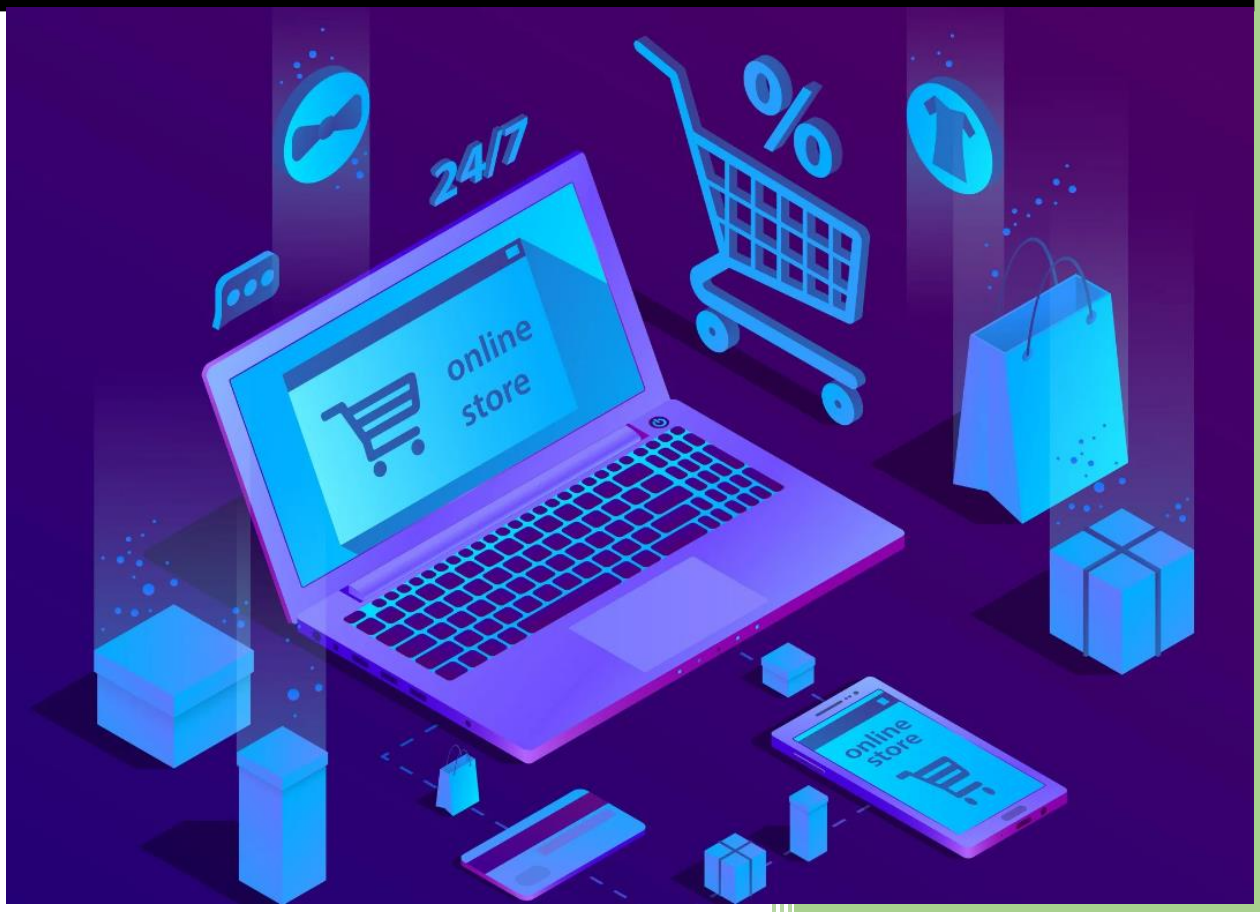
One Mount – Hieu Do

E-Commerce Data Analysis



Contents

Section 1: Data Exploration and Analysis

This section aims to explore and analyze the Brazilian dataset as deep as possible.

In general, this dataset pertains to Brazilian e-commerce transactions. Comprising data from 2016 to 2018, it encompasses details of 100,000 orders placed across various marketplaces in Brazil. Its comprehensive features enable the examination of orders from diverse dimensions, including order status, pricing, payment and shipping performance, customer location, product attributes, and customer reviews. Additionally, we have provided a geolocation dataset that maps Brazilian zip codes to corresponding latitude and longitude coordinates. For each of these dimensions, there will be a corresponding analysis, which can be found in the very next section.

It's important to note as a disclaimer that this report prioritizes the presentation of key findings derived from the Exploratory Data Analysis (EDA) procedure. Consequently, detailed information regarding data processing or manipulation is omitted herein. For a comprehensive examination of such aspects, readers are encouraged to refer to the accompanying "HieuDoDuc – One Mount – Technical Work." Additionally, it's worth highlighting that the exploration of this dataset has been strategically tailored to support the objectives to maximize Gross Merchandise Value and optimize spending.

