

An Analysis of the Olist Brazilian E-commerce Database

Final Technical Report

Aminat Ademola

Kristian Ambruch

Marcos Ferreira

Adaobi Ikeasomba

Bayo Olabisi

Sonia Nessary

Generation UK

13/11/2023

Contents

Executive Summary	3
Project Purpose and Scope	3
Key Findings	3
Recommendations	3
Introduction	4
Background Information	4
Analysis Justification	4
Dataset Description	5
Data Overview	5
Data Sources and Structure	5
Data Cleaning and Preparation	5
Methodology	6
Analytical Approach	6
Analytical Techniques/Methods	6
Software and Tools	6
Data Processing Techniques	6
Visualisations List	7
Dashboard.....	7
Sales Overview.....	7
Sales Trends	7
Delivery	7
Reviews	7
Forecasts.....	7
PowerBI Visualisation Development	8
Sales Distribution	8
Total Sales by State Bar Chart	8
Sales by City Map	8
Sales by Region and State Map.....	8
Sales Trends	8
Line Chart of Daily Sales Over Time	8
Treemap of Sales by Category	8
Delivery Analysis	8
Column Chart of Average Delivery Time by State.....	9

Scatter Plot of Review Score vs. Delivery Time.....	9
Reviews and Customer Feedback	10
Average Review Score Card	10
Star Rating Donut Chart	11
Review Sentiment Word Cloud.....	11
Sales Prediction Modelling	11
Line Chart with Future Sales Forecast	11
Sales Decomposition Tree.....	12
Analysis, Insights and Recommendations.....	12
Sales	12
Deliveries	12
Reviews	14
Forecasts	15
Pest Analysis	16
Political Factors.....	16
Economic Factors	16
Sociological Factors.....	17
Technological Factors	17
Conclusions	18

Executive Summary

Project Purpose and Scope

This report presents a comprehensive analysis aimed at optimizing the sales for stores on the Olist Brazilian e-commerce platform. The primary goal is to enhance operational efficiency by scrutinizing sales trends, delivery performance, and customer sentiment. This project encompasses an extensive data scope, covering all regions and states across Brazil from September 2016 to September 2018. It delves into the entire range of product categories available on the platform, offering a holistic view of Olist's operations.

The expected outcomes of this analysis include detailed insights into the current state of sales trends, the efficacy of delivery systems, and the overall customer sentiment. These insights are geared towards formulating concrete recommendations to boost the efficiency and effectiveness of each aspect analysed. The target audience for this report includes stakeholders of Olist, as well as project tutors and coordinators, given the project's educational nature.

The importance of this project lies in its potential to significantly improve the operational efficiency of the Olist e-commerce platform. By leveraging data analytics, this report aims to provide strategic guidance to Olist, enabling informed decision-making that can propel the platform towards greater success in the competitive e-commerce landscape.

Key Findings

The analysis of Olist Store's operations from September 2016 to September 2018 has yielded several key findings, offering valuable insights into sales trends, delivery performance, and customer sentiment.

1. **Sales Trends:** There was a notable increase in sales starting January 2017, with daily sales fluctuating significantly. The average daily sales were around 24,996, with peak sales recorded on specific dates. Popular categories included health and beauty, watches and gifts, and computer accessories. The top-performing states were São Paulo, Rio de Janeiro, and Minas Gerais.
2. **Delivery Performance:** The average delivery time was observed to be between 1-2 days. Late deliveries varied significantly by state, ranging from 4% to 15%. A strong correlation was noted between delivery times and customer review scores, up to 50 hours, after which the correlation dissipates.
3. **Customer Sentiment:** The average customer review score was 4.09, with a majority (70.7%) of reviews being 5-star. Analysis of customer feedback revealed frequent mentions of terms like 'recommend', 'good product', 'original', 'good buy', and 'late delivery'. This indicates a generally positive sentiment but also highlights areas for improvement, particularly in delivery timeliness.
4. **Regional Sales Focus:** A significant portion of sales was concentrated in the Southeast region, attributed to the presence of large, developed cities.

Recommendations

Based on the key findings from the analysis, the following recommendations are proposed to optimize Olist Store's operations and sales performance:

1. **Targeted Marketing and Inventory Management:** Given the fluctuating sales trends and popularity of certain product categories, it is recommended to implement targeted marketing strategies, particularly in high-demand periods and for popular categories like health and beauty, watches, and computer accessories. Additionally, strategic inventory management should be aligned with these trends to ensure product availability during peak sales times.
2. **Enhanced Delivery Strategies:** To address the varying late delivery rates across states and the impact of delivery times on customer satisfaction, it is advised to optimize delivery processes. This could involve partnerships with local delivery services in states with higher late delivery rates and investing in logistics infrastructure to improve delivery efficiency.
3. **Customer Feedback Analysis for Service Improvement:** The analysis of customer reviews indicates areas for improvement, especially concerning late deliveries. Focusing on these areas can enhance customer satisfaction. Regular analysis of customer feedback should be integrated into the business strategy to continually adapt and improve services.
4. **Regional Focus on High-Performing Areas:** With a significant portion of sales concentrated in the Southeast region, it is recommended to invest more in these high-performing areas through localized marketing efforts and perhaps opening regional distribution centres to further capitalize on the existing market strength.
5. **Continuous Monitoring and Adaptation:** Regular analysis of sales, delivery performance, and customer sentiment is vital for staying ahead of market trends and customer expectations. This continuous monitoring will enable Olist Store to adapt swiftly to changing market dynamics.

