

An Exploratory Visualization of Brazilian E-Commerce

Winnie Ekwegh

Average Delivery Time by Brazilian State

Based on deliveries made in 30 days or less

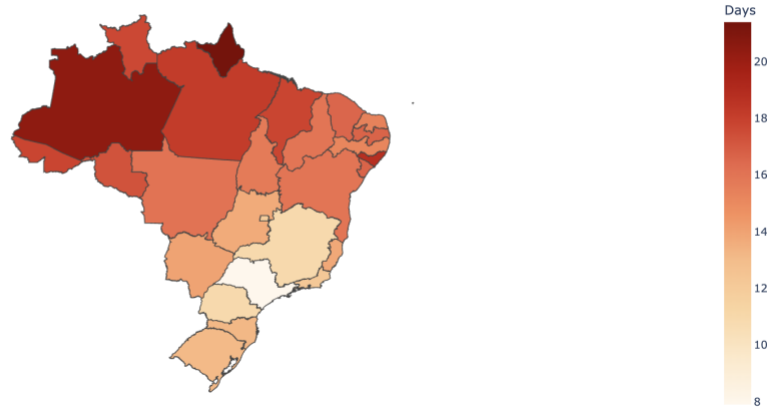


Fig1. Regional disparities in average delivery time across Brazilian states, based on e-commerce deliveries from 2016 to 2018.

Abstract – The growth of e-commerce in Brazil presents a unique opportunity to understand how digital marketplaces operate in a region marked by diverse geography, infrastructure, and consumer behavior. This project explores the Olist e-commerce dataset through interactive and visual analysis to uncover trends in order volume, delivery performance, payment preferences, and customer satisfaction across Brazilian states. Using animated choropleths, heatmaps, bar charts, and translated review sentiment, we reveal that São Paulo consistently dominates in activity, but smaller states like Espírito Santo and Paraíba show surprising engagement when normalized by population. We also find that most orders arrive early despite variation across product categories and sellers. Review analysis reveals recurring dissatisfaction with product accuracy and delivery reliability. These insights provide a multifaceted look at Brazil's online retail ecosystem and suggest areas for improving logistics, seller performance, and customer experience. Interactive visualizations and supplemental materials are available on Google Colab at: [\[Colab Link\]](#)

1 INTRODUCTION

In the last decade, Brazil has experienced a surge in e-commerce adoption, driven by increased internet access, smartphone use, and the rise of online marketplaces. As the largest economy in Latin America, Brazil offers a complex and varied digital retail landscape shaped by regional infrastructure gaps, diverse consumer behaviors, and uneven logistical capacity. Understanding how e-commerce functions across such a vast and heterogeneous country is critical for retailers, policymakers, and logistics providers looking to optimize operations and improve the customer experience.

This study uses the publicly available Olist dataset, which includes transactional data from a major Brazilian e-commerce platform that connects small and medium-sized sellers to nationwide marketplaces. The dataset contains rich information on orders, products, delivery times, customer locations, payment methods, and customer reviews from October 2016 to August 2018. Its comprehensive structure makes it an ideal resource for exploring macro-level patterns as well as more granular behaviors across regions, categories, and time.

Our goal is to uncover insights into how online shopping behavior and customer experiences vary across Brazilian states and product types. We investigate questions such as: Where are orders concentrated geographically? How do payment behaviors differ across regions? Are deliveries consistently on time? Which sellers and categories receive the most dissatisfaction? We also incorporate sentiment analysis of Portuguese-language customer reviews to identify common praise and complaint themes.

Through a series of visually driven analyses—including choropleths, bar charts, scatterplots, and word clouds, we highlight regional disparities in engagement, variations in delivery performance, installment behavior, and sentiment-laden friction points. By embedding these findings into a narrative structure, this paper demonstrates how exploratory data visualization can reveal critical dynamics within complex e-commerce systems, with implications for improving digital retail in Brazil and similar emerging markets.

2 RELATED WORK

Brazil's e-commerce market has grown quickly in recent years and is expected to pass \$200 billion by 2026 [1]. It leads Latin America, accounting for nearly half of the region's online retail, driven by more people getting internet access, the rise of mobile shopping, and the growth of social commerce platforms [2]. Efforts like Favela 3D are helping bring online shopping and delivery services to underserved areas, showing how infrastructure can shape who benefits from e-commerce [3].

One key difference in Brazil's e-commerce is how people pay. Pix, a real-time payment method launched by the central bank, is already more popular than credit cards for many online purchases and is expected to become the dominant method by 2025 [4]. New "buy now, pay later" (BNPL) services tied to Pix are also growing fast [5]. Before Pix, many people used installment payments on credit cards to spread out costs—especially in low-income or underbanked communities [6].

Delivery is another major factor in how people experience online shopping. Studies have shown that fast and reliable delivery makes customers more likely to come back and shop again [7]. In a country as large and diverse as Brazil, delivery depends on many things—like road infrastructure, how close warehouses are, and how sellers manage their orders. Some researchers have even built tools to measure the quality of delivery systems across different regions [8].

