

E-commerce Report

RFM

Introduction

E-commerce Data Analysis: Understanding Customer's Behavior

Online retail

Online retail, also known as e-commerce, has transformed the way consumers shop and businesses operate in the modern marketplace. With the proliferation of internet connectivity and the increasing adoption of mobile devices, online retail has experienced exponential growth, offering consumers unparalleled convenience, accessibility, and choice. Through online retail platforms, customers can browse and purchase a wide range of products and services from the comfort of their homes or on the go, eliminating geographical constraints and opening up global markets. Online retailers leverage advanced technologies such as data analytics, artificial intelligence, and personalized recommendation systems to enhance the shopping experience, tailor offerings to individual preferences, and optimize logistics and fulfillment processes. Additionally, the COVID-19 pandemic has accelerated the shift toward online shopping, further cementing its role as a dominant force in the retail landscape. As online retail continues to evolve, businesses must adapt to changing consumer behaviors and technological advancements to remain competitive and capitalize on the vast opportunities presented by this dynamic industry.

PREPARED BY

Saeed shirani

PREPARED FOR

Github

overview of the purpose of the analysis

The purpose of the analysis is to gain insights into various aspects of online retail and e-commerce, including customer behavior, purchasing patterns, satisfaction levels, and churn rates. By examining these factors, businesses can make informed decisions to improve their marketing strategies, optimize customer experiences, retain valuable customers, and drive long-term profitability. Additionally, the analysis aims to identify opportunities for enhancing customer engagement and implementing effective strategies for reacquiring lost customers.

Importance of customers in online retail

Customers are vital in online retail and e-commerce for several reasons. Firstly, they are the primary source of revenue, as their purchases drive sales and generate income for businesses. Additionally, satisfied customers are more likely to become repeat buyers, contributing to the establishment of a loyal customer base. Furthermore, customers also serve as brand ambassadors, sharing their positive experiences with others and potentially attracting new customers through word-of-mouth marketing. Understanding and catering to customer needs and preferences are essential for businesses to remain competitive and thrive in the ever-evolving digital marketplace.

Goals

- ⌚ To observe the behavior of customers, both on an individual and collective level.
- ⌚ To identify patterns and anticipate the potential future behaviors of our customers.
- ⌚ To determine the frequency with which our brand/company can maintain customer satisfaction.
- ⌚ To track the number of customers gained or lost on a monthly/yearly basis.
- ⌚ To investigate the reasons behind customer churn and gather their opinions on why they chose to leave.
- ⌚ What possible strategies can be employed to regain lost customers, and which customers are most valuable to target for re-engagement efforts?

Data Collection & Preparation

This dataset contains data on buying or selling stocks in an unknown market. This dataset is one of the best data sets I have found to intercept the behavior of customers, In sell rows, the 'Quantity' number is negative and I don't know the exact meaning of this negative sign on the other hand I just want to see how frequent is the interaction of the system and user, I used the absolute value of Quantity column.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

Overview of data cleaning, preprocessing, and any transformations applied

Our data includes 8 columns:

1. **InvoiceNo:** label/number of deal. (53628 unique values recorded.)
2. **StockCode:** represent the Stock name/Code. Unique value for each stock. (5305 unique value recorded)
3. **Description:** Description of product
4. **Quantity:** numbers of the stock unit
5. **InvoiceDate:** Data of Deal (Data is from the start of 2010 and until the end of 2011)
6. **UnitPrice:** price of each unit of data
7. **CustomerID:**
represents the ID of each user. Unique value for each user (5943 Values recorded)
8. **Country:** probably location of user

