Margarine Margins

24010577 Meyka Chan

24008144 Li Peng

24001157 Xie Yue

156.761 Customer Insights

Professor Malcolm Wright

21 August 2024

Table of Contents

Executive Summary

This report presents an analysis of consumer buying behavior for a margarine brand over five quarters, focusing on repeat purchase patterns based on the Negative Binomial Distribution (NBD) model.

Key Findings:

1. **Brand Performance Stability**: Across five quarters, key performance metrics—purchase rate, purchase frequency, and penetration rate—remained relatively stable. The brand showed consistent customer attraction but limited growth in penetration. The average purchase rate suggested that buyers typically purchase the product 2-3 times per quarter.
2. **Behavioral Segmentation**: Customers were segmented into non-buyers, light, medium, and heavy buyers. Analysis revealed that heavy buyers were the most stable group, while light and medium buyers exhibited more variability, indicating potential opportunities for growth by converting non-buyers and light buyers into more frequent purchasers.
3. **NBD Analysis**: The NBD model demonstrated strong predictive power, showing a good fit between theoretical and observed buyer distributions. The analysis underscored the potential to convert non-buyers into active buyers through targeted marketing strategies.
4. **Q5 Buyer and Sales Distribution**: In the fifth quarter, non-buyers constituted the largest segment (66%), while heavy buyers, though only 7% of the customer base (20% of the active buyer group excluding the nonbuyers), contributed nearly half of the total sales. The data deviates from the traditional 20/80 rule, suggesting that heavy buyers do not contribute as much as expected, highlighting the potential to boost sales by targeting non-buyers and light buyers.
5. **Buyer Flow Between Q4 and Q5**: The analysis of buyer transitions between Q4 and Q5 revealed significant stability among non-buyers and heavy buyers, with high levels of loyalty observed in the latter group. However, light and medium buyers showed considerable volatility, frequently transitioning between

categories, indicating susceptibility to churn and the influence of external factors like promotions.

6. **Marketing Insights:** The analysis supports several marketing principles, including the Double Jeopardy Law, which emphasizes the importance of brand penetration for growth, and the Pareto Law, which shows that heavy buyers contribute less to total sales than traditionally expected. The Law of Buyer Moderation and Duplication of Purchase Law were also observed, reinforcing the need for strategies to target non-buyers and light buyers.
7. **Recommendations:** focus on increasing purchase frequency among light and medium buyers through tailored incentives, while developing retention strategies for heavy buyers to prevent downgrading. Use advanced analytics to simulate the financial outcomes of different marketing strategies, optimizing resource allocation and maximizing return on investment. Strategies should include maintaining consistent branding, reinforcing category entry points, and targeting all category buyers to drive growth and increase brand awareness.

Introduction

In marketing, it is critical to understand consumer buying behavior to develop effective strategies. Companies can tailor their marketing efforts to gain a competitive advantage through analysis of consumer behavior patterns.

Repeat purchase behavior is the focus of this area. By analyzing the frequency of customer purchases, companies can segment their customers into different segments such as heavy buyers, medium buyers, light buyers, and non-buyers. This segmentation enables more targeted marketing strategies to achieve business goals, such as increasing customer loyalty (Knox, 1998). In addition, companies can improve inventory management, optimize the supply chain, minimize the risk of out-of-stock and overstock, and improve operational efficiency by understanding purchase frequency and forecasting future demand (Bala, n.d.).

The main objective of this analysis is to analyze the repeat purchase behavior of customers in the margarine brand using the Negative Binomial Distribution (NBD) model and to provide a reference for marketing strategies by identifying consumer behavior patterns.

To achieve this goal, we will first examine the brand's trends over five quarters and assess the brand's overall performance. Based on this, we will use the NBD model to generate a theoretical value as a benchmark for the final quarter, which will enable us to assess brand performance by comparing observed value and theoretical value. The distribution of different types of customers - none buyers, light buyers, medium buyers, and heavy buyers - will be examined in detail during this period. In addition, to uncover behavioral patterns, we will analyze how users have switched between purchase categories over the past two quarters. Finally, we will summarize the results of the analysis and provide strategic recommendations on how to improve marketing efforts in the future.

For this analysis, the NBD model is implemented using Excel, and certain data visualizations are created using Python to improve clarity and insight.

Methodology

The study first prepared the dataset to ensure it could be effectively and efficiently imported, read, and processed across different platforms or programs. While no data cleaning was required, the researchers used Excel to separate each transition matrix in a sheet. Using user-defined Python scripts, the brand's performance was assessed by gaining a general overview of all five quarters. The key metrics calculated were penetration rate, purchase rate, and purchase frequency. To further understand what contributed to the brand's growth, decline, or stability, customer segmentation according to purchase frequency was performed. Light, medium, and heavy buyers were defined by analyzing the cumulative frequency of the first quarter. By categorizing buyers, the researchers closely examined buyer distribution across all five quarters and gained insight into customer dynamics, over time. Focusing on the latest data and using Excel in conducting the next steps, further investigation was conducted on the buyer and sales distribution in the fifth quarter. This provided a deeper understanding of the current customer behavior and source of sales. A benchmark value for the expected number of individuals in each purchase class was then calculated using the Negative Binomial Distribution model. The theoretical and observed values were compared to assess whether the brand is under or over-performing. Lastly, a buyer flow analysis of the last two quarters was carried out to see how customers transition from one buyer category to another, revealing insights on customer loyalty and movement within the category's segments.

Results

This section of the study discusses the patterns and trends observed in the aggregated and disaggregated data of the brand's customers across all quarters. A more detailed investigation is done on the latest data, quarter 5 in the latter sub-sections.

Brand Performance Metrics

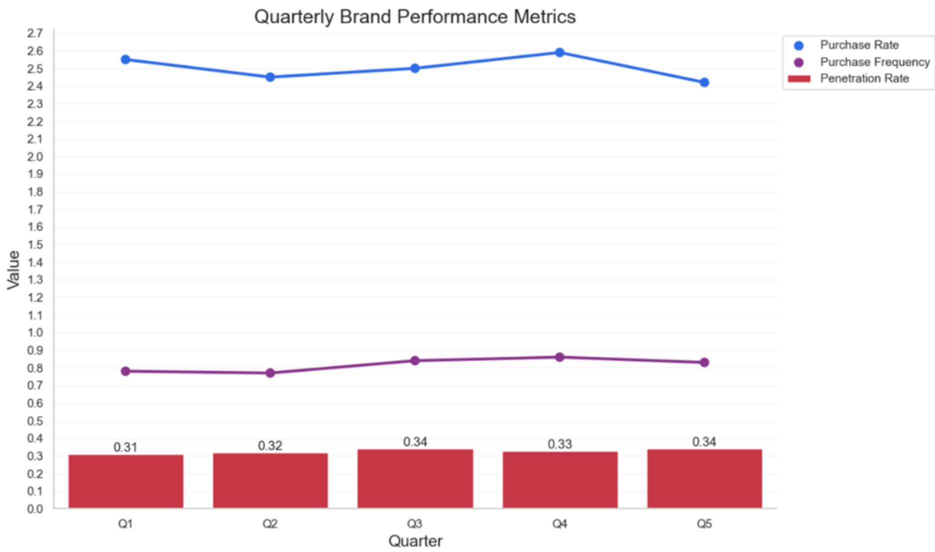


Figure 1. Brand performance metrics from Quarter 1 to 5

It can be observed in figure 1 that all three brand performance metrics – purchase rate, purchase frequency, and penetration rate across five quarters only fluctuate to at most 0.2, suggesting that the metrics are all quite stable. A stable penetration rate suggests that the brand is relatively consistent in attracting new buyers but has yet to achieve significant growth in penetration. A purchase rate of about 2.5 means that on average, brand buyers buy about 2 to 3 margarines every quarter. Lastly, a purchase frequency of less than 1 implies low customer loyalty and retention since only a few customers are repetitively buying the product in the next quarter.

It can also be seen that the purchase rate is more volatile than the penetration rate and purchase frequency. That said, since purchase frequency is the product of the other two metrics, it can be inferred that it is more affected by penetration rate. This implies that it is more effective to reach more new customers than trying to increase the number of purchases of brand buyers. Furthermore, the brand's performance is relatively better in quarters 3 and 4. While this may indicate potential seasonality buying habits, further investigation is needed to verify whether this is true or just a wobble in the data.

Behavioral Segmentation

Disaggregating the data according to purchase frequency will help the researchers gain further insights into the dynamics occurring from quarter to quarter. This was done by classifying buyers into none or zero, light, medium, and heavy buyers. While it was intuitive to classify buyers with zero purchases to none or zero buyers, the latter three categories required calculating the cumulative frequency of buyers per purchase class in quarter 1. The first quarter was chosen since it's the base quarter of the data provided. From the cumulative frequency, the top 15% were classified as heavy buyers while the bottom 25% and the 15th to 75th percentile were classified as light and medium buyers, respectively (Heavy, Medium and Light Shoppers, n.d.).

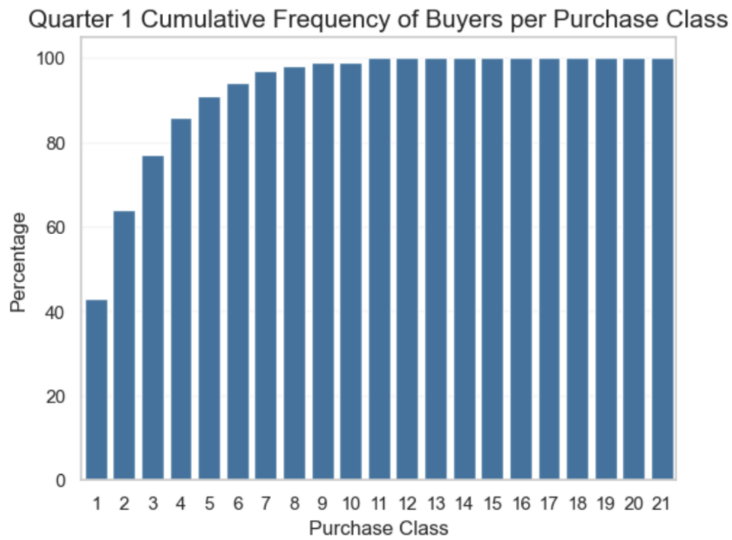


Figure 2. Cumulative frequency of brand buyers per purchase class in Quarter 1

As shown in figure 2, around 40% of all brand buyers in quarter 1 purchased the product once and approximately 75% purchased it less than or equal to three times. From the findings, it is clear that following the bottom 25%, top 15%, and between the 15th and 75th percentile in classifying buyers cannot strictly be followed. Nonetheless, it was used as a guide in classifying the buyers. It was then concluded that light buyers are customers who have purchased the product exactly once, medium buyers are those with two or three purchases, and heavy buyers are individuals who have purchased the margarine brand 4 or more times in a quarter.