Customer reviews also offer a lot of useful feedback. By analyzing what people write in reviews, we can understand what they like or don't like—whether it's delays, product problems, or communication issues. Some recent studies have used natural language processing (NLP) to study reviews written in Portuguese and have shown how important it is to account for language and cultural differences when analyzing sentiment [9], [10].

Finally, data visualization plays a big role in helping people understand patterns in large datasets. Interactive charts, maps, and dashboards have been used in other research areas like retail, logistics, and public health to explore trends and support decision-making [11]. This project builds on those ideas by using storytelling through interactive visualizations to make insights about Brazilian e-commerce more accessible and engaging for a broad audience.

3 PROJECT DESCRIPTION

This project explores key aspects of Brazil's online shopping landscape using the available Olist e-commerce dataset on Kaggle. It focuses on three main areas: how people pay, how fast and reliably items are delivered, and how satisfied customers are based on the reviews they leave.

To start, the data was cleaned and merged across multiple CSV files, combining information on orders, customers, payments, reviews, and products. Because many of the customer reviews were written in Portuguese, they were translated into English using Google Translate. This made it possible to apply sentiment analysis techniques to understand customer satisfaction more accurately. Each review was scored on a sentiment scale, from negative to positive, allowing us to

compare star ratings with how customers actually felt in their comments.

For the payment analysis, we looked at how many times people chose to split their payments into installments. These patterns were then visualized by region, showing where installment payments are most common. For delivery performance, we calculated delivery delays based on the difference between promised and actual delivery dates. These delays were mapped by state and region to highlight geographic disparities. For customer sentiment, we visualized both numerical scores and selected review snippets to reveal common themes related to dissatisfaction.

All visualizations were built using Python and Plotly. We used a mix of animated choropleths, bar charts, scatter plots, and interactive text elements to help tell a clear story. The goal was to create visuals that are easy to explore and understand, even for people who may not have a background in data science.

This project is designed to support both exploration and explanation. The intended audience includes students, educators, and decision-makers who want to better understand digital commerce in Latin America. The visualizations highlight where challenges exist—such as delivery delays in remote areas or mismatched sentiment in reviews—and open the door for deeper conversations about how infrastructure, payment access, and customer experience vary across Brazil.

4 PROJECT DETAILS

4.1 Mapping the E-Commerce Landscape: Where and when Are People Shopping?

To understand the growth and reach of online shopping in Brazil, we began with an animated choropleth map showing the monthly spread of e-commerce orders by state from 2016 to 2018. This visual illustrates the dominance of São Paulo from the start, with gradual adoption radiating outward into other regions. The animation emphasizes how digital retail grew over time but remained uneven across the country.

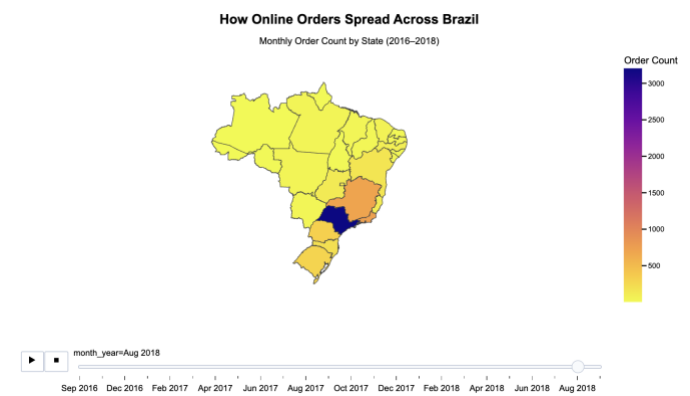


Figure 2. Monthly spread of online orders across Brazilian states (2016–2018).

Next, we used a heatmap to compress both time and geography into a single view, enabling easy comparison across states and months. Unlike the map, which emphasizes spatial spread, the heatmap makes it easy to spot temporal lags and inactivity, especially in northern and interior states. White cells highlight entire months with no recorded orders, revealing clear disparities in adoption. São Paulo shows consistent growth, while states like Acre and Roraima show sporadic bursts of activity rather than steady engagement.

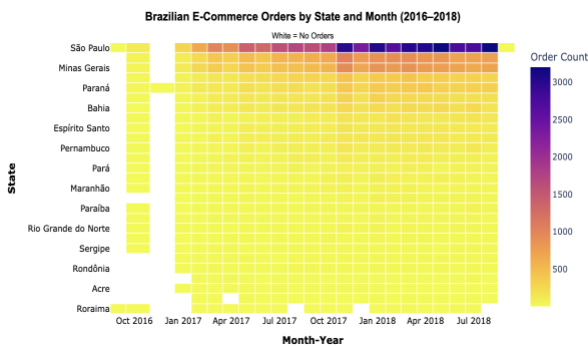


Figure 3. Heatmap of e-commerce order counts by state and month (2016–2018).

