



RETAIL STRATEGY AND ANALYTICS

AUTHOR: AHMED ELATWY

LAST UPDATED: 25-JAN-2025

TABLE OF CONTENTS

Title	Page
Executive Summary	3
Introduction	4
Methodology	5
Data Cleaning	7
Data Analysis	8
Discussion	33
Conclusion	40
Appendix	43

EXECUTIVE SUMMARY

This report aims to analyze customer purchasing behavior for chips and assess the impact of new trial store layouts.

By examining transaction and customer data, identifying key metrics, and comparing trial stores to control stores, we will provide data-driven recommendations to optimize chip sales and determine the feasibility of rolling out the new store layouts to all locations.

Key Findings:

- Insights into customer segments and their purchasing behavior.
- Analysis of sales drivers and top-selling regions.
- Evaluation of trial layouts' performance compared to control stores.
- Recommendations for targeting customer segments and implementing store layouts across all locations.

INTRODUCTION

Background:

Quantium's retail analytics team, in collaboration with the client (Category Manager for Chips), aims to better understand customer purchasing behavior within the chips category. Additionally, we will evaluate the impact of new trial store layouts to provide strategic recommendations for the next half year.

Objectives:

1. Analyze customer purchasing behavior for chips.
2. Identify key customer segments and sales drivers.
3. Compare the performance of trial stores with control stores.
4. Provide data-driven recommendations based on insights.

Scope:

- The analysis will focus on transaction and customer data from various stores, assessing sales performance, and determining the impact of new trial layouts.

METHODOLOGY

Data Examination:

- **Transaction Data:** Review inconsistencies, missing data, outliers, correctly identified category items, and numeric data across all tables. Data cleaning will involve making necessary changes and saving the dataset.
- **Customer Data:** Address similar issues, look for null values, and merge transaction and customer data ensuring the files are saved accordingly.

Data Analysis:

- **Metric Definition:** Analyze total sales, drivers of sales, and regions with the highest sales.
- **Visualization:** Create charts and graphs to explore interesting trends and insights.

METHODOLOGY

Top Performing Products

- Identify top performing products.
- Recognize Most Popular Brands.

Customer Segmentation:

- Define customer segments based on purchasing behavior.
- Identify target segments and evaluate the impact of packet sizes.

Control Store Selection:

- Define metrics for control stores, visualize drivers, and ensure appropriate selection.

Trial vs. Control Store Performance:

- Assess individual trial stores comparing them with control stores.
- Summarize findings and provide recommendations for each location based on insights

DATA CLEANING

Inconsistencies & Missing Data:

Identified and corrected anomalies.
Ensured no missing data was present
across transaction records.

Outliers: Removed outliers to maintain
data accuracy.

Category Items: Ensured all items were
correctly categorized as chips.

Numeric Data Verification: Verified
that numerical data (such as quantities
and prices) were accurate.

DATA ANALYSIS



Customer Segmentation: Understanding who shops by analyzing customer groups and their shopping behaviors



Sales Analysis: Digging into sales trends, transaction values, and top-selling products.



Trial and Control Store Analysis: Evaluating the impact of a trial intervention on sales and customer numbers.

DATA ANALYSIS

CUSTOMER SEGMENTATION

Lifestage Analysis.

