Capstone 3: E-Commerce Customer Segmentation Using K-means clustering

Springboard – JUNG, JIYOON

Introduction

In today's competitive eCommerce landscape, effectively capturing and utilizing customer behavior data is essential for optimizing marketing strategies and enhancing the shopping experience. Despite possessing extensive data on customer transactions, Olist requires a more structured approach to dissect and understand the multifaceted nature of its customer base.

This project aims to systematically segment customers based on their purchasing behaviors. Subsequently, a descriptive analysis of each group will be conducted, and marketing strategies will be generated based on their features.

Approach

First, the 'Olist' e-commerce dataset will be loaded and pre-processed for model development. During this process, outliers will be removed, and the dataset will be standardized for clustering.

Second, K-mean clustering will be conducted based on customer recency, frequency, and monetary (RFM) value analysis results. The ideal number of K will be chosen by comparing the Elbow and the Silhouette method.

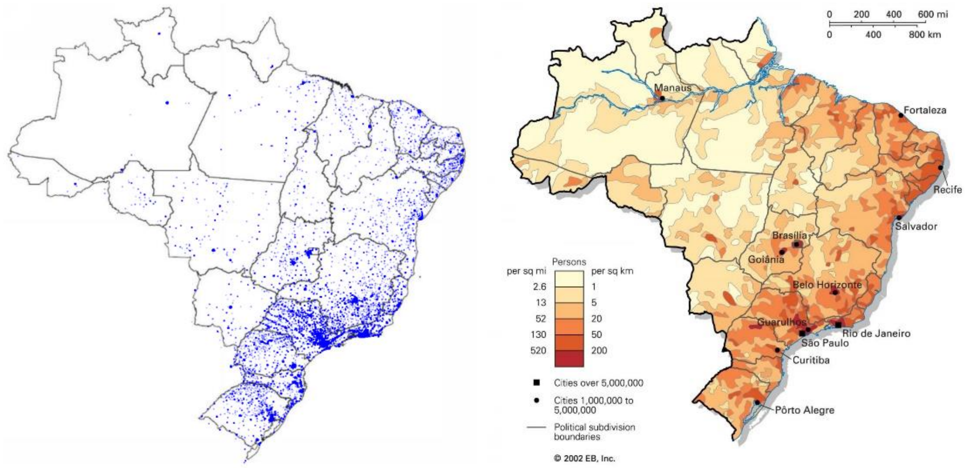
- Recency: days since the last customer transaction
- Frequency: number of transactions in the last 12 months
- Monetary Value: total spending in the last 12 months

Third, a descriptive analysis will be conducted on each segment to find the marketing strategy for each group based on features.

Finding

1. **Customer Zip Code**

   The customer zip codes were scattered and plotted on a map of Brazil. It was observed that most customers are located along the East Coast, with more in southern Brazil than in northern Brazil. When comparing the Olist customer locations with the Brazilian population, it is evident that they align with the population map.

