

Olist eCommerce Dataset Analysis Report

- **Prepared by:** Aniket Ashok Pawar
- **Guide by:** Vishal sir



Introduction:

The Olist eCommerce dataset is a collection of data from a Brazilian e-commerce company that connects small to medium-sized retailers with customers across the country.

Executive summary:

Key Findings:

The analysis of the Olist dataset reveals notable order trends, such as peak sales during weekdays and regional variations in customer preferences. Product performance varies significantly, with bedding, table ware & bath being top seller. Delivery efficiency generally meets customer expectations, though some regions experience delays.

Conclusion:

The insights highlight opportunities to optimize inventory based on regional demand, enhance product offerings, and improve logistics in underperforming areas. Addressing these areas can drive sales growth and customer satisfaction.

Recommendations:

- Tailor marketing and inventory strategies to regional preferences.
- Focus on improving delivery times in regions with frequent delays.
- Expand the range of high-performing product categories to maximize revenue.

Objective:

- To identify order trends
- Customer behaviour & regional analysis
- Product performance evaluation
- Delivery and logistics efficiency



Data overview:

➤ Data source:

The dataset is sourced from Olist, a Brazilian eCommerce platform and It is publicly available on Kaggle platform.

➤ **Data description:**

- The analysis focuses on following key columns: Order ID (unique order identifier), Customer unique ID (unique customer identifier), Order purchased date and Order delivered date (timestamps for order processing), Payment value (total price including freight), and Review score (customer rating from 1 to 5).
- **Records:** [Number of rows - 119143]
- **Features:** [Number of columns - 31]
- **Data records date:** [From – 04 sept, 2016]: [Till – 17 oct, 2018]

➤ **Data pre-processing:**

- **Handling Missing Values:**

Imputed missing values using median, mode values and calculated imputation based on data distribution.

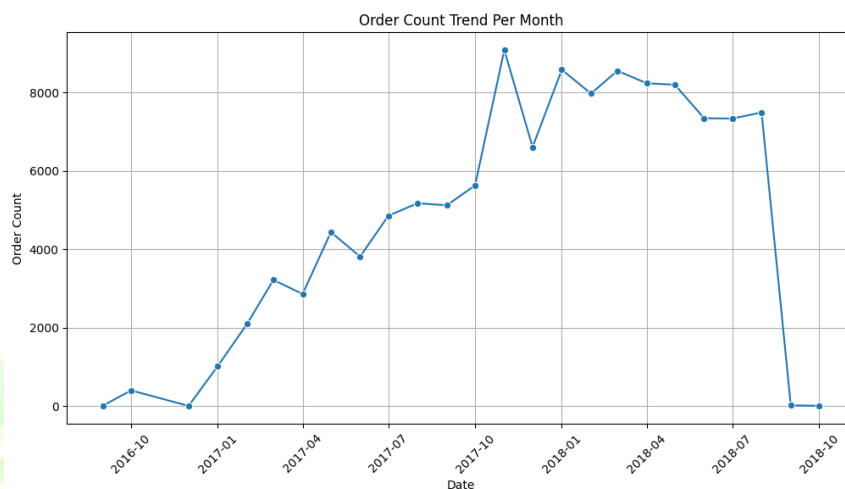
- **Remove duplicates rows:**
Removed duplicates rows using in-built pandas' function.
- **Data transformation:**
- Converted product category name from Brazilian language to English.

➤ **Methodology:**

The analysis was conducted using Python and key libraries such as Pandas for data manipulation, Seaborn and Matplotlib for data visualization, and Microsoft Word for report preparation. Statistical methods including mean, median, mode, and the five-number summary were applied to summarize data and identify key patterns and trends.