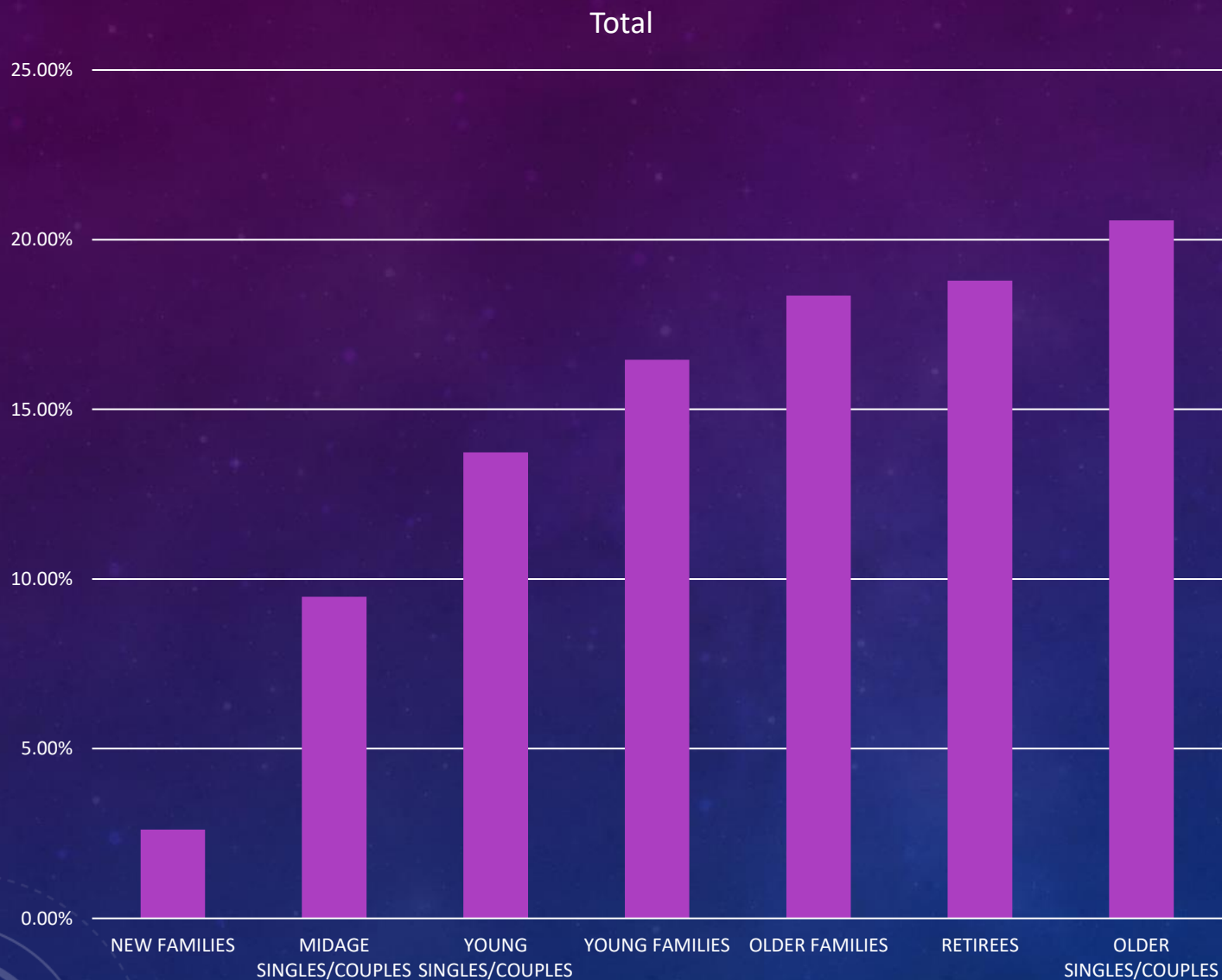
- **Shopping by Life Stages:** Retirees, older singles/couples, and young singles/couples make up the largest customer segments.

Membership Analysis.

- **Membership Breakdown:** Mainstream membership is the most popular, followed by Budget and Premium memberships.

Membership by Lifestage.

- **Membership Preferences:** Mainstream memberships are most popular with young singles/couples, while Budget memberships are favored by older families.