Quarterly Buyer Distribution

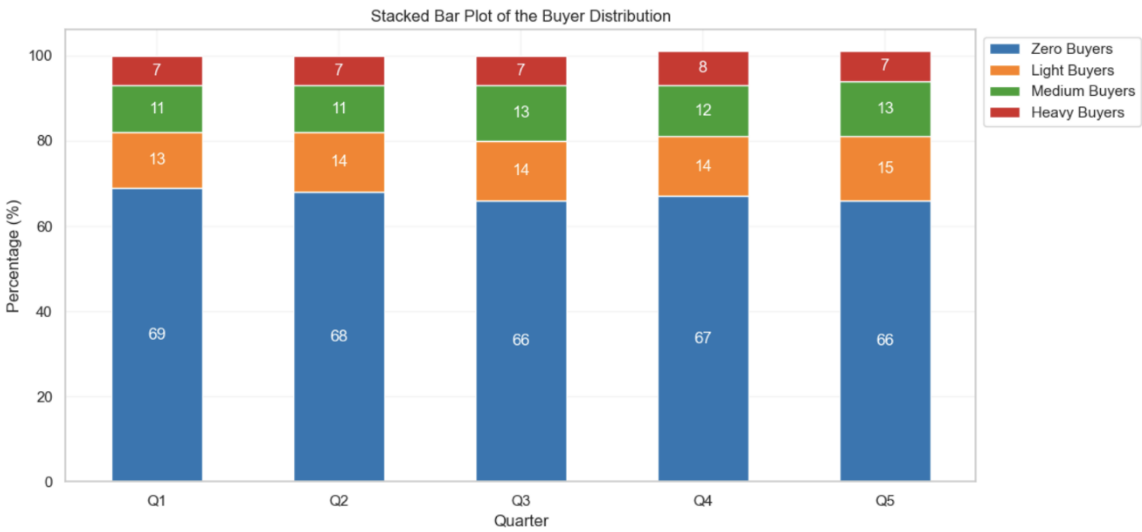


Figure 3. Buyer distribution from Quarter 1 to 5

The proportion of individuals belonging to each buyer category across all quarters is somewhat stable, as seen in figure 3. The zero buyer category remains the largest with the highest percentage in all quarters. Meanwhile, heavy buyers show the smallest portion of the customer base. Additionally, among the buyer categories, heavy buyers are relatively the most stable group while the zero, light, and medium are more relatively changing. Moreover, observe that in comparison to quarter 1, there is a small increase in light and medium buyers in quarter 5. This implies that there is some success in increasing penetration.

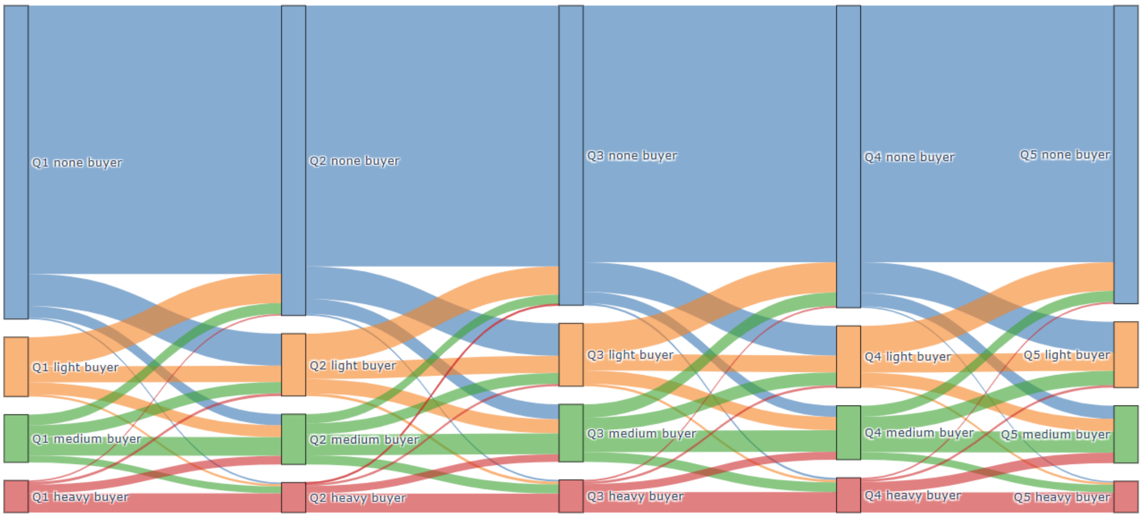


Figure 4. Sankey plot of buyer distribution from quarter 1 to 5

It can be observed in figure 4 how stochastic buyers are, where similar buying behavior can be seen in each quarter – regardless of the previous quarter's results, and how every group of individuals shifting from one category to another has a corresponding group that shifts into their place. This implies that buying behavior appears random despite having individually fixed underlying propensities and a regression to the mean.

Furthermore, notice that most of the buyers in the none and heavy buyer categories stay in their buyer category for the consecutive quarter. While more movement can be seen in the light and medium categories. This suggests that either zero buyers are highly satisfied with their non-purchasing behavior or consume other brands in the category. Meanwhile, stable heavy buyers reflect a relatively higher level of loyalty and consistent engagement with the product. On the other hand, buyers in the light and medium categories suggest that they are more influenced by factors such as price changes, special offers, or new product features. Notably, light buyers are particularly prone to shifting or churning, often transitioning into non-buyers.

In general, this means that consumers of the margarine brand are rarely loyal. This however is to be expected due to natural moderation, that is, heavy and medium buyers tend to buy less over time, and at the same time, none and light buyers purchase more. Thus, there is a big opportunity for the brand to target the zero buyers in hopes of them transitioning to light buyers and consistently penetrating other buyer categories so that they may transition to a higher buyer category, especially those in the light buyer category.

Quarter 5

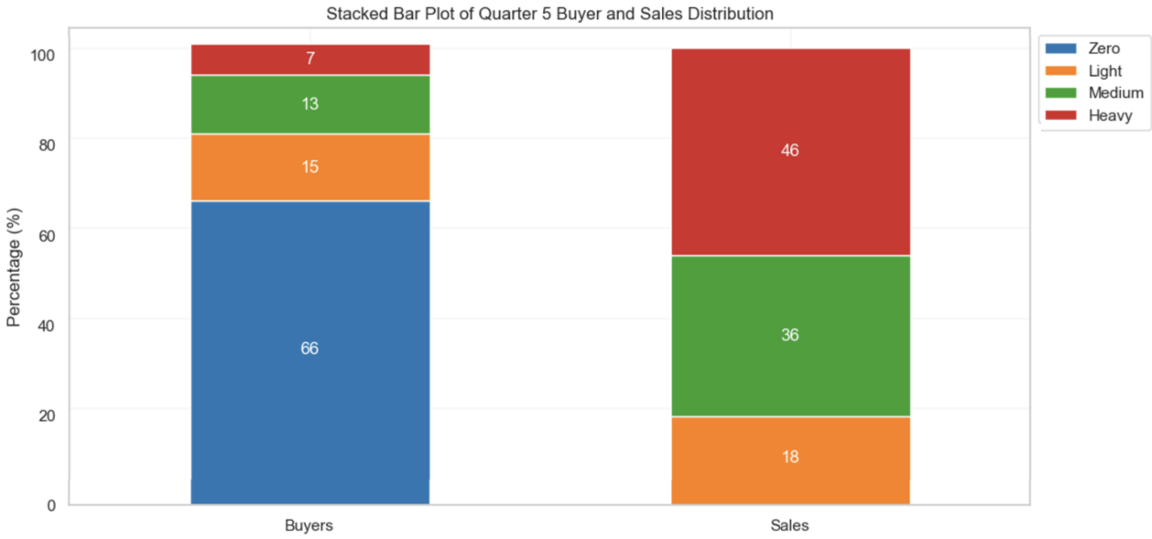


Figure 5. Quarter 5 buyer and sales distribution

As shown in figure 5, zero buyers have the largest buyer distribution and heavy buyers contribute to almost half of the sales percentage. Considering only the brand buyers, it can be observed that the data is quite close to Gerald Goodhart's 20:30:50 law. From the study's definition of buyer categories, it can be seen that the margarine brand has 43:37:20 buyers accounting for 18:36:46 purchases. In detail, the following observations and implications of the dataset are made:

Buyer Distribution (Left Column):

- Non-buyers make up 66% of the total buyers, which is about two-thirds.
- Light buyers account for around 15% of the total buyers, medium buyers make up approximately 13%, and heavy buyers constitute around 7%.

This indicates that non-buyers represent the largest group. Given their size, there is significant potential to convert them into light or even medium buyers.

Sales Distribution (Right Column):

- Heavy buyers contribute to approximately 46% of the total sales, nearly half of the overall sales.

- Light buyers contribute 18%, while medium buyers contribute around 36%.

Excluding non-buyers, heavy buyers make up about 20% (7/(7+13+15)) of the active buyer group but contribute only 46% of sales. This does not conform to the 20/80 Pareto rule but rather to a 20/50 pattern, indicating that heavy buyers in this data set do not contribute as much (80%) as traditionally expected. Conversely, while light buyers contribute only 18% of sales, their large numbers suggest a significant potential for growth. By converting non-buyers into light buyers, we could see a considerable increase in sales.

Negative Binomial Distribution Analysis

Recall the two key findings in figure 4, that is,

- purchases of buyers in the successive quarter are as-if random over time (Poisson distribution) and
- mean rates of buyer purchasing differ in the long run (Gamma distribution)

That said, an NBD analysis may be performed as the two assumptions of the model are met. This analysis is done to verify whether the buying behavior of the brand's consumer base is behaving normally.

		Count of Buyers		Count of Sales	
Number of Purchases (r)	Expected population proportion	Observed	Theoretical: Multiply by n=14471 for NBD Norm	Observed	Theoretical
0	0.6586	9531	9531	0	0
1	0.1576	2103	2281	2103	2281
2	0.0749	1161	1083	2322	2167
3	0.0414	674	600	2022	1799
4	0.0246	414	355	1656	1422
5	0.0151	244	219	1220	1095
6	0.0096	134	138	804	831
7	0.0062	70	89	490	623
8	0.0040	52	58	416	464
9	0.0026	38	38	342	344
10	0.0018	13	25	130	253
11	0.0012	17	17	187	186
12	0.0008	8	11	96	136
13	0.0005	6	8	78	100
14	0.0004	1	5	14	73
15	0.0002	1	4	15	53
16	0.0002	2	2	32	38
17	0.0001	0	2	0	28
18	0.0001	1	1	18	20
19	0.0001	0	1	0	15
20	0.0000	1	1	20	11
20+	0.0000	0	1	0	25
Sum	0.1576	14471	14471	11965	11962

Table 1. Observed count of buyers and sales in quarter 5 and its corresponding theoretical values using the NBD model

From table 1, it can be seen that the observed and theoretical values of the zero purchase class are the same with some discrepancies in all other purchase classes. It is to be expected that the zero purchase class has the same values as it's the base value the NBD model uses.