To further contextualize regional disparities, we included a bar chart normalized by population—measuring orders per 100,000 residents. This view adjusts for state size and population, uncovering unexpected leaders like Espírito Santo, Paraíba, and the Distrito Federal, where engagement is high relative to population. It also reshuffles our understanding of digital inclusion: states with large populations and raw order volumes, such as Bahia and Pará, fall lower in the rankings once normalized.



Figure 4. Orders per 100,000 residents by state, normalized by population.

Together, these visuals show that while Brazil’s e-commerce market is growing, its spread is far from uniform. Major urban hubs lead in both scale and consistency, but smaller states also demonstrate notable per-capita engagement, challenging assumptions about digital exclusion in less populous regions.

4.2 How Do Brazilians pay online?

Brazilian consumers overwhelmingly prefer credit cards when shopping online. As shown in Figure 5, nearly three-quarters (73.7%) of all online orders were paid via credit card, far outpacing boleto bancário (19.5%), vouchers, or debit cards. However, this national pattern masks important regional differences.

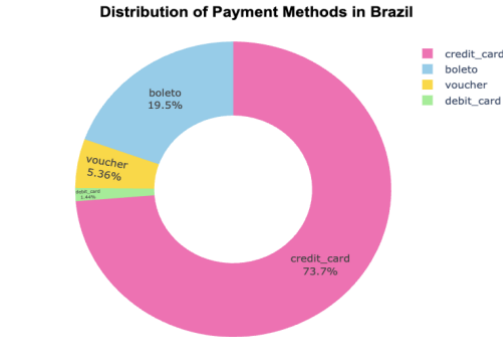


Figure 5. Distribution of payment methods in Brazil.

Figure 6 breaks payment behavior down by state, revealing that boleto usage is notably higher in northern states like Amazonas and Maranhão. In contrast, states like São Paulo and Santa Catarina show greater reliance on credit cards. This suggests that financial infrastructure, trust, or access to banking tools may vary significantly across regions.

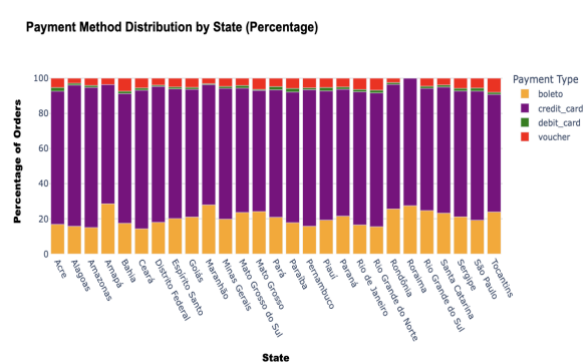


Figure 6. Regional breakdown of payment methods by state.

When it comes to how purchases are paid for, single payments are the norm. As shown in Figure 7, most orders were completed without using installments, even though Brazil is known for offering flexible financing options. However, the long tail of the bar chart reveals that many customers take advantage of installment plans—some stretching up to 24 payments.

Number of Orders by Payment Installment Type

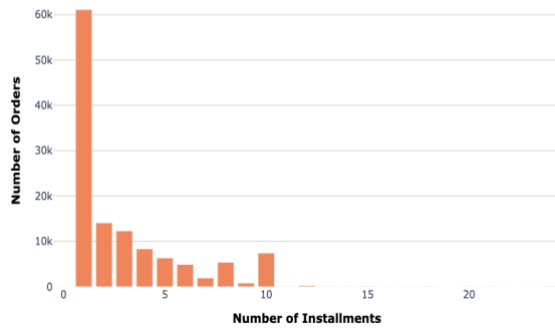


Figure 7. Distribution of installment counts per order.

These installment behaviors vary by state. In Figure 8, states like Paraíba and Alagoas average nearly 4 installments per purchase, while São Paulo and the Federal District tend toward just under 3. This indicates that installment usage may reflect both economic necessity and regional culture.

Average Number of Payment Installments by State

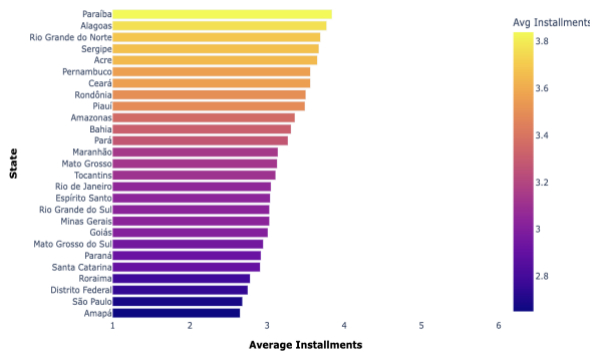


Figure 8. Average number of installments by state.

Interestingly, Figure 10 complicates the story: higher payment values do not necessarily lead to more installments. One might expect expensive items to be split across more payments, but the scatterplot reveals that many high-value purchases are paid in full. This hints at behavioral differences—some consumers may prefer financial flexibility, while others pay upfront regardless of cost.