DATA ANALYSIS

CUSTOMER SEGMENTATION

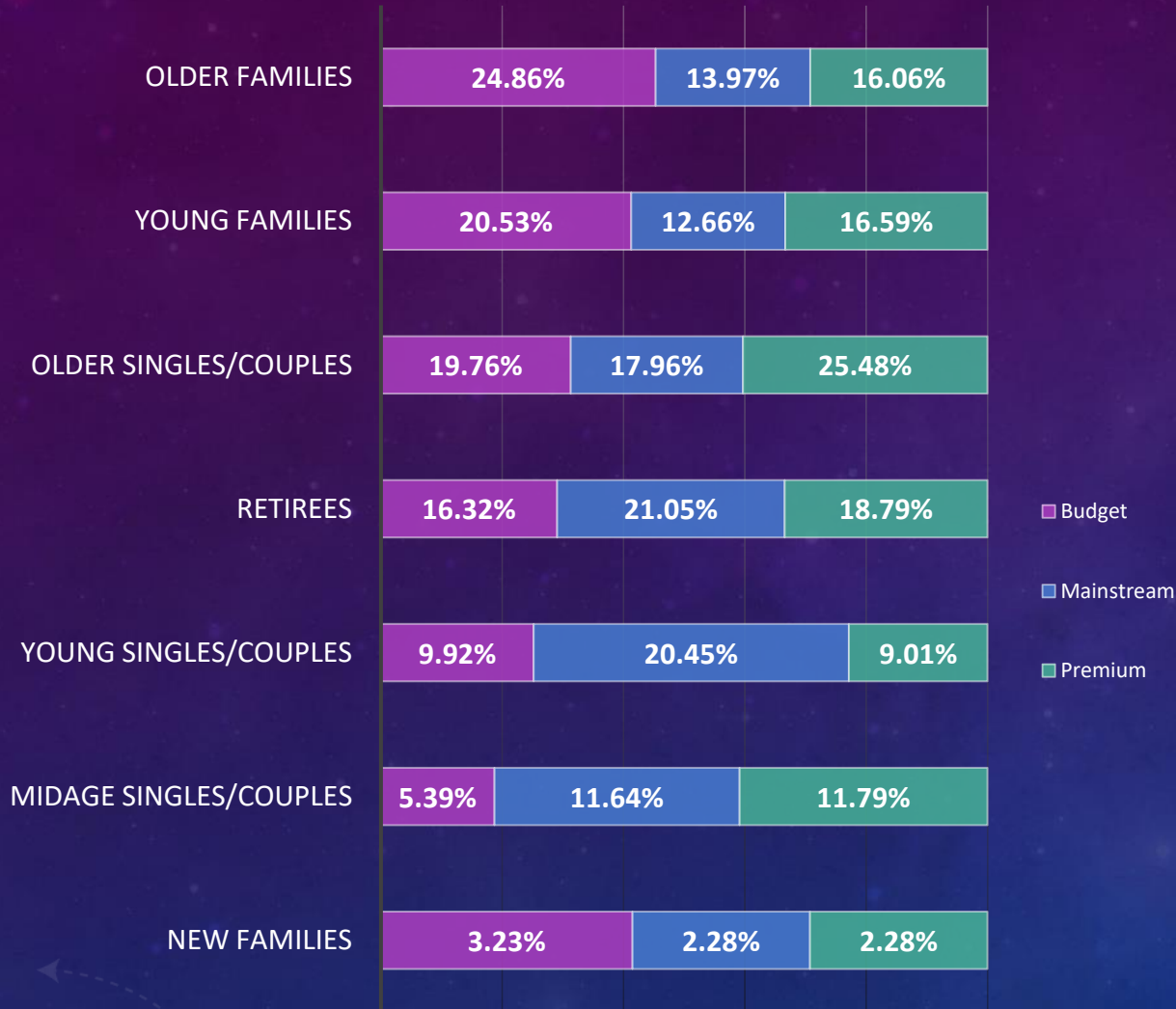
1- Life stage

Key Insights:

- Retirees represent the largest group, comprising 20.38% of the total customer base.
- Older Singles/Couples and Young Singles/Couples also form significant portions, each accounting for around 20%.
- Older Families and Young Families together make up about 26%, indicating a strong presence of families in your customer base.
- Midage Singles/Couples represent 10.02%.
- New Families are the smallest group, with only 3.51%.

DATA ANALYSIS

CUSTOMER SEGMENTATION



3- Membership Distribution

Key Insights:

1. Premium Preference:

- "Older Singles/Couples" show the highest preference for premium options (25.48%).

2. Budget-Conscious Groups:

- "Older Families" and "Young Families" prefer budget options (24.86% and 20.53% respectively).

3. Midstream Preferences:

- "Retirees" and "Young Singles/Couples" have significant preferences for midstream options (21.05% and 20.45% respectively).

4. New Families:

- They have the lowest preferences across all categories.

5. Diverse Preferences:

- Varied spread of preferences across all groups, highlighting the importance of catering to different market segments with tailored offerings.

DATA ANALYSIS

CUSTOMER SEGMENTATION

Key Observations:

1. New Families Lead:

The average price for New Families is the highest at 3.91. This indicates that this group might be willing to spend more, possibly due to their need for quality and quantity.