The NBD calculation yields the expected population proportions as shown in the second column of the table above. These proportions were then multiplied by the total number of category buyers resulting in the theoretical or expected number of buyers. Consequently, the theoretical count of sales per purchase class was calculated by getting the product of the purchase class and its corresponding theoretical buyer count estimates.

To further understand and visually see the magnitude of the difference between the observed and theoretical values, a bar chart was generated.

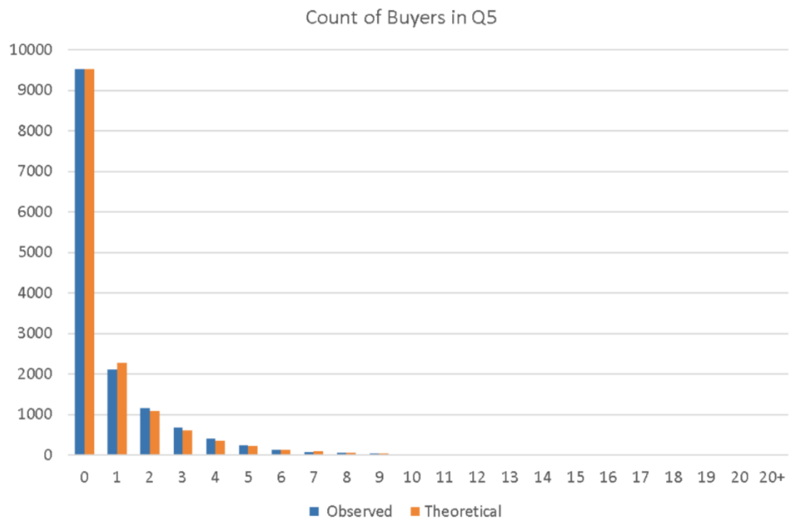


Figure 6. Observed vs. theoretical count of buyers in quarter 5

From the above chart of theoretical buyers compared to observed buyers, the following findings are found:

- There is a good fit between the theoretical and observed buyer distributions, with only minor differences, demonstrating the NBD model's strong predictive power.
- The proportion of non-buyers is significantly higher than other buyer frequencies—more than four times that of 1-time buyers—highlighting the

potential to convert non-buyers into 1-time or even 2-time buyers. Targeted marketing activities could facilitate this conversion.

- There is an overall decrease in the number of buyers as the number of purchases increases.

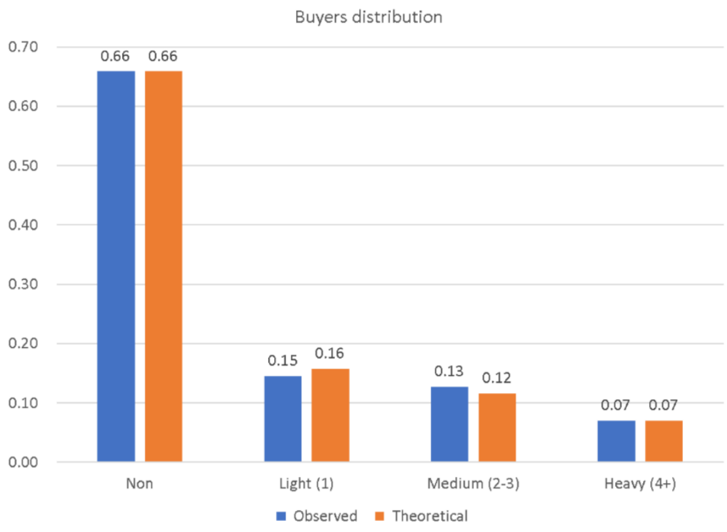


Figure 7. Buyer classification distribution in quarter 5

Aggregating the data by buyer category and comparing the distribution of theoretical and observed buyers, it can be observed that:

- Non-buyers represent the largest portion in both observed and theoretical distributions, at around 66%.
- For light buyers, observed buyers account for 15%, while theoretical buyers make up 16%, showing only a small difference.
- A similar situation is observed for medium buyers, where observed buyers constitute 13%, and theoretical buyers represent 12%, again with a small gap.
- Both observed and theoretical distributions show that heavy buyers account for 7%.

This analysis underscores the significant potential to grow sales by targeting non-buyers and converting them into active buyers, as well as the reliability of the NBD model in predicting buyer behavior within the data set. More than that, the NBD analysis is a manifestation that the brand is performing closely to its expected values.

This is an assurance that the activities done and performed in the past year and a quarter have been effective.

Buyer Flow Analysis

Lastly, an analysis was conducted in the final two quarters (Q4 and Q5) to understand the transitions between different buyer classes, including none or zero, light, medium, and heavy buyers.

		Q5			
	TOTAL	Q5 NB	Q5 L	Q5 M	Q5 H
Q4 NB	9661	8205	1294	150	12
Q4 L	3035	1164	1442	374	55
Q4 M	1051	138	422	361	130
Q4 H	724	24	106	203	391

Table 2. Buyer flow table from Q4 to Q5

As observed in table 2, most zero and heavy buyers in quarter 4 tend to stay in the same buyer category in quarter 5. Meanwhile, more than half of light and medium buyers in quarter 4 moved to other buyer categories but mostly to its neighboring categories. The diagonal values in the last four columns indicate the number of buyers who remained in the same category from quarter 4 to quarter 5. The other cells in each row indicate where the other buyers in the buyer category in quarter 4 shifted in quarter 5. For example, out of the 3035 light buyers in quarter 4, 1442 remained in the light buyer category while 1164 and 374 shifted to zero and medium buyers in quarter 5, respectively.

Generating the bar chart of the above table to see the proportion of buyers staying and shifting visually, we get the figure below.

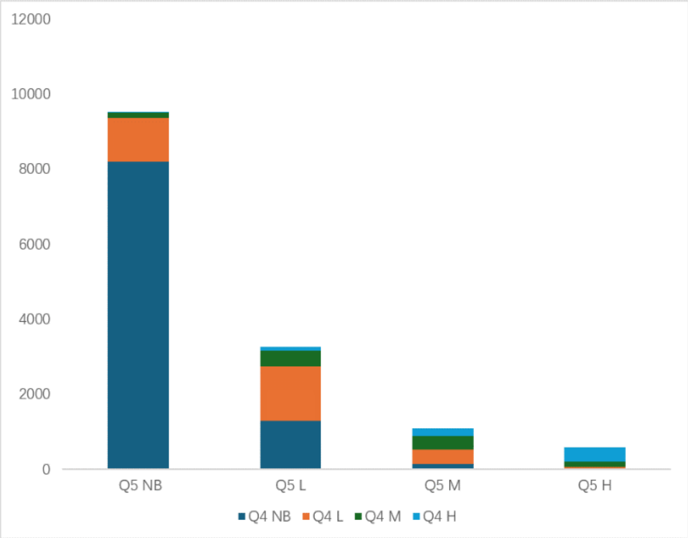


Figure 8. Transition of Buyer Classes from Q4 to Q5

Figure 8 shows the composition of each buyer class in Q5. This figure shows a detailed breakdown of each Q5 class by Q4 buyer category, including None Buyer (NB), Light Buyer (L), Medium Buyer (M), and Heavy Buyer (H).

From the chart, it is clear that the None Buyers group shows a stable nature. A large proportion of the None Buyers in Q5 were already None Buyers in Q4, meaning that many customers who didn't buy in Q4 continued to refrain from buying in Q5.

The Light Buyers in Q5 show a more diverse background, with contributions from all categories in Q4. This suggests that some customers who were previously Medium or Heavy Buyers reduced their purchase frequency, while some None Buyers from Q4 started to make occasional purchases in Q5.

The majority of Medium Buyers in Q5 have moved from the Light Buyers category in Q4. This shift indicates an improvement in buying behavior, suggesting that an increasing number of buyers are beginning to make purchases more frequently.

Finally, the majority of Heavy Buyers in Q5 came from the Heavy Buyers category in Q4, with a small contribution from Medium Buyers.

Repeat Buying

Q4 - Q5 Repeat Buyers (RB)		
#		3484
%		72%
Q5 Sales from RB		
# RB Sales		9816
% Sales from RB		82%

Table 3. Number and percentage of repeat buyers and sales in quarter 5

As observed in table 3, a significant amount of buyers and sales in quarter 5 comes from repeat purchases but the data also depicts the natural phenomenon of eroding customers over time. While it is not certain whether the customer churned or simply chose not to purchase in quarter 5, the brand must consistently bring new customers to buy the product to lower or stabilize the rate of erosion.

Discussion

This section of the study will discuss some of the law-like patterns discussed by Byron Sharp in his book entitled, *How Brands Grow*, and was observed in the margarine brand's data.

Double Jeopardy Law

Some believe that brand growth can be achieved in multiple ways – customer acquisition, retention, or a combination of the two. However, the Double Jeopardy law rebuts this and claims that brand growth stems from getting more customers to purchase from the brand, given at any period (Romaniuk et al., 2021). The margarine brand is not an exception to this law. While relative performance compared to its competitors cannot be verified, the findings in brand performance metrics indicate the driving force of growth – that is, through brand penetration. According to Brusselmans et al. (2014), since the brand's penetration rate is about 30%, the brand is considered to be one of the leading brands and will most likely have a significant market share. Moreover, the brand's purchase frequency and buyer flow further emphasize that while some buyers are loyal, most are not in terms of buying habits and attitude toward the brand. Thus, the brand's penetration rate must be kept or increased since this will also inadvertently affect retention (Sharp, 2010).

Pareto Law

From the findings in the bar chart of the buyer and sales distribution in figure 5, it was discovered that the Pareto's ratio of 80/20 law is a myth and an exaggeration of the true share of heavy buyers in most brands. For the margarine brand, only 46% of sales come from the heaviest 20% of brand buyers. The findings conform to the typical Pareto share where the Pareto share is expected to be around 50% over a year (Sharp, 2010). That said, this again emphasizes the importance of penetration over retention as light and medium buyers contribute to about half of sales and acquisition is less costly and more efficient.

Law of Buyer Moderation

By investigating the buyer distribution category across five quarters and analyzing the buyer flow of the latest two quarters, a regression to the mean phenomenon was observed. It showed how in subsequent periods, buyers tend to shift to its neighboring buyer categories. Some zero buyers become buyers, light buyers purchase more, and heavy buyers purchase less (Murrell, 2021). This reflects the high potential of zero and light buyers as well as the occurrence of erosion over time.

Duplication of Purchase Law

From the buyer distribution, it was seen that zero buyers make up more than half of the population proportion. So, while there is a large number of category buyers, the brand only has about a 30% share of the overall market due to the existence of other brands. Not only do brands share the same customer base but are also reliant on each brand's penetration (Sharp et al., 2002; Sacramento, 2024). More than that, it is also important to bear in mind that not only do brands belonging to a category compete but as well as brands of other categories may be a substitute. For instance, buyers of margarine and butter may interchangeably purchase from either category.

From the aforementioned scientific laws, it follows that even the margarine brand have margins or a guide as to how the brand's activity and performance will go. And the key to changing this margin and achieve growth is by mainly improving the brand's penetration rate in the market.