Payment Value vs. Number of Installments

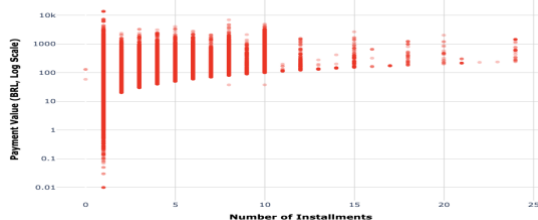


Figure 10. Scatterplot of payment value versus number of installments.

Lastly, Figure 11 shows how payment methods have changed over time. Credit card usage grew sharply through 2017, peaking around major sales events like Black Friday. Meanwhile, boleto and other methods remained relatively flat. These seasonal spikes and trends highlight how promotional periods influence not just what people buy, but how they choose to pay.

Monthly Contribution of Each Payment Method (2016–2018)

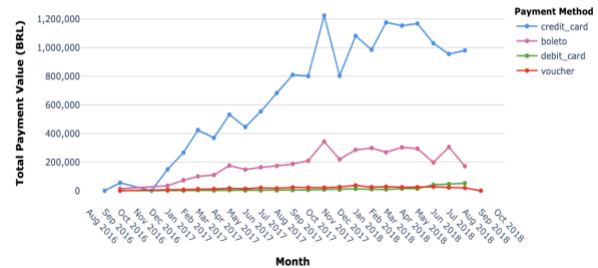


Figure 11. Monthly trend of payment methods.

Together, these visualizations illustrate a diverse and dynamic online payment landscape in Brazil—one shaped by geography, purchasing habits, and shifting consumer preferences.

4.3 Delivery Matters

Delivery performance is a crucial part of customer satisfaction in Brazilian e-commerce. While many orders arrive early, delivery experiences vary widely by region, seller, and product category.

Figure 12 shows the average number of days orders arrived early across Brazilian states. Northern states such as Amapá, Rondônia, and Amazonas lead with the earliest deliveries—arriving nearly 18 to 20 days before the estimated date. In contrast, states like Alagoas and Maranhão tend to deliver just 9 to 10 days early. This variation reveals how geography and logistics infrastructure influence delivery speed, often defying expectations.

How Early Are Orders Delivered? (By State)

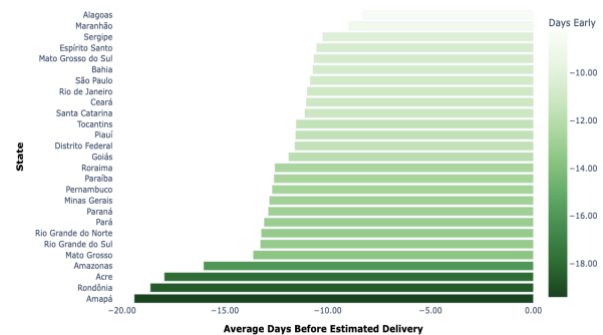


Figure 12. Average number of days orders arrived early by state.

Seller-level performance also plays a significant role. As shown in Figure 13, the top 30 sellers consistently deliver well before the estimated date, with early deliveries ranging from 9 to over 16 days. This level of consistency highlights the impact of logistics efficiency and seller reliability on customer experience.

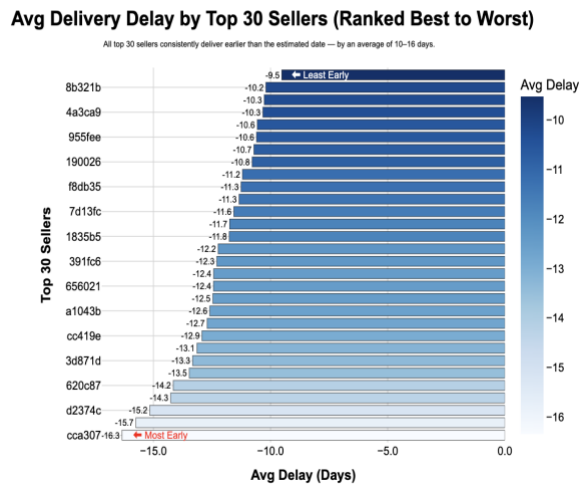


Figure 13. Average delivery delay by top 30 sellers.

Product categories also influence delivery outcomes. Figure 14 highlights the categories with the earliest deliveries on average. Items like *la_cuisine*, *security_and_services*, and *fashion_sport* consistently arrive 12–15 days ahead of schedule. These early arrivals may reflect efficient handling, low demand volatility, or favorable fulfillment processes.

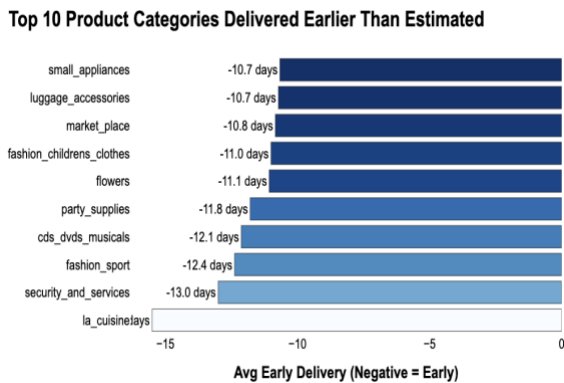


Figure 14. Top 10 product categories delivered earlier than estimated.

