

**Sales Data Analysis**

**Final Project Report**

**Presented By**

**Generation Australia**

**DA - Group 5**

Aqsa Shuja

Iris Engler

Stephy Zhu

Sujata Banerjee

**Sujata BSujata Banerjeeanerjee**

**Sujata Banerjee**

**Sujata BSujata Banerjeeanerjee**

**Table of Content**

[Project Overview 1](#_brentieg9hgr)

[Project Steps 2](#_1b06wjn7gj5p)

[Data Model Design 3](#_qp3wjj7bpapb)

[Business Overview 5](#_fj7p7pgbohfp)

[Sales 7](#_zajxrdlkb65f)

[Sales Forecast 9](#_7weuafv5pu5s)

[Logistics 13](#_saq32uqpxgf8)

[Quality Review 15](#_2s25cwxuqzil)

[Analysis and Recommendations 18](#_1igfv8hu2rm1)

# Project Overview

Olist is a Brazilian company founded in 2015 and operates as an online e-commerce site that connects shops to wider marketplaces.

The company is a pioneer in the business model called “marketplace of marketplaces”. It does not hold any inventory or sell products of its own, products are sent directly from merchant stores to clients around the country.

This report will be implemented in Power BI Desktop to analyze the e-commerce performance for the Financial Year 2016 , 2017 and 2018 starting from 15/09/2016 to 03/09/2018.

Objective:

To generate business insights and investigate improvements to the company’s processes.

# Project Steps

**Data Preparation**

**Business Insights**

**Visualization**

**Exploratory Data Analysis**

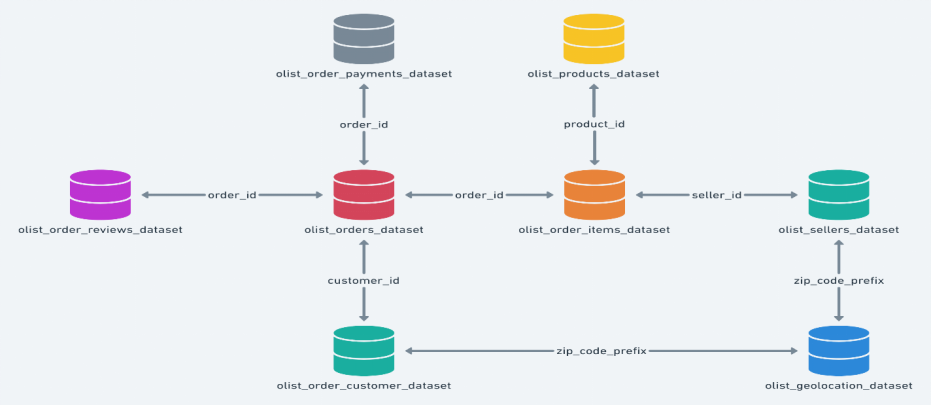
**Data Preparation**

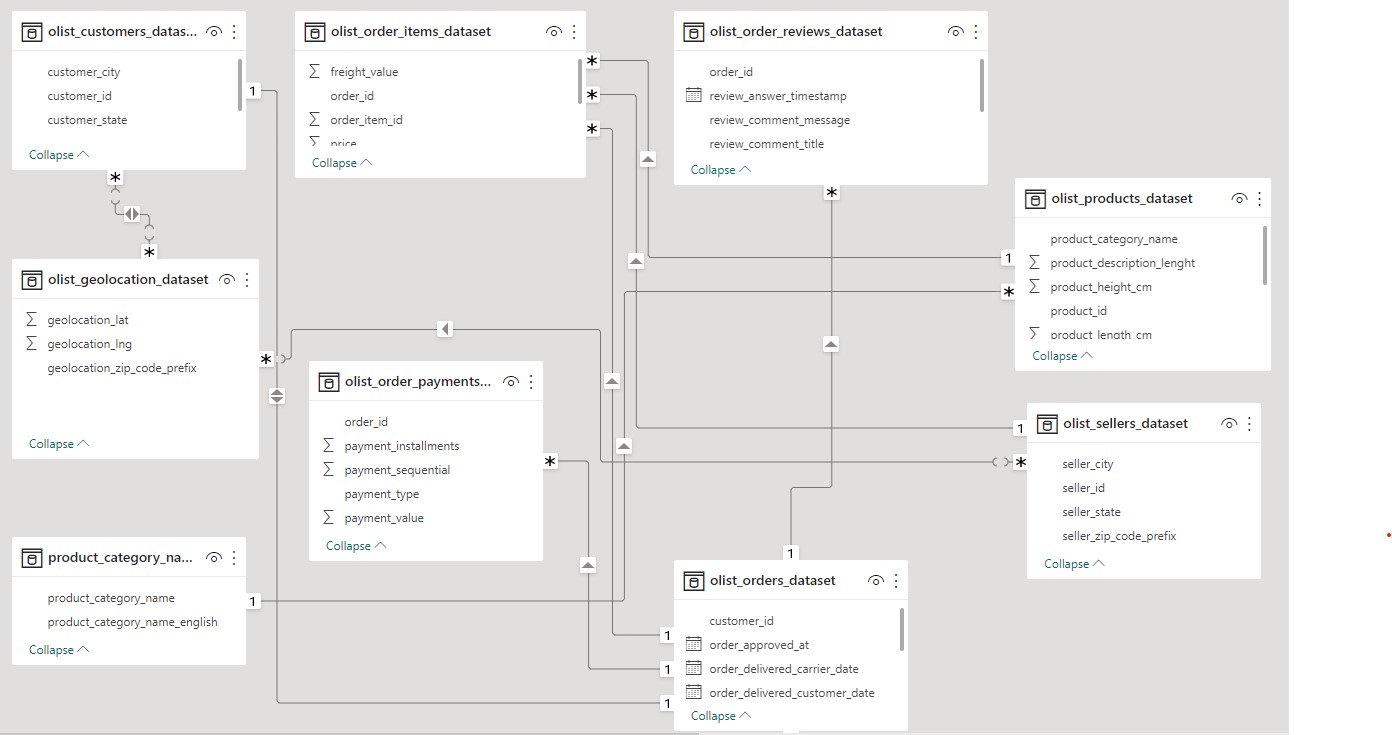
Pre-requisites:

* MS Power Bi Desktop - Version: 2.114.864.0 64-bit (February 2023)
* MS-Excel - Microsoft 365

# 

# Data Model Design





# Data Information

**Data Source:**



[Brazilian E-Commerce Public Dataset by Olist | Kaggle](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)

**Data Range:**

**15/09/2016 to 03/09/2018**

**Main Tables:**

* Orders
* Customers

**Reference Tables:**

* Order Items
* Products
* Sellers
* Payments
* Reviews

**Supporting Tables:**

* Geolocations
* Product Category Translations

# Business Overview

# 

Overview page contains statistics and visualizations about customer demographics, orders placed by customers and sales by year. It highlights performance in terms of number of orders placed and sales in 2016, 2017 and 2018.

For a business to thrive, it needs to know who their customers are. It affects important business decisions regarding pricing, packaging, and service offers. From the given data, we have generated the filled map visualization to show the states where customers are located. The Bar chart categorizes the customers count in terms of the number of orders placed. We found out that the top performing state in terms of number of orders placed is Sao Paulo. In second and third place we have Rio de Janeiro and Minas Gerais respectively.

The number of orders placed is also the number of products sold to customers. Increase in the number of orders over a period of time shows how successful the business is. Number of orders by year bar chart shows an increase in the number of orders from 60K in 2017 to 75K in 2018. Monthly order changes chart shows how the number of orders placed vary from one month to the other.

Sales by year waterfall chart shows an increase in sales. 0.1 million in 2016, 7.2 million in 2017 and 8.7 million in 2018 with a combined total of 16 million in sales over the period of 3 years.

# 

# Sales



The sales page displays the following metrics: sales by quarter, monthly sales versus freight, top and bottom product categories by sales, top selling states and total payment by payment type.

Sales by Quarter column chart: year and quarter are on the x axis and sales in millions of Brazilian Reals on the y axis. We can see that orders have increased steadily over time. There was an overall positive trend throughout 2017 which continued in 2018. There is very little data for 2016 and data exists only until early September 2018.

# 

Monthly sales vs freight line chart: months are on the x axis, sales on the y axis and freight value on the secondary y axis. There is a high correlation between sales total and delivery charges. The Increase in number of orders results in a correlated increase with both the delivery charges and sales.

Top categories by sales and bottom categories by sales: the top 4 categories are Bed Bath & Table, Watches & Gifts, Health & Beauty and Sports & Leisure. Some categories like computer accessories are seasonal, selling significantly more in the first couple of months of the year, perhaps customers purchasing back to school and work products. Bottom categories are more variable, although clothing categories do not seem to sell well.