Recommendations

Several important insights emerged from the buyer flow between Q4 and Q5. In particular, the None Buyers showed significant stability, with a large portion of the customer base consistently failing to purchase in either quarter. Similarly, Heavy Buyers demonstrated high levels of loyalty and engagement with the product. Meanwhile, the light and moderate shoppers showed considerable volatility, with many customers moving between different buyer classes. This instability is even more pronounced among the Light Buyers, who frequently switch to the Non-Buyer category, indicating that they are more likely to churn.

It is critical to implement a targeted marketing strategy based on these findings. For light and medium buyers, marketing efforts should focus on increasing the frequency of their purchases. These consumers have made initial purchases but have not yet established regular buying habits. we can offer tailored incentives such as discounts, personalized promotions, and special offers to encourage more consistent buying behavior. At the same time, for heavy buyers, it is critical to develop a retention strategy to prevent them from downgrading. The strategy could include offering premium services, exclusive offers, or loyalty rewards that not only reinforce their buying behavior but also increase their loyalty to the brand.

While several potential strategies have been proposed to increase buyer engagement and retention, with limited marketing budgets, it is difficult to prioritize and effectively implement these strategies without knowing their potential return on investment.

Follow-up analysis can incorporate larger, more diverse data sets, which can help provide a more comprehensive and robust analysis that leads to a more holistic strategy. In addition, using advanced analytics and modeling techniques to simulate the financial outcomes of different marketing strategies can provide valuable insights into which approaches are most cost-effective. By quantifying the expected return on investment for different strategies, organizations can better allocate limited marketing resources to ensure that their efforts are both efficient and economically viable. This strategic approach not only optimizes resource utilization but also paves the way for smarter, more effective marketing decisions.

Conclusion

Overall, the brand is performing well and conforms to the scientific laws discussed by Sharp in his book entitled, *How Brands Grow*. As emphasized throughout the report, brand growth comes from new and light buyers thus it is important to stabilize or increase the market penetration rate (Fallarme, 2020). At the same time, reactivating lapsed customers and nudging existing buyers should also be done since mental availability and repeat buying activity are also important concepts that contribute to product sales. Furthermore, the brand should also expect the shifting, churning, and/or erosion of customers, whether it be to acquire new customers to compensate lost customers, improve loyalty and retainment of the product, or to keep them as brand buyers, regardless of which buying category they fall into. Lastly, it is important to note that the brand is competing with other brands from the same and different categories. Therefore, consistent widespread market efforts should be done to increase the brand's awareness and reach.

To optimize marketing efforts and drive growth, the following strategies are recommended:

Consistency: Build and strengthen the memory structures of consumers by avoiding too much rebranding. Instead, focus on saliency and distinctiveness and continuously promote these.

Create or reenforce the brand's category entry point: To increase mental availability with the consumers, the brand must identify when they want buyers to remember the brand's margarine. For example, by associating margarine as a staple that can be used in cooking multiple breakfast dishes.

Target all Category Buyers: Implement sophisticated mass marketing strategies. Tailored incentives, such as discounts and personalized promotions, to increase purchase frequency among light and medium buyers, who have shown initial interest but lack consistent buying habits. Develop retention strategies to prevent downgrading, offering premium services, exclusive offers, or loyalty rewards to reinforce their purchasing behavior and strengthen brand loyalty.

Optimize Marketing Budget: Given limited resources, it is crucial to prioritize marketing strategies based on potential return on investment (ROI). Follow-up analyses should incorporate larger, more diverse datasets and advanced modeling techniques to simulate financial outcomes, guiding more effective resource allocation.

References

Bala, P. K. (n.d.). *Purchase-driven Classification for Improved Forecasting in Spare Parts Inventory Replenishment*. International Journal of Computer Applications, 10.

Brusselmans, G., Blasberg, J., & Root, J. (2014, March 19). *The Biggest Contributor to Brand Growth*. Bain & Company. https://www.bain.com/insights/the-biggest-contributor-to-brand-growth/

Fallarme, D. (2020, August 23). *How Brands Grow: A short summary*. The Marketing Student. https://www.themarketingstudent.com/how-brands-grow/

Knox, S. (1998). *Loyalty-based segmentation and the customer development process*. European Management Journal, 16(6), 729–737. https://doi.org/10.1016/S0263-2373(98)00049-8

Heavy, medium and light shoppers. (n.d.). NielsonIQ CPG Dictionary. https://microsites.nielseniq.com/cpg-dictionary/dictionary/heavy-medium-and-light-shoppers/

Murrell, A. (2021, August 16). *Byron Sharp: How Brands Grow*. Alex Murrell. https://www.alexmurrell.co.uk/summaries/byron-sharp-how-brands-grow

Romaniuk, J., Dawes, J., & Faghidno, S. (2021, May). *The Double Jeopardy Law in B2B shows the way to grow*. Ehrenberg-Bass Institute for Marketing Science. https://marketingscience.info/the-double-jeopardy-law-in-b2b-shows-the-way-to-grow/

Sacramento, J. (2024, February 22). *Scientific Laws of Marketing with Professor Byron Sharp*. Dreamdata. https://dreamdata.io/blog/scientific-laws-of-marketing-byron-sharp

Sharp, B. (2010). *How brands grow: what marketers don't know*. Oxford University Press.

Sharp, B., Wright, M., & Goodhardt, G. (2002). *Purchase Loyalty is Polarised into Either Repertoire or Subscription Patterns*. Australasian Marketing Journal (AMJ). 10. 7-20. 10.1016/S1441-3582(02)70155-9.

Appendix A

Python Scripts

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

#Loading the excel file
def load_data(file_name):
  try:
    data = pd.ExcelFile(file_name)
    if len(data.sheet_names) == 1:
      print("\nData loaded successfully!\n")
      return data
    else:
      print("\nPlease enter the name of the sheet you wish to use. The sheet names
available are the following:", data.sheet_names)
      sheet = input()
      data = pd.read_excel(file_name, sheet_name=sheet)
      print("\nData loaded successfully!\n")
      if data.iloc[22,1] == 0 or data.iloc[22,1] == 1:
        data.iloc[22,0] = int(21)
      elif data.iloc[22,1] == 2:
        data.iloc[22,0] = int(23)
      elif data.iloc[22,1] == 3:
        data.iloc[22,0] = int(25)
      elif data.iloc[22,1] == 4:
        data.iloc[22,0] = int(27)
      elif data.iloc[22,1] == 5:
        data.iloc[22,0] = int(29)
      return data
  except Exception as e:
    print(f"An error occurred: {e}")
    return None

def statistics(data):
  #total # of category buyers
  total_cat_buyers = data.iloc[0,1]
```

```
    #total # of brand buyers
    brand_buyers = data.iloc[2:,1]
    total_brand_buyers = brand_buyers.sum()

    #total # of purchases
    class_purchases = data.iloc[1:, 0:2].prod(axis=1)
    total_purchases = class_purchases.sum()

    #brand performance metrics
    b_zero = data.iloc[1,1]/total_cat_buyers
    b = total_brand_buyers/total_cat_buyers
    w = total_purchases/total_brand_buyers
    m = b*w

    #buyer distribution
    zero_buyers = round((data.iloc[1,1]/total_cat_buyers)*100)
    light_buyers = round((data.iloc[2,1]/total_cat_buyers)*100)
    medium_buyers = round((data.iloc[3:5,1].sum()/total_cat_buyers)*100)
    heavy_buyers = round((data.iloc[5:,1].sum()/total_cat_buyers)*100)

    #sales distribution
    zero_buyers_sales = round((class_purchases[1]/total_purchases)*100)
    light_buyers_sales = round((class_purchases[2]/total_purchases)*100)
    medium_buyers_sales = round((class_purchases[2:4].sum()/total_purchases)*100)
    heavy_buyers_sales = round((class_purchases[4:].sum()/total_purchases)*100)

    raw = {
        'parameters' : ['total_cat_buyers', 'total_brand_buyers', 'total_purchases', 'b_zero', 'b',
'w', 'm',
                'zero_buyers', 'light_buyers', 'medium_buyers', 'heavy_buyers',
                'zero_buyers_sales', 'light_buyers_sales', 'medium_buyers_sales',
'heavy_buyers_sales'],
        data.columns[1] : [total_cat_buyers, total_brand_buyers, total_purchases,
round(b_zero,2), round(b,2), round(w,2), round(m,2),
              zero_buyers, light_buyers, medium_buyers, heavy_buyers,
              zero_buyers_sales, light_buyers_sales, medium_buyers_sales,
heavy_buyers_sales]
    }

    info = pd.DataFrame(raw)
    return (info)
```

```
file_name = input("\nPlease enter the name of your dataset.xlsx: ")
Q1_data = load_data(file_name)
Q2_data = load_data(file_name)
Q3_data = load_data(file_name)
Q4_data = load_data(file_name)
Q5_data = load_data(file_name)

Q1_data['cumulative_freq'] =
round((Q1_data['Q1'].iloc[2:].cumsum()/Q1_data.iloc[2:,1].sum())*100)

#Specifying certain settings of some graph components
sns.set(font_scale = 1)
sns.set_style("whitegrid", {'grid.color': 'whitesmoke'})
#Specifying color palette details
enmax_palette = ["#1F77B4", "#f7022a", "#056eee", "#952e8f", "#FF7F0E", "#2CA02C",
"#D62728"]
sns.set_palette(palette=enmax_palette)

freq = sns.barplot(Q1_data.iloc[2:], x = 'r', y = 'cumulative_freq', color = 'C0')
#Adding labels in the bar graph
freq.set_xlabel('Purchase Class')
freq.set_ylabel('Percentage')
freq.set_title('Quarter 1 Cumulative Frequency of Buyers per Purchase Class',
fontdict={'size': 15})
plt.show()

Q1_stat = statistics(Q1_data)
Q2_stat = statistics(Q2_data)
Q3_stat = statistics(Q3_data)
Q4_stat = statistics(Q4_data)
Q5_stat = statistics(Q5_data)

#Creating one dataframe containing the summary of each quarter
q1q2 = pd.merge(Q1_stat, Q2_stat, on = 'parameters')
q1q2q3 = pd.merge(q1q2, Q3_stat, on = 'parameters')
q1q2q3q4 = pd.merge(q1q2q3, Q4_stat, on = 'parameters')
all_qtrs = pd.merge(q1q2q3q4, Q5_stat, on = 'parameters')
qtr_summary = all_qtrs.T[1:]
qtr_summary.columns = all_qtrs['parameters']
qtr_summary = qtr_summary.reset_index(names='quarters')
qtr_summary
fig, ax = plt.subplots(figsize=(12, 8))
```

```
sns.barplot(qtr_summary, x = 'quarters', y = 'b', label = 'Penetration Rate', color = 'C1')
sns.pointplot(qtr_summary, x = 'quarters', y = 'w', label = 'Purchase Rate', color = 'C2')
sns.pointplot(qtr_summary, x = 'quarters', y = 'm', label = 'Purchase Frequency', color =
'C3').legend(loc='upper left', bbox_to_anchor=(1, 1))

for container in ax.containers:
  ax.bar_label(container)

ax.set_ylabel('Value', size=15)
ax.set_xlabel('Quarter', size=15)
ax.set_title('Quarterly Brand Performance Metrics', size=18)
ax.margins(x=0.02) # less white space at the left and the right