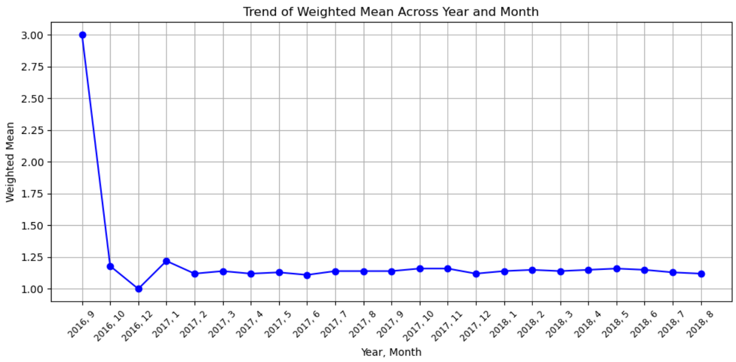


Image, https://maps-brazil.com/img/0/brazil-population-density-map.jpg

2. Order Size

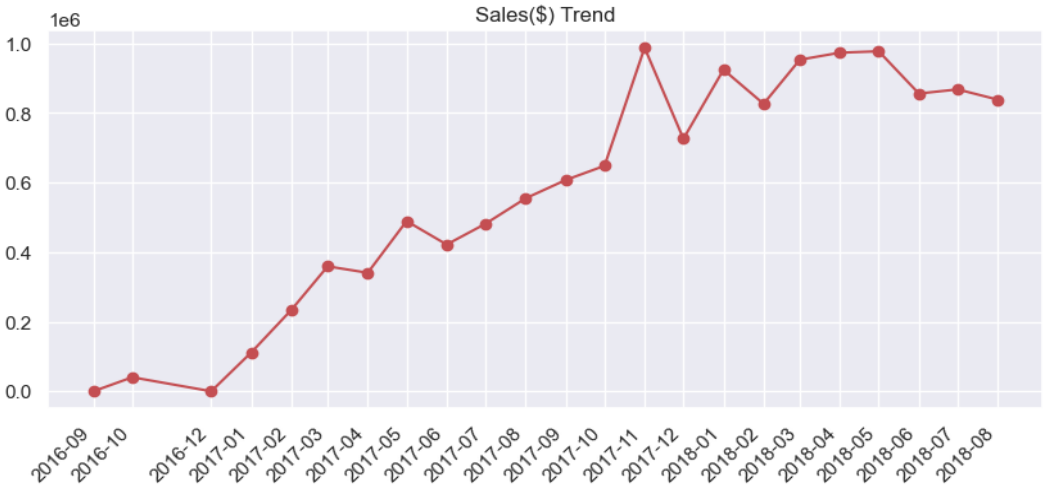
The order size table indicates that approximately 90% of total orders consisted of single-product orders. The line chart illustrates that the average order size fluctuated between 1.00 and 1.25 during the 2017-2018 period, with no noticeable seasonal patterns in order size from month to month.

	order_size	count	percentage
0	1	86843	90.013267
1	2	7392	7.661850
2	3	1306	1.353676
3	4	495	0.513070
4	5	193	0.200046
5	6	191	0.197973



3. Sales Trend

The sales consistently increased over time. In January 2017, the total sales amounted to $111,798, and by August 2018, they had risen to $838,576. There was a significant increase in sales in November 2017, which is common due to Black Friday sales promotions in the e-commerce and retail industry. The daily sales trend shows that Olist sales are generally stronger on weekdays, with the strongest sales occurring on Tuesdays and the weakest on Saturdays.

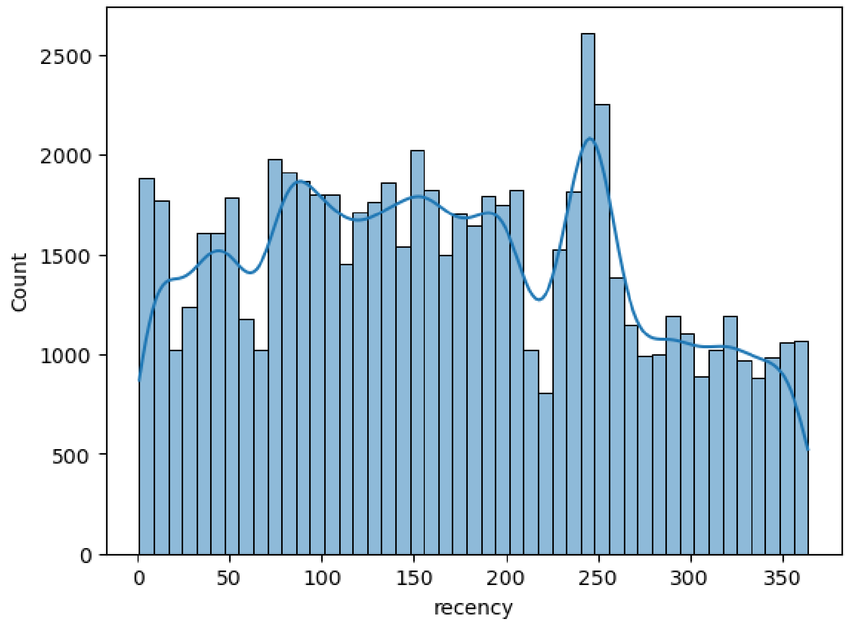


4. RFM Analysis

To conduct the RFM analysis, the dataset was subsetted from August 1, 2017, to July 30, 2018, to only include customers with one year of transaction history.