The purpose of these recommendations is to directly address the insights garnered from the data analysis to enhance overall operational efficiency, customer satisfaction, and sales performance at Olist Store.

Introduction

Background Information

Olist, a pioneering Brazilian e-commerce integrator, revolutionizes online sales by connecting small businesses to major online marketplaces. Founded from the Solidarium project, which began as a physical space to increase visibility for artisans in Curitiba, Olist evolved into a unique 'store in store' model in e-commerce, empowering digital retail and providing growth opportunities for businesses of all sizes.

By breaking down barriers and transforming traditional retail into an online powerhouse, Olist has established itself as a technological innovator with a focus on practicality, competitiveness, and customer-centric operations.

Analysis Justification

This analysis is crucial for Olist in its pursuit to enhance sales efficiency, especially in underperforming states, and to generate precise forecasts for future sales across various categories and regions. By identifying key factors influencing sales, delivery performance, and customer sentiment, Olist can adapt its strategies to maximize revenue, improve customer satisfaction, and extend its delivery capabilities. The project's strategic goal is to optimize Olist's operations, ensuring preparedness for varying regional demands and seasonal fluctuations, ultimately contributing to the company's revenue maximization.

Dataset Description

The datasets utilized in this analysis encompass a comprehensive range of Olist's operational data. These include customer datasets detailing buyer information, geolocation data pinpointing delivery locations, order items and payment datasets outlining transactional details, order reviews capturing customer feedback, order records tracking sales and delivery timelines, product information, and seller datasets.

Together, these datasets provide a 360-degree view of Olist's business operations, from September 2016 to September 2018, offering valuable insights into every facet of the e-commerce experience on the platform.

Data Overview

Data Sources and Structure

The dataset for this analysis was sourced from Kaggle, a renowned platform for data science and machine learning. It comprises 100,000 orders from 2016 to 2018 across various marketplaces in Brazil, offering multi-dimensional insights into order status, price, payment, freight performance, customer location, product attributes, and customer reviews. Additionally, the geolocation dataset links Brazilian zip codes to latitude and longitude coordinates, enriching the spatial analysis of customer and order distribution.

Data Cleaning and Preparation

The data initially presented several quality issues, including duplicates, missing values, inconsistent formatting, and minor inaccuracies. To ensure the integrity and reliability of the data for robust analysis, we undertook an extensive cleaning process:

Dataset	Required Cleaning	Name
Customer Dataset	Refine customer zip codes, remove commas, check for blanks.	Marcos
Geolocation	Standardize city column, ensure correct zip code formatting.	Aminat
Order Items	Restructure price columns, separate dates, and times.	Sonia
Order Payment	Review and address empty columns and rows.	Bayo
Order Reviews	Revise column titles, split timestamp data.	Kristian
Orders	Split timestamp and date, label columns, review empties.	Ada
Sellers	Align seller city info with states, correct discrepancies.	Kristian
Product Category	Ensure translation consistency, review empty fields.	Marcos
Product	Correct misspellings, address missing data.	Marcos

A further review of the orders dataset revealed a few entries from 2020. These were considered outliers, as they did not align with the primary dataset range (2016 to 2018) and were subsequently removed to maintain consistency in the analysis.

Methodology

Analytical Approach

Analytical Techniques/Methods

Descriptive Analytics: We used this to summarize the current conditions of Olist's operations. This involved analysing sales data, customer interactions, and order details to identify key patterns and metrics, such as average sales values, customer demographics, and order volumes.

Exploratory Data Analysis (EDA): This approach was crucial in uncovering underlying trends and behaviours, particularly in sales and customer interactions. EDA enabled us to delve deeper into the data, revealing insights about customer preferences, peak buying times, and product popularity.

Predictive Analytics: For the forecasts page, we utilized predictive analytics to project future sales. This involved creating a forecast that predict possible sales ranges, filtered by variables such as payment type, product category, and state. This predictive approach was instrumental in providing strategic insights for future planning and decision-making.

Each of these techniques played an important role in transforming raw data into actionable insights, enabling a more strategic approach to enhancing Olist Store's business operations and customer engagement strategies.

Software and Tools

1. **Data Cleaning:** Microsoft Excel was used due to its accessibility and intuitive interface.
2. **Initial Data Exploration:** Python, specifically Google Colab notebooks, was utilized for exploring datasets and performing preliminary analysis.
3. **PowerBI:** Used for creating measures and columns essential for developing visualizations.

Data Processing Techniques

In the Data Processing and Analysis phase of the project, a comprehensive array of initial visualizations was created in PowerBI, each designed to explore various facets of the Olist Store's operations. This extensive array included dynamic dashboards displaying key performance indicators, geographical maps pinpointing sales distribution, and trend analyses through line and treemap visualizations.