Numerical findings & analysis using Visualisation:

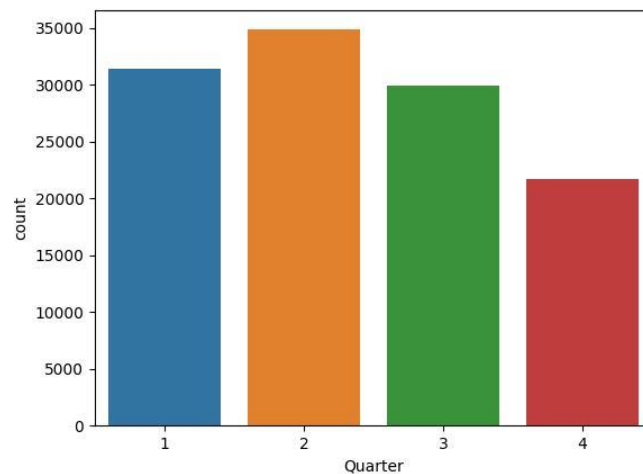
❖ **Overall orders trend:**



Key findings:

- From our analysis, we observed that the number of orders increased significantly over the years, with 409 orders in 2016, 54,549 orders in 2017, and 64,185 orders in 2018.

❖ Orders trend analysis:



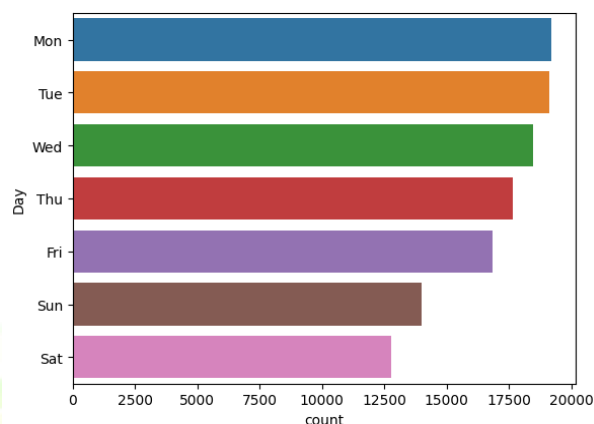
Key findings:

- Orders are significantly higher in the first three quarters of the year, with a noticeable decline in the fourth quarter.



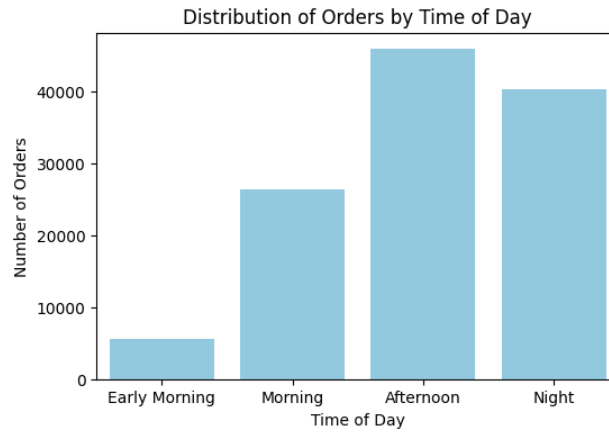
Key findings:

- Monthly analysis shows high order counts from March to August, with a sharp decline in September



Key findings:

- Weekly analysis reveals that orders are highest on Monday to Thursday, with a noticeable drop in order counts starting from Friday.



Key findings:

- Daily analysis indicates that most orders are placed during the afternoon and night
- Bedding, Tableware, and Bath is the most popular product category across all time periods
- Beauty and Health products consistently appear in the top 3 across all time periods
- Computer Accessories is particularly popular in the Morning comparatively
- Sports and Leisure products are in the top 3 only during the Afternoon
- Decorative Furniture ranks third in both the Early Morning and Night

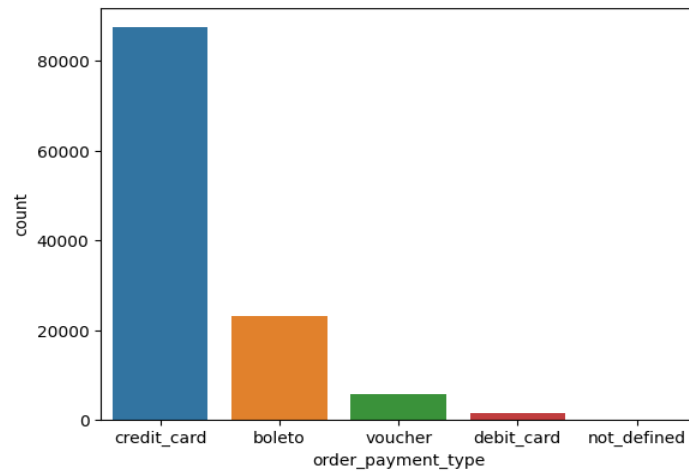
Conclusion:

- Orders are peak from March to August and are highest on weekdays (Monday to Thursday), with a drop starting Friday. Most orders are placed in the afternoon and night. There's a noticeable decline in orders during the fourth quarter, particularly from September onwards.
- The analysis shows that Bedding, Tableware, and Bath, Beauty and Health products dominate consumer interest at all times of the day, highlighting their importance in the Olist marketplace.
- The presence of Sports and Leisure in the Afternoon and Computer Accessories in the Morning reflects some variation in product preferences throughout the day.

Recommendations:

- Increase inventory levels during the first three quarters, particularly from March to August, to meet higher demand.
- Introduce special offers or discounts in September to counter the drop in order volumes and maintain customer engagement.
- Adjust workforce schedules to ensure adequate staffing during peak order times (Monday to Thursday, and afternoons/nights) for efficient order processing and customer service.
- Consider weekend-specific promotions or incentives to boost order volumes from Friday to Sunday.
- Implement loyalty programs or personalized offers during periods of low activity (Q4, weekends) to encourage repeat purchases and customer retention.
- Increase inventory for Bedding, Tableware, and Bath products, Beauty and Health products as they consistently rank as top sellers across all time periods
- Focus on Computer Accessories promotions in the morning and Sports and Leisure in the afternoon to align with peak interest.

❖ Customer behaviour & regional analysis:



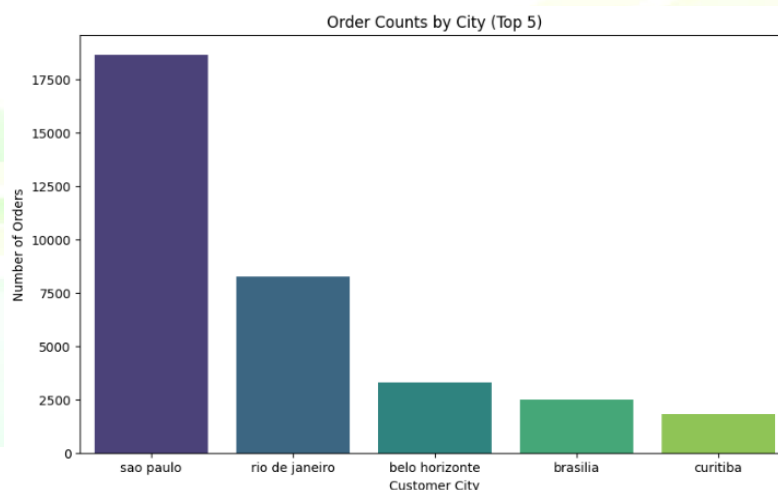
Key findings:

- Analysis shows that credit card users make up approximately 75% of transactions, while boleto users account for about 20%.



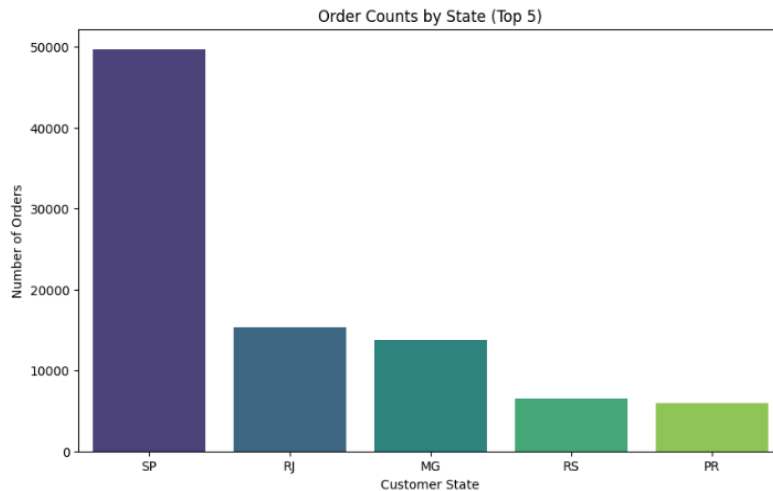
Key findings:

- Analysis of review scores by order status reveals that orders marked as 'Delivered' tend to receive higher review scores, with approximately 75% of customers giving scores of 4 or 5. In contrast, orders that are 'Cancelled' or 'Unavailable' frequently receive lower scores, with a high density of customers providing 1 as review score.
- The "missing" product category has the highest number of canceled and unavailable orders.
- Very few products fall under the "unavailable" order status, with categories like Babies, Perfumes, and Automotive showing only a handful of unavailable orders.
- For canceled orders, popular categories such as Household Utilities, Sports and Leisure, and Computer Accessories, Beauty and Health have relatively high cancellation counts



Key findings:

- The top cities by order counts are Sao Paulo with 18,648 orders, followed by Rio de Janeiro with 8,235 orders, and Belo Horizonte with 3,277 orders.
- Sao Paulo has the highest order count of 2078 for *Bedding, Tableware, and Bath* products.
- Belo Horizonte and Rio de Janeiro also show strong demand for Bedding, Tableware, and Bath, while Curitiba favors Decorative Furniture.



Key findings:

- Analysis of customer orders reveals that Sao Paulo (SP) accounts for 42.2% of the total orders, with 49,728 orders out of 118,018. Rio de Janeiro (RJ) follows with 13.0% of the total, totalling 15,391 orders, while Minas Gerais (MG) contributes 11.6% with 13,714 orders.

Conclusion:

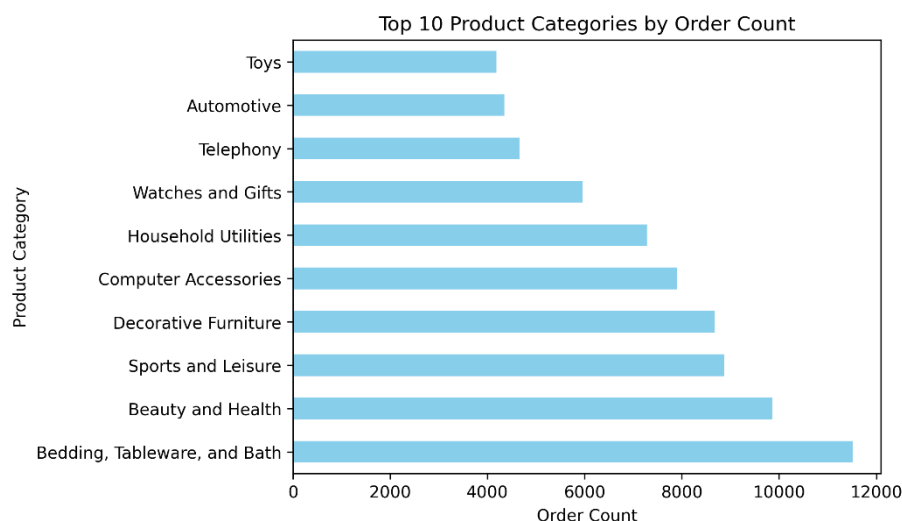
- The customer behaviour analysis shows that credit cards dominate transactions, accounting for 75%, while boleto payments make up 20%. Orders marked as 'Delivered' receive high satisfaction, with 75% of reviews scoring 4 or 5, whereas 'Cancelled' or 'Unavailable' orders often receive low scores. Geographically, Sao Paulo leads with 42.2% of total orders, followed by Rio de Janeiro and Minas Gerais.
- The "missing" product category has the most canceled and unavailable orders, highlighting significant gaps in product data.
- Popular categories such as Household Utilities and Computer Accessories also face higher cancellation rates, indicating potential issues in order processing or inventory management.
- Curitiba and Brasilia demonstrate interest in Decorative Furniture and Beauty and Health, respectively.

