Title: Analyzing Amul Butter Sales and Customer Preferences

Introduction:

Amul Butter is a flagship product of Amul, a renowned dairy cooperative based in India. Established in 1946, Amul has since grown into one of the largest dairy brands in the country, known for its high-quality dairy products and innovative marketing strategies.

Amul Butter holds a significant place in Indian households and is recognized for its rich taste, creamy texture, and nutritional value. Made from fresh and pure milk sourced from dairy farmers associated with the Amul cooperative, Amul Butter is renowned for its superior quality and consistency.

Amul Butter has become a staple ingredient in Indian cuisine and is commonly used for cooking, baking, and spreading on bread and other snacks. It is known for its versatility and is preferred by consumers for its authentic taste and natural ingredients.

Over the years, Amul Butter has garnered a loyal customer base and has earned accolades for its commitment to quality, affordability, and innovation. With its iconic packaging featuring the Amul girl mascot, Amul Butter has become a symbol of trust and reliability in the Indian dairy industry.

Overall, Amul Butter continues to be a household favorite, cherished for its rich flavor, nutritional benefits, and association with the heritage and tradition of Indian dairy farming.

Objective:

The reason for performing this analysis is as follows:

1. Which brand of butter do the consumers mostly buy?
2. What is the most appealing thing due to which people buy Amul butter?
3. Primary reason for buying butter?
4. Is the packaging of Amul butter convenient and durable?
5. How is the price of Amul butter compare to other brands?
6. What type of Amul butter are usually purchased by the consumers?

7. If a new variant of Amul butter was introduced why would the consumers want to buy it? (What should be the focus when producing a new variant of butter?)
8. Are ingredients important to the consumers when buying Amul butter?
9. What are the factors that influence customers decision to purchase the butter?
10. What do the consumers usually eat butter with? (What would they mostly buy along with butter?)
11. Is the quality of Amul butter up to the mark?
12. Would the consumers recommend the product to others?
13. Which age group usually consume butter?
14. What are the ratings of the butter? (Are the customers satisfied?)

Methodology:

I collected data through a survey that I made in google forms. I followed a structured process designed to gather insights into consumer preferences, behaviors, and perceptions related to Amul Butter. Here's a brief description of how I have conducted the survey:

a) Data Collection:

Participants were invited to complete the survey by providing their responses to the questionnaire. Respondents have completed the survey online. The survey process have included measures to ensure the validity and reliability of the data, such as attention checks or quality control measures.

b) Data Analysis:

Once the survey responses were collected, I cleaned and prepared the data for analysis. This involved checking for completeness, removing any duplicate or invalid responses, and coding the data for analysis. Statistical techniques and software tools have been used to analyze the survey data and derive meaningful insights.

c) Visualization:

After conducting the analysis, you interpreted the findings to identify trends, patterns, and key insights related to consumer preferences for Amul Butter. The results were then

compiled into a report or presentation, which may have included visualizations (e.g., charts, graphs) to effectively communicate the findings. Recommendations or implications based on the analysis were also likely included in the report to inform decision-making and future strategies related to Amul Butter.

Result and Analysis:

1. Code & Output:

```
library(ggplot2)
library(dplyr)
#Import data
data = read.csv("Customer Feedback.csv", header = TRUE)
head(data)
#removing null values
data[data == ""] = NA
#Remove rows with NA values
data = na.omit(data)
#Butter brand consumers mostly buy
mode = function(x){
  uniq = unique(x)
  uniq[which.max(tabulate(match(x, uniq)))]
}
Brand = mode(data$Butter_Brand_Preference)
Type = mode(data$Variant_Purchased)
```

```
> Brand
[1] "Amul"
```

```
> Type
[1] "Regular Salted Butter"
```

Conclusion: The most commonly bought butter variant is Amul's standard salted butter.

2. Code & Output:

```
library(tidyr)
library(dplyr)
# Split the consumption methods into separate rows
data=data %>%
  separate_rows(Appealing_Aspect, sep = ";")
#Create a frequency table of Appealing_Aspect
appealing_freq = table(data$Appealing_Aspect)
#Create a contingency table of Appealing_Aspect vs. Purchase_Decision
contingency_table = table(data$Appealing_Aspect, data$Purchase_Frequency)
#Perform chi-square test
chi_sq_test = chisq.test(contingency_table)
#Print frequency table
print("Frequency table for Appealing_Aspect:")
print(appealing_freq)
#Print contingency table
print("Contingency table for Appealing_Aspect vs. Purchase_Decision:")
print(contingency_table)
#Print chi-square test results
print("Chi-square test results:")
print(chi_sq_test)
```

```
                                 Creaminess                                      Flavor
                                        117                                         108
          Like old picture on the packaging                          Nutritional value
                                          1                                          48
                                  Packaging                                Spreadability
                                         31                                         113
When u eat it with other things it taste nice
                                          3
```

```
> print(contingency_table)

                                          Daily Monthly Never Occasionally Weekly
  Creaminess                                 15      48     4           22     28
  Flavor                                     14      47     4           18     25
  Like old picture on the packaging           0       1     0            0      0
  Nutritional value                           5      18     4            6     15
  Packaging                                   7       9     4            5      6
  Spreadability                              12      47     4           21     29
  When u eat it with other things it taste nice   0       3     0            0      0
```

```
        Pearson's Chi-squared test

data:  contingency_table
X-squared = 19.173, df = 24, p-value = 0.7427
```

H_0: There is no association between the "Appealing_Aspect" and "Purchase_Frequency" variables.