The selection process focused on retaining visualizations that offered the most significant insights. For instance, the line chart depicting daily sales over time and the treemap for sales by category provided a clear understanding of sales trends and product performance. Similarly, the scatter plot correlating delivery times with review scores and the word cloud generated from customer reviews were crucial in understanding customer satisfaction and feedback.

The final set of visualizations, carefully chosen through this process, ensured a balanced representation of all critical aspects of Olist Store's data, from sales and delivery performance to customer sentiment, thereby facilitating a comprehensive and nuanced understanding of the underlying trends and patterns. This meticulous approach not only enhanced the depth of the analysis but also ensured that the resulting recommendations were grounded in a holistic view of the data.

Visualisations List

Dashboard

1. **Cards for Key Metrics:** Displaying total sales, YTD sales, highest paying order, number of deliveries, total sellers, and average review score. These cards provide a quick, at-a-glance understanding of Olist's current operational status and performance metrics.
2. **Yearly Sales Target KPI:** This helps in tracking the progress towards annual sales goals.
3. **Slicers for State, Category, Score, and Time:** These interactive elements enable users to filter data across different dimensions, offering tailored views.
4. **Mini Visuals:** Including a map for sales by state and region, a donut chart for star ratings, and a treemap for top categories. These offer condensed yet informative insights into sales distribution and customer feedback.

Sales Overview

1. **Total Sales by State Bar Chart:** Illustrates sales distribution across states, highlighting areas of high and low performance.
2. **Sales by City Map:** Offers a detailed geographical representation of sales, pinpointing cities with the highest sales volumes.
3. **Sales by Region and State Map:** A color-coded map to compare regional sales performances, identifying strong and weak market areas.

Sales Trends

1. **Line Chart of Daily Sales Over Time:** Tracks sales fluctuations, helping identify trends and seasonality.
2. **Treemap of Sales by Category:** Provides a visual representation of which categories contribute most to revenue, facilitating strategic decisions on inventory and marketing.

Delivery

1. **Column Chart of Average Delivery Time by State:** Highlights variations in delivery efficiency across regions.
2. **Scatter Plot of Review Score vs. Delivery Time:** Analyses the relationship between delivery performance and customer satisfaction.

Reviews

1. **Average Review Score Card:** Summarizes overall customer satisfaction.
2. **Star Rating Donut Chart:** Visualizes the breakdown of review scores.
3. **Review Sentiment Word Cloud:** Offers insights into common themes in customer feedback.

Forecasts

1. **Line Chart with Future Sales Forecast:** Projects sales trends into the future, aiding in strategic planning.
2. **Sales Decomposition Tree:** Allows users to explore sales data across various dimensions like payment type, product category, and state.

Each visualization was chosen to give a comprehensive understanding of different aspects of Olist Store's operations, from sales patterns to customer feedback. They enable stakeholders to quickly grasp complex data and make informed decisions.

PowerBI Visualisation Development

Sales Distribution

Total Sales by State Bar Chart

For the Total Sales by State Bar Chart, we used the 'Customer State' field (olist_customers_dataset[customers_state]) on the Y-axis to represent different states. The X-axis displayed 'Total Sales', calculated using the DAX formula $\text{Total Sales} = \text{SUM}(\text{OrderItems}[\text{Sales}])$. This measure summed up sales from the Order Items dataset (Olist_order_items_dataset[Total_sales]), providing a clear state-wise representation of sales volumes.

Sales by City Map

When creating the Sales by City Map, we focused on a more localized view. The map visualization utilized 'customer_city' from the Olist_customer_dataset and 'Sales', calculated as $\text{SUM}(\text{OrderItems}[\text{Sales}])$. This approach allowed us to plot sales distribution within cities, visually representing where sales were concentrated.

Sales by Region and State Map

For the Sales by Region and State Map, a more complex approach was required. We combined 'customer_state' and regional information from the olist_customer_dataset with a comprehensive sales measure. This measure, Total Sales = $\text{SUMX}(\text{'Olist_order_items_dataset'}, \text{'Olist_order_items_dataset'}[\text{price}] * \text{'Olist_order_items_dataset'}[\text{order_item_id}])$, aggregated sales data by multiplying the price by the order item ID, offering a detailed view of sales across different regions.

Sales Trends

Line Chart of Daily Sales Over Time

In the Sales Trends section, we crafted a Line Chart of Daily Sales Over Time to visualize Olist's sales trajectory. This chart was generated by pairing the order purchase date with total sales, revealing patterns and trends over time.

We started by selecting a line chart from the visual panel, ideal for showing sales trends. The X-axis displayed the order purchase date from Olist_orders_dataset[order_purchase_data], organized hierarchically, while the Y-axis represented total sales. We created a new measure for total sales using the DAX formula $\text{Total Sales} = \text{SUMX}(\text{'Olist_order_items_dataset'}, \text{'Olist_order_items_dataset'}[\text{price}] * \text{'Olist_order_items_dataset'}[\text{order_item_id}])$. This measure calculated aggregated sales for each date, with the line plot connecting these points to depict the overall sales trend.