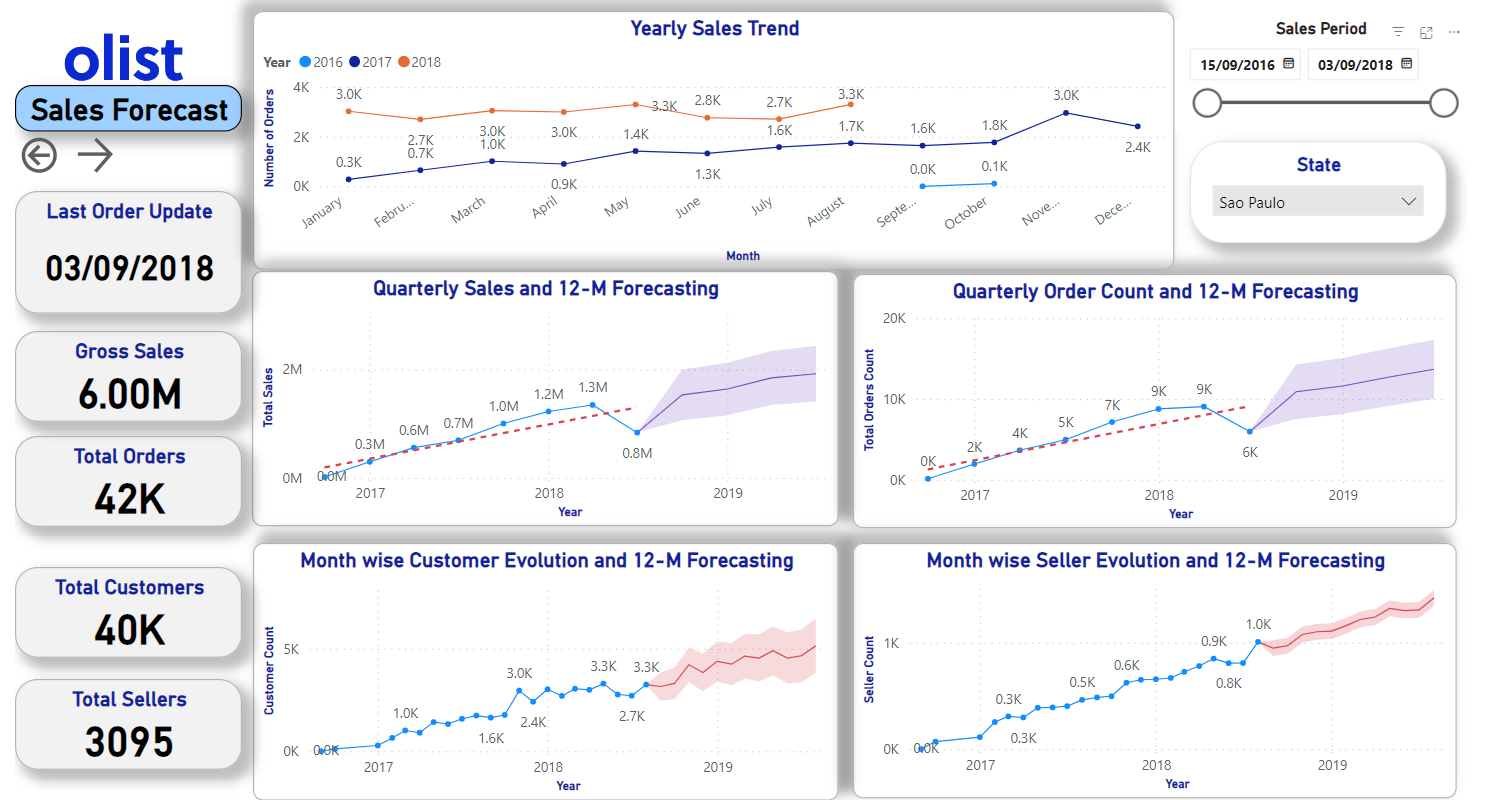
Top selling states map: Olist customers are heavily skewed toward South Brazil which is not surprising as it’s the richest and most populated area in the country. Top selling states are São Paulo, Rio de Janeiro, and Espírito Santo. São Paulo (the richest city in South America) is the capital of Sao Paulo, the state with the highest sales (over 40%).

Payments by payment type donut chart: credit card and Boleto payments account for over 95% of sales. Most customers use credit cards, but Boleto is also common. It’s a Brazilian method of payment that requires the user to go to the nearest bank to make a money transfer and is popular as it’s regarded as more secure.

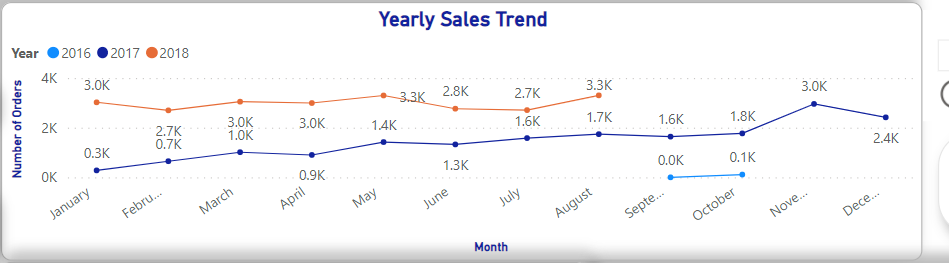
# 

# Sales Forecast

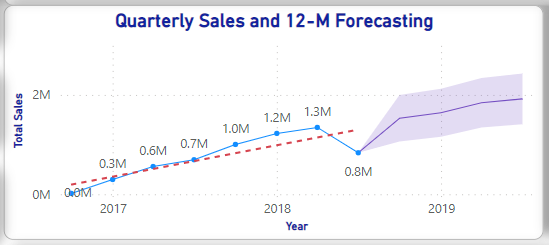
* Forecasting is a technique for making predictions of the direction of the future trends based on analysis of past and present data.Businesses use forecasting to anticipate how to allocate their budgets or plans for an upcoming period of time.It is a decision making tool that helps businesses cope up with the impact of the future's uncertainty by examining historical data and trends.A planning tool that enables businesses to chart their next moves.
* The analysis is based on the 2 years of available historical data from Olist from 2016 to 2018.