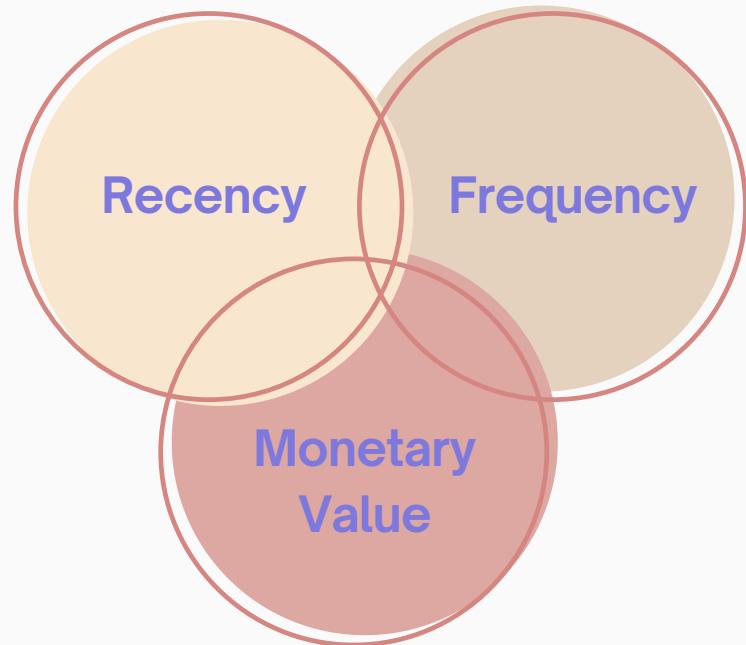
RFM Customer Analysis

RFM segmentation approach

There are various methods for segmenting customer behavior, with the most effective being the "RFM" technique. Through the calculation of RFM scores and applying different approaches to these scores, businesses can evaluate their performance about customer behavior and strive to maintain customer satisfaction.

RFM analysis categorizes customers based on Recency, Frequency, and Monetary Value. It helps businesses identify and prioritize customer segments, enabling targeted marketing strategies. This approach enhances customer engagement, satisfaction, and retention.

RFM Segments



Recency:

“How recently a customer made a purchase or interacted with the business, indicating their level of engagement or responsiveness.”

Frequency:

“How often a customer makes purchases or engages with the business over a specific period, reflecting their loyalty or level of activity.”

Monetary Value:

“The total amount of money a customer has spent on purchases or transactions with the business, indicating their overall contribution to revenue or profitability.”

RFM Factors in our dataset

The dataset encompasses user interactions recorded during the years 2009, 2010, and 2011. Within this timeframe, numerous deals were captured, allowing us to examine the most recent interactions users had with our system.

The dataset reveals a significant variation in interaction frequency. The highest number of deals for a user is 12,638, while the lowest is just one. This broad spectrum of interaction frequencies provides valuable insights into user engagement with the system.

Calculating the total value each user contributed to the system involves multiplying the unit price by the number of items purchased and summing them up. This monetary analysis allows us to gauge the financial impact of each user on the system.

3 year
TIME INTERVAL

53628
NUMBER OF DEALS

5943
NUMBER OF CUSTOMERS

43
NUMBER ON COUNTRIES

Overview to Recorded Factors in dataset

Customer Segmentation

Overall look at customer's scores

For an overview of our customers' total RFM scores, we categorize them into four groups as follows:

- Customers with an RFM score greater than 1.7 are classified in the **Gold** group.
- Customers with an RFM score between 1.7 and 1 are classified in the **Silver** group.
- Customers with an RFM score between 1 and 0.5 are classified in the **Bronze** group.
- Customers with an RFM score between 0.5 and 0 are classified in the **Tin** group.

Defining intervals for this segmentation is subjective and can vary based on individual preferences and business requirements.

DATASET

	frequency	monetary	days_since_deal	rfm_score	Segment
customer_id					
14911.0	0.905516	0.535981	0.998645	2.4401	Gold
12748.0	0.527103	0.101666	1.000000	1.6288	Silver
12478.0	0.002770	0.001149	0.995935	0.9999	Bronze
13108.0	0.002691	0.001263	0.495935	0.4999	Tin

Now we sum the RFM scores of each segment. We aim to visualize the segmentation by displaying the total RFM scores of all users for each segment in a diagram.



About 40% of our customers are located in the Silver group.

32% of our customers are in the bronze group.

14% Gold and 14% Tin.

A high score in the silver and bronze categories indicates that the company's performance is average, or they may have recently initiated their operations and have not yet had the chance to acquire gold customers.

Analyze customer behavior & implement new strategies accordingly.

In the previous section, customers were grouped based on their RFM score. In this section, I intend to segment them based on each factor of RFM. It is common to divide them into 11 segments. Here are the segments:

-
- 1. CHAMPIONS
 - 2. LOYAL CUSTOMERS
 - 3. POTENTIAL LOYALISTS
 - 4. RECENT CUSTOMERS
 - 5. PROMISING CUSTOMERS
 - 6. CUSTOMERS NEEDING ATTENTION
 - 7. ABOUT TO SLEEP
 - 8. AT RISK
 - 9. CAN'T LOSE THEM
 - 10. HIBERNATING
 - 11. LOST
-

Here is a definition and explanation of each segmentation:

Champions: These are your best customers who buy recently, frequently, and spend a lot. They are loyal, engaged, and responsive to your offers. You should reward them, retain them, and encourage referrals.