A. Recency

The "Recency" metric, as described by the statistics, measures the number of days since a customer's last purchase. Analyzing the data for 69,749 customers, we observe that the average recency is approximately 169 days, indicating that on average, customers made their last purchase about half a year ago. The standard deviation of around 99 days suggests a wide variability in customer behavior, with some customers having made a purchase very recently and others having not made a purchase in a longer period. Specifically, the recency ranges from 1 day (indicating very recent engagement with the company) to 364 days (suggesting it has been nearly a year since the last purchase). The interquartile range, from the 25th to the 75th percentile, spans from 88 to 248 days, highlighting that the middle 50% of the customer base made their last purchase between about 3 months to slightly over 8 months ago.

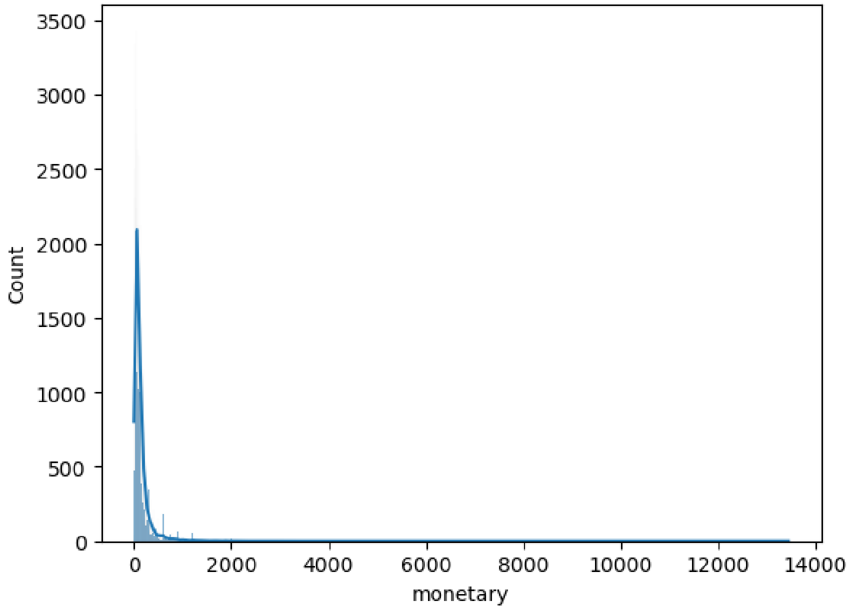


B. Monetary(M)

The "Monetary" metric quantifies the total spending of a customer. Based on the provided statistics for 69,749 customers, the average total spend is approximately $141.35, indicating the typical monetary value a customer brings to the company. However, there's a significant spread in the spending behavior as shown by a high standard deviation of about $210.78. This wide variability suggests that while some customers spend modestly, others may spend substantially more, leading to a broad range of total monetary contributions to the company.

The distribution of spending is skewed towards lower monetary values, with the minimum spend being as low as $0.85, and 25% of customers spending $47.99 or less. The median spend of $89.90 reflects that half of the customer base spends less than this amount, while the other half spends more. The third quartile at $153.96 indicates that 75% of customers spend up to this amount, and beyond this point, the higher spenders contribute significantly more, up to the maximum observed spend of $13,440.