H_1: Reject H_0

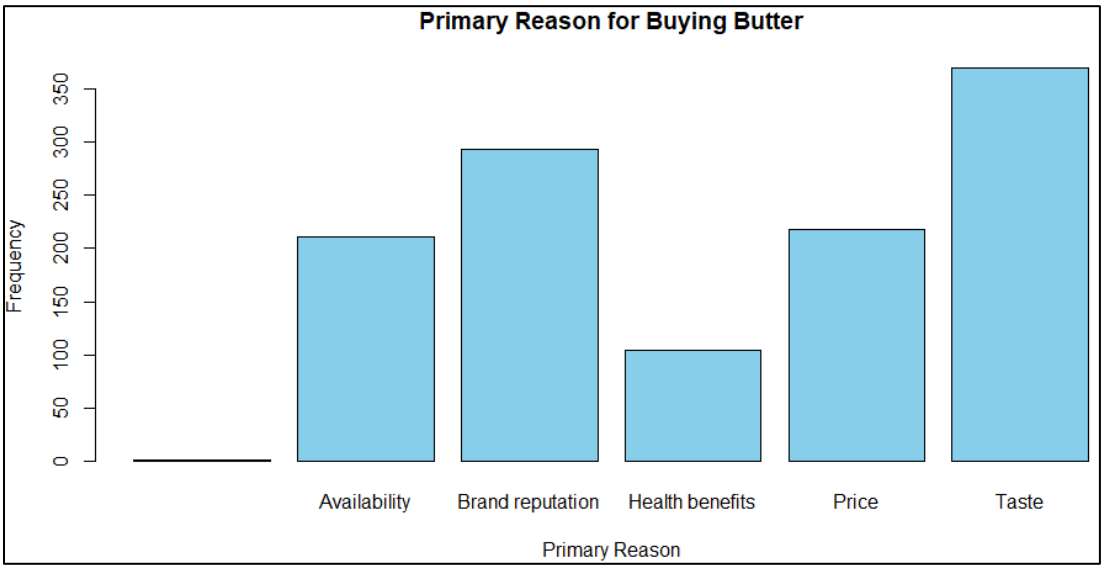
Conclusion: This analysis helps to understand if there's any significant association between the appealing aspect of Amul Butter and how frequently it's purchased. The chi-square test performed indicates that there is no significant association between the variables "Appealing_Aspect" and "Purchase_Frequency." The p-value obtained (0.7427) suggests that there is insufficient evidence to reject the null hypothesis, implying that these variables are likely independent of each other.

3. Code & Output:

```
# Split the consumption methods into separate rows
data=data %>% separate_rows(Appealing_Aspect, sep = ";")
Creating a frequency table of Primary_Reason of buing butter
reason_freq = table(data$Primary_Reason)
#Print frequency table
print("Frequency table for Primary_Reason:")
print(reason_freq)
# Plot a bar plot of Primary_Reason
```

barplot(reason_freq, main = "Primary Reason for Buying Butter", xlab = "Primary Reason", ylab = "Frequency", col = "skyblue")

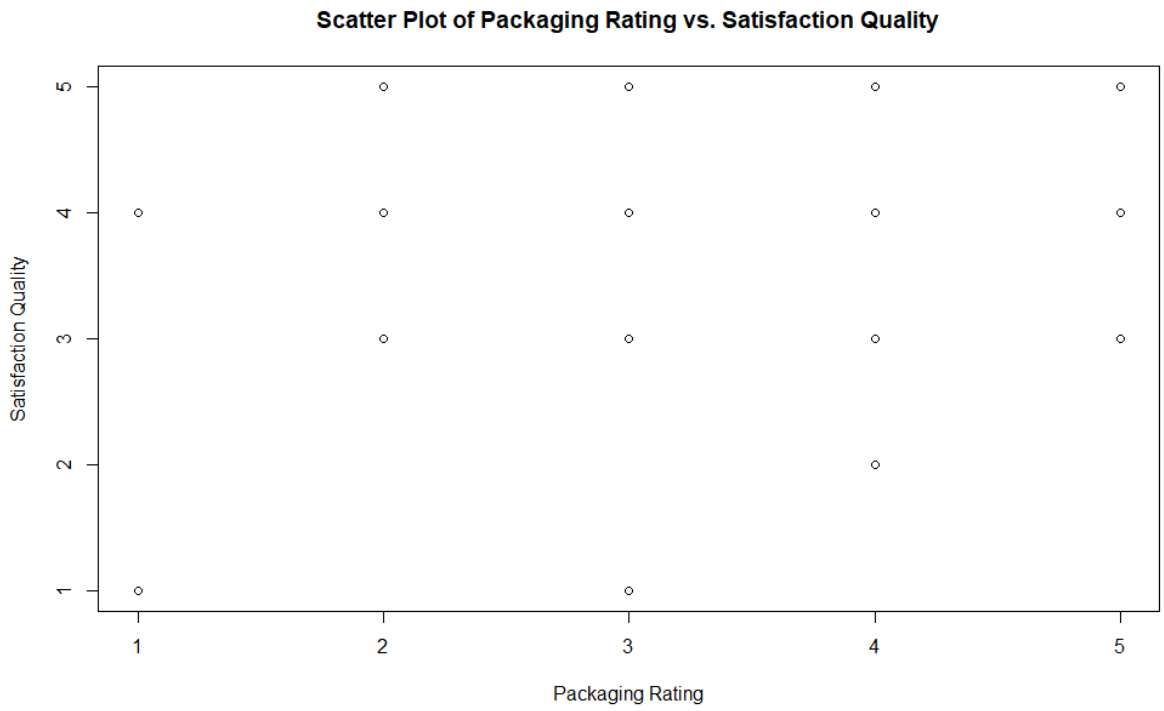
```
[1] "Frequency table for Primary_Reason:"
> print(reason_freq)

                  Availability Brand reputation  Health benefits           Price           Taste
             2             211              293              105             218             369
```



Conclusion: A large number of consumers buy Amul butter because of it's taste.

4. Code & Output:

```
plot(data$Packaging_Rating, data$Satisfaction_Quality,
    xlab = "Packaging Rating", ylab = "Satisfaction Quality",
    main = "Scatter Plot of Packaging Rating vs. Satisfaction Quality")
# Perform correlation analysis
correlation = cor(data$Packaging_Rating, data$Satisfaction_Quality)
print(paste("Correlation between Packaging Rating & Satisfaction Quality:", correlation))
cor_test <- cor.test(data$Packaging_Rating, data$Satisfaction_Quality)
print(cor_test)
```




```
> print(paste("Correlation between Packaging Rating & Satisfaction Quality:", correlation))
[1] "Correlation between Packaging Rating & Satisfaction Quality: 0.51467422712618"
```

```
        Pearson's product-moment correlation

data:  data$Packaging_Rating and data$Satisfaction_Quality
t = 5.7577, df = 92, p-value = 1.116e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3483877 0.6495545
sample estimates:
      cor
0.5146742
```

H_0: There is no correlation between "Packaging Rating" and "Satisfaction Quality".