Loyal customers: These are customers who buy frequently and have a good relationship with your brand. You should reward them, upsell and cross-sell to them, and ask for feedback.

Potential loyalists: These are customers who have bought recently and spent a good amount, but not frequently. They have the potential to become loyal customers if you nurture them and offer incentives.

Recent customers: These are customers who have bought only once recently. They may be first-time buyers or reactivated customers. You should provide onboarding support, follow up, and cross-sell to them.

Promising customers: These are customers who have bought once recently, but not spent much. They may be testing your products or services. You should offer them relevant products, discounts, and free trials to increase their engagement.

Customers needing attention: These are customers who have not bought recently, but bought frequently and spent a lot in the past. They may be at risk of churning or switching to competitors. You should re-engage them with personalized offers, feedback surveys, and loyalty programs.

About to sleep: These are customers who have not bought recently, frequently or spent much. They may be inactive or uninterested in your brand. You should wake them up with reactivation campaigns, product reviews, and win-back offers.

At risk: These are customers who have bought frequently and spent a lot, but not recently. They may have been loyal customers before, but have stopped buying. You should try to retain them with loyalty programs, satisfaction surveys, and special offers.

Can't lose them: These are customers who have bought frequently and spent a lot, but a long time ago. They may have been your best customers, but have switched to other brands. You should try to win them back with high-value offers, personal outreach, and feedback requests.

Hibernating: These are customers who have not bought recently, frequently, or spent much. They may have forgotten about your brand or have no need for your products or services. You should try to reawaken their interest with reactivation campaigns, product updates, and newsletters.

Lost: These are customers who have not bought for a long time, and have low frequency and monetary value. They may have no intention of buying from you again. You should minimize your marketing costs for them, or remove them from your database.

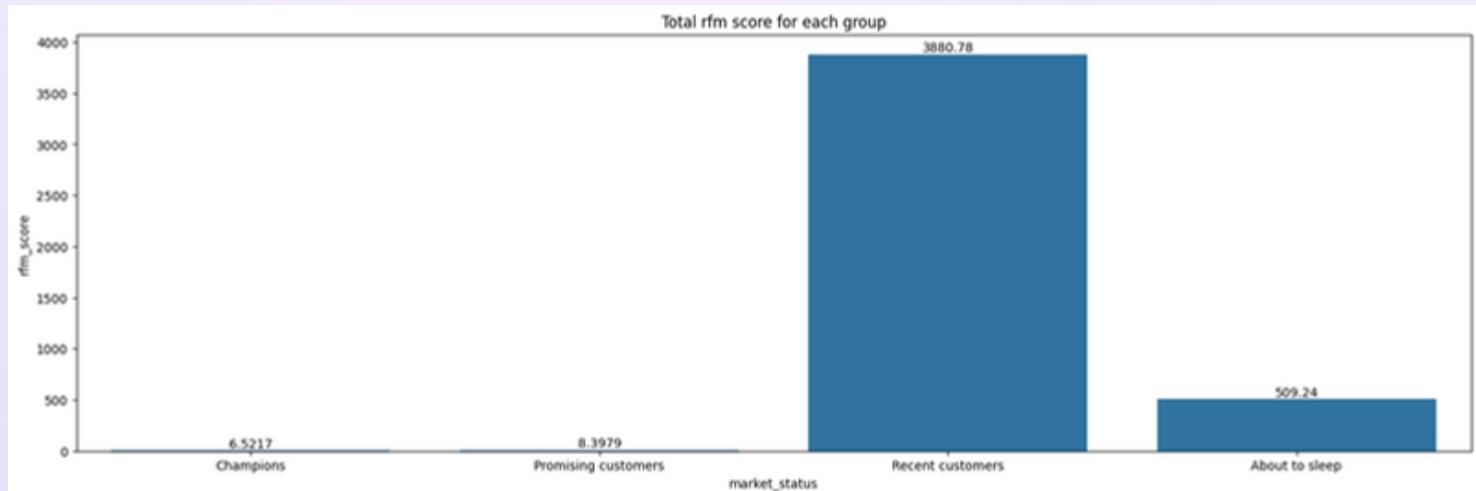
After applying these rules to our dataset, here is what it looks like:

	frequency	monetary	days_since_deal	rfm_score	Segment	market_status
customer_id						
14911.0	0.905516	0.535981	0.998645	2.4401	Gold	Champions
14646.0	0.307747	0.902415	0.998645	2.2088	Gold	Champions
17841.0	1.000000	0.118160	0.998645	2.1168	Gold	Promising customers
18102.0	0.083010	1.000000	1.000000	2.0830	Gold	Recent customers
14156.0	0.325789	0.559187	0.987805	1.8728	Gold	Champions
12748.0	0.527103	0.101666	1.000000	1.6288	Silver	Promising customers
16446.0	0.000317	0.569564	1.000000	1.5699	Silver	Recent customers
14606.0	0.514283	0.051900	1.000000	1.5662	Silver	Promising customers
15311.0	0.362270	0.199787	1.000000	1.5621	Silver	Promising customers
14096.0	0.405713	0.123746	0.994580	1.5240	Silver	Promising customers

visualize the total RFM score for each segment

From All 11 segments, our brand only has customers in 4 of them:

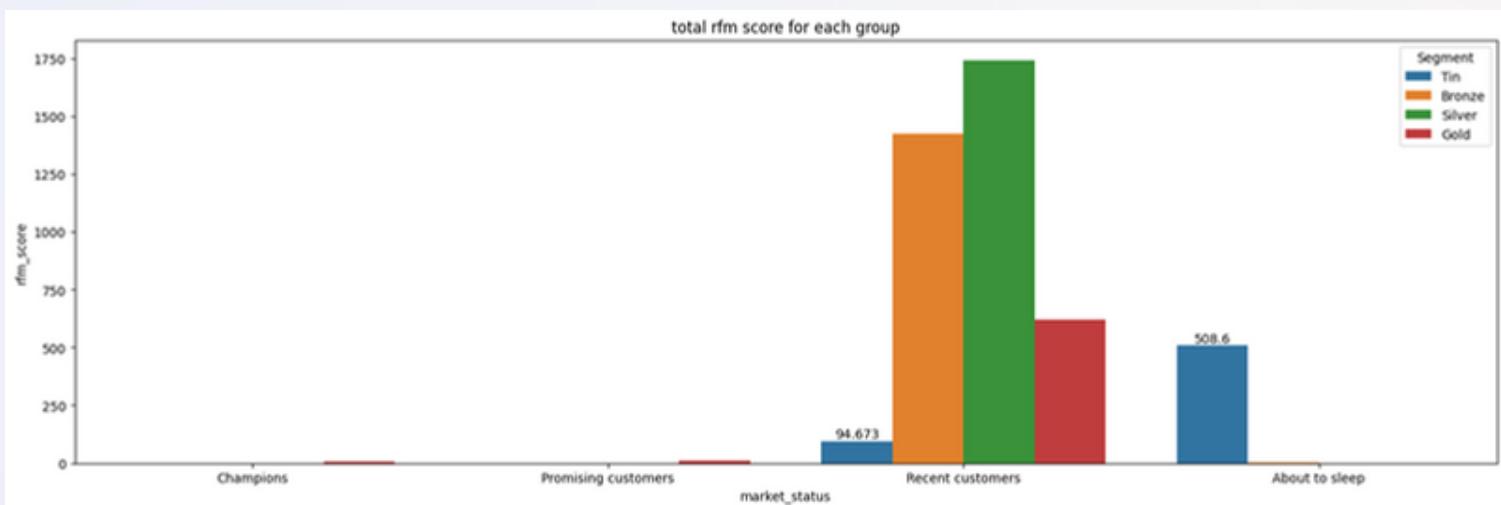
Champions, Promising Customers, recent customers, about to sleep



- Most customers fall into the "Recent customers" segment, indicating that our marketing strategies and advertising efforts have been effective. We should continue using these strategies and consider intensifying them. Additionally, we must implement strategies to convert these recent customers into loyal ones.
- The second highest score in this diagram is for the "About to sleep" segment. That suggests that many customers have only made one or two purchases and are not engaged with our brand. We need to investigate why these customers were dissatisfied. Is the issue with our brand or do they simply not require our products? Conducting a survey of these customers could provide valuable insights. Furthermore, offering them a suitable discount plan may help re-engage them.

- The third group consists of "Promising customers". To ensure satisfaction among these customers, we must offer them the best products and services. They represent our potential future loyal customers.
- The fourth group comprises "Champions", but their number is alarmingly low. This indicates a need to convert our recent customers into groups like champions, loyal customers, or potential customers. Additionally, we should analyze the products purchased by champions and gather data on factors such as age, gender, location, occupation, education level, marital status, etc. Understanding these factors will help us identify our target audience and elevate more customers to the champion category.

As I said above, our recent customers are a lot! It could be helpful to see what segments are in this group.