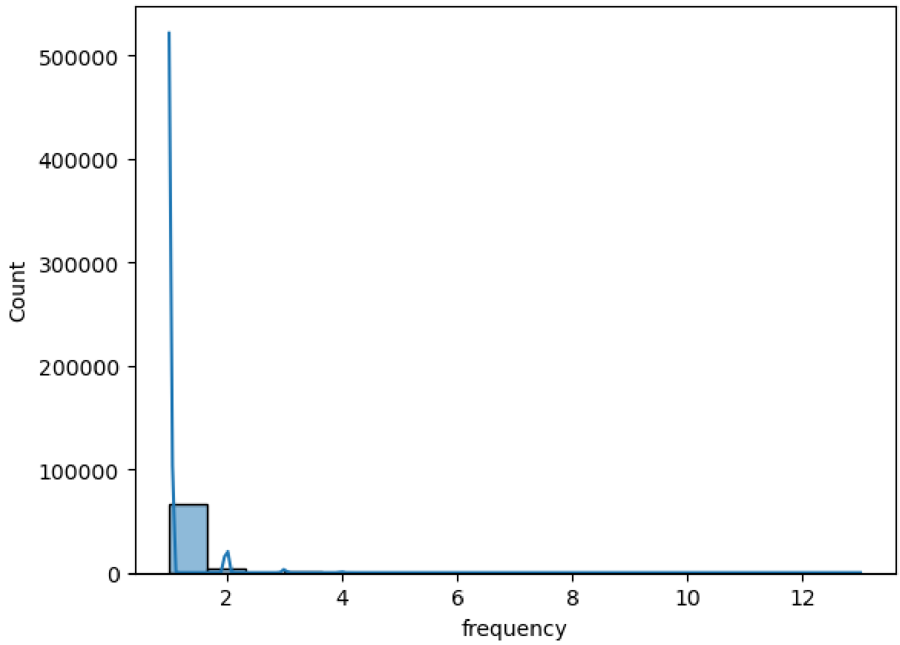
The large gap between the 75th percentile and the maximum value suggests the presence of outliers or a small percentage of customers who contribute exceptionally high monetary values, which significantly influences the average spending. This insight could be crucial for tailoring marketing strategies and loyalty programs to not only boost spending among lower and median spenders but also to recognize and reward the high-value customers who contribute significantly to the revenue.



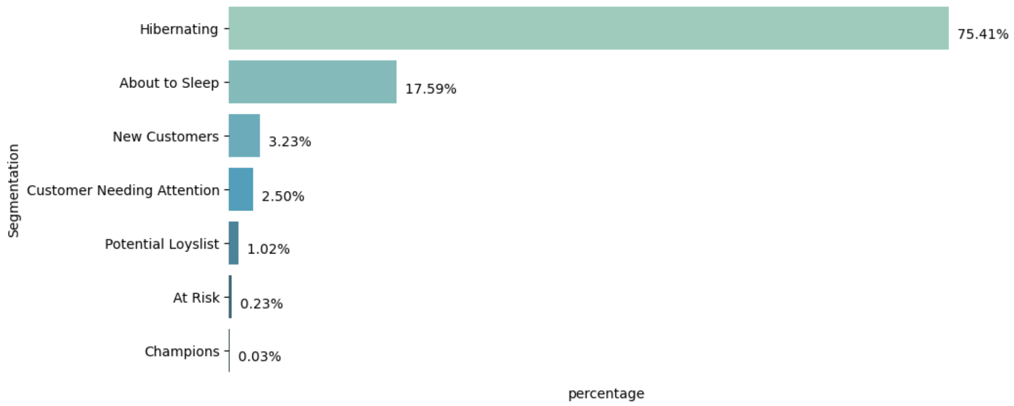
C. Frequency(F)

The "Frequency" metric captures how many times a customer has made a purchase. Analyzing data for 69,749 customers, it's observed that on average, customers have made about 1.07 purchases, with a standard deviation of 0.31. This relatively low standard deviation, alongside the quartiles, suggests that the vast majority of customers have made only one purchase, as evidenced by the 25th, 50th (median), and 75th percentiles all being at 1 purchase. The maximum frequency recorded is 13 purchases, indicating a few customers have engaged with the company multiple times.

Interestingly, when looking at customers who have made more than one purchase, they constitute approximately 5.69% of the total customer base. This highlights that a small portion of the customer base is responsible for repeat business, emphasizing the importance of understanding and catering to this segment to encourage customer loyalty and repeat purchases. The data underscores the potential opportunity for targeted marketing strategies aimed at increasing customer engagement and encouraging more frequent purchases among the wider customer base.