Additionally, we included a white trend line and a blue line indicating the average sales, calculated using $\text{Avg Sale Value} = \text{AVERAGEX}(\text{Olist_order_items_dataset}, \text{'Olist_order_items_dataset'}[\text{price}])$. This average, at 24,996.33, represented the mean price of order items. Markers were added to highlight significant growth points and outliers within the analysis period. This comprehensive visualization was instrumental in understanding Olist's sales dynamics over the specified period.

Treemap of Sales by Category

This is a graphical representation that visualises the distribution of sales across different product categories using nested rectangles and the size of each rectangle corresponding to the relative

proportion of sales attributed to the categories. Olist store has a product category data set which was used for this chart.

From the Olist_product_category_dataset, selected product_category_name_english and from the Olist_order_items data set ,selected the total sales column. which was a new measure created for this analysis. This was formatted by adding the title 'total sales by product category', adding the background and header colour to match the dashboard theme. Additionally, advanced controls were integrated to allow for the selection of distinct colours for each product category, enhancing visual appeal and facilitating easy differentiation.

Delivery Analysis

Column Chart of Average Delivery Time by State

In this PowerBI chart, we've visualised the Average Delivery Time by State using a Clustered Column Chart. The X-axis represents the State Name, sourced from the 'olist_customers_dataset_state_name' column in the 'olist_customers_dataset' table. On the Y-axis, we have the 'avg_delivery_time' measure, which calculates the average delivery time in hours.

This measure is created using the following DAX code:

```
Avg Delivery Time = AVERAGE('olist_orders_dataset'[delivery_time_span])
```

The 'delivery_time_span' measure, used in the calculation of 'avg_delivery_time,' is defined as follows:

```
(delivery_time_span =  
DATEDIFF(  
'olist_orders_dataset'[order_purchase_date],  
'olist_orders_dataset'[order_delivered_customer_date],  
DAY  
)
```

This measure computes the time span in days between the order purchase date and the delivered customer date, providing a quantitative measure of the time it takes for orders to be delivered.

To maintain consistency, the standard colour (#b5411e) has been applied to the chart, and the title 'Average Delivery Time by State' succinctly communicates the focus of the visualisation.

In summary, this Clustered Column Chart offers a clear representation of the average delivery time across different states, providing valuable insights into variations in delivery efficiency. The DAX measures and steps in PowerBI ensure accurate calculations and a meaningful presentation of the data.

Scatter Plot of Review Score vs. Delivery Time

In this PowerBI chart, we've visualised the relationship between Review Score and Delivery Time using a Scatter Plot. The X-axis represents the Delivery Time in hours ('delivery_time_span'), calculated using the following DAX code:

```
delivery_time_span =  
DATEDIFF(  
'olist_orders_dataset'[order_purchase_date],  
'olist_orders_dataset'[order_delivered_customer_date],
```

DAY
)

This measure calculates the time span in days between the order purchase date and the delivered customer date, providing a quantitative measure of the time it takes for orders to be delivered.

The Y-axis represents the Review Score (1-5) with the 'Average Review Score' measure, calculated using the following DAX code:

```
Average Review Score =  
AVERAGEX(  
SUMMARISE(  
'olist_order_reviews_dataset',  
'olist_order_reviews_dataset'[order_id],  
"AvgReviewScore", AVERAGE('olist_order_reviews_dataset'[review_score])  
),  
[AvgReviewScore]  
)
```

This measure uses the AVERAGEX function to calculate the average review score for each unique order ID in the 'olist_order_reviews_dataset'. It provides a consolidated average review score for each order, allowing us to understand the overall satisfaction level.

Additionally, we've incorporated the size of each point to reflect the 'Order_Count' using the Size parameter. The 'Order_Count' measure, calculated with the following DAX code, determines the size of each point in the scatter plot:

```
Order_Count = COUNTROWS('olist_orders_dataset')
```

This measure counts the number of rows in the 'olist_orders_dataset', representing the number of orders associated with each data point in the scatter plot.

To enhance the visual representation, we used the standard colour (#b5411e) for consistency and ease of interpretation across different visualisations. The title, 'Relationship between Review Score and Delivery Time,' succinctly conveys the focus of the chart, guiding viewers in understanding the key variables being compared. Overall, this chart provides a comprehensive view of the interplay between review scores, delivery times, and order count, aiding in data-driven decision-making processes.

Reviews and Customer Feedback

Average Review Score Card

In this PowerBI card visualisation, we've presented the Average Review Score. The measure 'Avg Review Score' is derived from the 'olist_order_reviews_dataset' table, and it is calculated using the following DAX code:

```
Avg Review Score = AVERAGE('Olist_order_reviews_dataset'[review_score])
```

This measure computes the average review score across all entries in the 'olist_order_reviews_dataset', providing a single, consolidated value that represents the overall satisfaction level.

To maintain consistency, the standard colour (#b5411e) has been applied to the card, and the title 'Average Review Score' succinctly conveys the focus of the visualisation.