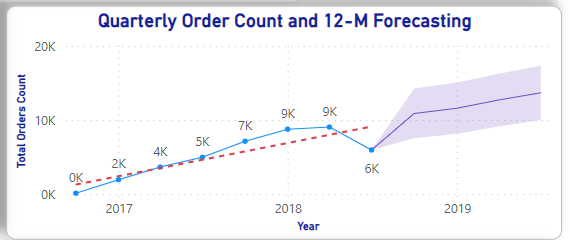


**Yearly Sales Trend**



* A line chart is used to represent the Yearly Sales Trend in terms of the Yearwise total order count.
* Sales trends analysis is done using historical data for a Sales period between 15/09/2016 to 03/09/2018 from Kaggle data source and comparing it with previous timeframes to identify patterns and project future trends.
* 2017-2018 had an overall **positive monthly trend** for the total number of orders.
* 2017 shows a steady positive trend with the highest number of orders placed (7395) in Q4. This was due to the Holiday period.
* 2018 shows a flat trend with a decrease in sales in the Q3 as only 8 months of data is available.



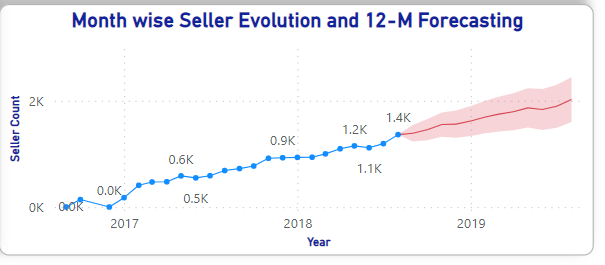


# **Sales and Orders 12-M Forecast:**

* A Lines chart is used to show the Quarterly sales based on historical data. The markers represent the consistent increase in sales in Brazilian Rial(R$).
* The Forecast is done using Seasonality point 4 as its is showing next 4 quarters Forecast using a 95% confidence interval.
* 2017-2018 showed a **positive quarterly trend** in Sales growth and order counts.
* Decrease in sales and order in 2018 Q3 is due to only 8 months available historical data.
* **27% increase in sales predicted in the 12 Month forecast from 2018 Q2 from 3345864.65R$ to 4642562.75R$.**
* **23% increase in order counts predicted in the 12 Month forecast from 2018 Q2 from 20008 to 26120.**