Recommendations:

- Since 75% of transactions are made with credit cards, ensure that this payment method is optimized for a seamless user experience.
- Maintain and potentially enhance support for boleto payments, as they account for 20% of transactions, ensuring accessibility for a broader customer base.
- Given that 'Delivered' orders receive the highest customer satisfaction, focus on improving delivery processes to reduce cancellations and unavailability, thereby boosting overall customer satisfaction.

- For orders that are cancelled or unavailable, implement follow-up strategies, such as re-engagement offers or personalized communication, to reduce negative customer experiences.
- **Focus on Sao Paulo, Rio de Janeiro, and Minas Gerais:** Since these regions contribute the highest order volumes, enhance marketing campaigns and promotions to these key areas to maximize reach and effectiveness.
- **Focus on regions with less existence:** Consider regional preferences and behaviour's in these areas to create more personalized and relevant marketing messages that resonate with local customers.
- To address the gaps in the "missing" category, improve product data accuracy and collaborate with vendors to reduce cancellations.
- Enhance inventory management for popular categories like Household Utilities and Computer Accessories to prevent order cancellations. Collect customer feedback to identify issues causing high cancellations in specific categories and optimize communication regarding stock availability and delivery timelines to enhance customer experience.
- Focus on optimizing stock and marketing for Bedding, Tableware, and Bath in cities like São Paulo, Belo Horizonte, and Rio de Janeiro. Additionally, strengthen the supply chain for Decorative Furniture in Curitiba and target health-conscious consumers in Brasilia by promoting Beauty and Health products more effectively.

❖ Product performance evaluation:



Key findings:

- Bedding, Tableware, and Bath leads with 9.76% of total orders, followed by Beauty and Health at 8.36%. Sports and Leisure, Decorative Furniture, and Computer Accessories contribute 7.51%, 7.35%, and 6.69%, respectively.
- **The Top 5 highly rated product:**
Top-rated product categories include Music CDs and DVDs (4.64), Fashion Youth Clothing (4.50), Imported Books (4.42), General Interest Books (4.41), and Technical Books (4.33). These categories receive the highest customer satisfaction scores in the dataset.
- **The Top 5 worst rated product:**
The worst-rated product categories are "Insurance and Services" with an average rating of 2.5, and "Gaming PCs" with 3.1. Other low-rated categories include "Diapers and Hygiene" and "Portable Kitchen Appliances and Food Preparers," both with average ratings just above 3.0.

➤ **Peak months for daily uses product category:**

1. **Bedding, Tableware, and Bath:** July
2. **Household Utilities:** June
3. **Beauty and Health:** August
4. **Sports and Leisure:** March
5. **Telephony:** March

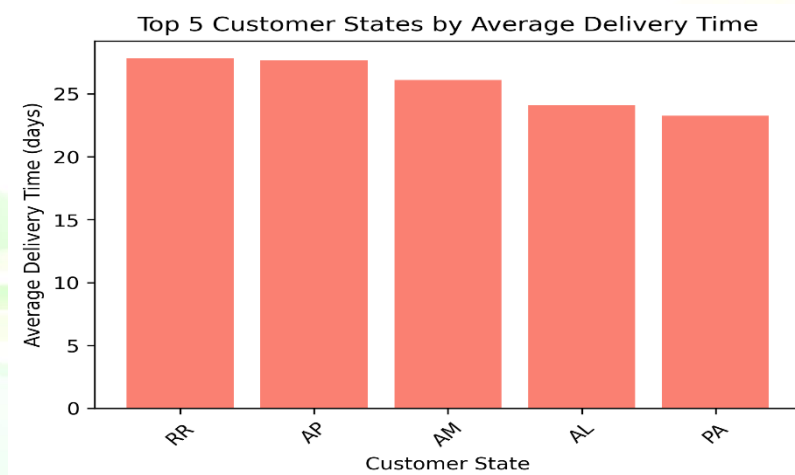
Conclusions:

- Bedding, Tableware, and Bath are the top product categories, each making up a significant portion of total orders, with Beauty and Health also performing well.
- The highest customer satisfaction scores are for Music CDs and DVDs, Fashion Youth Clothing, and various books. In contrast, "Insurance and Services" and "Gaming PCs" are the lowest-rated, receiving notably lower scores, along with "Diapers and Hygiene" and "Portable Kitchen Appliances."