# Customizing y-axis ticks
plt.yticks(np.arange(0, 2.8, 0.1))

sns.despine()
plt.show()
# Plotting
fig, ax = plt.subplots(figsize=(12, 6))

# Create a stacked bar plot
qtr_summary[['quarters', 'zero_buyers', 'light_buyers', 'medium_buyers',
'heavy_buyers']].plot(kind='bar', stacked=True, ax=ax, color=['C0', 'C4', 'C5', 'C6'])

#Adding labels and title
ax.set_xticklabels(qtr_summary['quarters'], rotation = 0)
ax.set_ylabel('Percentage (%)')
ax.set_xlabel('Quarter')
ax.set_title('Stacked Bar Plot of the Buyer Distribution')

#Annotating the bars with their values
for container in ax.containers:
  ax.bar_label(container, label_type='center', size = 12, color = 'white')

# Custom legend labels
custom_labels = ['Zero Buyers', 'Light Buyers', 'Medium Buyers', 'Heavy Buyers']

#Moving the legend outside the plot
ax.legend(custom_labels, loc='upper left', bbox_to_anchor=(1, 1))

plt.show()
```

```
#Assigning row numbers to each DataFrame to ensure alternating rows in
concatenating
data1 = qtr_summary[['quarters', 'zero_buyers', 'light_buyers', 'medium_buyers',
'heavy_buyers']][qtr_summary['quarters'] == 'Q5'].copy()
data2 = qtr_summary[['quarters', 'zero_buyers_sales', 'light_buyers_sales',
'medium_buyers_sales', 'heavy_buyers_sales']][qtr_summary['quarters'] == 'Q5'].copy()
data1['RowID'] = range(0, 2 * len(data1), 2)
data2['RowID'] = range(1, 2 * len(data2) + 1, 2)
# Concatenate the DataFrames and sort by the RowID
combined = pd.concat([data1, data2]).sort_values(by='RowID').drop('RowID',
axis=1).reset_index(drop=True)

# Plotting
fig, ax = plt.subplots(figsize=(12, 6))

# Create a stacked bar plot
combined.plot(kind='bar', stacked=True, ax=ax, color=['C0', 'C4', 'C5', 'C6'])

#Adding labels and title
ax.set_xticklabels(['Buyers', 'Sales'], rotation = 0)
ax.set_ylabel('Percentage (%)')
ax.set_title('Stacked Bar Plot of Quarter 5 Buyer and Sales Distribution')

#Annotating the bars with their values
for container in ax.containers:
   ax.bar_label(container, label_type='center', size = 12, color = 'white')

# Custom legend labels
custom_labels = ['Zero', 'Light', 'Medium', 'Heavy']

#Moving the legend outside the plot
ax.legend(custom_labels, loc='upper left', bbox_to_anchor=(1, 1))

plt.show()

#Define a function, converts the dataset into a plottable sankey chart format.

def process_and_create_sankey_data(df, source_quarter, target_quarter):
   # Aggregate and create new columns
   df_aggregated = df.copy()
```

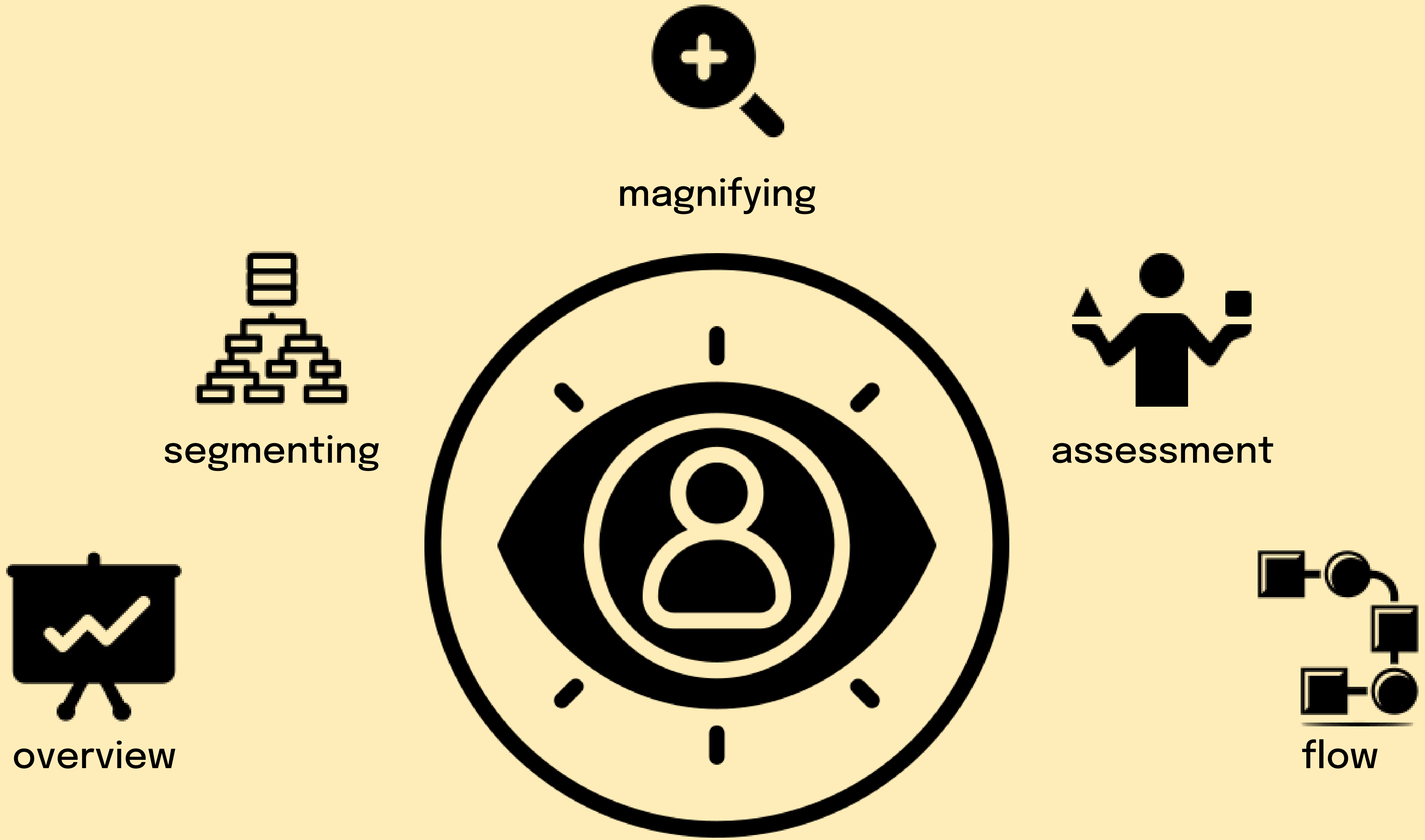
```
    df_aggregated['none buyer'] = df_aggregated[0]
    df_aggregated['light buyer'] = df_aggregated[1]
    df_aggregated['medium buyer'] = df_aggregated[2] + df_aggregated[3]
    df_aggregated['heavy buyer'] = df_aggregated.loc[:, 4:'20+'].sum(axis=1)

    # rename
    index_map = {0: 'none buyer', 1: 'light buyer', 2: 'medium buyer', 3: 'medium buyer'}
    df_aggregated.index = df_aggregated.index.map(lambda x: index_map.get(x, 'heavy
buyer'))

    df_numerical = df_aggregated[['none buyer', 'light buyer', 'medium buyer', 'heavy
buyer']]
    df_grouped = df_numerical.groupby(df_numerical.index).sum()

    index_order = ['none buyer', 'light buyer', 'medium buyer', 'heavy buyer']
    df_final = df_grouped.reindex(index_order)

    # Creating source-target-weight lists
    sources = []
    targets = []
    weights = []

    for source, row in df_final.iterrows():
        for target, value in row.items():
            if value > 0:
                sources.append(f"{source_quarter} {source}")
                targets.append(f"{target_quarter} {target}")
                weights.append(value)

    # Creating a DataFrame for Sankey
    sankey_data = pd.DataFrame({
        'Source': sources,
        'Target': targets,
        'Weight': weights
    })

    return sankey_data

import plotly.graph_objects as go

# Using Functions to Generate Data
sankey_data_Q1Q2 = process_and_create_sankey_data(Q1_data, 'Q1', 'Q2')
```

```
sankey_data_Q2Q3 = process_and_create_sankey_data(Q2_data, 'Q2', 'Q3')
sankey_data_Q3Q4 = process_and_create_sankey_data(Q3_data, 'Q3', 'Q4')
sankey_data_Q4Q5 = process_and_create_sankey_data(Q4_data, 'Q4', 'Q5')

# Merging data
combined_sankey_data = pd.concat([sankey_data_Q1Q2, sankey_data_Q2Q3,
sankey_data_Q3Q4, sankey_data_Q4Q5])

# Create unique lists of nodes and connections
all_nodes = pd.concat([combined_sankey_data['Source'],
combined_sankey_data['Target']]).unique()
node_dict = {node: i for i, node in enumerate(all_nodes)}

# Update source and target to index
combined_sankey_data['Source'] = combined_sankey_data['Source'].apply(lambda x:
node_dict[x])
combined_sankey_data['Target'] = combined_sankey_data['Target'].apply(lambda x:
node_dict[x])

# Define colors for each buyer type
buyer_type_colors = {
  'none': "rgba(31, 119, 180, 0.8)",   # blue
  'light': "rgba(255, 127, 14, 0.8)",  # orange
  'medium': "rgba(44, 160, 44, 0.8)",  # green
  'heavy': "rgba(214, 39, 40, 0.8)"    # red
}

# Generate link colors
link_colors = []
for index, row in combined_sankey_data.iterrows():
  source_node = list(node_dict.keys())[int(row['Source'])]
  buyer_type = source_node.split()[1]
  link_colors.append(buyer_type_colors[buyer_type.split()[0]])