2. Close Contenders:

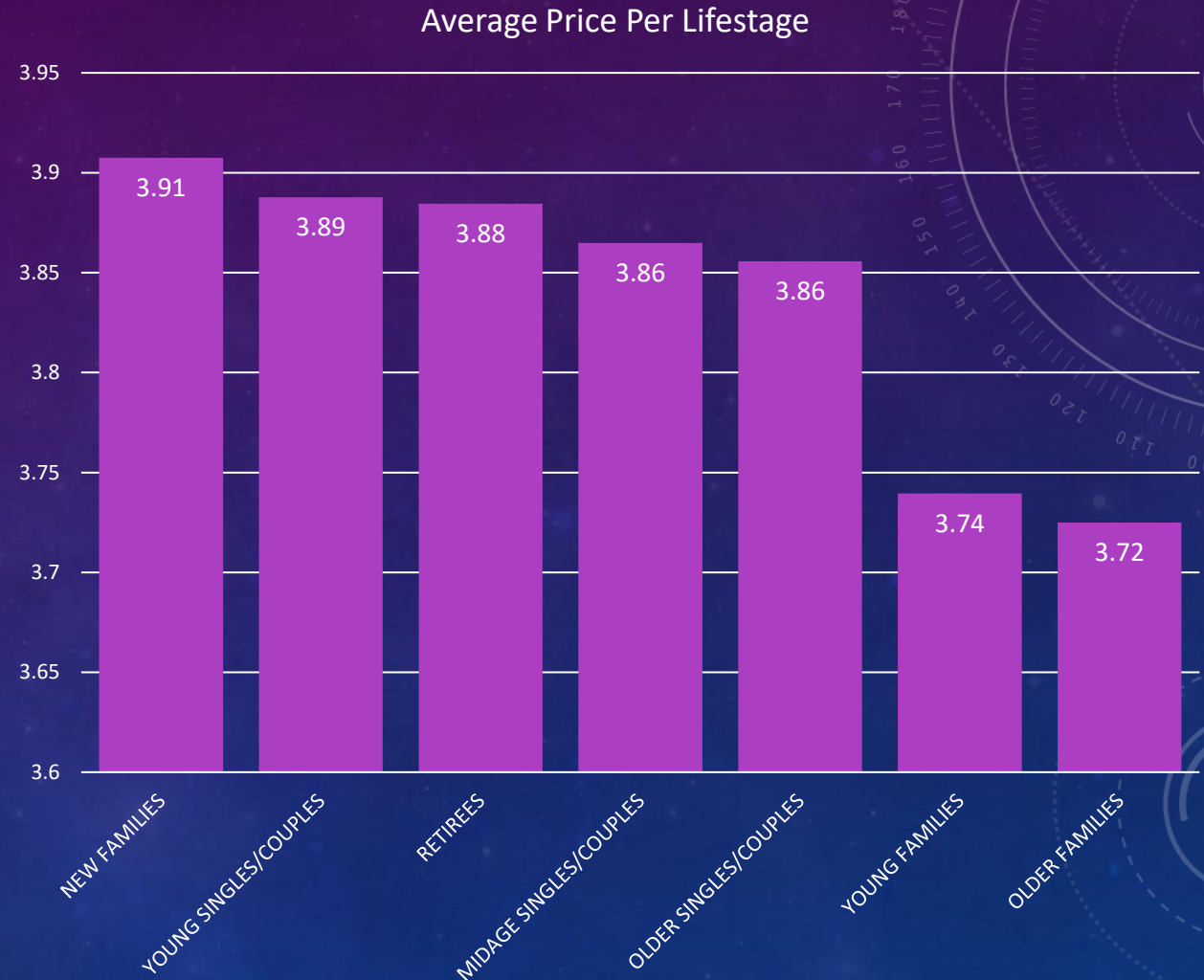
Young Singles/Couples (3.89) and Retirees (3.88) follow closely behind. Young Singles/Couples might be spending more due to lifestyle choices, while Retirees may have the disposable income and prefer higher-priced options.

3. Middle Ground:

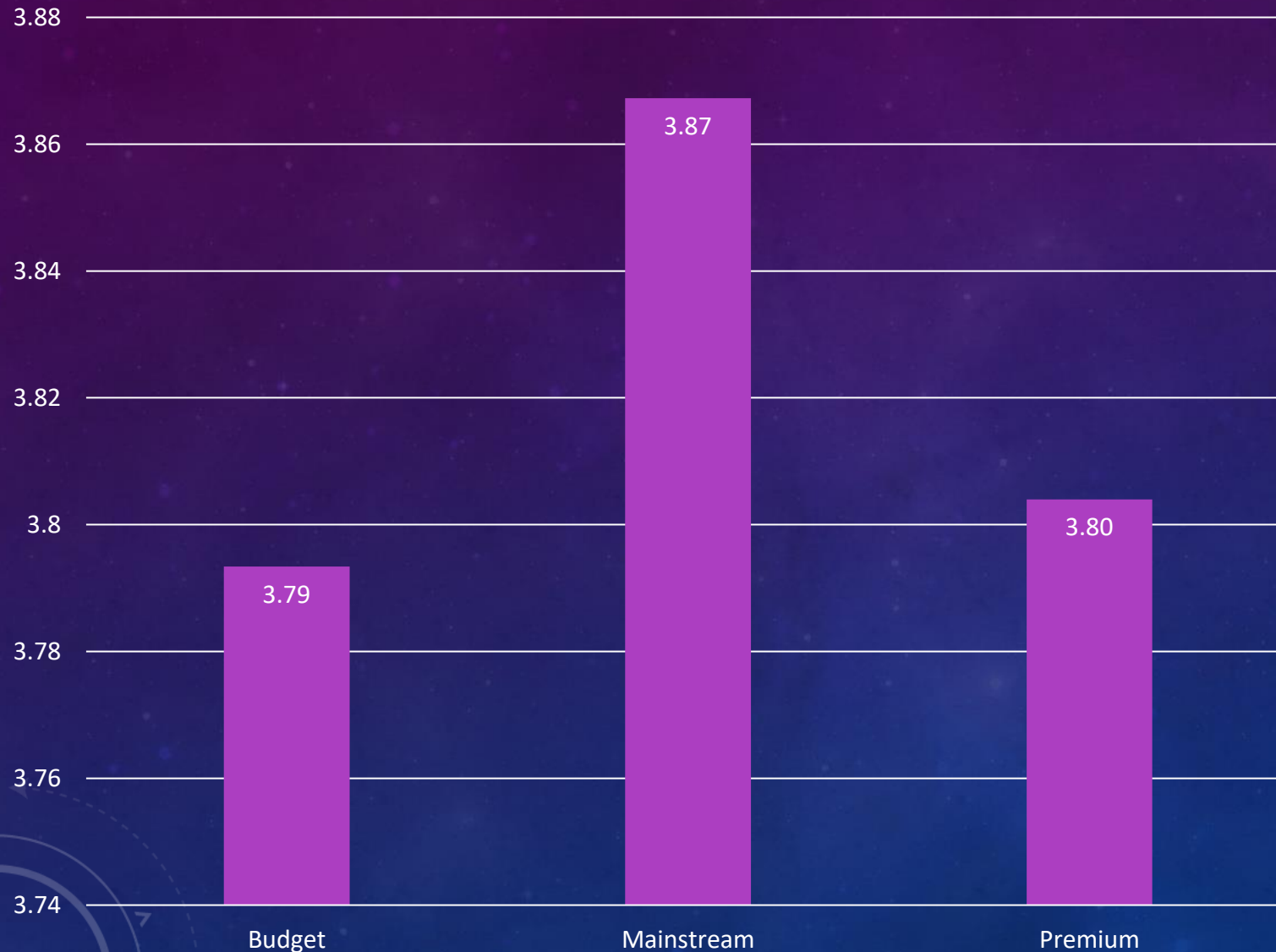
Midage Singles/Couples and Older Singles/Couples both have an average price of 3.86, suggesting a similar spending pattern.