Recommendations:

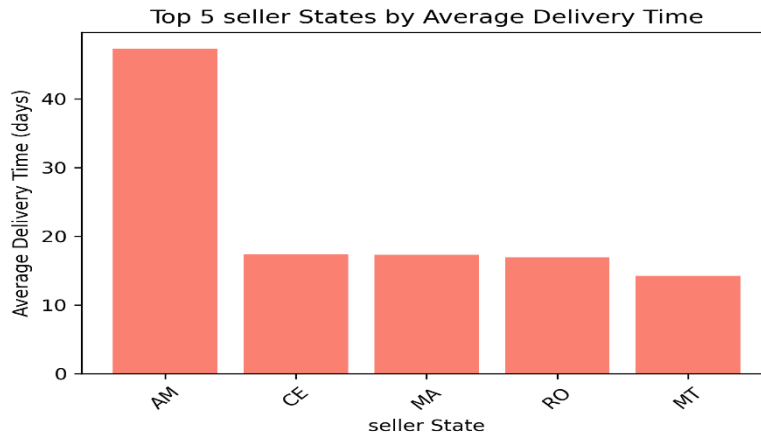
- Continue to invest in and promote high-performing categories like Bedding, Tableware, Bath, and Beauty and Health to maintain strong sales.
- Promote top-rated products like Music CDs, DVDs, and books in your marketing campaigns to take advantage of their high customer satisfaction. This can help boost sales and popularity.
- Consider improving product quality or customer service in low-rated categories to enhance satisfaction and potentially boost ratings.
- Regularly review product performance metrics to adapt strategies and focus on categories showing growth or decline.
- Focus on increasing inventory and launching targeted promotions during peak months, such as Bedding and Bath in July and Beauty and Health in August. Leverage seasonal trends to boost sales, especially for Sports and Telephony products in March.

❖ **Delivery & logistics efficiency:**



Key findings:

- The customer states with the highest average delivery times are Roraima (RR) at 27.83 days, Alagoas (AL) at 27.66 days, Amazonas (AM) at 26.10 days, Amapa (AP) at 24.12 days, and Para (PA) at 23.25 days.
- While average delivery time is 12 days for all orders.



Key findings:

- The seller states with the highest average delivery times are Amazonas (AM) with 47.33 days, Ceara (CE) with 17.39 days, Maranhao (MA) with 17.29 days, Rondonia (RO) with 16.93 days, and Mato Grosso (MT) with 14.23 days.
- **top 5 product products categories by longest avg. delivery time:** The top 5 product categories with the longest average delivery times are Office Furniture at 16.71 days, Christmas Items at 15.28 days, Insurance and Services at 15.00 days, Fashion Footwear at 14.89 days, and Home Comfort 2 at 13.81 days.
- Percentage of late deliveries: 6.37%, Percentage of early deliveries: 89.51%, Percentage of on-time deliveries: 1.26%

Conclusion:

- Delivery times vary significantly across regions, with customer states Roraima, Alagoas, Amazonas, Amapa, and Para experiencing the longest average delivery times, all exceeding 23 days. In contrast, the overall average delivery time for all orders is 12 days.
- Seller states also show considerable variation, with Amazonas having the highest average delivery time at 47.33 days, followed by Ceara, Maranhao, Rondonia, and Mato Grosso.

Recommendations:

- Enhance logistics infrastructure and distribution networks in high-delay states to accelerate deliveries. Additionally, optimize route planning and increase local warehousing to improve order fulfilment speed.
- Strengthen coordination between sellers and logistics providers to ensure faster processing and shipping.
- Use customer and seller feedback to address specific problems or challenges and further refine logistics processes for improved efficiency.
- enhance customer communication by notifying them of early arrivals and offering rescheduling options to improve overall satisfaction.