# plot sankey
fig = go.Figure(data=[go.Sankey(
  node=dict(
    pad=15,
    thickness=20,
    line=dict(color="black", width=0.5),
    label=list(node_dict.keys()),
    color="gray"
```

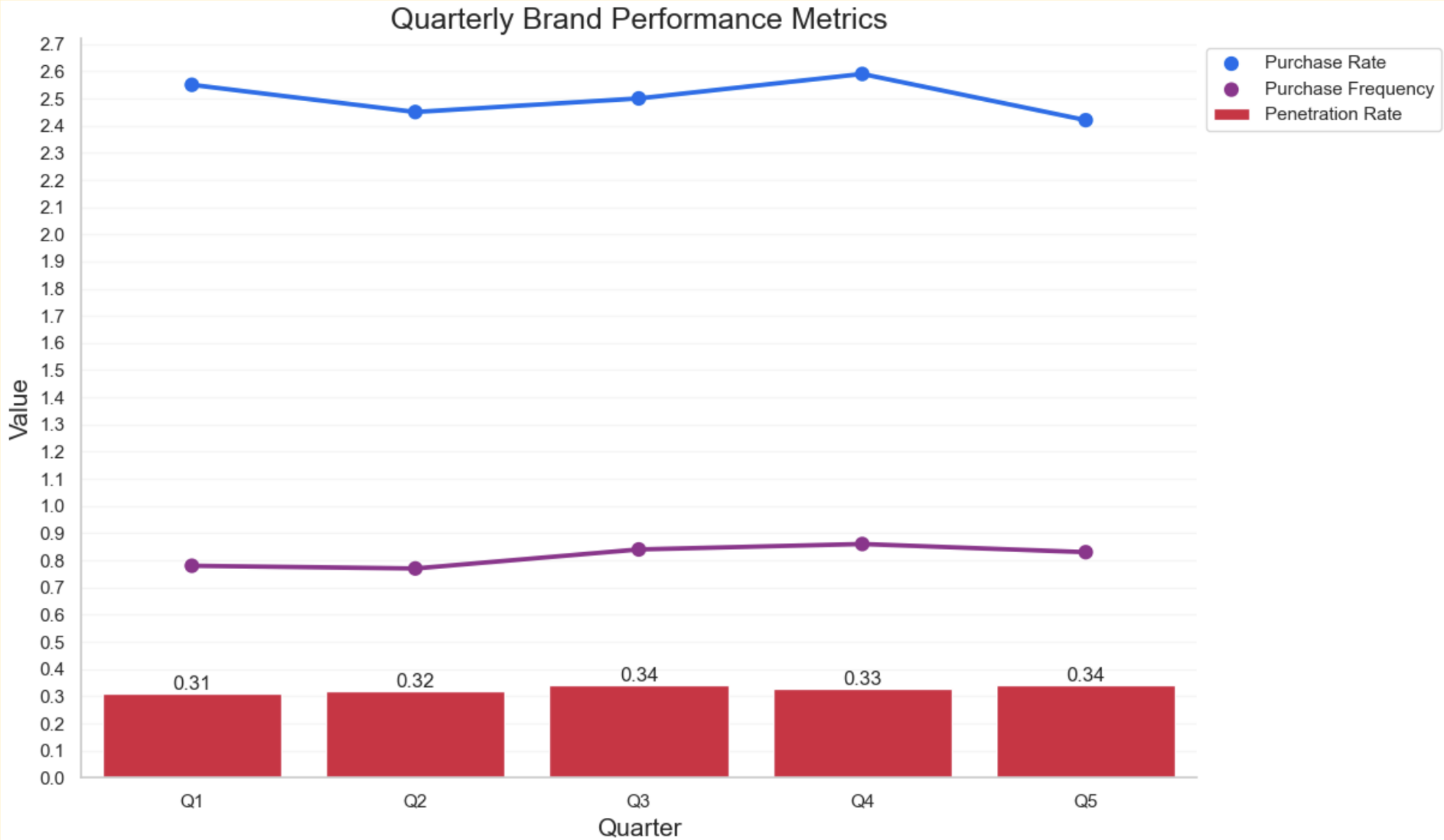
```
  ),
  link=dict(
    source=combined_sankey_data['Source'],
    target=combined_sankey_data['Target'],
    value=combined_sankey_data['Weight'],
    color=link_colors
  )
)])