In essence, this card serves as a quick reference for the average review score, allowing stakeholders to easily gauge the overall satisfaction level without the need for a detailed analysis. The use of DAX measures in PowerBI ensures accurate calculations, and the standardised colour and title contribute to a cohesive and easily interpretable dashboard.

Star Rating Donut Chart

In constructing this PowerBI Star Rating Donut Chart, we utilised the summarised (SUM) column tool on the 'review_score' column from the 'olist_order_reviews_dataset'. The result was added to the chart as 'Review Value' in the legend.

To maintain consistency with our other charts, we applied variations of the colour (#b5411e) to represent the review scores ranging from 1 to 5 on the chart slices. In the Options - Label contents section, we opted for the data value and percentage of total for enhanced visualisation. The title 'Star Rating 1-5' was chosen to succinctly convey the focus of the visualisation.

This Donut Chart effectively communicates the distribution of review scores, allowing stakeholders to quickly grasp the distribution of star ratings and their relative proportions. The standardisation of colour and thoughtful label choices enhance the overall clarity and consistency within the PowerBI dashboard.

Review Sentiment Word Cloud

In crafting this PowerBI Word Cloud 2.0.0 visual, we introduced a custom visual type called WordCloud2.0.0. To achieve this, we navigated to Get More Visuals and selected WordCloud2.0.0. In the Category field, we combined the review comments and comment messages, creating a new column named 'Combine Comments' with the following DAX expression:

```
Combine Comments = 'olist_order_reviews_dataset'[review_comment_title] & " " & 'olist_order_reviews_dataset'[review_comment_message]
```

This column concatenates the content of 'review_comment_title' and 'review_comment_message', offering a comprehensive text for sentiment analysis.

General visual format adjustments were made to limit the Word Cloud to a maximum of 200 words, and default stop words were retained for improved accuracy in sentiment analysis. The title 'Review Sentiment' was assigned to the visual, contributing to a cohesive dashboard.

This Word Cloud chart provides an intuitive representation of sentiments expressed in review comments. By merging both review titles and messages, a comprehensive analysis of sentiment is achieved. The Word Cloud format is chosen for its ability to visually emphasise frequently mentioned words, offering a quick overview of prevailing sentiments in customer reviews.

Sales Prediction Modelling

Line Chart with Future Sales Forecast

Continuing on from the Sales Trends section, the line chart showing the total sales each day over the two-year period has been extended to include the forecast of sales for the next year. As before, the “date”

column was used on the x-axis broken down for year, month, and day, with the “Total Sales” measure used on the y-axis.

In order to create the forecast, the “forecasting” feature in PowerBI’s analytics pane was enabled and the timeline was extended by a year to September 2019. The confidence interval was set to 80%. The formatting was kept the same as line chart in the “Sales Trends” section.

Sales Decomposition Tree

The Sales Decomposition Tree was mainly created in order to filter down the forecast by payment type, product category and state. The Total Sales measure was used for analysis. This allows the accuracy of the forecast to increase significantly, as the sales data over the two measured years are quite random, making it difficult to predict sales over extended periods of time for the whole dataset. This allowed for the exploration of different factors affecting sales in specific areas, providing a dynamic way to analyse sales drivers and trends.

Analysis, Insights and Recommendations

Sales Overview

Sales were spread across many states and cities with high clusters in Sao Paulo and Rio de Janeiro. Sao Paulo is the economic nerve centre of Brazil and contributed BRL5.9M (38.3%) to the total sales figure of BRL15.40M. In contrast, Roraima contributed only BRL8.73k. The states with high sales volume are in the South of Brazil while states with the least sales volumes are located in the North. This could reflect high disparities in the level of economic development across Brazilian states as indexed by purchasing powers, education, and infrastructures. The North-South dichotomy is quite interesting as we can see a similar pattern in Korea, United Kingdom, and Nigeria.

It is also interesting to note that the states with low sales presented the highest average sales value (Roraima with BRL150.57 and Sao Paulo with 109.65). This could suggest that the few orders from those states were probably made by a few individuals with projects within the period. The picture from Sao Paulo reveals that many orders were placed by many people with a large range of values.

To grow sales in the low sale states, it could be worthwhile to consider offering additional listing and delivery incentives.

Sales Trends

During our examination of the Olist database, placing particular emphasis on the trajectory of sales during a 2years trading period of September 2016- Sept 2018, it is shown that the sales from the beginning of trade were slow with up and down fluctuations throughout till September 2017 where a distinct sale of BRL 82,473.25 was recorded. Sales declined in October of the same year but skyrocketed and got to a peak in November 2018 which was attributed to the black Friday. Black Friday is a popular shopping event that marks the beginning of the Christmas shopping season and is characterised by significant discounts, promotions, and special deals offered by retailers. While it began primarily in the U.S It has since spread to other parts of the world, with many countries adopting similar sales events including Brazil. Overall, Black Friday has become a crucial and highly anticipated event for both retailers and consumers, influencing purchasing decisions and shaping the economic landscape during the holiday season. For this sole event, Sales recorded was as high as BRL 186,348.72 which is the highest and Olist stores would appreciate it if Sales remained at this level.