4. Lower Spending:

Young Families (3.74) and Older Families (3.72) have the lowest average prices. This could be due to budget constraints or a focus on value-for-money products.



AVERAGE PRICE PER MEMBERSHIP



DATA ANALYSIS

CUSTOMER SEGMENTATION

Key Insights:

1. Mainstream Membership:

Average Price: 3.87

This membership type has the highest average price among the three. It suggests that the Mainstream membership is positioned as the mid-tier option, possibly offering a balanced mix of features and benefits.

2. Premium Membership:

Average Price: 3.80

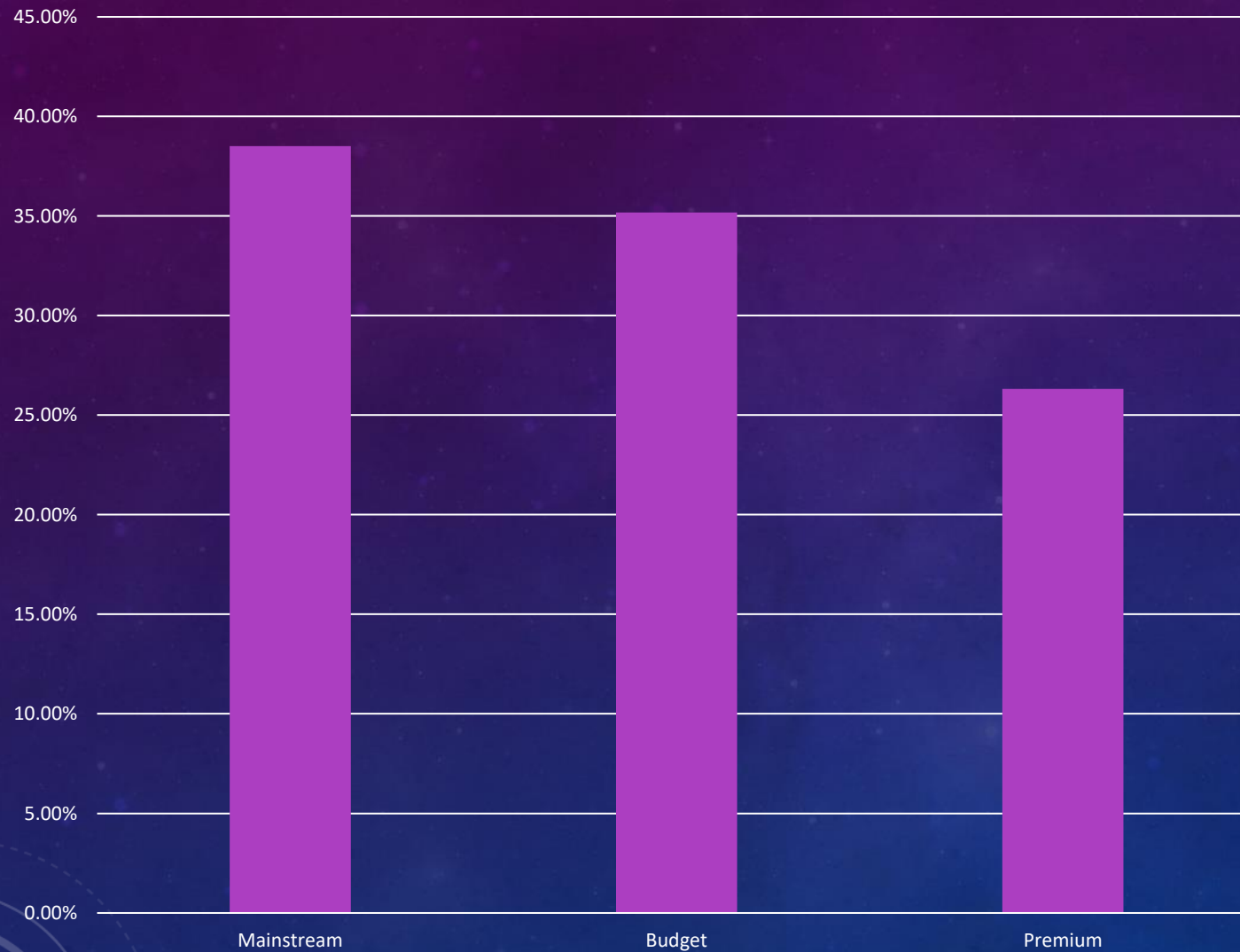
The Premium membership has the second-highest average price. Despite being labeled "Premium," its average price is slightly lower than the Mainstream membership. This might imply value-driven premium features that appeal to a broader audience.