As evident from the diagram, most of our recent customers fall into the silver and bronze categories. There are four possible strategies to address this:

1. The timeframe of our analysis might be too short, and we may need to wait a little longer to observe how these customers behave in the future.
2. One possible reason for this phenomenon could be that our brand is not well-known, and we need to introduce it to our target audiences more effectively.
3. Another possible reason could be the performance of our brand. This suggests that only individuals with lower or average financial abilities are our customers, and we must improve the quality of our offerings.
4. As previously mentioned, we need to identify societies of gold customers and develop effective advertising, discounting, and marketing strategies to target them.

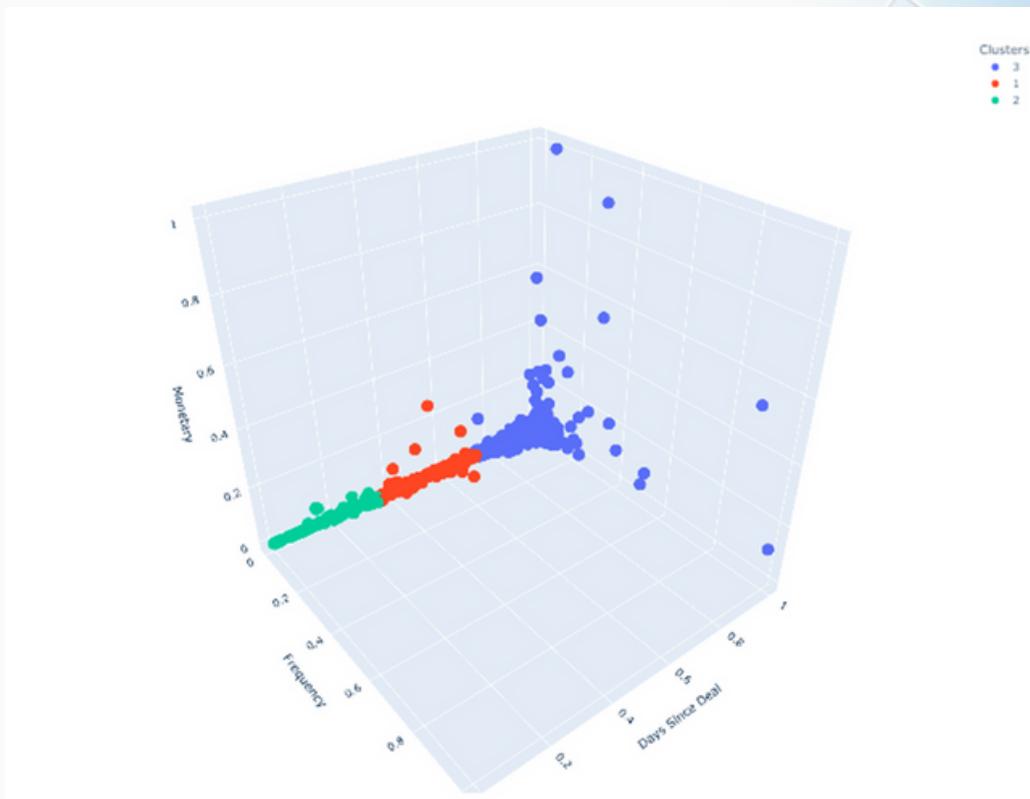
Next Step

Data Clustering

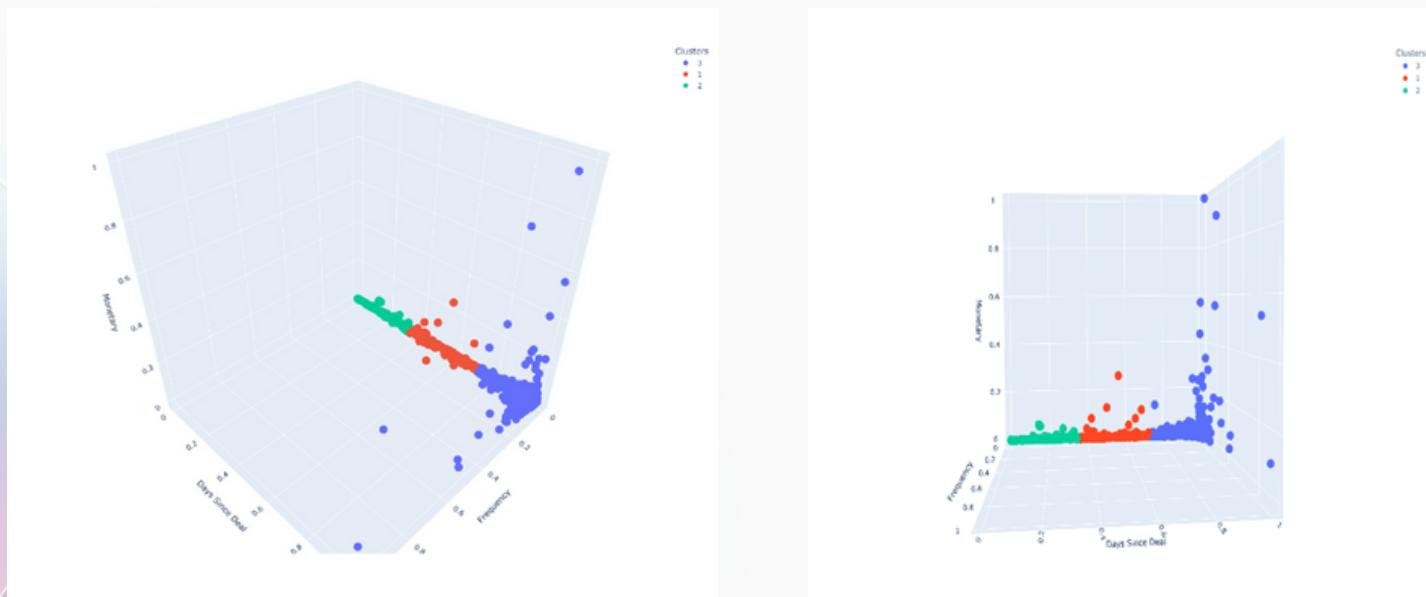
Now that we have an overview of our customer's interactions with our system, we need to identify customers with similar characteristics to gain a comprehensive understanding of our brand's performance in handling customers.

In RFM analysis, once the RFM scores have been computed for each customer, the data is then clustered to group customers with similar characteristics together. This process is essential for identifying distinct customer segments based on