To understand delivery timing overall, Figure 15 presents the distribution of delays. While the majority of orders arrive on time or early, there is a long right-skewed tail—representing substantial delays that affect a smaller portion of orders.

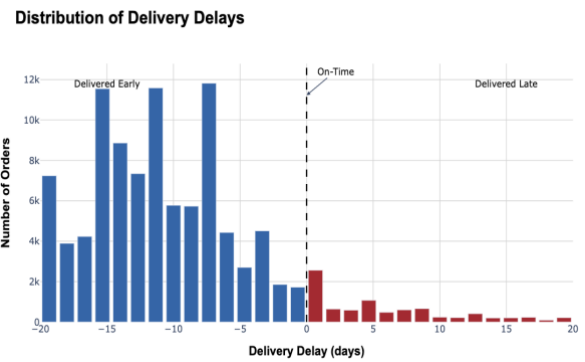


Figure 15. Distribution of delivery delays.

For deeper insight into delivery anomalies, Figure 16 separates typical orders from outliers. The left panel shows the majority of orders, which are delivered within 0–40 days and have delays between –20 and +20 days. The right panel highlights outlier orders that fall outside these thresholds—either arriving extremely early or late or taking over 40 days to be delivered. These outliers represent rare but severe disruptions in the delivery process.

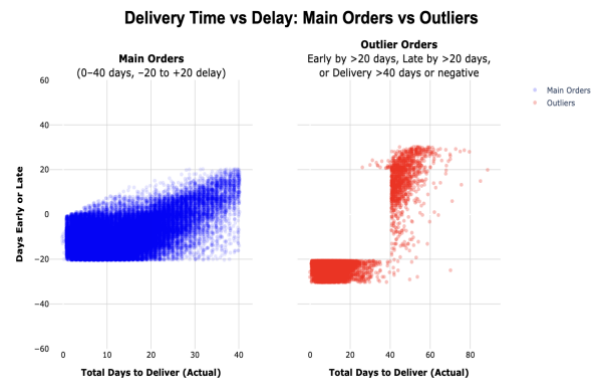


Figure 16. Scatterplot of delivery time vs. delay.

Finally, **Figure 17** incorporates both time and product category dimensions, showing the average delivery delay across months for each category. Several categories exhibit seasonal variability. For instance, *computers_accessories* shows significant spikes in delays—particularly in late 2016 and early 2017. Other categories such as *fashion_male_clothing* and *home_appliances_2* also display periodic delays, while some, like *garden_tools*, appear more stable across time. This matrix-like view reveals which categories suffer from timing volatility and may benefit from supply chain interventions.

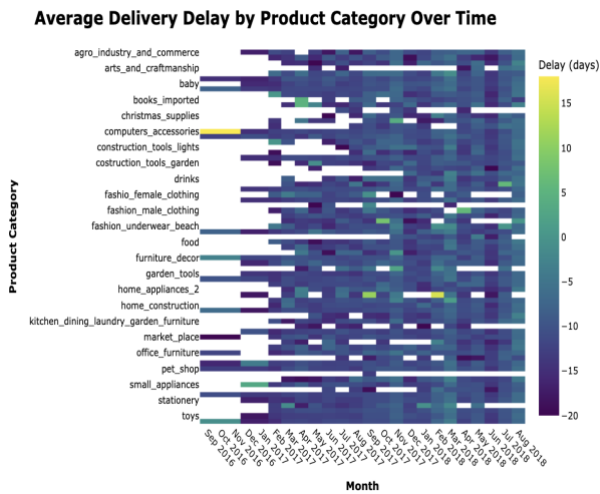


Figure 17. Heatmap of average delivery delay by category and month.

These visualizations highlight the multifaceted nature of e-commerce delivery performance in Brazil. From regional delivery disparities to variations across product categories and time, the data reveals that delivery outcomes are shaped by geographic, temporal, and categorical factors. Such patterns can inform targeted improvements in logistics and fulfillment strategies.

4.4 What are people buying? Are they happy?

The most frequently purchased category by far is bed_bath_table, followed by health_beauty, sports_leisure, and computers_accessories. These four categories dominate the Brazilian e-commerce market. Categories beyond the top five see a significant drop in purchase volume, revealing a highly concentrated demand profile.