H_1: The true correlation coefficient (ρ) between "Packaging Rating" and "Satisfaction Quality" is equal to 0.

Conclusion: The alternative hypothesis states that the true correlation is not equal to 0. This suggests that there is a significant correlation between "Packaging Rating" and "Satisfaction Quality". The confidence interval for the correlation coefficient ranges from

0.348 to 0.650. This means that we are 95% confident that the true correlation coefficient falls within this interval.

In summary, based on the very small p-value and the confidence interval that does not include 0, we can conclude that there is a significant positive correlation between "Packaging Rating" and "Satisfaction Quality". The correlation coefficient of approximately 0.515 suggests a moderate positive relationship between these two variables.

5. Code & Output:

```
# Create a frequency table of pricing perception by brand
freq_table <- table(data$Butter_Brand_Preference, data$Pricing_Perception)
# Print frequency table
print("Frequency table of Brand vs. Pricing Perception:")
print(freq_table)
# Perform chi-square test
chi_sq_test <- chisq.test(freq_table)
print("Chi-square test results:")
print(chi_sq_test)
```

	Affordable	Expensive	Neutral	Very affordable	Very expensive
Amul	34	11	28	10	1
Britannia	0	0	1	0	0
Delight	0	0	0	0	1
Gowardhan	2	0	3	0	0
Nutralite	2	0	1	0	0

```
[1] "Chi-square test results:"
> print(chi_sq_test)

        Pearson's Chi-squared test

data:  freq_table
X-squared = 51.89, df = 16, p-value = 1.141e-05
```

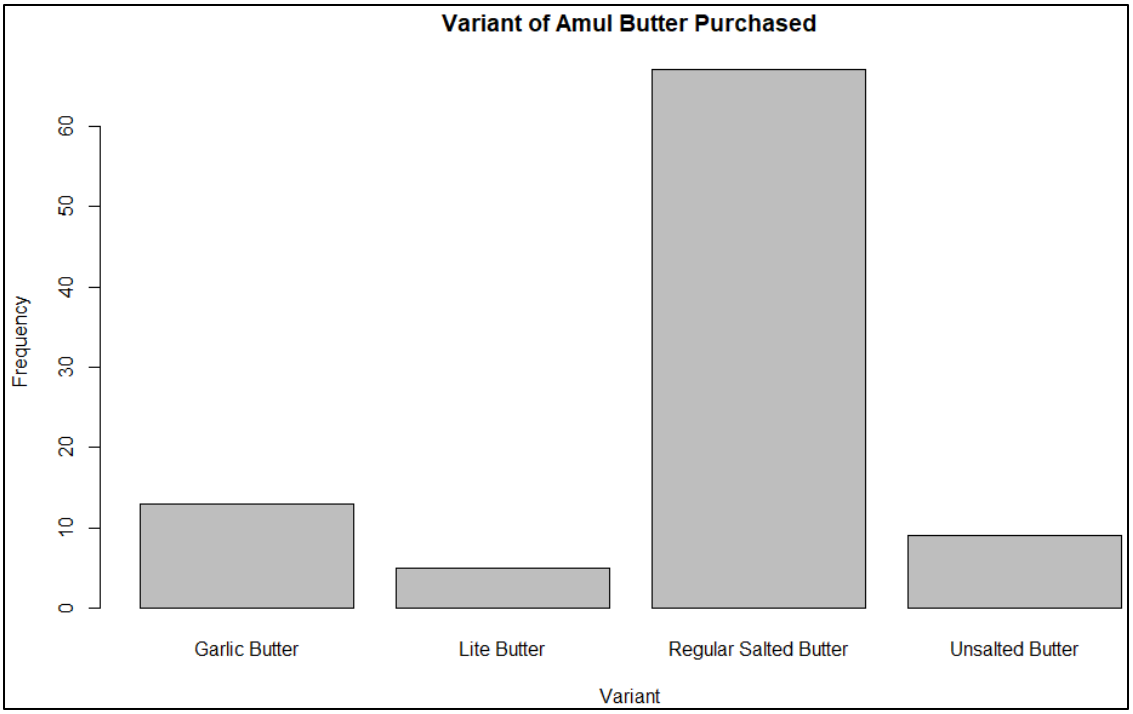
Conclusion: The p-value associated with the chi-squared test is approximately 1.141e-05 (or 0.00001141 when expressed in decimal notation). This p-value is very small, indicating strong evidence against the null hypothesis.

The interpretation of this result would be that there is a significant association between the brand and pricing perception variables. In other words, the perception of pricing differs significantly depending on the brand. The small p-value suggests that it is highly unlikely to observe such a strong association between the variables if there were no true association in the population. Therefore, you would reject the null hypothesis of independence between brand and pricing perception.

6. Code & Output:

```
# Calculate mean of variant purchased
mean_variant <- mean(data$Variant_Purchased)
print(paste("Variant Purchased:", mode_variant))
# Create a frequency table of variant purchased
variant_freq <- table(data$Variant_Purchased)
# Plot a bar plot to visualize the distribution of variant purchased
barplot(variant_freq,
        main = "Variant of Amul Butter Purchased",
        xlab = "Variant",
        ylab = "Frequency")
```

```
> print(paste("Variant Purchased:", mode_variant))
[1] "Variant Purchased: Regular Salted Butter"
```

