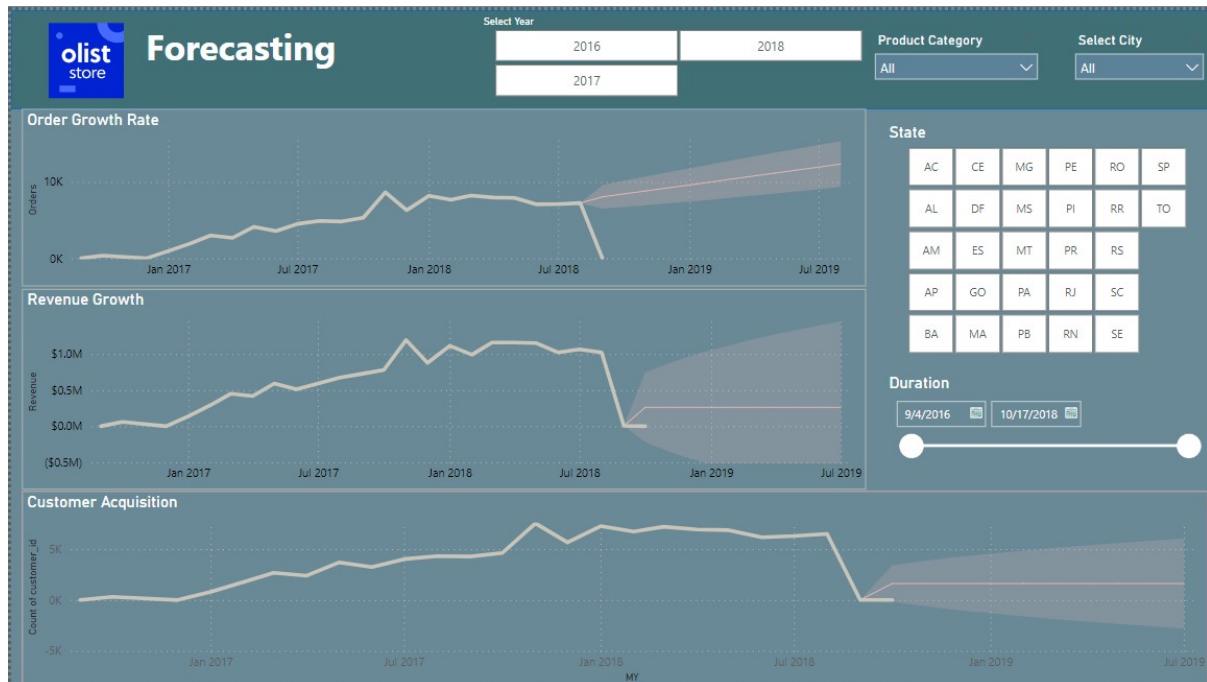
Delivery Analysis Dashboard



The delivery dashboard provides insights into key aspects such as delivery status, average delivery time, and freight prices. An impressive 97.02% of orders were successfully delivered, with only 625 orders reported as canceled. In terms of the count of delivered orders, Monday stands out with a total of 15,701 orders, which is 48.75% higher than Saturday, the day with the lowest count of delivered orders at 10,555.

An interesting trend is the negative correlation between Average Delivery Time and total Average Rating. These two factors diverged the most in February when the Average Delivery Time was 11.80 higher than the Average Rating. From a business perspective, the freight value appears to be acceptable, as it constitutes just 14% of the total cost.

Forecasting Dashboard



We can see an increasing trendline in number of orders and as per the data set, the number of orders generally increase from Jan to July hence projecting the same. Similarly for revenue, the projections are having stable revenue since revenue trended up, resulting in a 133.77% increase between September 2016 and October 2018.

Count of customer_id remained relatively constant between September 2016 and October 2018 but it started trending down on June 2018, falling by 99.94% (6,163) in 4 months.

Insights:

- Overall customers are satisfied with our services, however, there are around 11k customers who have rated us 1 star, this is concerning for business as that makes around 11% customer who are not happy with service.
- Revenues, total sales, total orders and customer acquisition have increased year on year basis. These numbers present a positive picture but considering that frequent customer numbers are only 9803 which is only 10% can be concerning. This low number could be a result of unsatisfied customers giving a rating of 1 star only.
- It is observed that the customer average payment value is almost same over the years, other than a few hikes, one of those hikes is observed in Sep 2018, for which the data is not complete

which means that it could be a single order that was worth more than average value of the orders.

- This can also be seen in the data that quantity of orders over the years has increased and is in range of 7000 – 8000 orders.
- Most of the customers have ordered only 1 item.
- Year on year new customers are acquired, in 2018 54k customers were acquired.
- No seasonality is observed in the customers buying patterns. However, if dig down on product category level we may see some seasonality for products.
- Also since new customers were being constantly acquired during year 2017 and 2018, the seasonal data on buying is also not very valid, given that new customers were constantly being added which increase orders over the year.
- One good thing observed is that the sellers with whom we are doing the most of business are all highly rated or the products delivered form the to customers were rated good.
- It is visible that most of our orders are generating from Saint Paulo, around 41% of the orders are coming from this city. This could primarily because of Saint Paulo being a big city or may be our reach or acceptance level is not much in other cities compared to Saint Paulo.
- We are paying huge sums of money on freight cost.
- Most of our deliveries are made late. Late deliveries are sometimes delivered 200 days later than the estimated date of delivery. This is affecting customer experience very badly, and probably because of these the rating for 11% customers is 1 star.
- On average there is a difference of around 12 days of delivery time versus the estimated delivery time.
- One key thing observed is that rating by customer is directly linked to delivery time, in Nov 2107 the average rating dropped to 3.91 which was the lowest for that year, and it was associated with highest average delivery time of 14 days.
- Good news is that approval time for the order delivery is quite good, just 0.69 days and 97% of the orders were delivered.
- Forecasting shows positive signs of increase in order, revenue and new customer acquisition growth.

Recommendations:

- Need to conduct a detailed survey with the customers who have rated us 1 star, to find out the core reasons for this and improve on those points, this will also help to increase the number of frequent customers.

- Cross selling opportunities should be created to increase the quantity of items ordered by customers. This will increase average bucket size of the customer and increased revenue for the company.
- 73.92% of our customers are paying with credit cards, this can be used to create promotional packages for the customers influencing their buying patterns and again increasing average bucket size for each customer.
- Data on customer acquisition is available with us, however, we don't have any data on the customer churning, we need to conduct a study to understand how many of our customers have churned.
- To analyse patterns and areas for improvement, customer reviews should be frequently tracked. This can entail utilizing data analysis tools to find recurring themes in customer feedback and then applying this knowledge to alter and enhance the customer experience.
- Since security and services receive the fewest orders, a plan of action should be developed to determine how to increase sales in this category while maintaining consistency in bed and bath, which receives the most orders.
- Cancelled orders should also be focused on and underlying reason should be identified for the reason of cancellation and should be done by customer service department by calling the customers and taking their feedback.
- Frequent customer data base should be maintained, and they should be offered a loyalty card to avail discounts and since they are already buying frequently hence it would keep them consistent and would give them an incentive as well.
- Freight costs should be revisited and realigned with stakeholders to improve the overall revenue numbers.
- Categories should be thoroughly analysed through market trends, particularly those affecting women, have the potential to increase sales in the health and beauty area which has the second highest orders.
- Delivery times need to be improved these are highly linked with customer satisfaction and this should also be ensured that estimated time for delivery is met and orders are not delivered late.