**Customer Acquisitions and Sellers subscription**

# 

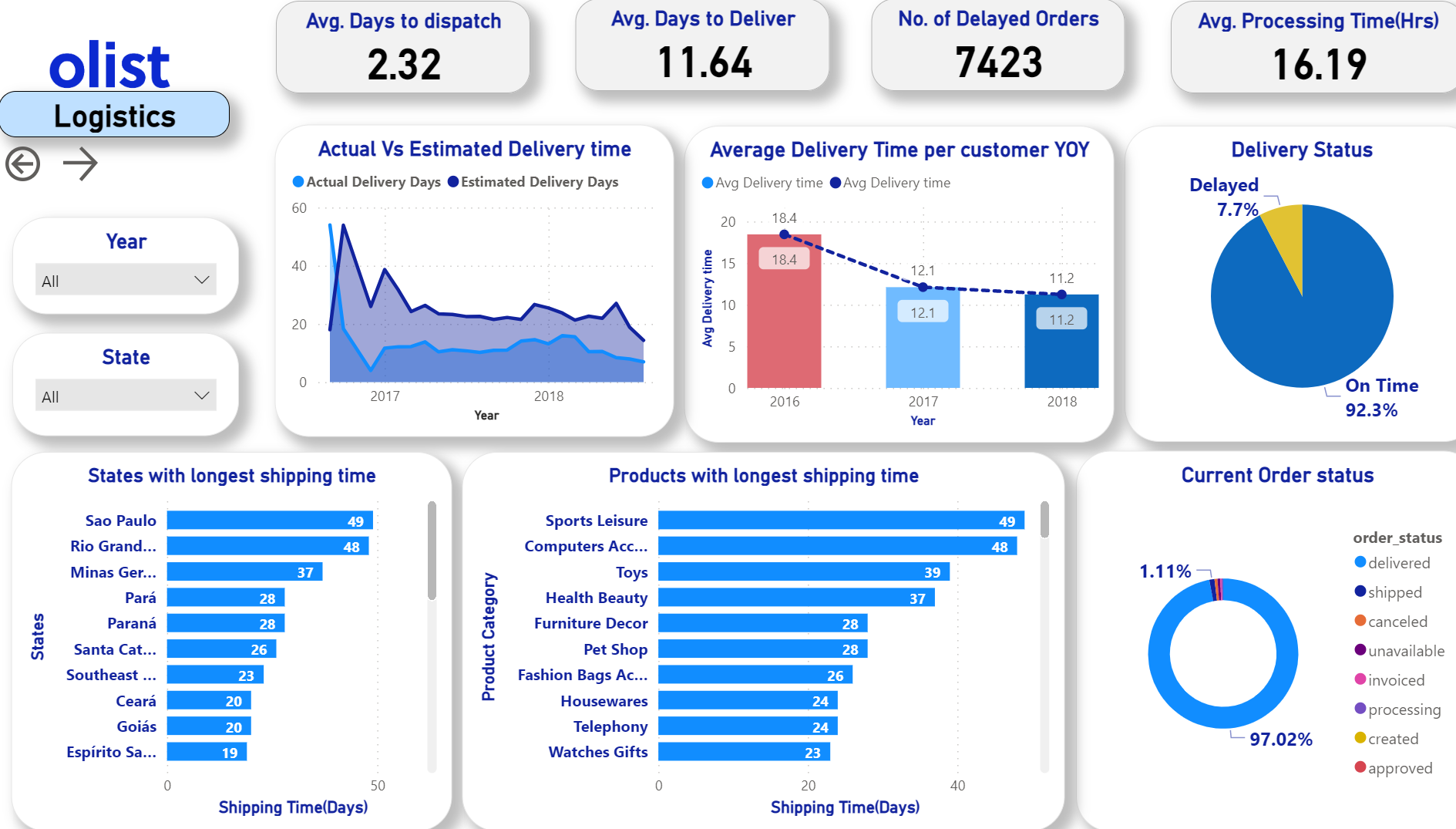


* A Lines chart is used to show the Quarterly sales based on historical data. The markers represent the consistent increase in sales in Brazilian Rial(R$).
* The Forecast is done using Seasonality point 12 as it is showing next 12-months forecast using a 95% confidence interval.
* 2017-2018 had a Positive monthly trend in customer acquisition and Seller subscription.