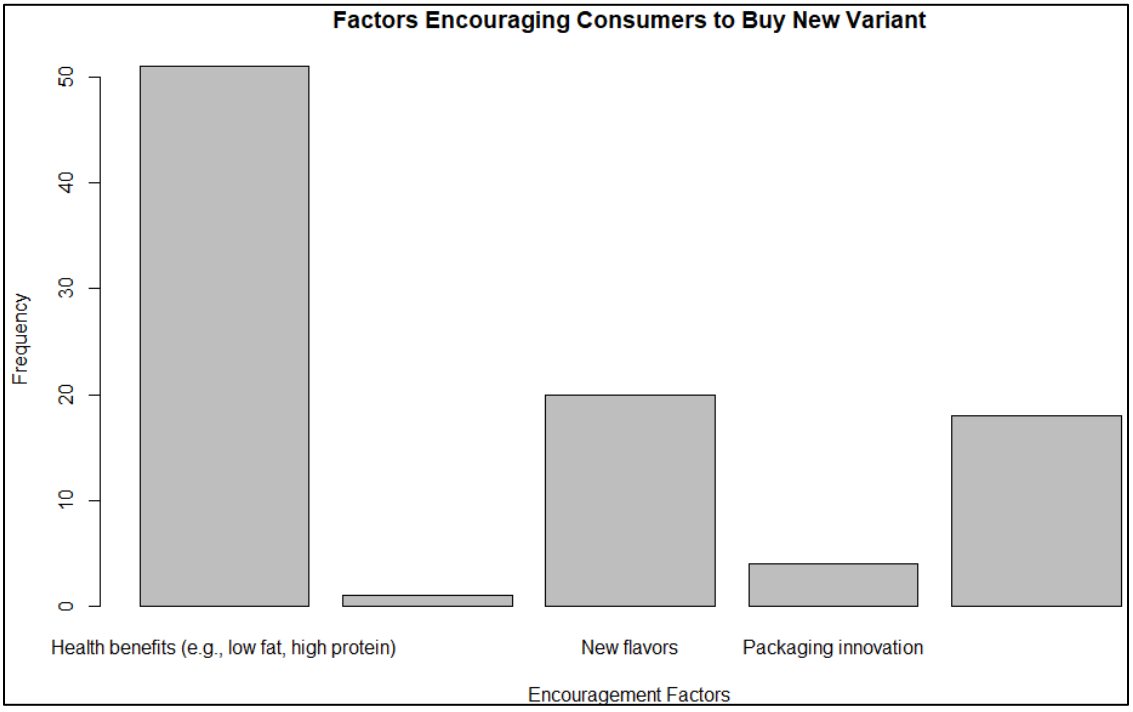


Conclusion: The most popular variant of butter is Amul's Regular salted butter.

7. Code & Output:

```
# Create a frequency table of encouragement factors
encouragement_freq = table(data$Encouragement_Factors)
# Print frequency table
print("Frequency table for Encouragement Factors:")
print(encouragement_freq)
# Plot a bar plot to visualize the distribution of encouragement factors
barplot(encouragement_freq,
        main = "Factors Encouraging Consumers to Buy New Variant",
        xlab = "Encouragement Factors",
        ylab = "Frequency")
```

```
[1] "Frequency table for Encouragement Factors:"
> print(encouragement_freq)

Health benefits (e.g., low fat, high protein)   Like old address, picture on the packaging
                                            51                                            1
                                     New flavors                     Packaging innovation
                                            20                                            4
                  Special promotions or discounts
                                            18
```



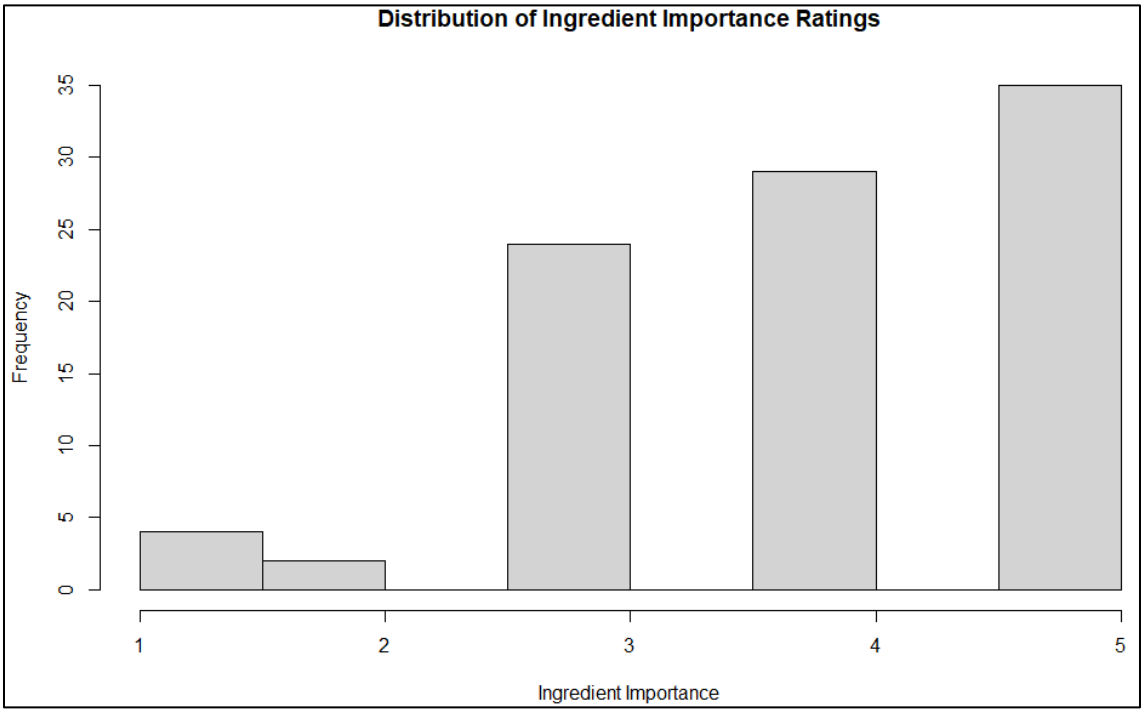
Conclusion: The primary incentive for consumers to experiment with a new variation of Amul butter is likely the perceived health advantages it offers.