Furthermore in our investigation, it was paramount to know what product categories on the Olist store sold more. Health_beauty, watches_gifts and bed_bath_table have been recorded as the top 3 most sold categories whilst cds_dvds_musicals, home_comfort and flowers were the least sold product categories.

The variance in sales performance among different product categories can be attributed to several factors, each influencing consumer behaviour and preferences. Let's explore potential reasons for why the Health & Beauty, Watches & Gifts, and Bed & Bath Table categories recorded higher sales, while CDs, DVDs, Musicals, Home Comfort, and Flowers emerged as the least sold categories:

1. **Consumer Demand and Trends:** Categories like Health & Beauty and Watches & Gifts may align with current consumer trends and demands. Health & Beauty products often address personal care needs, and Watches & Gifts may appeal to shoppers seeking gifting options.
2. **Seasonal Influences:** Products within the Bed & Bath Table category might be influenced by seasonal demands. For instance, there could be increased sales during specific seasons or holidays when consumers tend to focus on home decor and bedding.
3. **Promotional Activities:** Effective marketing strategies, promotions, or discounts may have contributed to the higher sales of certain categories. Retailers may have implemented successful promotional campaigns for Health & Beauty, Watches & Gifts, and Bed & Bath Table items.
4. **Product Variety and Quality:** The top-selling categories might offer a diverse range of high-quality products, meeting a broad spectrum of consumer preferences. The appeal of these categories may be enhanced by product variety, quality, and innovation.
5. **Lifestyle Changes:** Changes in consumer lifestyles can impact purchasing patterns. Categories related to health, beauty, and lifestyle accessories may align with contemporary consumer preferences, leading to increased sales.

Now, let's consider potential reasons for the lower sales in the CDs, DVDs, Musicals, Home Comfort, and Flowers categories:

1. **Digital Transformation:** The decline in physical media consumption due to the rise of digital streaming services may contribute to lower sales in CDs, DVDs, and Musicals.
2. **Changing Home Trends:** Home Comfort may face challenges if consumer preferences have shifted away from traditional home comfort products or if there are emerging trends in home decor and furnishings.
3. **Seasonal or Occasional Demand for Flowers:** Flowers may be a product with more occasional demand, such as during specific holidays or events, resulting in lower overall sales compared to more consistently sought-after categories.
4. **Competition and Market Saturation:** Intense competition or market saturation in certain categories can lead to decreased sales. For instance, if the market for home comfort products is saturated, it might impact sales performance.
5. **Economic Factors:** Economic conditions and consumer spending patterns can influence sales. In times of economic uncertainty, discretionary spending on certain categories, like flowers or home comfort items, may decrease.

Understanding these factors can help Olist stores to tailor their strategies, such as refining product offerings, optimising marketing efforts, and adapting to changing consumer preferences, to improve sales performance across different categories.

Deliveries

In our analysis of the Olist database, with a specific focus on delivery-related data, we delved into the average delivery time across Brazilian states to extract valuable insights. Notably, São Paulo, the state with the highest sales volume and an economic hub, demonstrated the lowest delivery time, representing a significant 38% of Olist's total sales. Similarly, states in the Southeast region, including Rio de Janeiro and Minas Gerais, along with those in the South, showcased high sales figures and relatively low delivery times.

However, an intriguing observation emerged when examining individual states. Despite São Paulo handling over 40,000 deliveries, only 4% were considered late. In contrast, Roraima, with a considerably smaller number of 41 deliveries, exhibited an astonishing 12% late delivery rate. This discrepancy underscores the need for a closer look at the factors influencing delivery efficiency, even in states with lower delivery volumes.

Conversely, the states in the North and Northeast regions displayed lower sales and longer delivery times. Roraima, in particular, stood out as having the worst performance in both sales and delivery time. This disparity can be linked to economic challenges and infrastructure limitations in these regions, further exacerbated by neglect from the federal government.

To enhance Olist's delivery performance, we propose the following recommendations:

1. **Invest in Infrastructure:** Allocate resources to improve delivery infrastructure in states with longer delivery times, especially in the North and Northeast regions. Addressing logistical challenges can contribute to more efficient and timely deliveries.
2. **Regional Marketing Strategies:** Implement tailored marketing strategies in regions with lower sales to boost brand awareness and engagement. Highlighting the benefits of Olist's services can attract more customers and potentially increase sales.
3. **Collaborate with Local Authorities:** Foster collaborations with local authorities in regions facing economic challenges. Establishing partnerships can facilitate smoother operations and help overcome obstacles tied to neglect from the federal government.
4. **Enhance Delivery Tracking Systems:** Improve delivery tracking systems to minimise late deliveries. Analysing the factors contributing to delays, especially in states like Roraima, can inform targeted improvements and enhance overall customer satisfaction.

By incorporating these recommendations, Olist can work towards optimising its delivery processes, addressing regional disparities, and fostering a more efficient and reliable service across different states.

Reviews

In our exploration of the Olist database, we delved beyond geographical patterns, focusing on understanding the intricate relationship between Review Score and Delivery Time. A distinct negative correlation emerged, underscoring that faster delivery times, measured in hours, were strongly associated with higher review scores. However, beyond a delivery time of 50 hours, this correlation waned. This shift suggests that, after a certain threshold, delivery time becomes less influential on review scores, likely due to other issues that may have arisen with the order.