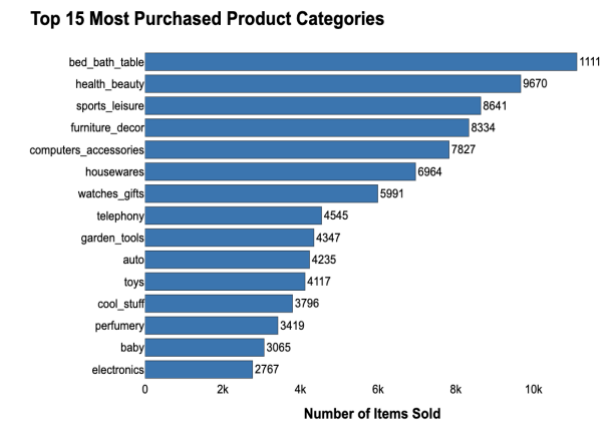


Figure 18. Top 15 most purchased product categories in Brazil.

A Pareto analysis shows that just 10 product categories account for more than 70% of total purchases. This distribution confirms the dominance of a few categories and reflects a long-tail pattern in the remaining product variety.

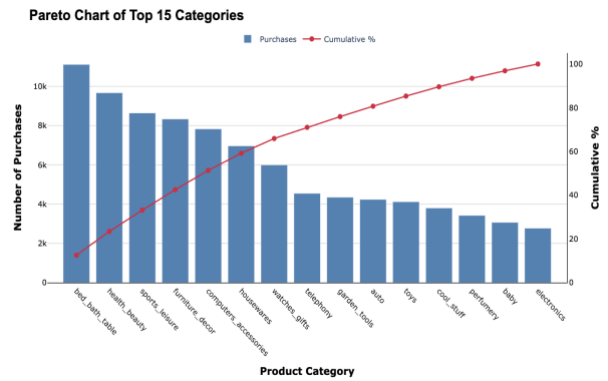


Figure 19. Pareto chart of purchase volume by category.

Figure 20 shows that office_furniture has the longest average delivery time at 20.39 days, significantly higher than all other categories. christmas_supplies and security_and_services follow, each averaging around 15 days. The remaining categories, including fashion_shoes and home_comfort_2, range between 13 and 15 days, indicating moderate but notable delivery delays.

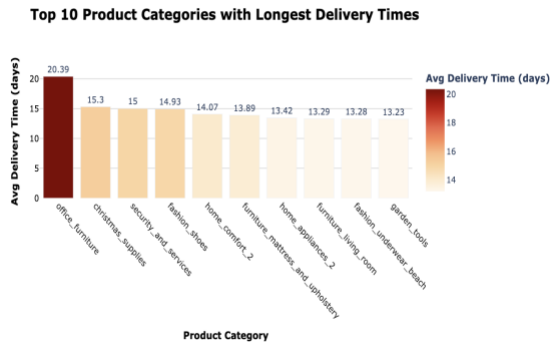


Figure 20. Top 10 product categories with the longest average delivery times.

Fashion_underwear_beach, furniture_decor, and construction_tools_lights have the highest rates of low-rated reviews. These categories may involve issues of fit, assembly, or durability, which are common points of dissatisfaction reflected in customer feedback.

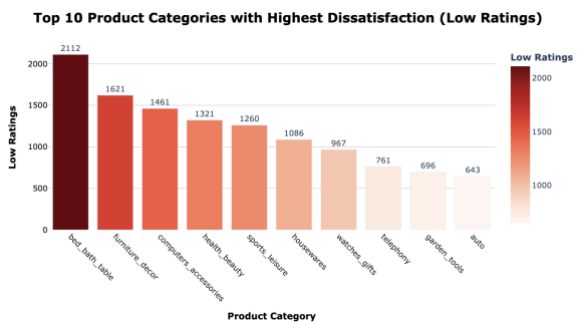


Figure 21. Categories with the highest proportion of negative reviews.

Bed_bath_table maintains consistent volume year-round, with notable peaks in Q4. sports_leisure and toys exhibit stronger seasonality, with demand surging during mid-year and holiday periods. This indicates event-driven or seasonal shopping behavior in specific segments.

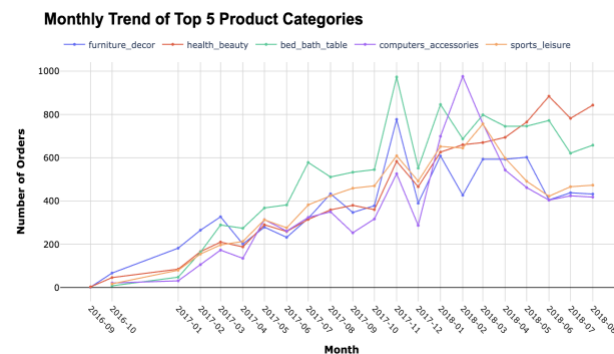


Figure 22. Monthly order trends for the top five categories.

4.5 Voices of the Customers: Text and Sentiment Analysis

While numerical ratings provide a broad measure of satisfaction, analyzing the language used in reviews offers deeper insight into customer sentiment. As shown in Figure 22, review scores are heavily skewed toward 5-star ratings, but this national pattern masks key behavioral differences.