* The 12-M forecast from 2018 Q3 till 2019 Q3 predicted an increase in Customer growth of upto **10833** i.e **39.5% more** than the existing count (**6549**).
* The 12-M forecast predicted an increase in Seller subscription upto **2022** i.e **32.5% more** than the existing 2018 Q3 count(**1363**).

# 

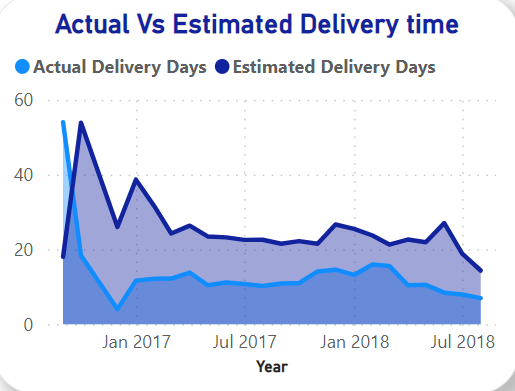
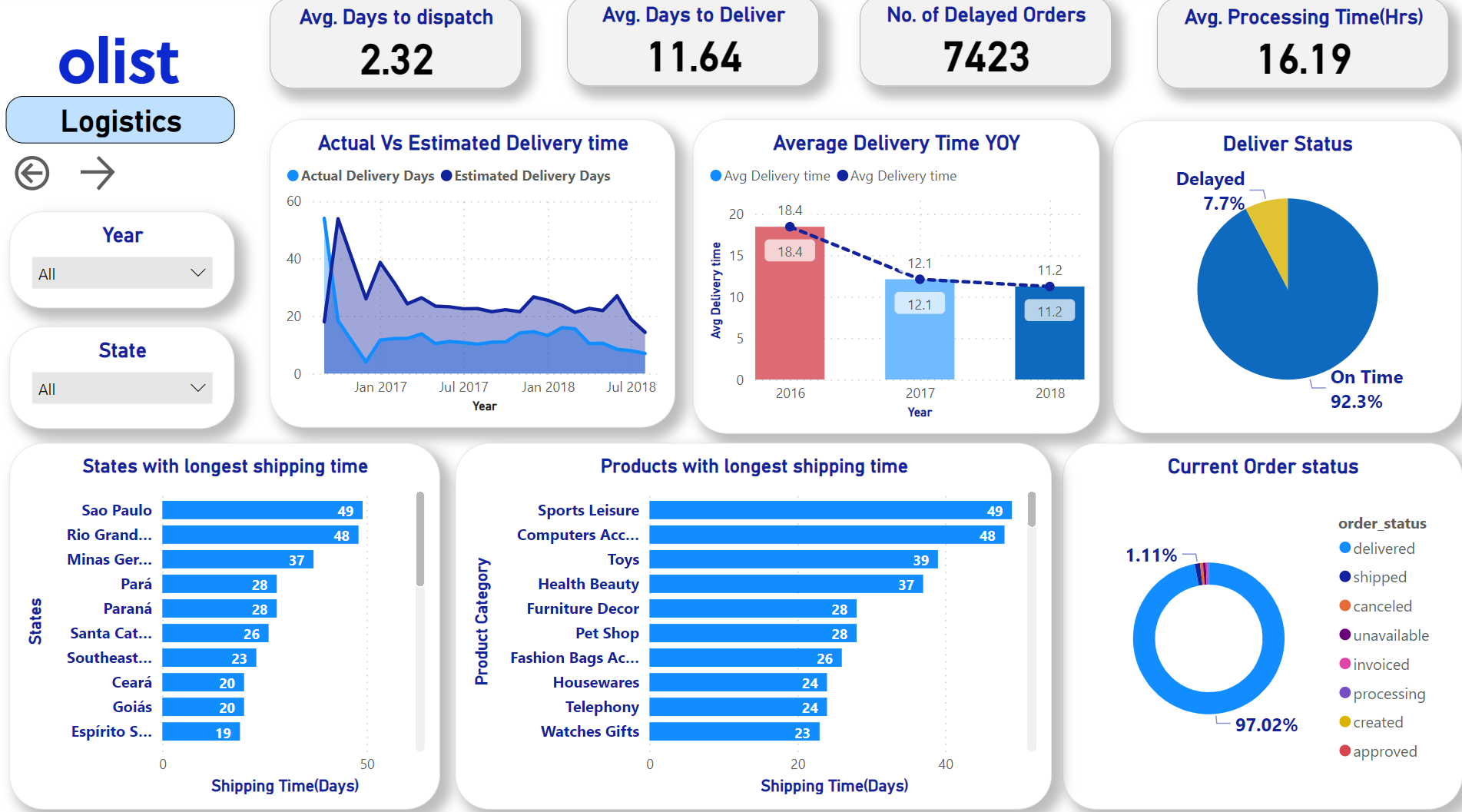
# Logistics