fig.update_layout(title_text="Comprehensive Sankey Diagram from Q1 to Q5",
font_size=10)
fig.show()
```

Margarine Margins

Method



Overview



Small fluctuations in the metrics

relatively better performance in quarters 3 and 4

Behavioral Segmentation

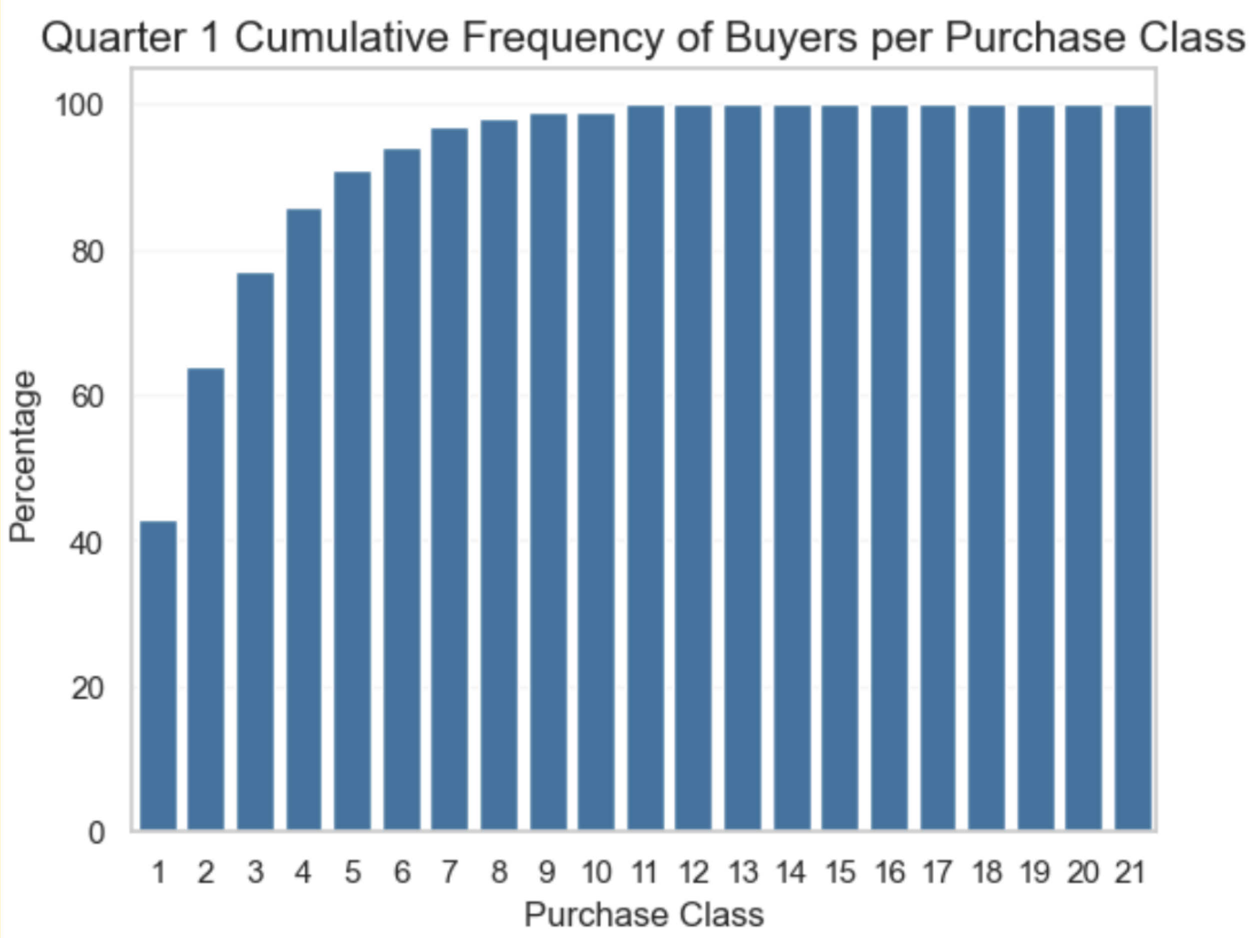
according to

PURCHASE FREQUENCY

TOP 15%	BOTTOM 25%	15% TO 75%
HEAVY	LIGHT	MEDIUM

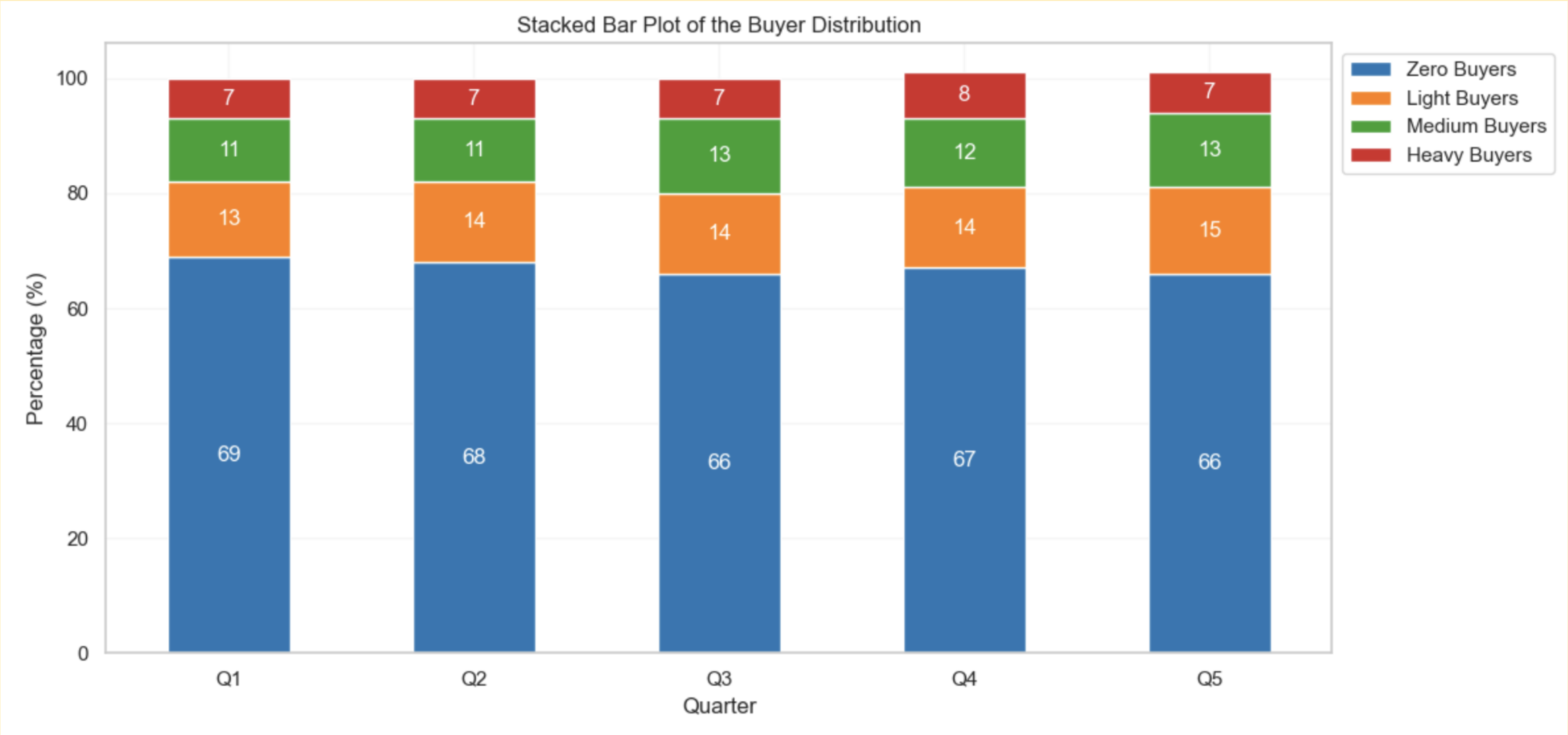
purchases:

4+	1	2-3



Behavioral Segmentation

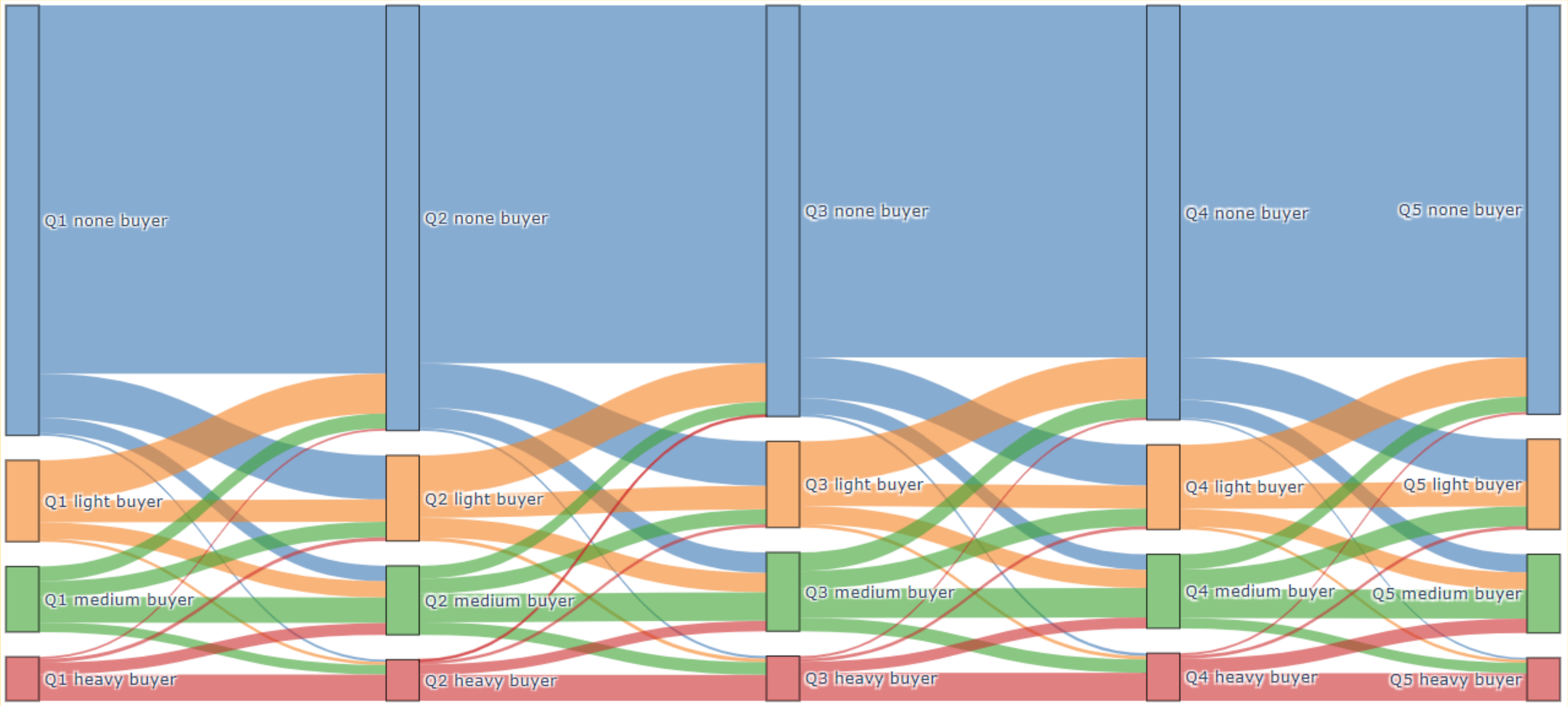
Margarine Margins



somewhat **STABLE** buyer distribution across all quarters

Behavioral Segmentation

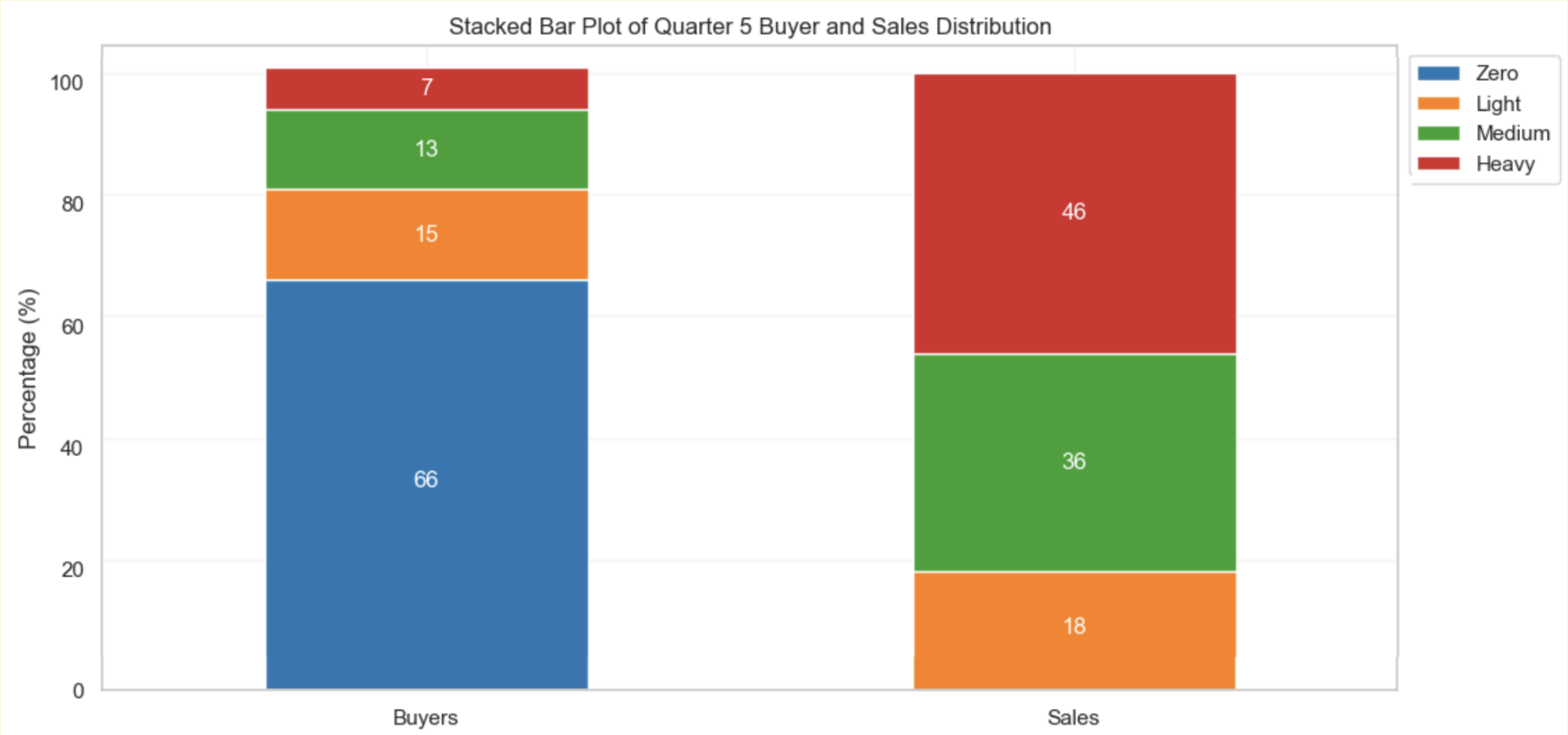
Margarine Margins



Buyers typically **follow similar buying behaviors** across different quarters

Regression to the mean

Quarter 5



None buyers have the **largest distribution**

Heavy buyers contribute **most** to **sales**

Performance Assessment Q5

Margarine Margins

		Count of Buyers		Count of Sales	
Number of Purchases (r)	Expected population proportion	Observed	Theoretical: Multiply by n=14471 for NBD Norm	Observed	Theoretical
0	0.6586	9531	9531	0	0
1	0.1576	2103	2281	2103	2281
2	0.0749	1161	1083	2322	2167
3	0.0414	674	600	2022	1799
4	0.0246	414	355	1656	1422
5	0.0151	244	219	1220	1095
6	0.0096	134	138	804	831
7	0.0062	70	89	490	623
8	0.0040	52	58	416	464
9	0.0026	38	38	342	344
10	0.0018	13	25	130	253
11	0.0012	17	17	187	186
12	0.0008	8	11	96	136
13	0.0005	6	8	78	100
14	0.0004	1	5	14	73
15	0.0002	1	4	15	53
16	0.0002	2	2	32	38
17	0.0001	0	2	0	28
18	0.0001	1	1	18	20
19	0.0001	0	1	0	15