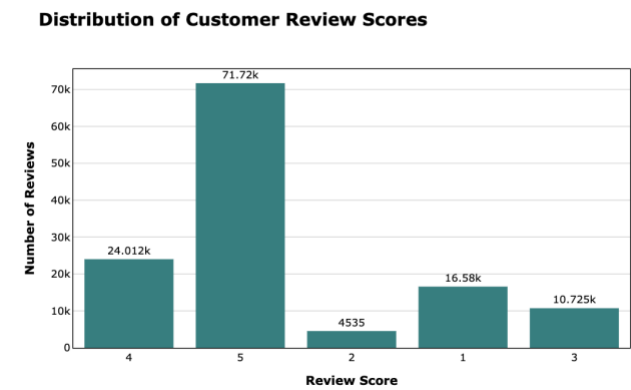


Figure 22. Distribution of customer review scores.

Figure 23 compares new and repeat customers, revealing that repeat buyers are more likely to leave 5-star reviews and less likely to leave 1-star ratings. This suggests that loyalty may be driven by more consistent or satisfactory experiences.

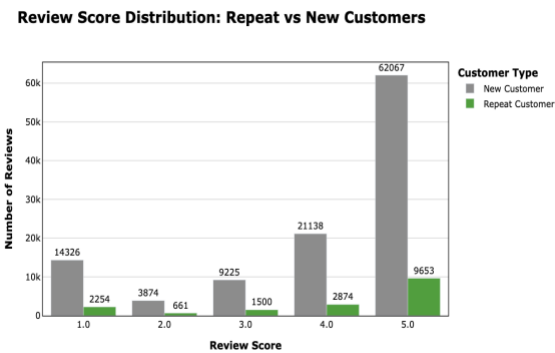


Figure 23. Review score distribution by new vs. repeat customers.

Delivery delays emerge as a common frustration. Figure 24 displays the distribution of delivery times across review scores, showing that lower ratings are associated with longer and more variable delivery windows.

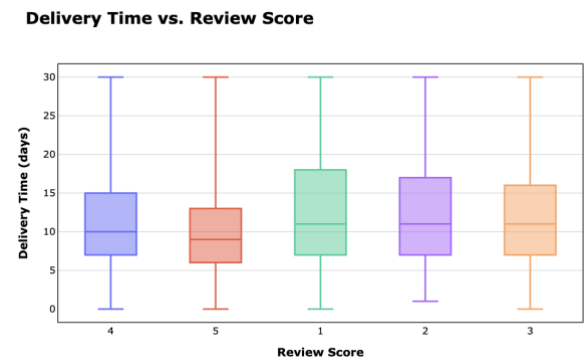


Figure 24. Delivery time distribution across review scores.

To better understand customer sentiment, we applied natural language processing to the review corpus. Figure 25 contrasts the most frequent words in positive and negative reviews using word clouds. Positive reviews frequently reference timely delivery and product satisfaction—for example, “*chegou antes*” (*arrived early*) and “*bem embalado*” (*well packaged*). Conversely, negative reviews often mention trust and fulfillment issues, such as “*não recebi*” (*did not receive*) and “*produto errado*” (*wrong product*).

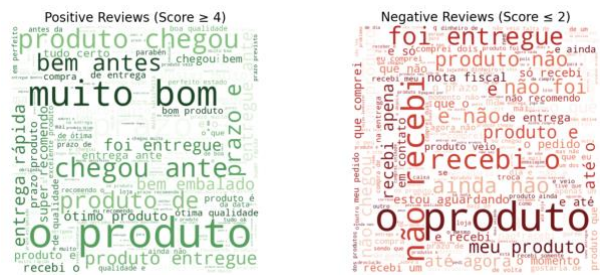


Figure 25. WordClouds comparing positive and negative customer reviews.

This is reinforced in Figure 26, where a bar chart highlights the top 20 words in low-rated reviews, dominated by terms like “não,” “recebi,” and “produto.” These findings suggest that missing, faulty, or incorrect items are key drivers of dissatisfaction.



Figure 26. Top 20 most frequent words in low-rated seller reviews.

Finally, Tables 1 and 2 present translated excerpts from actual customer reviews. These voices reveal not only what went wrong (Table 1), but also what went right (Table 2). While negative reviews highlight emotional frustration, delivery failures, and product issues, positive reviews often focus on satisfaction with prompt service or product quality. Together, these excerpts provide rich, contextual insight that complements the broader patterns observed in the data – offering a more human perspective on customer satisfaction and dissatisfaction.

Portuguese (BR)	English (US)
O produto tem um preço legal, mas não entregaram...	The product has a good price, but they did not deliver ...
Enviaram o produto errado pra mim!	They sent the wrong product to me!
RECEBI UM PRODUTO F-A-L-S-I-F-I-C-A-D-O. Ñ SABE...	I received a F-A-K-E product. Don't know...
O relógio chegou sem funcionar, e não é original...	The watch arrived non-functional and is not original...
Recebi um relógio falsificado. Vou devolver e q...	I received a counterfeit watch. I will return and...