Efficient logistics plays a crucial role in the e-commerce industry. The logistics dashboard provides key insights into the performance of Olist's logistics operations.



Out of the almost 100 thousand orders analyzed, the on-time delivery rate stands at 92.3%, suggesting that around seven thousand orders were delivered after the estimated timeframe.

The area chart illustrates that, except for the initial period of the financial year 2016, the actual average delivery time consistently remains shorter than the estimated delivery time. Additionally, there is a noticeable reduction in delivery time each year, indicating a positive trend of improved efficiency.

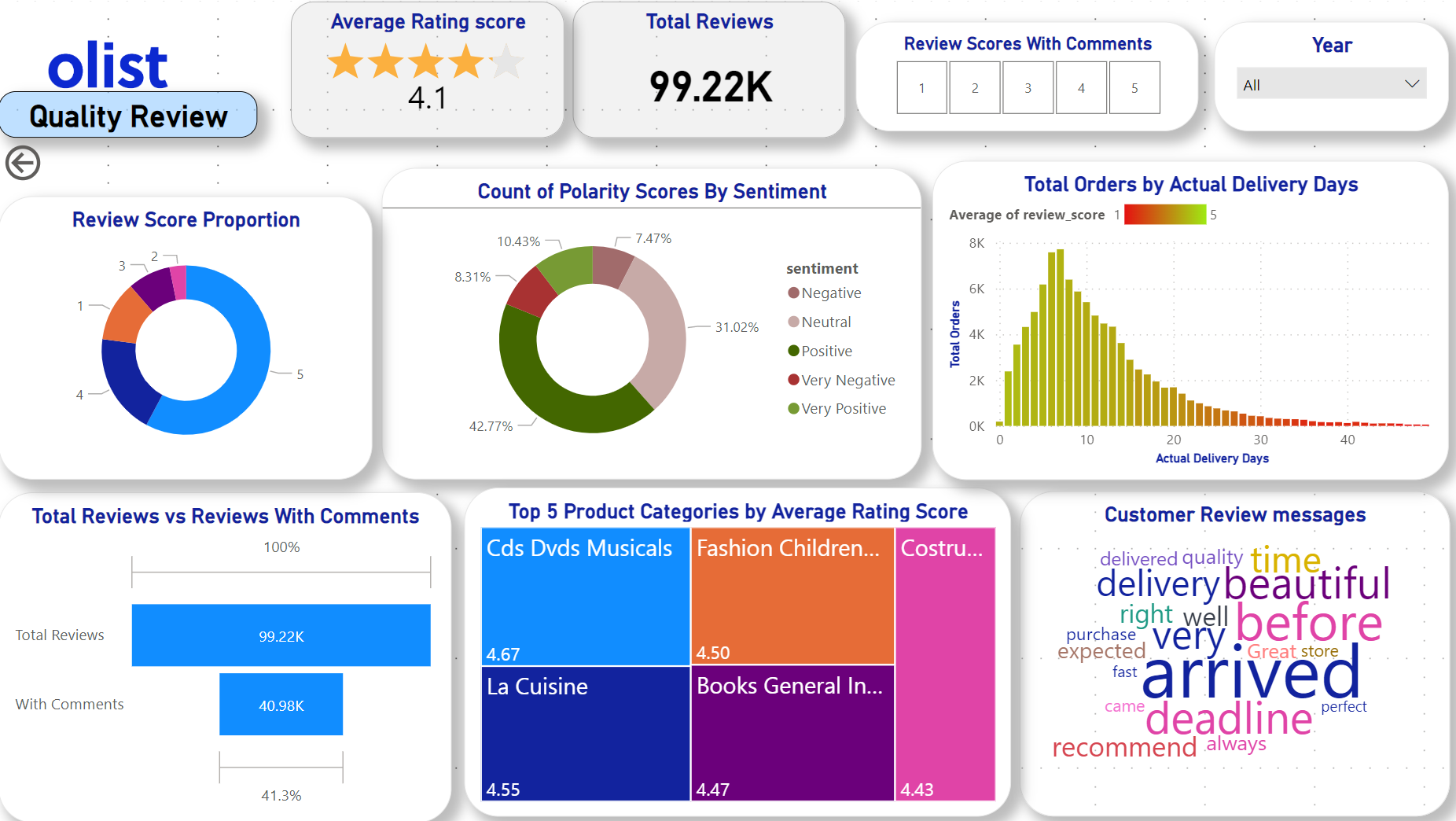
 

The dashboard also provides insights into the states and products experiencing the longest shipping times. This visibility allows the organization to identify and address any weaknesses in their logistics chain, leading to potential improvements in overall efficiency and performance.

# 

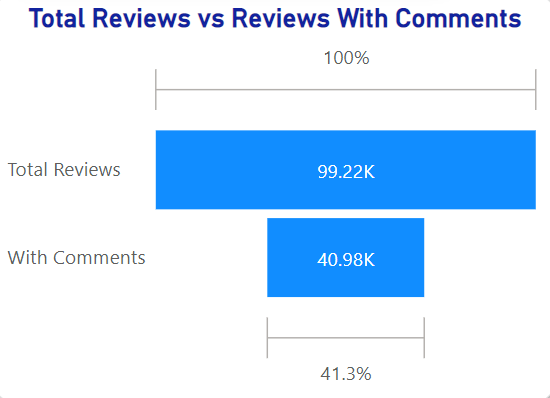
# Quality Review



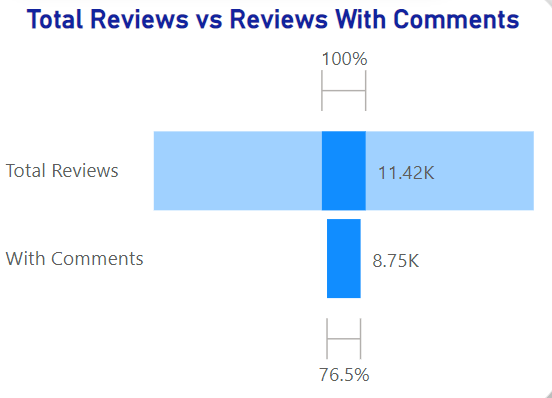
Out of the 99,220 reviews in the dataset, the average customer rating score is 4.09. Although most of the reviews are 4 or 5 stars, it is worth noting that 1-star reviews rank third in terms of quantity. This suggests the possibility of product quality issues in certain product categories.

The top 5 product categories by average ratings are: audiovisual products, children clothes, la cuisine, books of general interest, and construction tools.

It is also noteworthy that only 41.3% of the reviews include comments regarding the shopping experience and product. However, for reviews with a 1-star rating, this proportion significantly increases, with 76.5% of negative reviews having a comment.

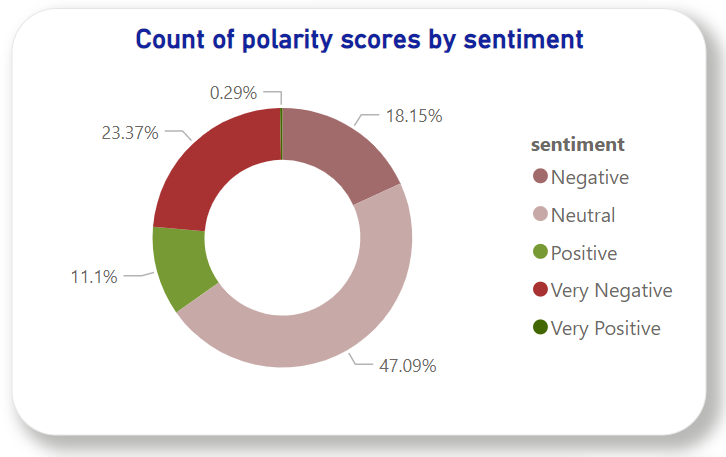
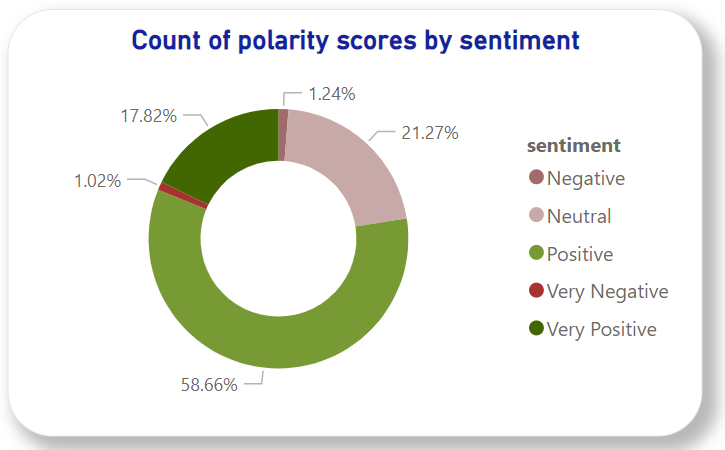