their purchasing behavior. By clustering the data, businesses gain valuable insights that enable them to tailor their marketing strategies and offerings more effectively. This includes the ability to personalize marketing campaigns, provide targeted product recommendations, and implement loyalty programs, all of which contribute to maximizing customer satisfaction and driving revenue growth. Moreover, clustering facilitates the identification of high-value customers, the re-engagement of lost customers, and the optimization of resource allocation for marketing efforts. Overall, the clustering of RFM data serves as a powerful tool for generating actionable insights and guiding strategic decision-making to enhance customer relationships and improve business performance within the e-commerce and retail sectors.

I clustered the data into three clusters. The graphical representation of these clusters is as follows:



Let's examine the plot from two additional perspectives:

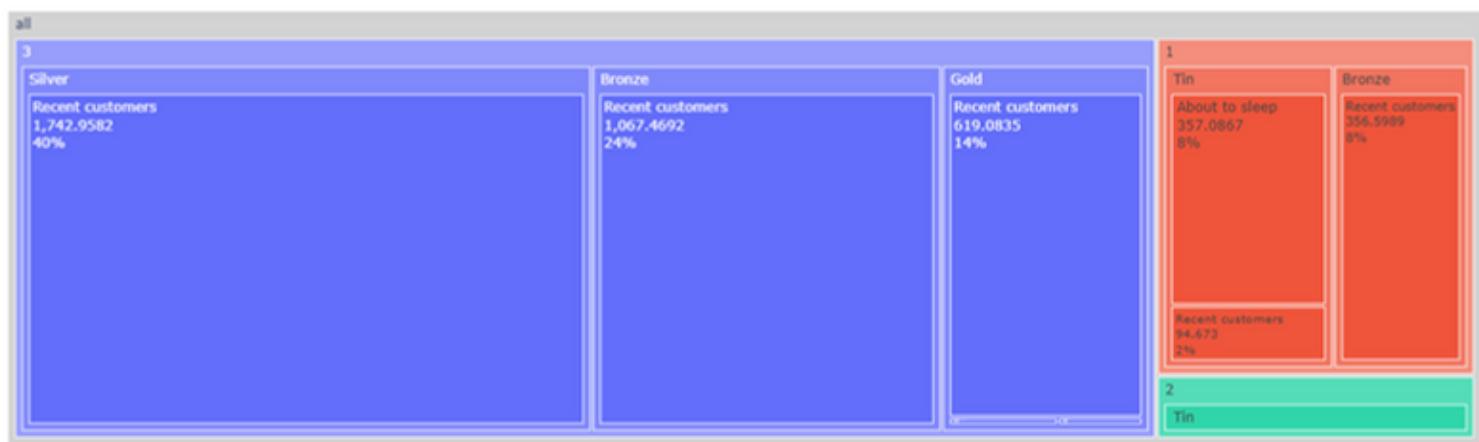


There is significant residuality in the blue cluster, and we can also observe light residuality in the red cluster. Our outliers are primarily in the blue cluster. However, it's important to note that in this analysis, outliers may indicate the start of a sales paradigm or a loss of sales paradigm. To determine this, we need to identify which customers are in the blue cluster.