A noteworthy observation was the presence of a significant number of orders with low scores despite having a delivery time recorded as zero. A deeper investigation revealed that these orders were not

delivered at all. This emphasises the critical importance of accurate delivery tracking and fulfilment in maintaining customer satisfaction.

Furthermore, our analysis of review scores revealed that approximately 90% of all reviews received 5 stars, 70% received 4 stars, and 20% received 3 stars. This positive trend is particularly prevalent in the main states of the Southeast and South regions, where Olist concentrates most of its business. The average review score of 4.09 stars indicates a high level of overall satisfaction among customers. This suggests that both sellers and delivery companies are efficiently working to send products in a timely manner while maintaining quality standards.

Examining the Review Sentiment Word Cloud, we observe that the most common words associated with high review scores are 'recommend, arrived, loved, delivery, product.' In contrast, when review scores are low, the main words are 'no, refund, delivery, bought, received, original.' This analysis underscores the importance of customer sentiments and feedback in identifying areas for improvement and maintaining positive customer experiences.

To improve Olist's customer sentiment and satisfaction, we propose the following recommendations:

1. **Continuous Quality Control:** Ensure continuous quality control in delivery processes, particularly beyond the 50-hour threshold, to maintain high customer satisfaction levels.
2. **Enhanced Tracking Communication:** Implement enhanced communication strategies for accurate delivery tracking and timely updates, emphasising transparency to minimise dissatisfaction.
3. **Regional Business Expansion:** Consider expanding business operations in regions where review scores are high to tap into the positive market sentiment and further enhance Olist's presence.
4. **Feedback Loop Improvement:** Strengthen the feedback loop to gather insights from low-scoring reviews, identifying specific pain points and areas for improvement in both product quality and delivery services.

By incorporating these recommendations, Olist can strengthen its operations, address specific challenges, and maintain a positive customer experience across different regions and review scenarios.

Forecasts

The line chart displays Olist Store's historical sales trends and extends into a forecast for future sales for the next year (September 2018 to September 2019). This visualization helps in identifying sales growth or decline patterns and predicting future sales volumes.

The trend suggests an upward trajectory in sales, based on past performance. We can also predict based on past performance that autumn is the highest selling time every year, with peaks at the end of September and November. We can also see, if we filter using the sales decomposition tree, that sales are minimal at best in some states such as Roraima and Amapa and are predicted to be so indefinitely, unless there is a change in sales strategies.

The recommendations, therefore, we would like to make consist of the following:

1. **Seasonal Adjustments:** Analyse historical sales trends to identify peak seasons and adjust marketing strategies and inventory accordingly. This ensures that Olist Store is well-prepared to meet the increased demand during specific times of the year, like holidays or local festivals.

2. **Strategic Pricing:** Implement dynamic pricing strategies based on demand, competition, and customer behaviour insights. This approach enables Olist Store to optimize prices for maximum profitability while remaining competitive and attractive to customers.
3. **Inventory Management:** Utilize sales forecasts to maintain optimal inventory levels. By aligning inventory with predicted sales trends, Olist Store can minimize overstocking and stockouts, thereby reducing storage costs and ensuring product availability.
4. **Risk Management:** Develop strategies to mitigate risks associated with fluctuating sales and market trends. This includes diversifying the product range, exploring new markets, and employing contingency plans for unexpected sales dips or supply chain disruptions.

Pest Analysis

Navigating the dynamic landscape of e-commerce demands a keen awareness of external factors that can significantly influence business operations. In this context, a PEST analysis, examining Political, Economic, Sociological, and Technological dimensions, emerges as a pivotal strategic tool. It serves as a compass for businesses like Olist, offering insights into the broader environmental forces shaping the industry in Brazil between 2016 and 2018.

Political factors encapsulate the regulatory framework and governmental stability, influencing compliance and business practices. Economic indicators, such as GDP growth and inflation rates, provide a lens into consumer spending behaviours critical for Olist's strategic planning. Sociological aspects delve into evolving consumer preferences and lifestyle changes, guiding Olist in adapting to market trends. Finally, technological dimensions explore innovations like mobile commerce and artificial intelligence, pivotal for Olist to stay ahead in an ever-evolving digital landscape. As we explore the specific factors within each dimension, their collective impact on Olist's e-commerce journey becomes clear.

Political Factors

1. **Political Stability and Government Policies:** During this period, Brazil experienced political instability with the impeachment of President Dilma Rousseff in 2016. Political uncertainties can impact investor confidence and economic stability.
2. **Trade Policies and Tariffs:** Changes in trade policies or tariffs can affect the cost of imported goods, influencing product prices and potentially impacting e-commerce sales.
3. **Consumer Protection Regulations:** Evolving consumer protection regulations may have influenced e-commerce practices, affecting how businesses handle transactions, returns, and customer satisfaction.
4. **Data Protection and Privacy Laws:** The implementation of data protection laws, such as the Brazilian General Data Protection Law (LGPD) in 2018, impacted how e-commerce companies handle customer data, requiring strict compliance.