Total Reviews vs Review with Comments



Total 1-star Reviews vs 1-star Review with Comments

By categorizing the sentiment expressed in the comments as positive, neutral, or negative, we can gain a deeper understanding of customer sentiments beyond the numerical rating. This information can help identify areas of strength or improvement, address customer concerns, and enhance overall customer satisfaction.



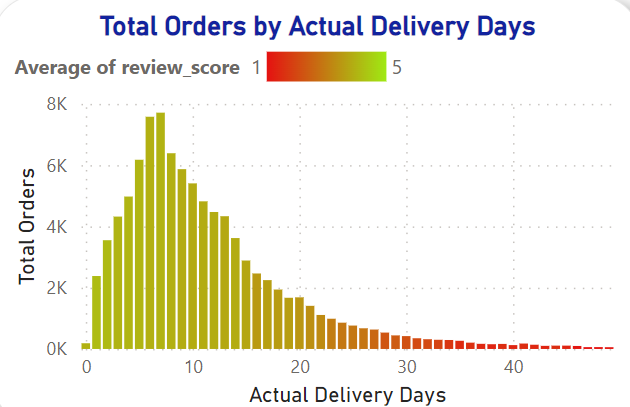
Sentiment proportion for 5-star reviews Sentiment proportion for 1-star reviews

Upon examining the word cloud, which displays keywords based on their frequency in the review comments, a clear pattern emerges. Regardless of whether the review is positive or negative, the word "delivery" consistently appears as the most frequent. This observation strongly suggests that customer satisfaction is closely tied to the performance of the delivery service.



Word Cloud for Review Rate 5 Word Cloud for Review Rate 1

The bar chart indicates that delivery performance has a strong impact on review scores. The color representation in the chart shows that lower review scores are associated with a red color, while higher review scores are associated with a greener color. It is apparent that longer delivery periods tend to result in lower average review scores.



# Analysis and Recommendations

* 60% of yearly sales are generated from only the top 10 products.
* Average delivery time is a period of 11 days.
* Long shipping time from major states like Sao Paulo.
* Low review score for product delivery due to long shipping time or lost deliveries.

Based on the analysis of the dashboard, we have several recommendations for the organization to improve its sales and logistic performance.

1. Analyze the sales patterns and demand forecasting within the popular categories. Ensure that the sellers have sufficient inventory levels to meet customer demand, while also avoiding excessive inventory holding costs. Implement effective inventory management practices to optimize stock levels and reduce stockouts.
2. To address the issue of slow delivery times, a strategic approach would be to consider establishing distribution centers in regions with longer delivery times. By doing so, products that typically require extended delivery durations can be stored in these locations. This approach aims to improve overall efficiency in the transportation process and reduce the time it takes for customers to receive their orders.
3. Providing a proper shipment tracking system aids in having clear and concise communication between customers, sellers and couriers. Regular updates can set proper expectations among customers and specific delivery instructions of customers can be properly accommodated. By establishing trust and communication, both parties can work together to resolve any issues that may arise.

Challenges faced:

* Merging of individual .pbix files into one file.
* Publishing Dashboard on Power BI Service due to domain issue.
* Generating map visuals.

Documents submitted with Final Project Report:

* Final Project PPT-DA-Group5

[Final Project PPT-DA-Group5](https://docs.google.com/presentation/d/1gWTuUIFdX9YOfezlCjP011k5-AHwqZm6xs7_o1dur-k/edit#slide=id.gc6f9e470d_0_43)

* Final Project Proposal -DA-Group 5

[Final Project Proposal -DA-Group 5](https://docs.google.com/document/d/1d5_iVnG2RHxgJEn7BX4ZysEy8YXqSlGf2FX0-6JQHQI/edit)

* Final Project Pbix file

Final-Project-Olist-Data-Analysis.pbix

* Additional files

(Jupyter Notebook):

Review\_cleaning\_sentimental\_test\_Code.ipynb

(Updated Dataset files)

Olist\_order\_reviews\_with\_comment\_translated (English Translated)

Olist\_geolocation\_dataset (Updated States Name for the States code)