20	0.0000	1	1	20	11
20+	0.0000	0	1	0	25
Sum	0.1576	14471	14471	11965	11962

NBD calculation:

comparison of Observed results to Theoretical benchmarks

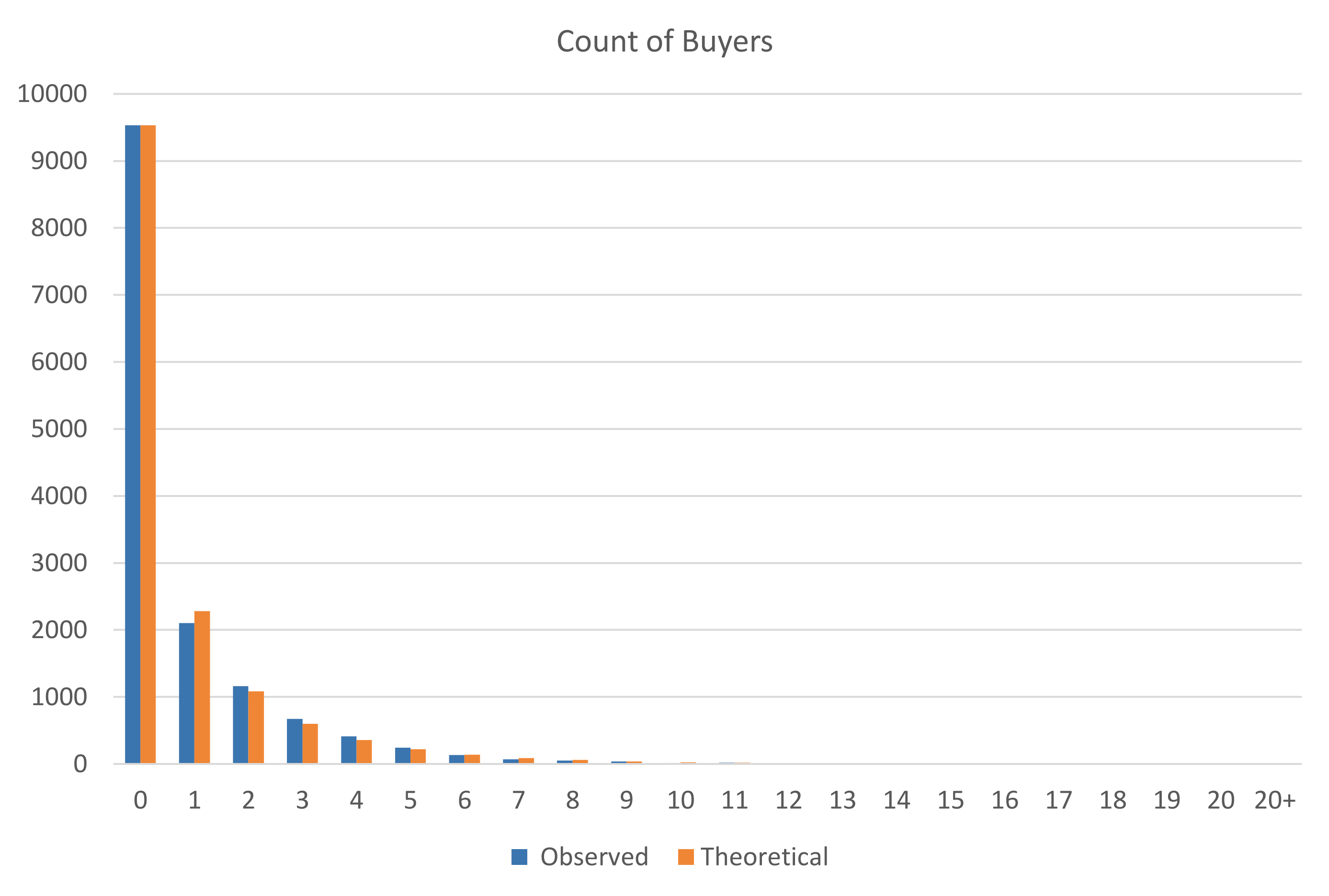
Performance Assessment Q5

Margarine Margins

Good fit

Huge portion
of zero buyers

Overall decrease

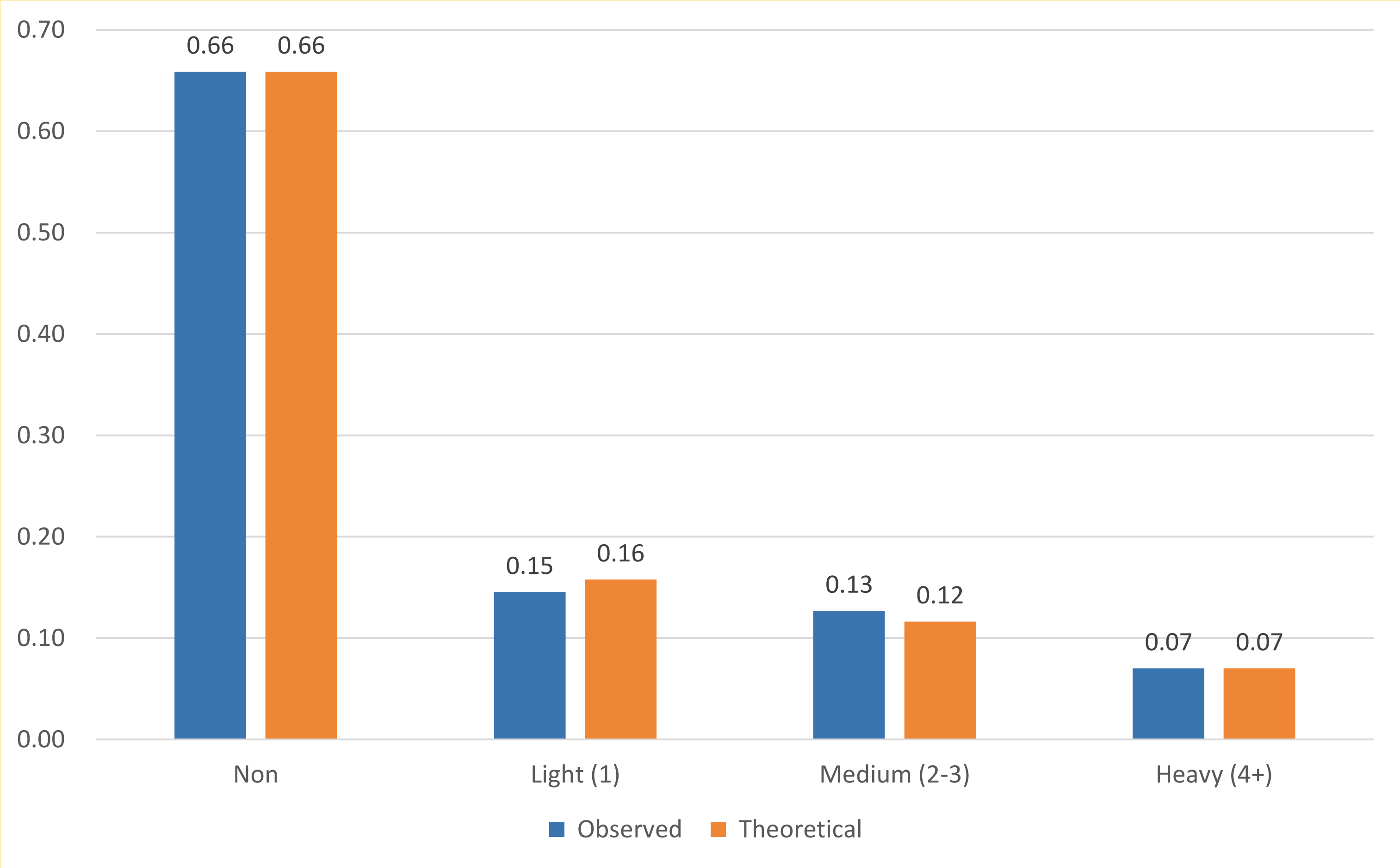


Performance Assessment Q5

Definition according to sales contribution

Large portion of zero buyers

Small gap between observed and theoretical



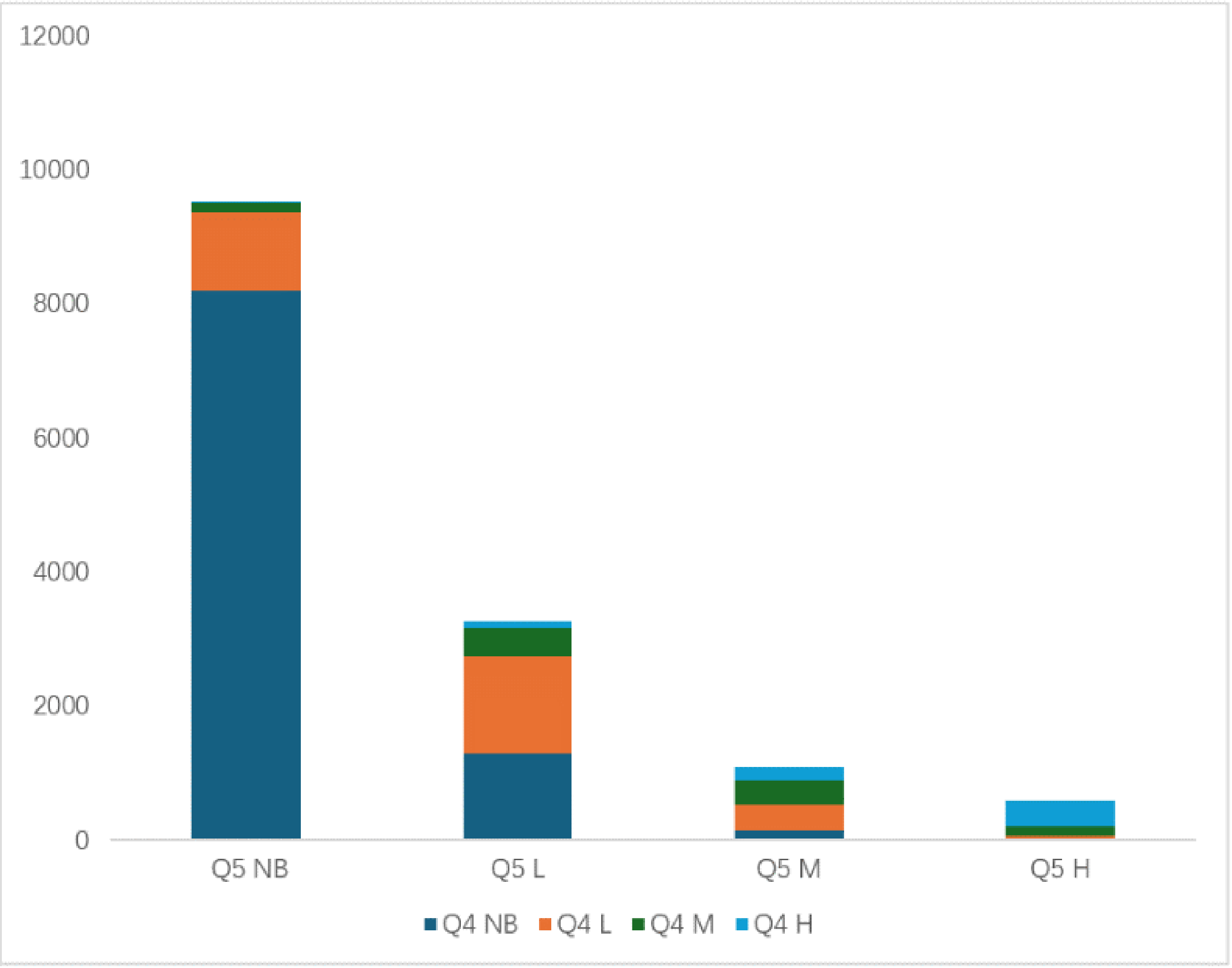
Buyer Flow Analysis

		Q5			
	TOTAL	Q5 NB	Q5 L	Q5 M	Q5 H
Q4 NB	9661	8205	1294	150	12
Q4 L	3035	1164	1442	374	55
Q4 M	1051	138	422	361	130
Q4 H	724	24	106	203	391

A considerable amount of repetitive buyers

Buyer Flow Analysis

Margarine Margins



Stability in Extremes

Volatility of Light Buyer

Indecisive Mid-Tiers

Heavy Buyer **Downgrading**

Conclusion

High value of Light Buyers

increase the **purchase frequency** of light buyers:

- personalized marketing efforts
- loyalty programs
- targeted promotions aimed at encouraging repeat purchases

None Buyers ↔ Light Buyers

retention strategies:

- follow up communications
- 2nd purchase special offers
- improve content for mental availability

Conclusion

Heavy Buyers' **High Loyalty**

maintain & enhance **relationship** with heavy buyers.

- exclusive benefits
- new product early access
- Personalized services

Dynamic Transition

monitor **buyer behaviour** and update strategies.

- use data analytics to forecast and detect shifting or churning of buyers

Thank You

Margarine Margins