Treemap

A Different Approach to Classification

Now that we have identified the residuals of our customers using three segmentation methods, we can combine them to gain a comprehensive overview.



Percentages on each rectangular is the percentage of each group of data from the whole dataset.

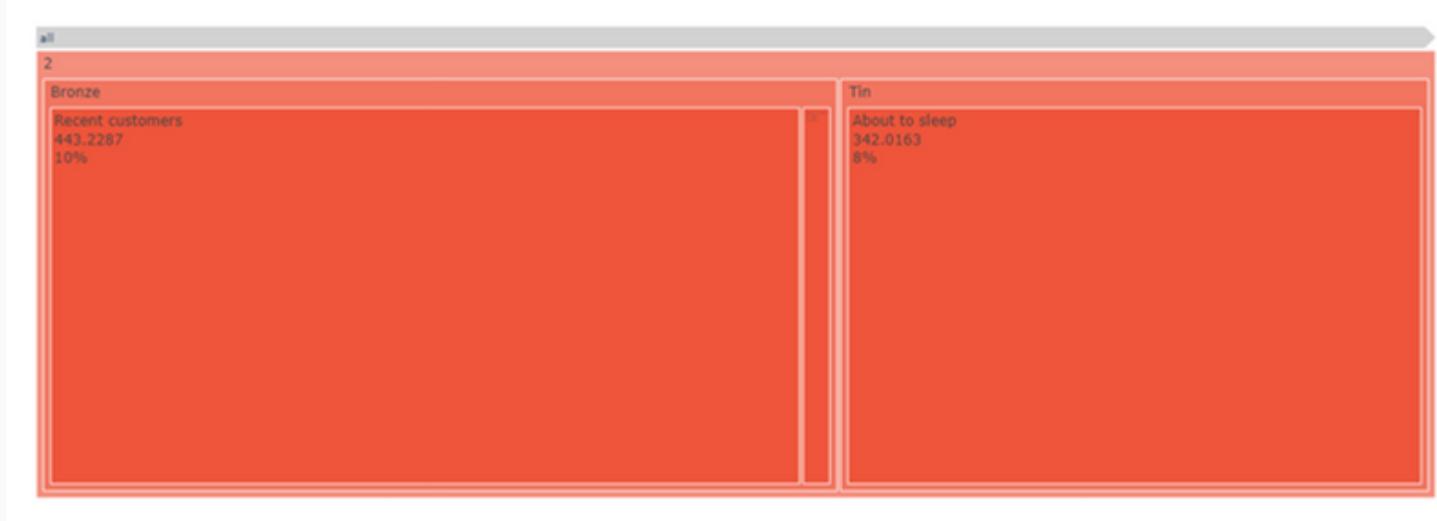
Let's look at each rectangular separately:

CLUSTER 1:



The blue cluster, likely our most significant cluster, comprises 78% of our data. This cluster includes both bronze and silver customers, indicating that 78% of our customers are recent customers, with the majority (approximately 57%) having less impressive RFM scores. Both segments within this cluster consist solely of recent customers. Therefore, the best initial step for the company could be to focus on converting bronze customers to silver customers.

CLUSTER 2:

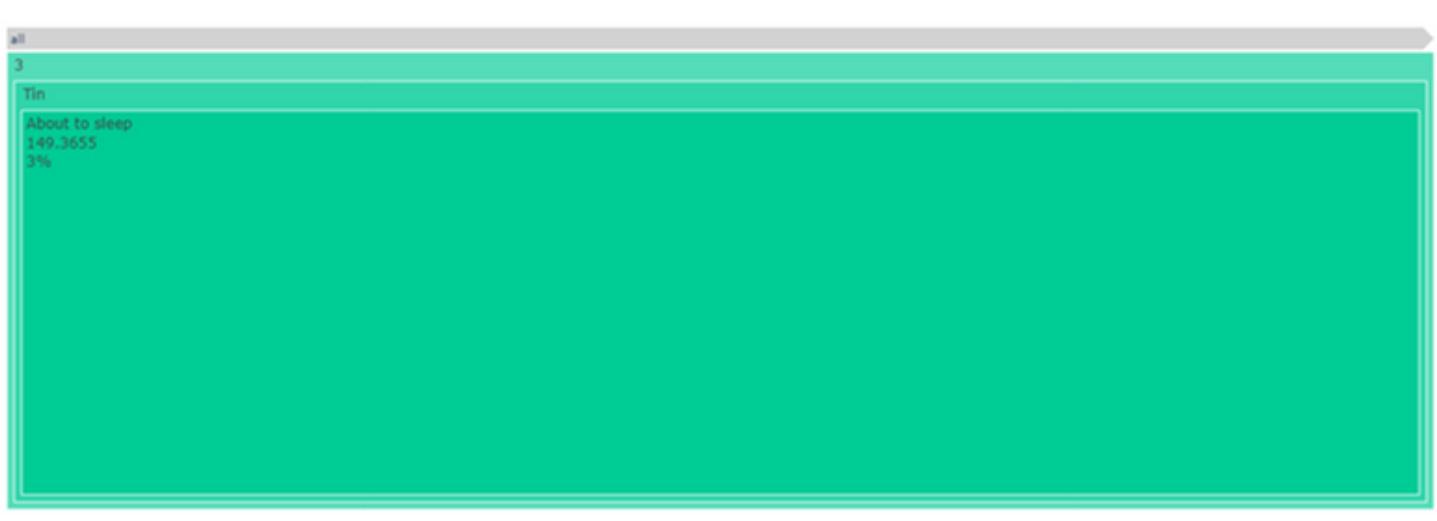


This cluster comprises 18% of our dataset and includes the Bronze segment with recent customers, as well as the Tin segment with customers categorized as "about to sleep".

For customers in the Bronze segment, the company must identify members of this group and strive to increase their RFM score by implementing advertising and discount plans.

Regarding the Tin segment, all individuals are at the "about to sleep" stage. Therefore, we should continue collecting data from these customers and allow them time to potentially change their condition and behavior with more targeted strategies.

CLUSTER 3:



The third cluster is less significant, comprising only 3% of the total RFM score. Similarly, all individuals in this cluster are at the "about to sleep" stage. Therefore, we should continue collecting data from these customers and give them time to potentially change their condition and behavior with more targeted strategies.

KPI and metrics

keep track of our results



Defining KPI:

When a company initiates a campaign or advertising effort, it typically targets a specific demographic. It's essential to monitor the fluctuations in the number of customers within each category, namely Gold, Silver, Bronze, and Tin, to gauge the impact of the advertising on our brand and customer base. A useful Key Performance Indicator (KPI) in this regard could be the monthly changes in segment distribution or **Customer Segment Movement (CSM)**.

To calculate **CSM**, we need to subtract the number of members in each segment in each month from those in the same segments in the previous month.

By comparing the **Customer Segment Movement (CSM)** for each category of customers, we can evaluate the effectiveness of the company's new strategies. For instance, if the company implements a three-month advertising plan targeting Bronze customers to attract more customers within this segment, we should observe an increase in the CSM of the Bronze segment. Meanwhile, CSM in other segments should remain relatively unchanged.

Discussion and Suggestions

Overall, there are a lot of new customers among our customer base, indicating several **positive factors**:

1. The company's products have good enough quality.
2. The company started its operations at an opportune time and in a favorable market.
3. The marketing strategies of the company were effective.

However, there are several significant **problems**:

1. Our customers are not in the gold segment, which could be due to at least three reasons:
 - 1.1. People don't trust our brand and prefer not to make high-value purchases.
 - 1.2. The target audience of our brand and market consists of poor or average-income individuals.
 - 1.3. The nature of our products and services may contribute to this issue. Even if a person buys regularly from our brand, the pricing of our products/services may lead to a low value of cash flow.
2. The company must focus on retaining existing customers. Sudden changes in company strategies might cause churn and lead to the loss of some customers.

Ultimately, the company must continue collecting data from sales and each customer's properties (such as age, gender, etc. especially for "about to sleep" group)). This is essential for designing targeted and effective marketing plans for each customer group.

Thank you!

Thank you for taking the time to read this report. If you have any questions or would like to discuss our findings further, please don't hesitate to reach out to me.

 Isfahan

 09105996314

 saeedshirani7878@gmail.com