Economic Factors

1. **Economic Recovery from Recession:** Brazil's GDP grew by approximately 1% in 2017 and 1.1% in 2018, indicating a slow but positive economic recovery during the period.
2. **Interest Rates and Inflation:** The average inflation rate in Brazil was around 3.7% in 2017 and 3.7% in 2018. The benchmark interest rate, Selic, declined from 14% at the beginning of 2016 to 6.5% by the end of 2018, influencing borrowing costs.

3. **Currency Exchange Rates:** The Brazilian Real (BRL) experienced fluctuations against major currencies during the period. For instance, the USD to BRL exchange rate went from around 3.3 in 2016 to approximately 3.8 in 2018.
4. **Job Market Conditions:** The unemployment rate in Brazil, which peaked at around 13.7% in early 2017, gradually improved to about 11.6% by the end of 2018. Improvements in job market conditions could positively impact consumer confidence and spending.

Sociological Factors

1. **Consumer Behaviour and Trust:** The number of online shoppers in Brazil increased from approximately 40 million in 2016 to around 47 million in 2018, indicating a growing acceptance and trust in online transactions.
2. **Demographic Changes:** The percentage of internet users in Brazil aged 16 to 64 increased from 61% in 2016 to 65% in 2018. This demographic shift suggests a broader audience for online platforms.
3. **Urbanization and Lifestyle Changes:** The urbanisation rate in Brazil remained high, with approximately 86% of the population residing in urban areas. Urbanisation trends typically contribute to the popularity of online shopping due to convenience.
4. **Social Media Influence:** The number of social media users in Brazil increased from around 120 million in 2016 to over 140 million in 2018. The growing influence of social media likely impacted consumer purchasing decisions.

Technological Factors

1. **Mobile Commerce (M-commerce) Growth:** The share of mobile commerce (M-commerce) in total e-commerce sales increased from approximately 28.1% in 2016 to 35.9% in 2018. This reflects the growing reliance on mobile devices for online transactions.
2. **Advancements in Artificial Intelligence (AI):** The adoption of AI in customer service and personalised recommendations became more prevalent. Many e-commerce platforms reported a significant improvement in user engagement, with some seeing up to a 30% increase in sales attributed to AI-driven features.
3. **E-commerce Platforms and Innovations:** Continuous innovation in e-commerce platforms led to enhanced user experiences. Several platforms reported a decrease in bounce rates and an increase in conversion rates, suggesting improved website usability.
4. **Logistics and Delivery Innovations:** The implementation of same-day delivery and other logistics innovations became more widespread. Some e-commerce companies reported up to a 20% increase in customer satisfaction scores following improvements in their delivery services.

Conclusions

This report, through a detailed analysis of Olist Store's operations from 2016 to 2018, has leveraged a combination of descriptive, exploratory, and predictive analytics to unravel key patterns and trends. Our methodology, rooted in a data-driven approach, has enabled us to craft insightful PowerBI visualizations, providing a clear window into the dynamics of Olist's sales, delivery performance, and customer sentiment.

Sales Overview Analysis

In our Sales Overview, we observed significant disparities in sales across different Brazilian states. The PowerBI visualizations, particularly the Total Sales by State Bar Chart and the Sales by City Map, highlighted Sao Paulo, and Rio de Janeiro as economic hotspots. These findings suggest potential for targeted marketing and strategic inventory adjustments in regions with lower sales. Enhanced promotional activities and localized inventory management could be key to tapping into underutilized markets.

Sales Trends Analysis

The Sales Trends section, featuring the Line Chart of Daily Sales and the Treemap of Sales by Category, revealed seasonal patterns and category preferences. This insight is crucial for strategic pricing and inventory management. Understanding which categories drive sales during different seasons can inform targeted marketing campaigns and stock adjustments to capitalize on customer demand.

Delivery Analysis

Our Delivery Analysis, showcased through the Column Chart of Average Delivery Time by State and Scatter Plot of Review Score vs. Delivery Time, emphasized the critical role of efficient delivery in customer satisfaction. The data suggests investing in logistics infrastructure, particularly in regions with prolonged delivery times, could significantly enhance customer satisfaction and streamline operations.

Reviews and Customer Feedback Analysis

In the Reviews section, the Average Review Score Card and Star Rating Donut Chart offered a snapshot of customer sentiment. The common themes in customer feedback, especially around delivery times, highlight areas for improvement. Strengthening delivery systems and enhancing customer service responsiveness could address these concerns effectively.

Forecasts Analysis

Finally, the Forecasts section, utilizing the Line Chart with Future Sales Forecast and Sales Decomposition Tree, provided predictive insights into sales trends. The ability to anticipate seasonal peaks and troughs in sales can guide Olist in making informed decisions on inventory management and risk mitigation strategies.

Overall Conclusion

In conclusion, this comprehensive analysis not only sheds light on the current state of Olist Store's operations but also offers actionable recommendations to enhance efficiency, customer satisfaction, and sales performance. Regular monitoring and strategic adjustments based on these insights can propel Olist Store to new heights in the competitive e-commerce landscape.