8. Code & Output:

```
summary(data$Ingredient_Importance)
# Histogram to visualize the distribution of ingredient importance ratings
hist(data$Ingredient_Importance,
    main = "Distribution of Ingredient Importance Ratings",
    xlab = "Ingredient Importance",
    ylab = "Frequency")
correlation = cor(data$Ingredient_Importance, data$Satisfaction_Quality)
```

print(paste("Correlation between Ingredient Importance and Satisfaction Rating:", correlation))

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   4.000   3.947   5.000   5.000
```




```
> print(paste("Correlation between Ingredient Importance and Satisfaction Rating:", correlation))
[1] "Correlation between Ingredient Importance and Satisfaction Rating: 0.207093429413305"
```

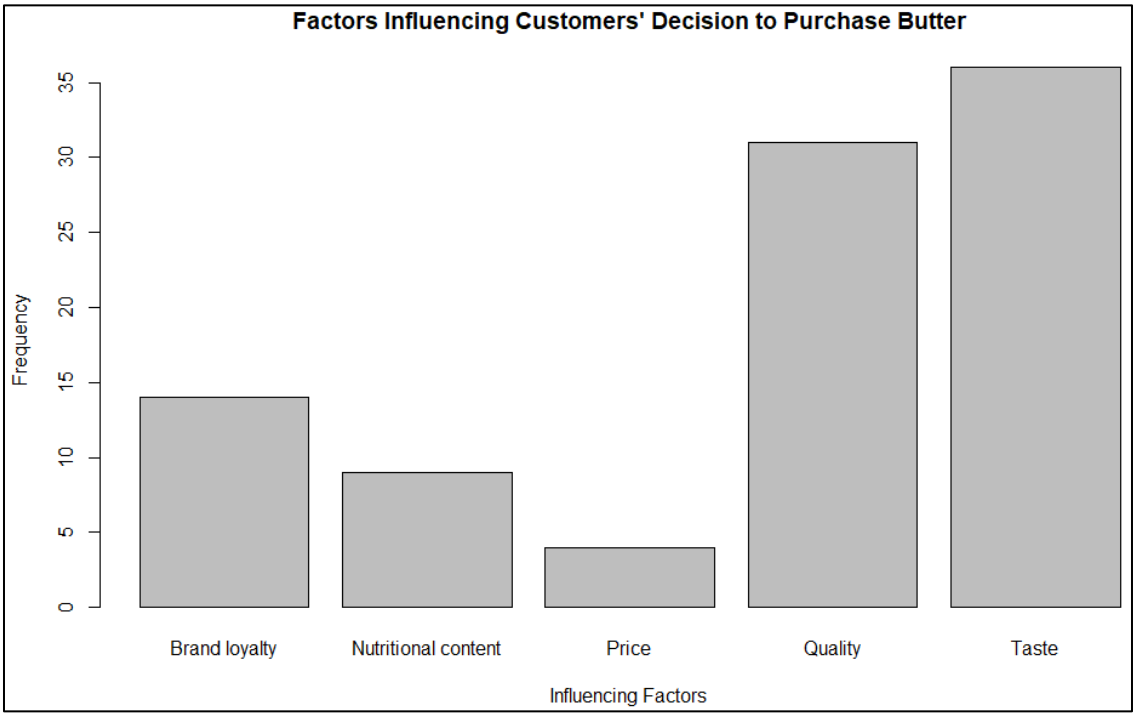
Conclusion: the ingredients used to make the butter is very important to the consumers. The positive sign indicates that as ingredient importance increases, satisfaction rating tends to increase as well. The magnitude of the correlation coefficient (0.207) suggests a weak to moderate level of association between the two variables.

9. Code & Output:

```
# Create a frequency table of influencing factors
factor_freq <- table(data$Influencing_Factor)
# Print frequency table
```

```
print("Frequency table for Influencing Factors:")
print(factor_freq)
# Plot a bar plot to visualize the distribution of influencing factors
barplot(factor_freq,
        main = "Factors Influencing Customers' Decision to Purchase Butter",
        xlab = "Influencing Factors",
        ylab = "Frequency")
```

```
[1] "Frequency table for Influencing Factors:"
> print(factor_freq)

      Brand loyalty Nutritional content               Price             Quality               Taste
                 14                   9                   4                  31                  36
```

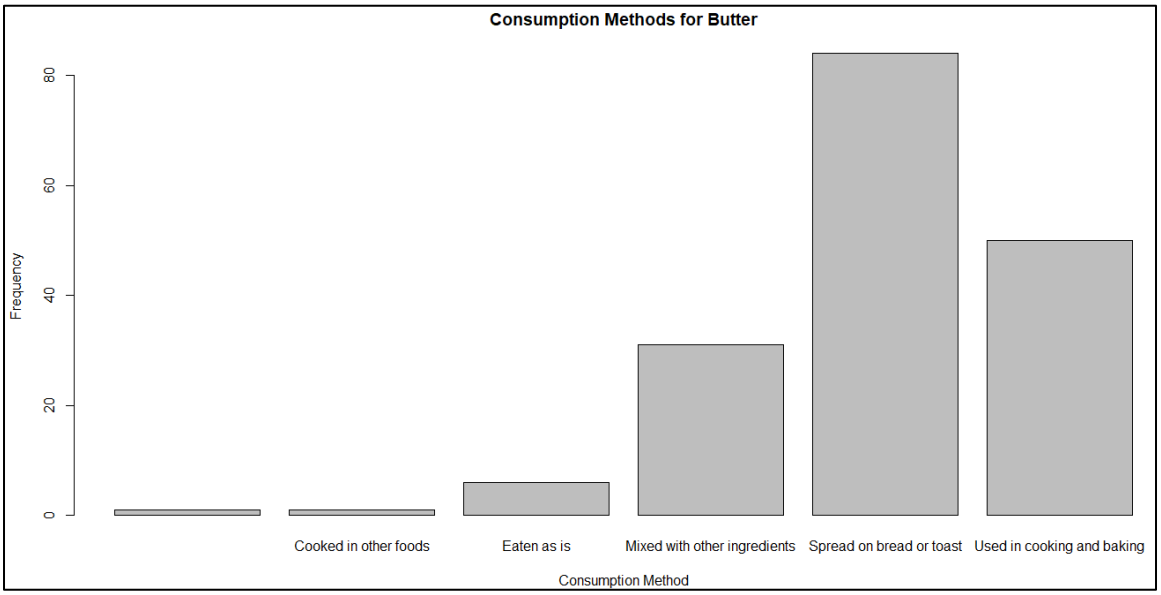


Conclusion: The factors influencing the customers to purchase Amul butter is the taste and quality of the butter.

10. Code & Output:

```
# Split the consumption methods into separate rows
butter_data=data %>%
  separate_rows(Consumption_Method, sep = ",")
# Create a frequency table of consumption methods
consumption_freq=table(butter_data$Consumption_Method)
# Print frequency table
print("Frequency table for Consumption Methods:")
print(consumption_freq)
# Plot a bar plot to visualize the distribution of consumption methods
barplot(consumption_freq,
        main = "Consumption Methods for Butter",
        xlab = "Consumption Method",
        ylab = "Frequency")
```

```
[1] "Frequency table for Consumption Methods:"
> print(consumption_freq)

                            Cooked in other foods                 Eaten as is Mixed with other ingredients
                                                1                             1                             6                            31
   Spread on bread or toast   Used in cooking and baking
                                               84                            50
```

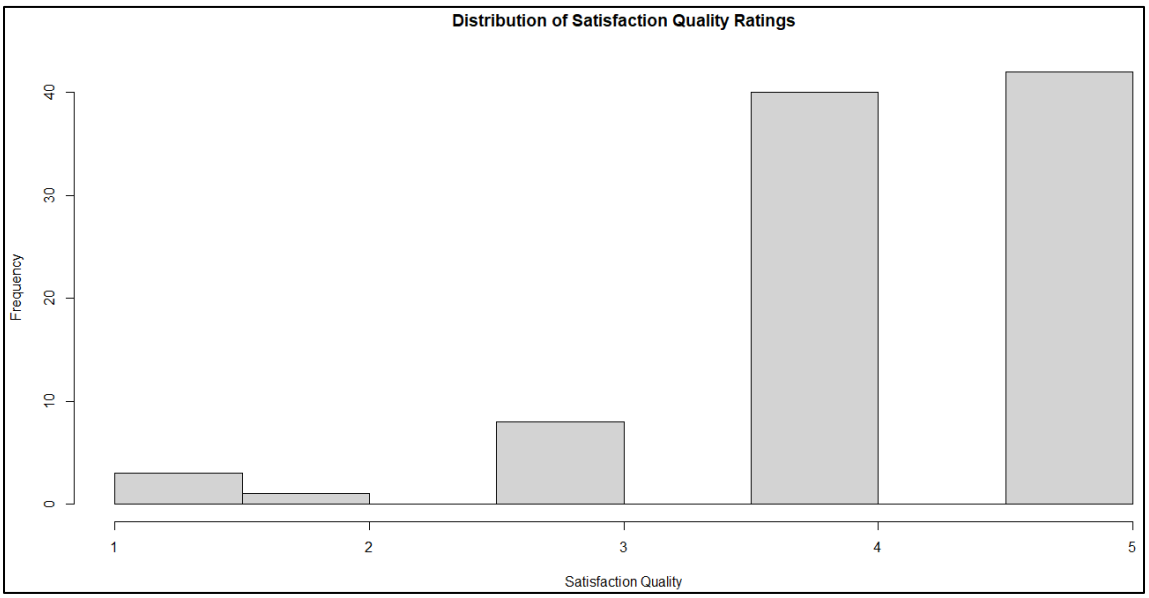


Conclusion: The chart indicates that the majority of consumers prefer consuming butter with bread. Thus, by stocking butter alongside bread, there's a possibility of boosting sales.

11. Code & Output:

```
# Descriptive statistics
summary(data$Satisfaction_Quality)
# Histogram to visualize the distribution of satisfaction quality ratings
barplot(data$Satisfaction_Quality,
    main = "Distribution of Satisfaction Quality Ratings",
    xlab = "Satisfaction Quality",
    ylab = "Frequency")
# Perform hypothesis testing (e.g., t-test) against a reference value
# For example, if 3 is considered "upto the mark"
t_test = t.test(data$Satisfaction_Quality, mu = 3)
print(t_test)
```

```
> summary(data$Satisfaction_Quality)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   4.000   4.000   4.245   5.000   5.000
```



```
> print(t_test)

        One Sample t-test

data:  data$Satisfaction_Quality
t = 13.405, df = 93, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 4.060291 4.429070
sample estimates:
mean of x
 4.244681
```

H_0: The true mean satisfaction quality rating is equal to 3.

H_1: True mean is not equal to 3.

Conclusion: The p-value is extremely small, indicating strong evidence against the null hypothesis. Since the p-value is less than the significance level (commonly 0.05), we reject the null hypothesis. Therefore, there is strong evidence to suggest that the true mean satisfaction quality rating is not equal to 3. The 95% confidence interval for the mean satisfaction quality rating (4.060 to 4.429) does not include the reference value of 3, further supporting the rejection of the null hypothesis.

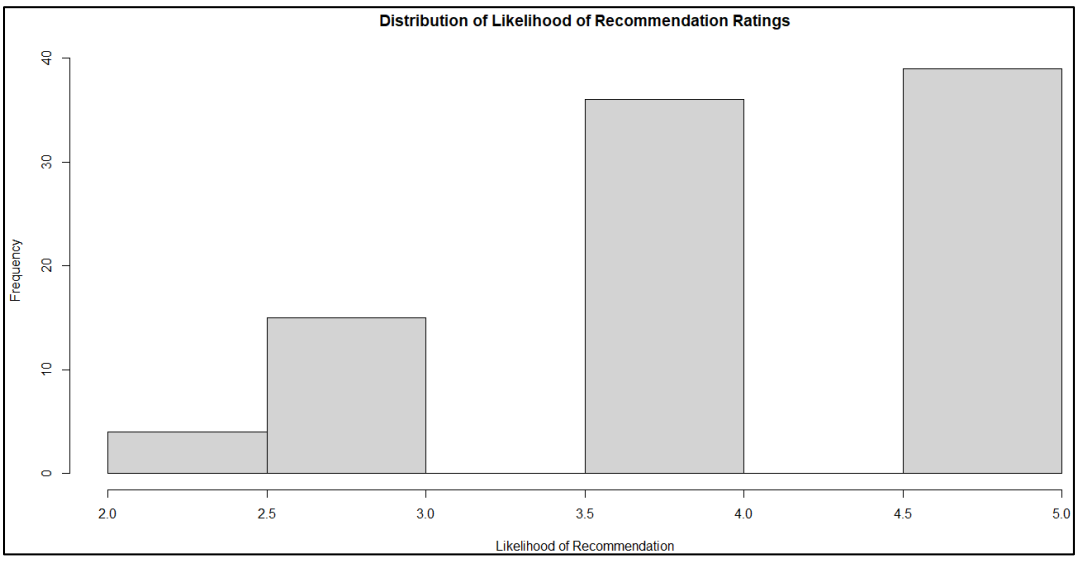
In conclusion, based on the one-sample t-test, there is strong evidence to suggest that the quality of Amul Butter, as indicated by the satisfaction quality ratings, is significantly different from 3, indicating that it is not up to the mark.

12. Code & Output:

```
summary(data$Likelihood_Recommendation)
# Histogram to visualize the distribution of likelihood of recommendation ratings
hist(data$Likelihood_Recommendation,
    main = "Distribution of Likelihood of Recommendation Ratings",
    xlab = "Likelihood of Recommendation",
    ylab = "Frequency")
# Perform hypothesis testing (e.g., t-test) against a reference value
# 4 is considered "high likelihood"
t_test = t.test(data$Likelihood_Recommendation, mu = 4)
```

print(t_test)

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.00    4.00    4.00    4.17    5.00    5.00
```




```
        One Sample t-test

data:  data$Likelihood_Recommendation
t = 1.9403, df = 93, p-value = 0.05537
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 3.996007 4.344419
sample estimates:
mean of x
 4.170213
```

H_0: The true mean likelihood of recommendation rating is equal to 4.

H_1: Reject H_0.

Conclusion: The p-value is 0.05537, which is greater than the significance level (commonly 0.05).Since the p-value is greater than the significance level, we fail to reject the null hypothesis. Therefore, there is insufficient evidence to suggest that the true mean likelihood of recommendation rating is different from 4. The 95% confidence interval for

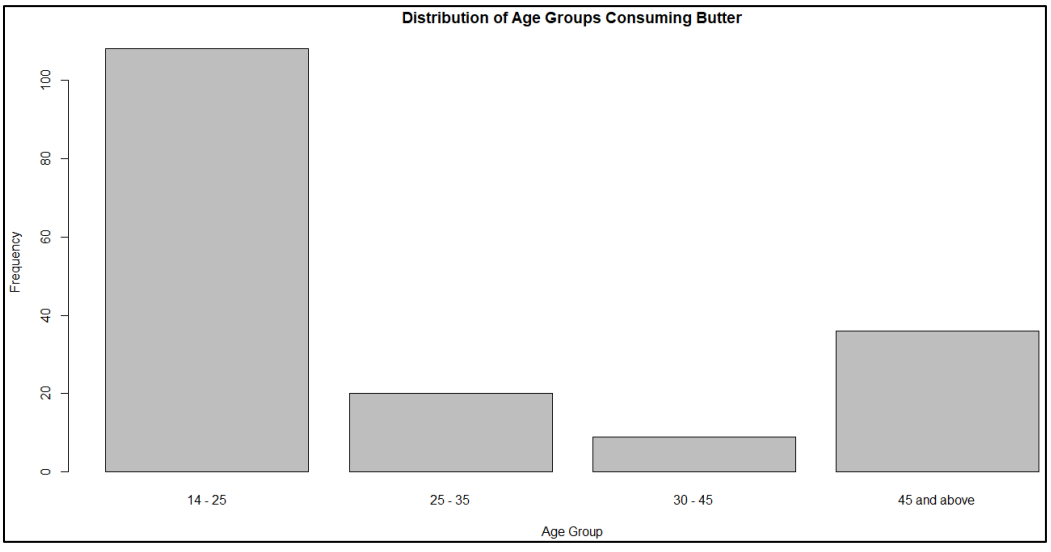
the mean likelihood of recommendation rating (3.996 to 4.344) includes the reference value of 4, further supporting the failure to reject the null hypothesis.

In conclusion, based on the one-sample t-test, there is insufficient evidence to conclude that consumers' likelihood of recommending the product to others differs significantly from a mean value of 4.

13. Code & Output:

```
# Create a frequency table of age groups
age_group_freq <- table(butter_data$Age_Group)
# Print frequency table
print("Frequency table for Age Groups:")
print(age_group_freq)
# Plot a bar plot to visualize the distribution of age groups
barplot(age_group_freq,
    main = "Distribution of Age Groups Consuming Butter",
    xlab = "Age Group",
    ylab = "Frequency")
```

```
[1] "Frequency table for Age Groups:"
> print(age_group_freq)

     14 - 25      25 - 35      30 - 45 45 and above
         108           20            9           36
```

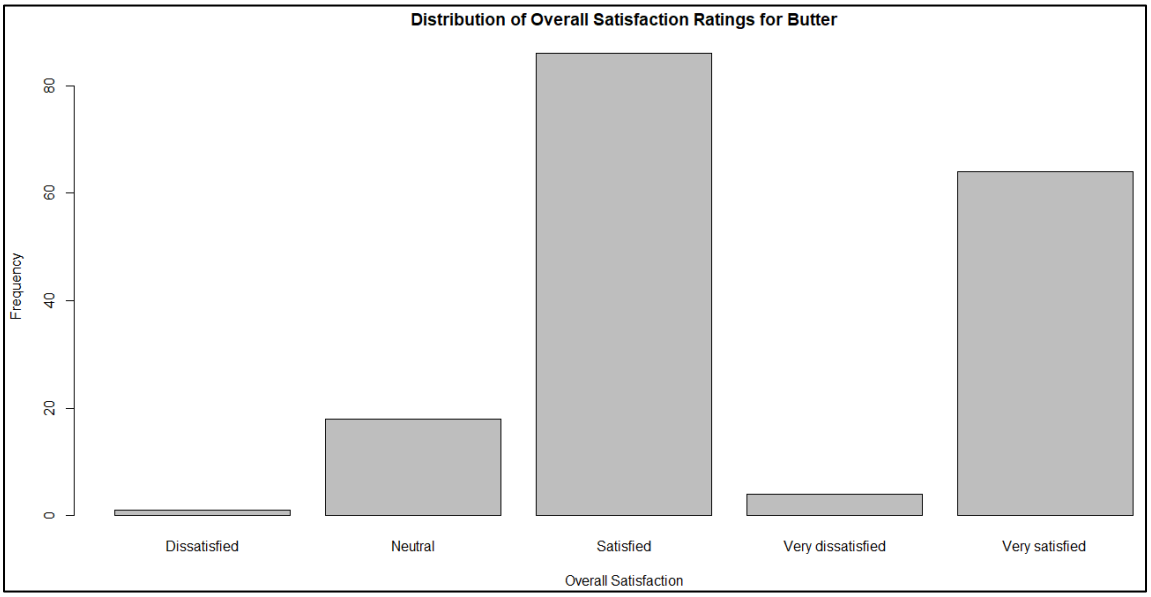


Conclusion: Individuals aged between 14 and 25, as well as those aged 45 and older, typically include butter in their diet.