Table 1. Translated excerpts from negative customer reviews

Portuguese (BR)	English (US)
Um verdadeiro absurdo hj já são 09 de agosto e ...	A real nonsense, today is already August 9 and ...
Comprei a base dermacol cor 207, pela cor mostr...	I bought the Dermacol foundation color 207, by the color shown...
O cartão sd apresentou problemas e não consigo ...	The SD card had problems and I can't ...
As toalhas muito ruins , finas	The towels are very bad, thin

Portuguese (BR)	English (US)
Já faz algum tempo que a compra ocorreu e ainda...	It's been a while since the purchase occurred and still...

Table 2. Translated excerpts from positive customer reviews.

5 CONCLUSION AND REFLECTIONS

This project provides a multifaceted view of Brazil’s growing e-commerce ecosystem through visual exploration of the Olist dataset. Our analysis reveals important regional, behavioral, and operational patterns shaping online retail experiences from 2016 to 2018. While São Paulo dominates in overall volume and infrastructure, states like Paraíba and Espírito Santo show high engagement when normalized for population—challenging assumptions about digital exclusion in less-developed areas.

Payment behavior shows strong national reliance on credit cards, but regional variations persist, particularly in boleto usage. Installment practices also differ by state, reflecting financial culture and consumer flexibility rather than just price sensitivity. Delivery performance emerges as a key driver of satisfaction, with delays varying widely by region, seller, and product type. Interestingly, while many orders arrive ahead of schedule, a long tail of severe delays undermines consistency.

Our sentiment and text analysis reinforces the importance of delivery reliability and product accuracy. Most reviews are positive, but when issues occur, they are intensely emotional—often tied to failed expectations around delivery, faulty items, or poor service. These insights are especially valuable for improving seller accountability and targeting infrastructure gaps.

More broadly, this study highlights how interactive visualizations can uncover hidden dynamics and support evidence-based decisions. Future work could integrate more recent data, explore emerging payment systems like Pix, or build predictive models to anticipate dissatisfaction. As Brazil’s e-commerce landscape continues to evolve, understanding regional patterns and customer voices will remain essential for equitable and efficient digital commerce.

6 REFERENCES

[1] C. Pereira da Veiga, D. Ribeiro, F. Matos, and M. Lima, “E-Commerce in Brazil: An in-depth analysis of digital growth and strategic approaches for online retail,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, no. 2, pp. 1559–1580, 2024. doi: 10.3390/jtaer19020076

[2] U.S. International Trade Administration, “Brazil – eCommerce,” *Country Commercial Guide*, Dec. 2023. [Online]. Available: <https://www.trade.gov/country-commercial-guides/brazil-ecommerce>

[3] Wired Staff, “Brazil Dreams of Favela 3D,” *Wired*, Aug. 2023. [Online]. Available: <https://www.wired.com/story/brazil-dreams-favela-3d/>

[4] Reuters Staff, “Brazil’s Pix to overtake credit cards in e-commerce as soon as 2025,” *Reuters*, Sep. 2024. [Online]. Available: <https://www.reuters.com/business/finance/brazils-pix-overtake-credit-cards-e-commerce-soon-2025-study-shows-2024-09-09/>

[5] Research and Markets, “Brazil Buy Now Pay Later Business Report 2025,” *BusinessWire*, Feb. 2025. [Online]. Available: <https://www.businesswire.com/news/home/20250228118181/en/>

- [6] R. Fonseca and A. Matray, "Financial inclusion, credit, and savings: Evidence from Brazil," *Princeton GCEP Working Paper*, no. 308, Sept. 2022. [Online]. Available: https://gceps.princeton.edu/wp-content/uploads/2023/03/wp308_FonsecaMatray_Brazil_unbanked.pdf
- [7] N. N. Gomes and M. L. R. de Oliveira, "Artificial intelligence applied to assess perceptions of e-commerce logistics quality: A case study in Brazil," *Brazilian Business Review*, vol. 19, no. 3, pp. 250–269, 2022. doi: 10.15728/bbr.2022.19.3.4
- [8] R. H. M. Oliveira, A. N. Silva, and J. M. L. Carvalho, "Determinants of e-commerce logistics service quality in emerging markets," *Sustainability*, vol. 16, no. 11, pp. 1–24, 2024. doi: 10.3390/su16114743
- [9] A. Silva, L. G. Nascimento, and L. R. T. de Freitas, "Embedding generation for text classification of Brazilian Portuguese user reviews: From bag-of-words to transformers," *arXiv preprint*, arXiv:2212.00587, 2022. [Online]. Available: <https://arxiv.org/abs/2212.00587>
- [10] N. Shrestha and F. Nasoz, "Deep learning sentiment analysis of Amazon.com reviews and ratings," *arXiv preprint*, arXiv:1904.04096, 2019. [Online]. Available: <https://arxiv.org/abs/1904.04096>
- [11] J. Heer, B. Shneiderman, and C. Stolte, "A taxonomy of visualization techniques," *ACM Queue*, vol. 4, no. 2, pp. 1–24, Mar. 2006.