With all that in mind, we will proceed to the analysis's findings:

1. Product Analysis

The product dataset comprises approximately 32,000 rows, each representing a distinct product categorized into one of 72 categories. Through a left join operation with tables related to orders, we can obtain a holistic perspective of the selling performance of these products over the study period. Figure 1 illustrates this performance:

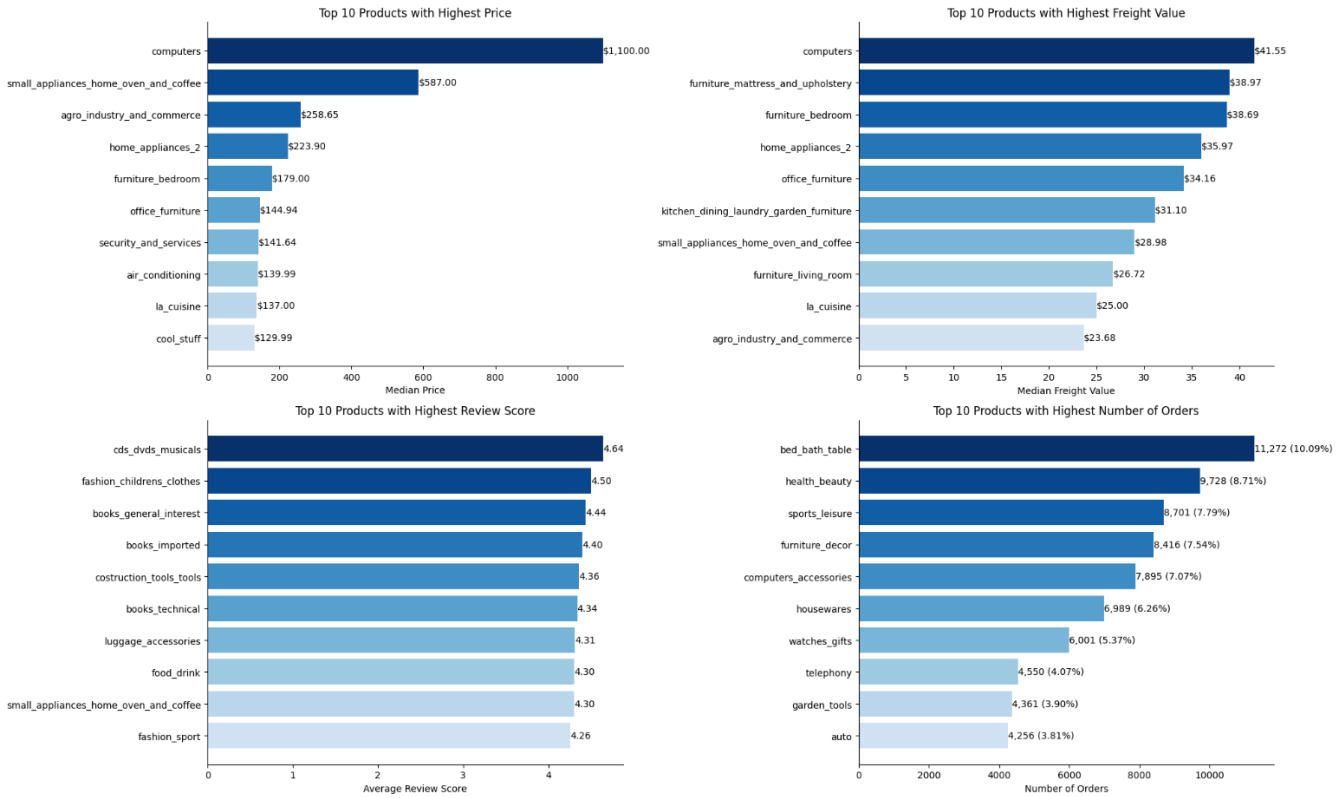


Figure 1: Product Analysis by Price, Freight Value, Customer Review and Number of Orders

In general, Computers stand out as both the most expensive and having the highest freight value. Moreover, the notably higher freight value for computers compared to heavier items like mattresses and furniture indicating potential challenges in production and logistics. Possible reasons for this difference may include inefficient shipping methods, long shipping distances due to manufacturing location, or other suboptimal insurance strategies.

On a positive note, small appliances such as home ovens and coffee makers are performing exceptionally well. Despite their costs, these products are consumer favorites, highlighting their value and quality. We should aim to capitalize on this success by refining pricing strategies and logistics to drive growth and strengthen our position in the small appliance market.

Besides, our top-selling products span a wide range of consumer interests, underscoring the importance of catering to diverse preferences for long-term market relevance.

Zooming out, it can be observed that 7 out of 10 most expensive product also appear in top 10 product with highest freight value. As correlations between these product attributes is suspected, a correlation analysis will be adopted next:

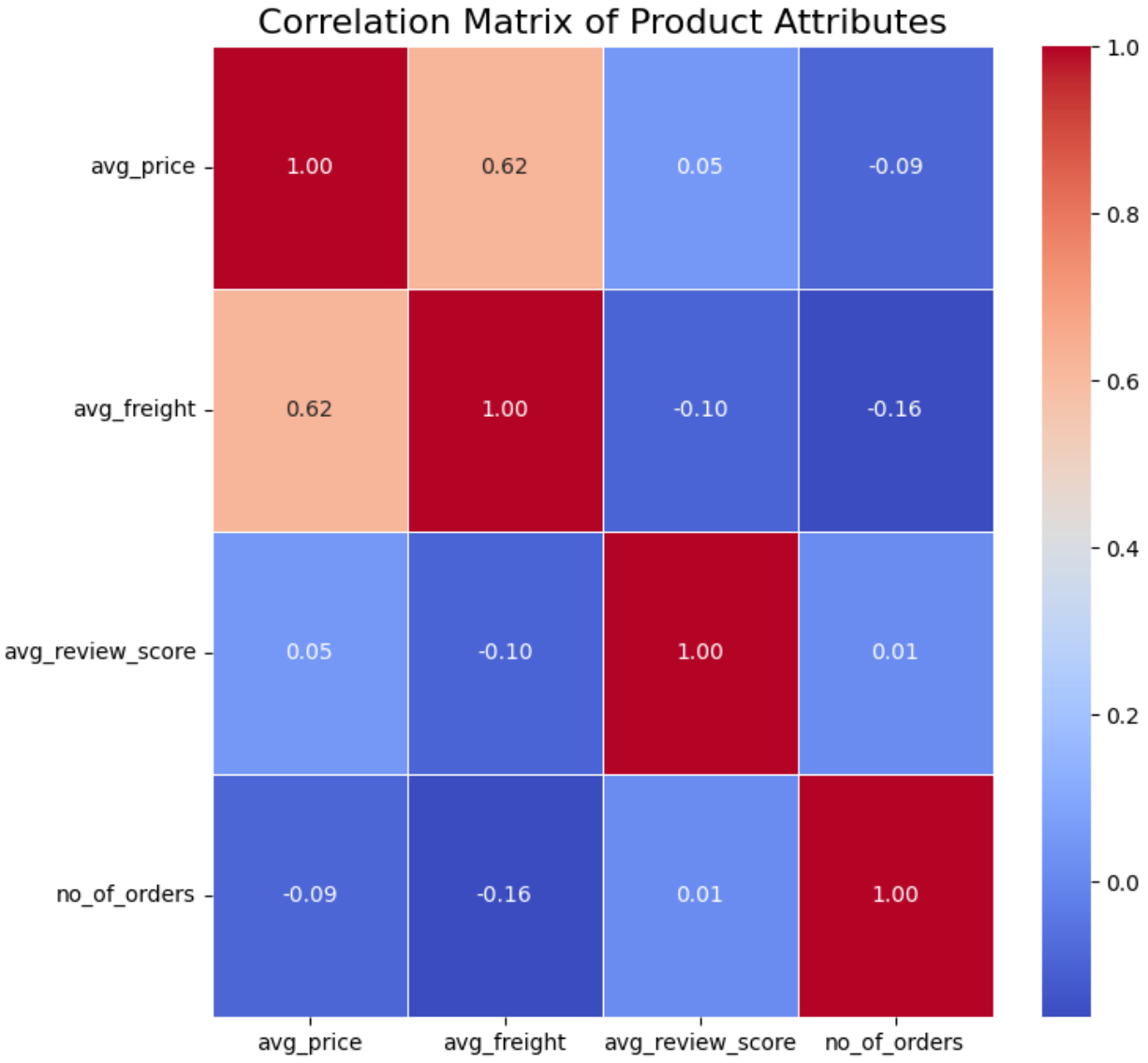


Figure 2: Correlation matrix between products attributes.

As expected, there is a strong correlation between product price and their freight value. Higher freight costs can contribute to higher overall product costs, particularly if those costs are passed on to consumers. For instance, the unexpectedly high freight value for computers may pose challenges in reaching customers at competitive prices.

Interestingly, our analysis reveals that neither price nor freight cost significantly impacts the review scores of products. This suggests that while freight costs are an essential consideration for businesses, they may not directly influence customer satisfaction as reflected in review scores.

Regarding customer reviews, it witnessed diverse range of product types being popular among customers, which is a positive indicator of market demand. However, optimizing our logistics procedures to lower freight and total product costs for high-value items can further enhance profitability.

2. Delivery Analysis

As its name has stood for, this analysis focusing on the delivery performance of product to customer. There are 7 different delivery status: Delivered, Shipped, Invoiced, Processing, Canceled, Unavailable, Approved. Although there are many directions to investigate delivery performance, such as proportion of delivery status; we would only investigate the performance of delivery with status is Delivered, as they are most direct contributor to the GMV of the section 2.

Keep that in mind, orders dataset was imported and give us the performance of delivered status, specifically if they were delivered on time or be late:

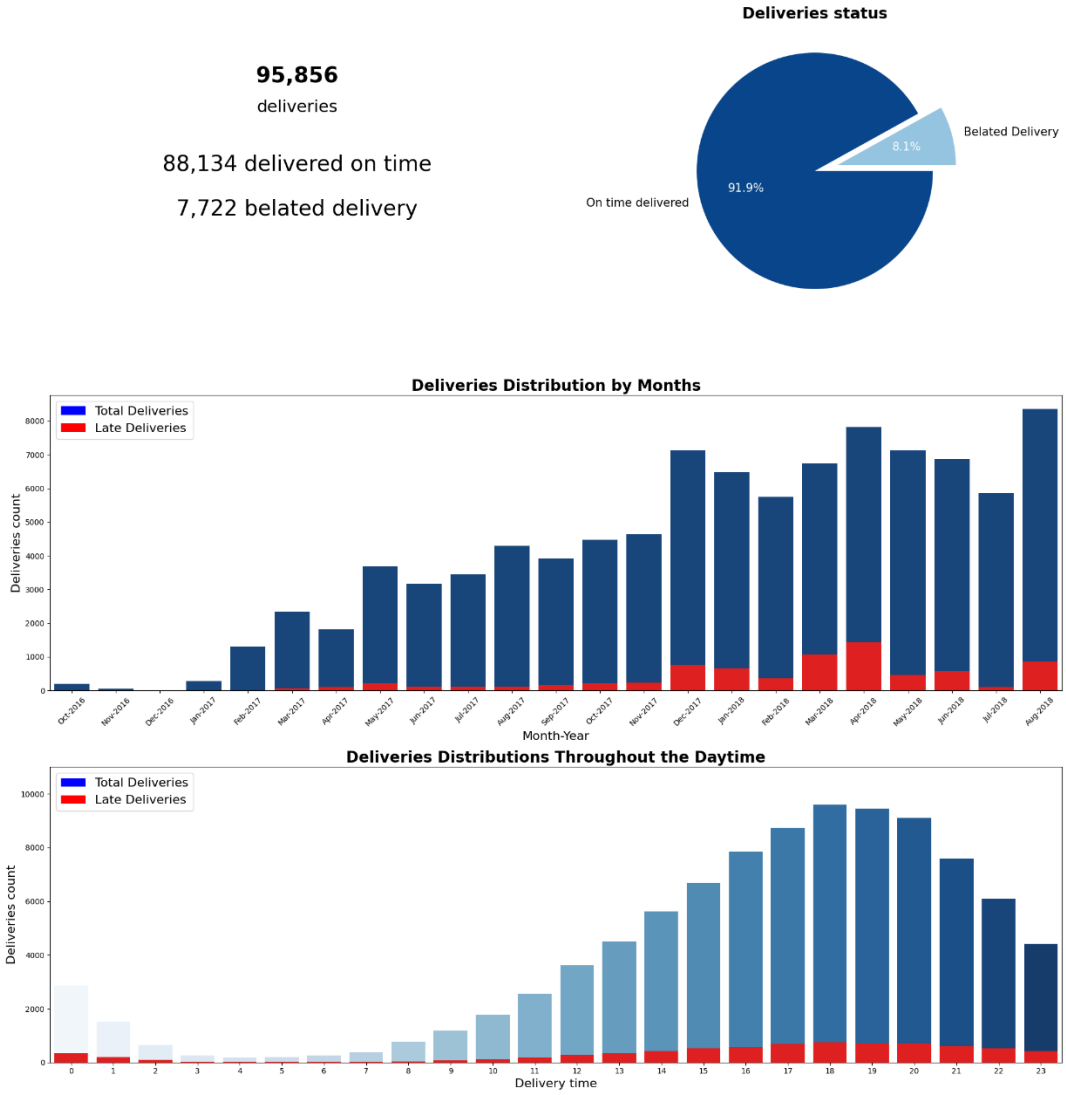


Figure 3: Delivery Performance Analysis: On-Time vs. Late Deliveries

In the period from October 2016 to September 2018, *95,856* deliveries were accomplished. Of these, the on-time delivery rate is *91.9% (88,134* orders), while *8.1%* of orders (*7,722* orders) experienced delays.

Analysis shows a consistent upward trend in total deliveries, particularly during peak demand periods such as December 2017 to January 2018 (Christmas), April 2018 (End of Semester), and August (back-to-school season). However, these periods also saw a higher incidence of late deliveries.

Further examination of order submission times revealed a gradual increase from sunrise to a peak at 18:00, followed by a decline.

In summary, while on-time delivery performance is commendable, there are opportunities for improvement, especially during peak demand periods. Addressing these challenges will bolster customer satisfaction and company's market position.

3. Payment Analysis

This analysis delves into customer preferences regarding their preferred payment methods. Further details are provided below:

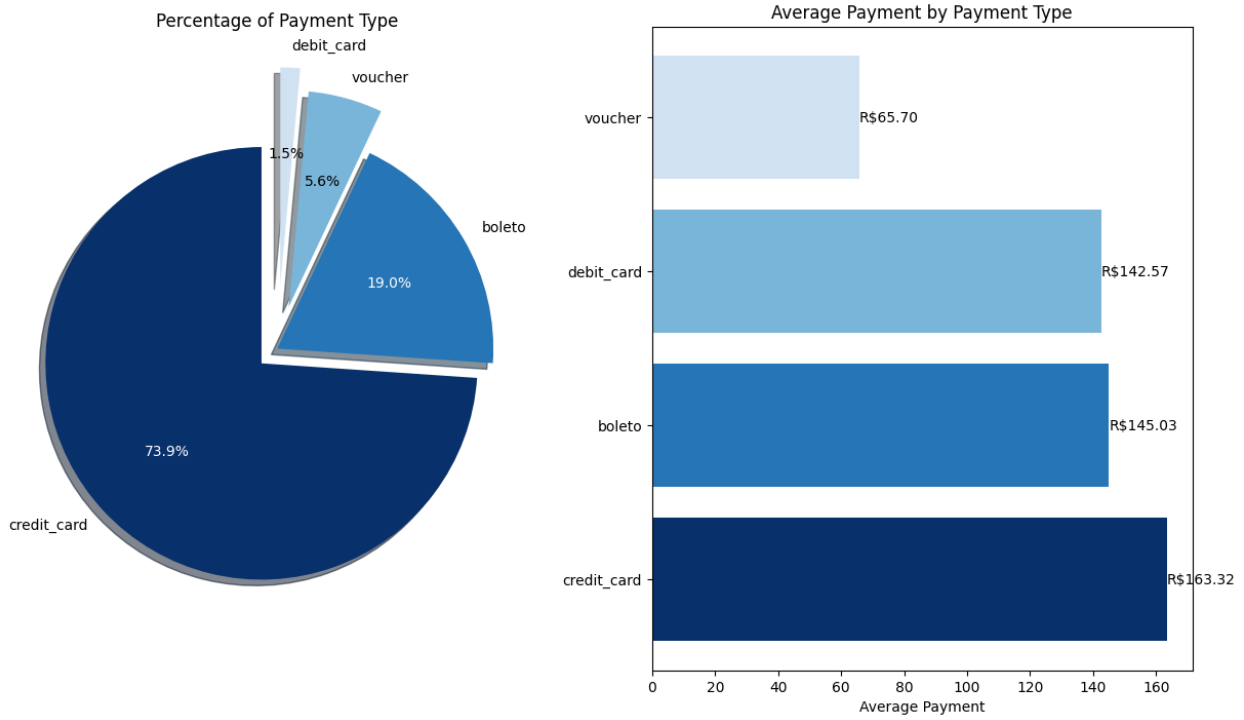


Figure 4: Payment Performance

The analysis indicates that *Credit Cards* dominate payments, comprising 73.9% of transactions, with *Boleto* as the second, scores way lower with only 19%. On the other hand, average *Credit card* payments amount to 163.3 R$, surpassing *Boleto*'s 145.0 R$. This underscores customer preference for *Credit Cards* and their inclination towards higher-value transactions compared to other payment method.

Interestingly, *Boleto* and *Debit Card* transaction values are comparable, suggesting factors beyond product price influence payment choice.

Finally, given the dominance of *Credit Cards* and their higher transaction values, we can brainstorm some several strategies to maximize Gross Merchandise Value (GMV), such as tailored marketing campaign to credit card users or optimizing credit card payment process. All those strategies will contribute to attract more customer due to the simplicity of the company's payment system. Additionally, the company can also explore the factors driving *Boleto* and *Debit Card* transactions to optimize payment methods and enhance customer experience.

4. Geolocation Analysis

This analysis illustrates the geographical distribution discrepancy between customer versus sellers. This understanding can inform decisions related to logistics, marketing strategies, and resource allocation to better align with the market landscape and improve overall operational efficiency.

Figure 5: Geographical distribution discrepancy between customers and sellers.

Based on above geospatial analysis result, it's vital to note that there's an imbalance in the distribution of customers and sellers. While both groups are predominantly clustered in major cities of southeastern Brazil like Curitiba, Sao Paulo, and notably Rio De Janeiro, sellers are primarily concentrated in this area, whereas customers extend to eastern regions such as Salvador and Recife. This indicates potentially high freight costs for customers due to the scarcity of sellers in their vicinity.

As a recommendation to address this issue, we should forge partnerships with more sellers in the east and northeast, particularly in key urban centers. Concurrently, developing marketing strategies to penetrate the western region would be advantageous. By expanding our seller network and strategically targeting new markets, we can mitigate shipping expenses and enhance accessibility for customers across Brazil.

5. Orders and GMV Analysis

This part examines the sales performance, specifically analyzing the monthly/daily distribution of total orders and their corresponding GMV. Please take this note as disclaimer that, within this analysis, the formula of calculating GMV is:

Gross Merchandise Value = (Sales Price of the products + Freight value) x Number of products sold

The reason for this is that there is no declaration within the dataset documentation that price of products has covered freight value of that product or not.

With that in mind, below Figure 6 illustrates the sales performance:

99,421 Orders

were made from Sep 2016 to Aug 2018

With total purchase of

15,843,387 R$

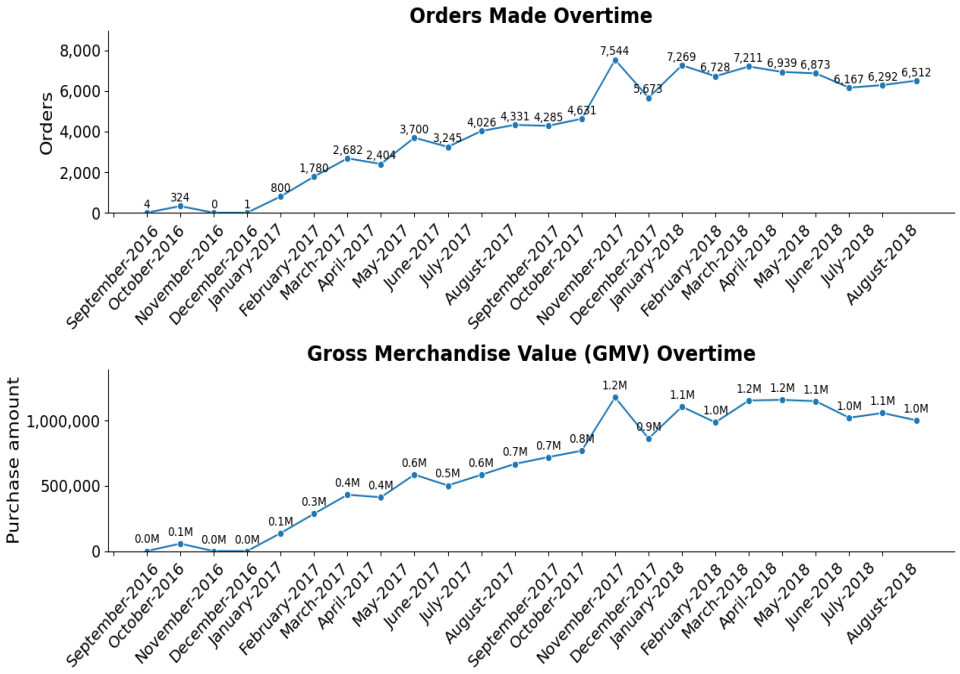


Figure 6: Order/Sales performances.

Based on the analysis depicted in the provided figure, it is evident that there has been a consistent upward trend observed over the study period in both the Number of Orders and the Gross Merchandise Value (GMV). This trend indicates a progressive increase in business activity over time. Moreover, since the conclusion of 2017, the observed trend has stabilized, with minimal significant fluctuations noted in both metrics.

Considering these findings, the next step involves implementing seasonal analysis techniques to glean additional insights. By examining potential patterns in the evolution of GMV over time, we aim to uncover further actionable intelligence that may contribute to strategic decision-making processes.

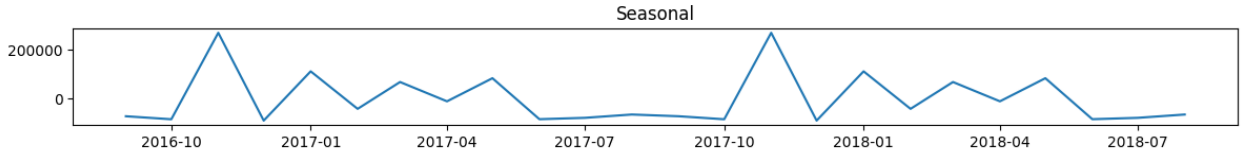


Figure 7: Seasonal analysis of GMV.

The seasonal decomposition of the GMV data reveals distinct patterns throughout the observed period. Notably, there is a noticeable spike in November, likely attributable to holiday sales and promotions. This is followed by a decrease in December as holiday shopping subsides, with a subsequent rebound in January. Further, fluctuations in GMV persist throughout the year, indicating ongoing variability influenced by factors such as economic status and e-commerce industry-specific trends.

As the next action, which can be deprived from our findings, to maximize GMV, series of targeted strategies to capitalize on peak period and mitigate fluctuation as below for example:

- Develop targeted marketing campaigns and promotions to leverage the spike in GMV observed in November, ensuring visibility during peak shopping seasons and capitalizing on holiday-driven consumer spending.
- Introduce seasonal products or exclusive deals tailored to the observed trends, aligning inventory with anticipated demand fluctuations to meet customer needs and preferences.
- Implement personalized marketing strategies and loyalty programs to drive repeat purchases and maintain engagement throughout the year, leveraging insights from seasonal patterns to deliver relevant and timely messaging.
- Explore opportunities to expand sales channels or partnerships to reach broader audiences and mitigate the impact of seasonal fluctuations, such as tapping into new markets or optimizing online sales platforms to capture off-peak demand.

6. Repeat Customer Analysis

Within the last analysis of this report, two types of customers will be adopted for a comparison of their performance: Repeat and One-Time customer.

Result of this analysis can be illustrated in Figure 8 below:

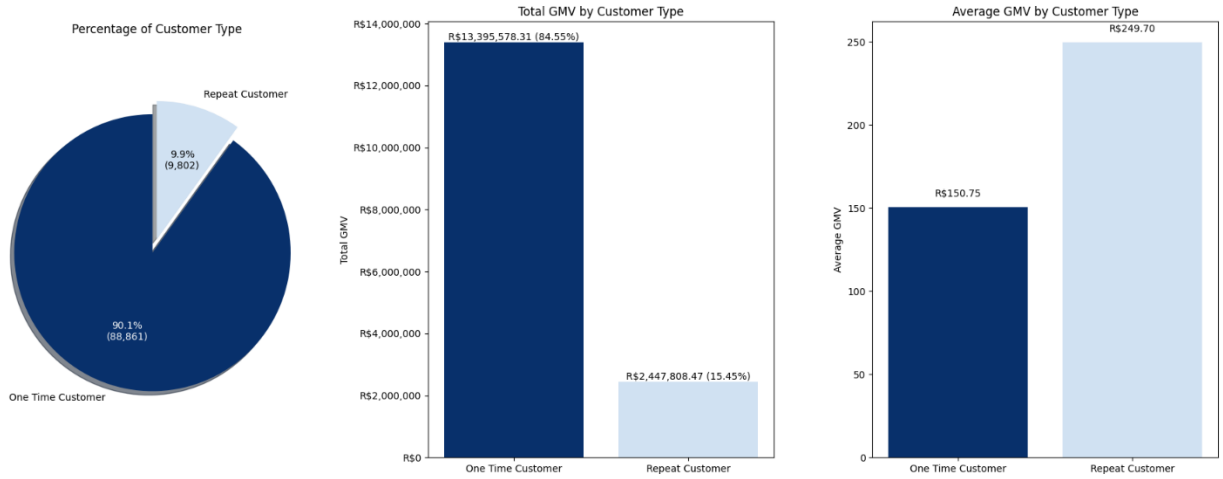


Figure 8: Repeat versus One-time Customer

The repeat customer reveals a significant contrast in revenue contribution between repeat and one-time customers. Despite comprising only 9.9% of our customer base, repeat customers contribute a substantial 15.45% to our total Gross Merchandise Volume (GMV). Conversely, most of buyers, accounting for 90.1%, are one-time customers, yet they contribute 84.55% of GMV.

Moreover, delving deeper into the data, we observe that repeat customers exhibit a higher average order value compared to one-time customers. On average, repeat customers spend 249.7 R$ per transaction, while one-time customers spend 150.75 R$.

These findings underscore the critical importance of prioritizing strategies aimed at retaining repeat customers. Not only do they contribute significantly to our revenue stream, but their propensity to spend more per transaction further emphasizes their economic value to the business. Therefore, retaining repeat customers is not only vital for fostering customer loyalty but also directly correlates with higher GMV. Repeat customers tend to demonstrate higher purchase frequency, larger average order

values, and lower acquisition costs. By focusing on enhancing customer experiences, incentivizing repeat purchases, and building lasting relationships, we can ensure a stable revenue stream and drive sustainable growth while capitalizing on the higher lifetime value that repeat customers inherently offer.

Section 2: Business Acumen

As a requirement of this section, the goal is to investigate series of metrics to evaluation company performance on how to maximize GMV and optimize spending.

Before moving to that part, first we need to revise the formula of GMV, as has been introduced in Orders and GMV Analysis section as below:

1. Formula Breakdown

Gross Merchandise Value = (Sales Price of the products+Freight value) x Number of products sold ***(1)***

As can be seen, GMV is constructed by 3 components:

- ***Sales Price of the products:*** The sales price of our products is predominantly determined by sellers rather than being directly controlled by the company. Insights gleaned from our Product Analysis reveal week correlations between average price and the volume of orders. This finding contradicts the assumption that price adjustments alone significantly impact sales. In fact, the week correlation would mean customers would willing to pay for the production, if the price is reasonable compared to the quality of that product. Furthermore, Figure 9 provides visual clarity on these relationships:

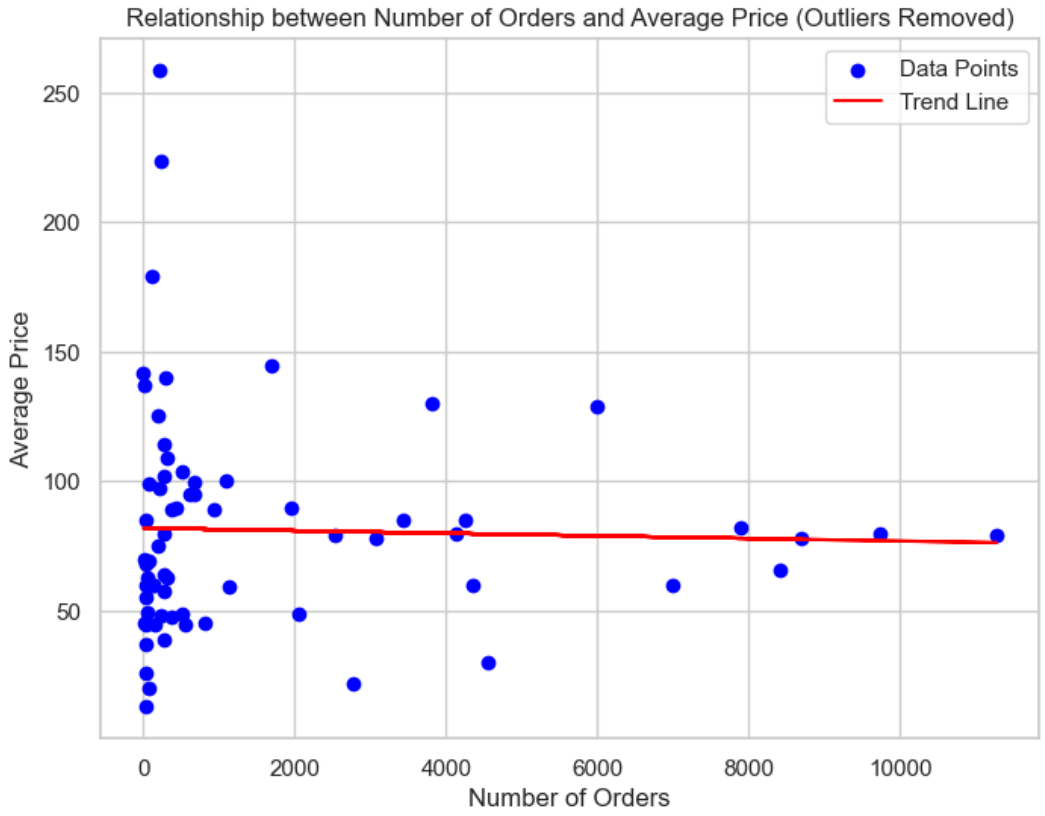


Figure 9: Relationship between No. of Order and Avg Price

Hence, despite its role as the primary driver of Gross Merchandise Value (GMV), our analysis advises investment in price management or dynamic pricing initiatives at this juncture. However, to implement these strategies effectively, it is imperative to gather additional data, which will be elaborated later within next section.

- ***Freight value:*** Gross merchandise value (GMV) is a metric that measures the total value of goods sold on our company platform, representing the full amount paid by customers before any deductions such as fees, discounts or returns. Meanwhile, freight value is the monetary cost associated with shipping and delivery services that will be paid by the customers.

  One may argue that, according to Formular (1) above, it suggests that both the sales price of products and the freight value influence the overall GMV positively. However, the relationship between freight value and GMV depends on various factors such as the pricing strategy.

  For example, if a company increases the freight value without a corresponding increase in sales price or product demand, it could potentially lead to a decrease in GMV as higher shipping costs may deter customers from making purchases. Conversely, optimizing freight value by reducing shipping costs or offering free shipping promotions could potentially lead to increased sales volume and thus a higher GMV.

Therefore, while freight value is a component of GMV, its impact on overall GMV depends on how it is managed in conjunction with other factors affecting sales and pricing strategies.

- ***Number of products sold:*** This metric stands out as the most transparent Key Performance Indicator (KPI) for gauging the company's performance in relation to GMV, besides the GMV itself. The number of products sold, synonymous with the volume of orders, serves as a direct reflection of customer activity and engagement on the ecommerce platform. A surge in sales volume indicates heightened market demand for the available products, thereby potentially amplifying the GMV. In Orders and GMV Analysis, we have specifically seen the correlation between number of products sold and GMV.

2. Measures Selections

Now we have the constructive components behind GMV, we can start designing list of metrics that can help us reflect the company performance regarding this specific. These measures can be categorized into direct and indirect:

Direct Measures:

These are quantifiable measures directly related to the performance of our company's GMV:

- **GMV (Gross Merchandise Value):** The total value of goods sold over a specific period. See Orders and GMV Analysis.
- **Number of Products Sold**: The total quantity of products sold within a given timeframe. See Orders and GMV Analysis.
- **Repeat Purchase Rate:** The percentage of customers who return to make additional purchases. See Repeat Customer Analysis to see importance of customer retention strategies in maximizing GMV.

In-direct Measures:

These are measures that indirectly impact or reflect the performance of our company regarding GMV. In the given context, the list of in-direct measures are selected because they reflect the performance of the logistics procedure within the company, which will then affect the customer satisfaction and eventually GMV:

- **Freight Value:** As has been explained in previous section, in spite of appearing in the formula of the GMV, it would be very difficult to interpret directly the impact of Freight Value against GMV. However, as can be seen from Product Analysis, this measure still gives us indicator of potential logistics challenges, such as the unexpectedly high freight cost for Computer. Together with Geolocation Analysis, Freight Value will be a KPI that helps logistics department debug their on-going issues and plan for upcoming improvements of a better spending.
- **On-Time Delivery Rate**: This metric measures the percentage of deliveries that are completed on time. A higher on-time delivery rate indicates efficient logistics and fulfillment processes, leading to greater customer satisfaction and potentially higher GMV. See Delivery Analysis.

3. Next step and additional data suggestion

The dataset provided offers a robust foundation for comprehensive analysis, particularly in assessing the company's performance, notably in terms of Gross Merchandise Volume (GMV). An avenue for further exploration involves leveraging key customer metrics such as recency, frequency, and monetary value (RFM) to construct a segmentation model. By employing this approach, we can discern customer segments and ascertain metrics like Churn Rate, providing valuable insights for our Marketing Team to tailor strategies towards personalized engagement, thereby enhancing effectiveness over generalized approaches.

Furthermore, by delving into historical transactional data, we can implement a Recommender System to anticipate and proactively suggest products aligning with individual customer preferences. This proactive outreach strategy aims to augment customer satisfaction and drive additional sales.

In considering additional data collection efforts, I propose tracking customer interactions within our e-commerce platform. This enriched dataset would empower data analysts to monitor user behavior encompassing actions such as sign-ups, purchases, and page views. Such granular insights facilitate the quantification of Conversion Rate, indicating the platform's efficacy in converting visitors into active customers. This metric serves as a barometer for assessing the performance of various initiatives spanning marketing endeavors, user experience enhancements, and product offerings, thereby enabling informed decision-making and strategic refinements.

Moreover, a system to trace marketing department effectiveness like salesforce should also be considered. With this, we can link together with conversion rate or repeat customer rate after every new campaign to quantify the performance of the send out campaign.

End of Assessment