3. Budget Membership:

Average Price: 3.79

The Budget membership has the lowest average price, which is expected. It's designed to be the most affordable option, attracting cost-conscious consumers.

Membership Percentage



DATA ANALYSIS

CUSTOMER SEGMENTATION

2- Membership

Key Insights:

- Mainstream Membership is the most popular among customers, comprising 40.26% of the total customer base.
- Budget Membership follows, representing 33.69% of the customers.
- Premium Membership is the least common, with 26.05% of the customers.

DATA ANALYSIS

SALES ANALYSIS

Average Transaction Value

- **Pack Size Analysis::** The most common pack size for chips.

Total Sales Distribution

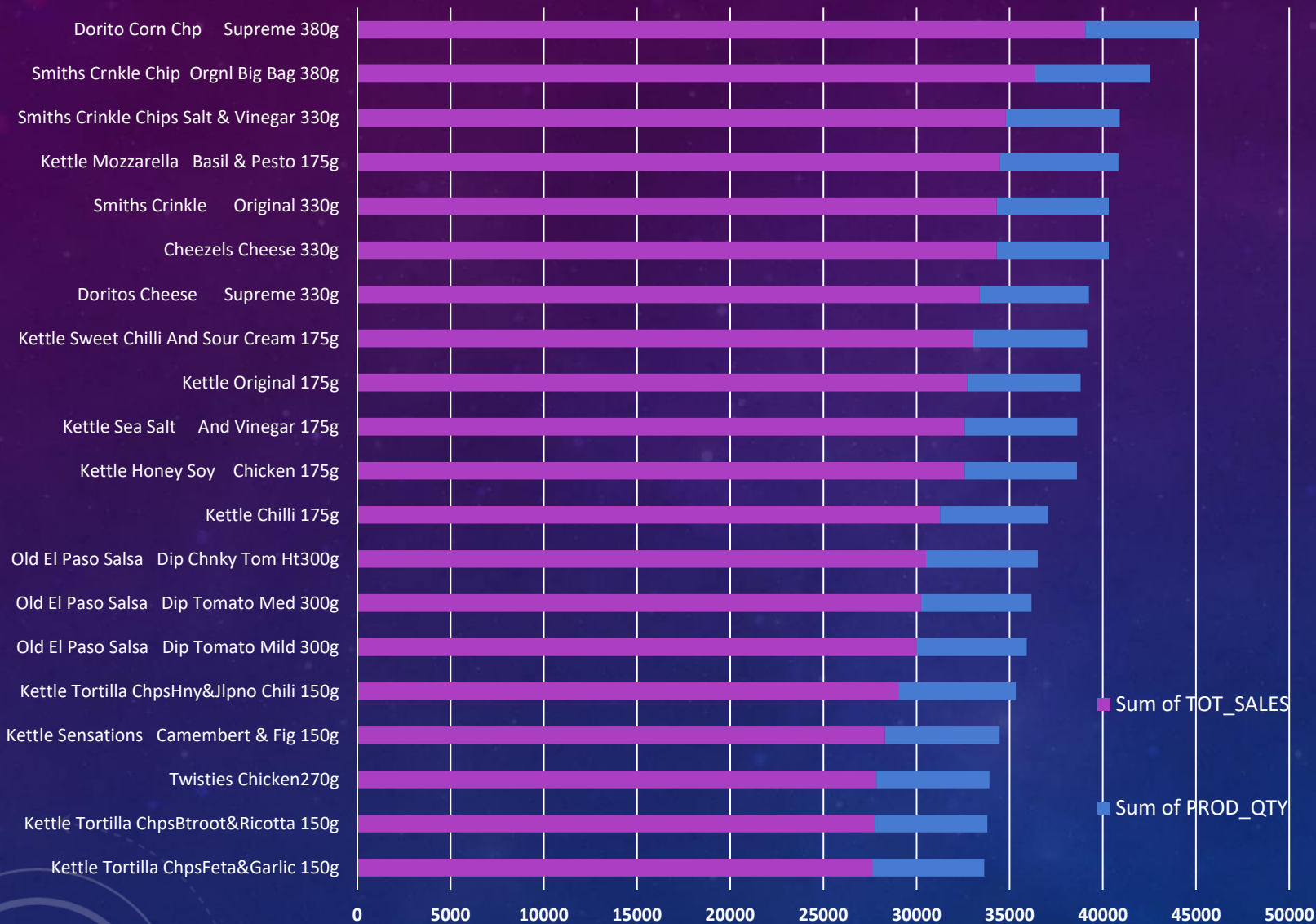
- **Sales Snapshot:** Identifying Total Sales Distribution.