5. **Customer Segmentation**
   A. New Customers: Customers who have interacted recently (R>=3) with an average frequency and monetary value between 1 and 2. These are customers who have started purchasing recently. They have the potential to become more valuable, so it's a good strategy to encourage them to purchase again.
   B. Potential Loyalists: Customers with recent interactions and an average frequency and monetary value between 2 and 3. These customers are on the path to becoming loyal to the brand. They've shown a promising value and frequency of purchase, indicating a growing affinity towards the brand.
   C. Champions: Customers with the most recent interactions and high average frequency and monetary values (between 3 and 4). These are your best customers, who buy often and spend a lot. They are likely to advocate for your brand and should be recognized and rewarded to encourage continued loyalty.
   D. Customers Needing Attention: Customers who have interacted fairly recently (R between 1 and 3) with good frequency and monetary values (between 2 and 4). They have shown potential but are not as engaged as Champions. Tailored marketing efforts can help to increase their interaction and spending.
   E. About to Sleep: Customers who have interacted fairly recently (R between 1 and 3) but with lower average frequency and monetary values (between 1 and 2). These customers are at risk of losing interest. Engaging them with relevant offers or reminders can help rekindle their interest in the brand.
   F. At Risk: Customers who have not interacted recently (R between 0 and 1) but had good frequency and monetary values in the past (between 2 and 4). These were once valuable customers who have stopped purchasing recently. It's crucial to reconnect with them before they turn to competitors or lose interest entirely.
   G. Hibernating: Customers who have not interacted recently (R between 0 and 1) and have lower average frequency and monetary values (between 1 and 2). These customers are inactive and may need significant incentives to return.
   H. Lost: Any customers that do not fall into the above categories. This implies they have low recency, frequency, and monetary scores and may no longer be reachable or interested in the brand.



Recommendation

Conclusion:¶

75.41% - Hibernating:

These are customers who have bought a long time ago, only a few times and have spent little.

CRM Strategy:

- Launch re-engagement campaigns that highlight major changes or updates to your products or services since they last purchased. Use storytelling to reconnect emotionally.
- Standard communication for sending offers;
- Offer relevant products and good deals.

17.59% - Hibernating:

Customers who have interacted fairly recently but with lower average frequency and monetary values. These customers are at risk of losing interest.

CRM Strategy:

- Engaging them with relevant offers or reminders can help rekindle their interest in the brand.
- Implement win-back campaigns with a sense of urgency, such as "We miss you! Come back within 48 hours for a special offer." Personalize these offers based on past purchase history to make them more appealing.

3.23% - New Customer:

Customers who have interacted recently with an average frequency and monetary value between 1 and 2. These are customers who have started purchasing recently. They have the potential to become more valuable, so it's a good strategy to encourage them to purchase again.

CRM Strategy:

- Create an onboarding sequence that educates them about the brand, including values and what sets you apart. This can help in building a strong brand relationship early on.
- Encouraging customers with discounts that expire after a certain period can help increase their likelihood of revisiting.

2.5% - Customer Needing Attention:

These are customers who have recently purchased, however are still in doubt whether they will make their next purchase from the company or a competitor.

CRM Strategy:

- Use data analytics to predict products they might be interested in based on their browsing behavior and purchase history. Personalized emails or app notifications can make them feel valued and increase engagement.
- Promotional campaigns for a limited time;
- Product recommendations based on their behavior;
- Show the importance of buying with the company.

1.02% - Potential Loyalist:

These are recent buyers, spend a good amount and have bought more than once.

CRM Strategy:

- Offer a loyalty program;

- Keep them engaged;
- Personalized and other product recommendations.
- Encourage user-generated content by inviting them to share their experiences with your brand on social media. This not only engages them but also acts as social proof to attract new customers

0.03% - Champions:

These are customers who have bought recently, buy often, and spend a lot.

CRM Strategy:

- Special offers, products and discounts for these customers so they feel valued;
- Ask for reviews and feedbacks constantly;
- Avoid sending massive amounts of offers;
- Personalized communication;
- Give rewards.
- Consider implementing an exclusive VIP club or offering early access to new products. This reinforces their special status and can encourage continued loyalty.