14. Code & Output:

```
# Create a frequency table of overall satisfaction ratings
satisfaction_freq <- table(butter_data$Overall_Satisfaction)
# Print frequency table
print("Frequency table for Overall Satisfaction Ratings:")
print(satisfaction_freq)
# Plot a bar plot to visualize the distribution of overall satisfaction ratings
barplot(satisfaction_freq,
    main = "Distribution of Overall Satisfaction Ratings for Butter",
    xlab = "Overall Satisfaction",
    ylab = "Frequency")
```

```
[1] "Frequency table for Overall Satisfaction Ratings:"
> print(satisfaction_freq)

     Dissatisfied           Neutral         Satisfied Very dissatisfied    Very satisfied
                1                18                86                 4                64
```



Conclusion: The customers are content with Amul butter.

Additional Analysis:

1. Code & Output:

   # ANOVA for "Satisfaction_Quality" across different "Age_Groups"

   anova_model <- lm(Satisfaction_Quality ~ Age_Group, data = data)

   summary(anova_model)

```
> summary(anova_model)

Call:
lm(formula = Satisfaction_Quality ~ Age_Group, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-3.00868 -0.60000 -0.00868  0.99132  2.27041

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.00868    0.04568  87.749  < 2e-16 ***
Age_Group25 - 35    -1.27909    0.10329 -12.383  < 2e-16 ***
Age_Group30 - 45     0.18362    0.25843   0.711    0.478
Age_Group45 and above  0.59132    0.10946   5.402 7.95e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.297 on 1194 degrees of freedom
Multiple R-squared:  0.1532,    Adjusted R-squared:  0.1511
F-statistic:    72 on 3 and 1194 DF,  p-value: < 2.2e-16
```

   Conclusion: Based on the output, we can conclude that age group significantly influences satisfaction quality, as indicated by the significant coefficients and the highly significant F-statistic. Additionally, the adjusted R-squared value suggests that age group explains a modest proportion of the variability in satisfaction quality.

2. Code & Output:

   lm_model=lm(Satisfaction_Quality~ Purchase_Frequency + Consumption_Frequency + Brand_Reputation_Importance + Packaging_Rating, data = data)

   summary(lm_model)

```
Call:
lm(formula = Satisfaction_Quality ~ Purchase_Frequency + Consumption_Frequency +
    Brand_Reputation_Importance + Packaging_Rating, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6913 -0.3233  0.0000  0.3659  1.9478

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      1.53080    0.16141   9.484  < 2e-16 ***
Purchase_FrequencyMonthly        0.58334    0.08083   7.217 9.47e-13 ***
Purchase_FrequencyNever         -1.15147    0.14933  -7.711 2.63e-14 ***
Purchase_FrequencyOccasionally   0.23801    0.11443   2.080  0.03774 *
Purchase_FrequencyWeekly        -0.14910    0.08471  -1.760  0.07866 .
Consumption_FrequencyMonthly    -0.84410    0.08341 -10.120  < 2e-16 ***
Consumption_FrequencyOccasionally -0.09884    0.10815  -0.914  0.36093
Consumption_FrequencyWeekly      0.18306    0.06564   2.789  0.00537 **
Brand_Reputation_Importance      0.08205    0.04013   2.045  0.04112 *
Packaging_Rating                 0.53861    0.02066  26.068  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7208 on 1188 degrees of freedom
Multiple R-squared:  0.7397,    Adjusted R-squared:  0.7378
F-statistic: 375.2 on 9 and 1188 DF,  p-value: < 2.2e-16
```

Conclusion: In conclusion, the results suggest that factors such as purchase frequency, consumption frequency, brand reputation importance, and packaging rating significantly influence satisfaction quality with Amul Butter. Consumers who purchase and consume Amul Butter more frequently tend to have higher satisfaction quality, while positive perceptions of brand reputation and packaging also contribute to overall satisfaction. These findings can inform marketing strategies and product development efforts aimed at enhancing consumer satisfaction with Amul Butter.

Final Conclusion:

Based on the analysis conducted through the survey on consumer preferences for Amul Butter, several key insights have been identified:

1. Consumer Satisfaction: The majority of respondents expressed high levels of satisfaction with Amul Butter, with a significant proportion rating it positively.
2. Brand Loyalty: There appears to be a strong level of brand loyalty towards Amul Butter among consumers, as indicated by the high satisfaction ratings and likelihood of recommendation.
3. Age Group Preferences: Analysis of age group preferences revealed that consumers across various age groups consume Amul Butter, with no significant preference observed for any particular age group.
4. Overall Satisfaction Ratings: Categorical overall satisfaction ratings indicate that a large proportion of consumers are either "very satisfied" or "satisfied" with Amul Butter.
5. Price Perception: The perception of pricing compared to other brands was favorable, with a majority of consumers considering Amul Butter to be competitively priced or affordable.
6. Encouragement Factors: Consumers expressed interest in trying new variations of Amul Butter products, indicating potential opportunities for product innovation and expansion.
7. Importance of Ingredients: While ingredients are considered important in the purchasing decision, they may not be the sole determining factor for consumers when choosing Amul Butter.
8. Packaging Convenience: The majority of consumers found the packaging of Amul Butter to be convenient and durable, which contributes positively to overall satisfaction.

In conclusion, the survey results suggest that Amul Butter enjoys a strong reputation among consumers, with high levels of satisfaction, brand loyalty, and perceived value. However, there are still opportunities for Amul to further enhance its product offerings and marketing strategies to meet evolving consumer preferences and expectations. By continuing to prioritize quality, innovation, and customer satisfaction, Amul Butter can maintain its position as a preferred choice among consumers in the dairy market.