Top-Selling Products

- **Popular Picks:** Acknowledge The Top Performing Products.

Sales Over Time

- **Sales Trends:** Observing sales with slight fluctuations over time.



DATA ANALYSIS

SALES ANALYSIS

Top Performing Products.

- **Dorito Corn Chip Supreme 380g:**
 - **Highest Sales:** This product significantly outpaces others in both total sales and quantity sold.

Insights:

- **Pricing and Popularity:** The relationship between total sales and product quantity highlights variations in pricing strategies and product popularity.
- **Inventory Management:** This data can inform inventory management decisions, ensuring top-selling products are adequately stocked.

DATA ANALYSIS

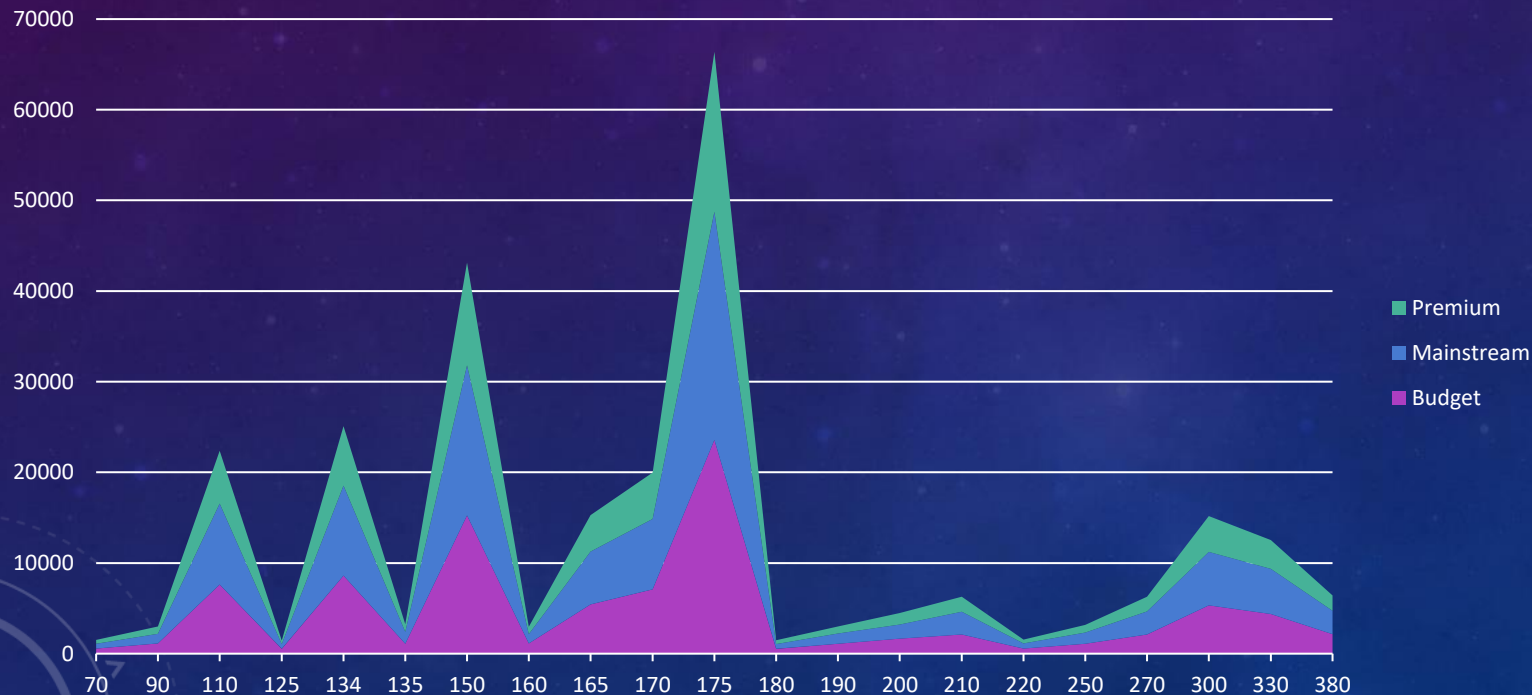
SALES ANALYSIS

Key Insights

Insights:

- **Popular Sizes Dominance:** 175 and 150 account for 41.4% of the total count.
- **Underutilized Sizes:** Sizes 180, 70, 125, and 220 have very low counts.
- **Optimization Potential:** Focus on high-performing sizes and consider phasing out underperforming ones.

PACK Size Distribution



Most Popular Pack Sizes:

175: 66,390 (25.1% of total count)

150: 43,131 (16.3% of total count)

110: 22,387 (8.5% of total count)

134: 25,102 (9.5% of total count)

These four pack sizes account for 59.4% of the total count, indicating a strong preference for these sizes.

Least Popular Pack Sizes:

180: 1,468 (0.6% of total count)

70: 1,507 (0.6% of total count)

125: 1,454 (0.5% of total count)

220: 1,564 (0.6% of total count)

These pack sizes account for only 2.3% of the total count, indicating low demand or niche usage.

Mid-Range Pack Sizes:

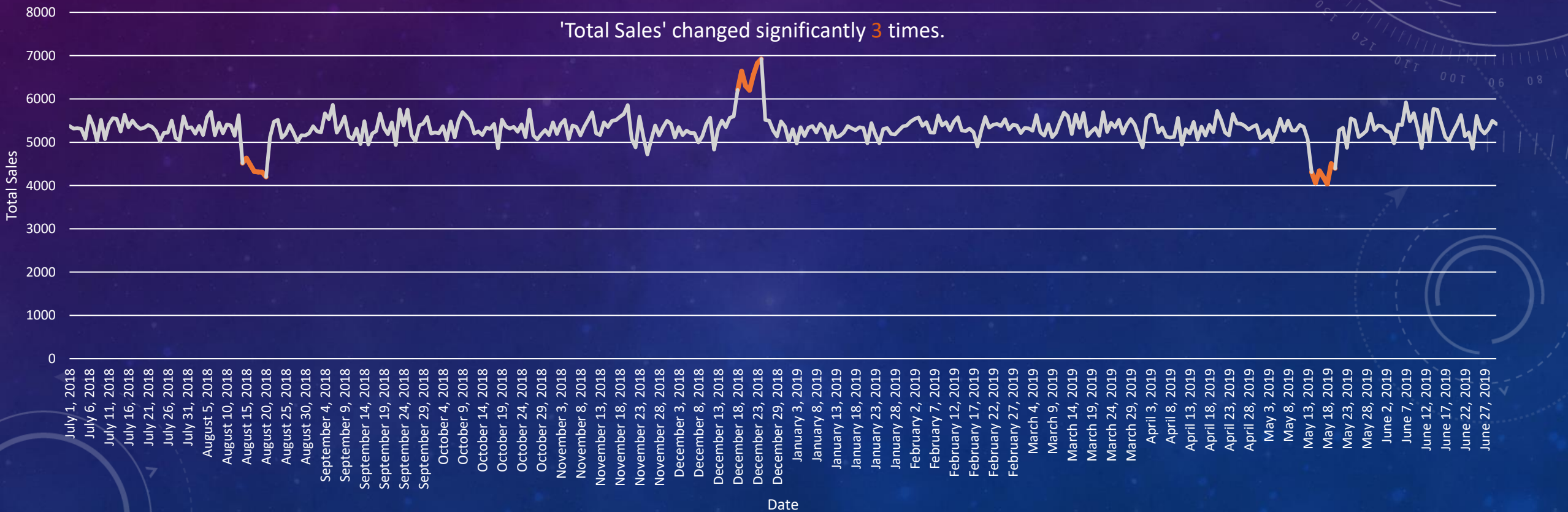
Pack sizes like **90, 135, 160, 190, 200, 210, 250, 270, and 380** have moderate counts, ranging from 2,970 to 6,285.

DATA ANALYSIS

SALES ANALYSIS

Key Observations:

- Fluctuations:** Total sales varied between **4000** and **7000** over the period.
- Peak:** In January 2019, total sales peaked at approximately **6900**, indicating a **20% increase** compared to the preceding month, likely due to seasonal promotions or holidays.
- Dip:** In May 2019, total sales dropped to around **4200**, reflecting a **39% decrease** from the January peak. This suggests a period of low demand or other external factors affecting sales.

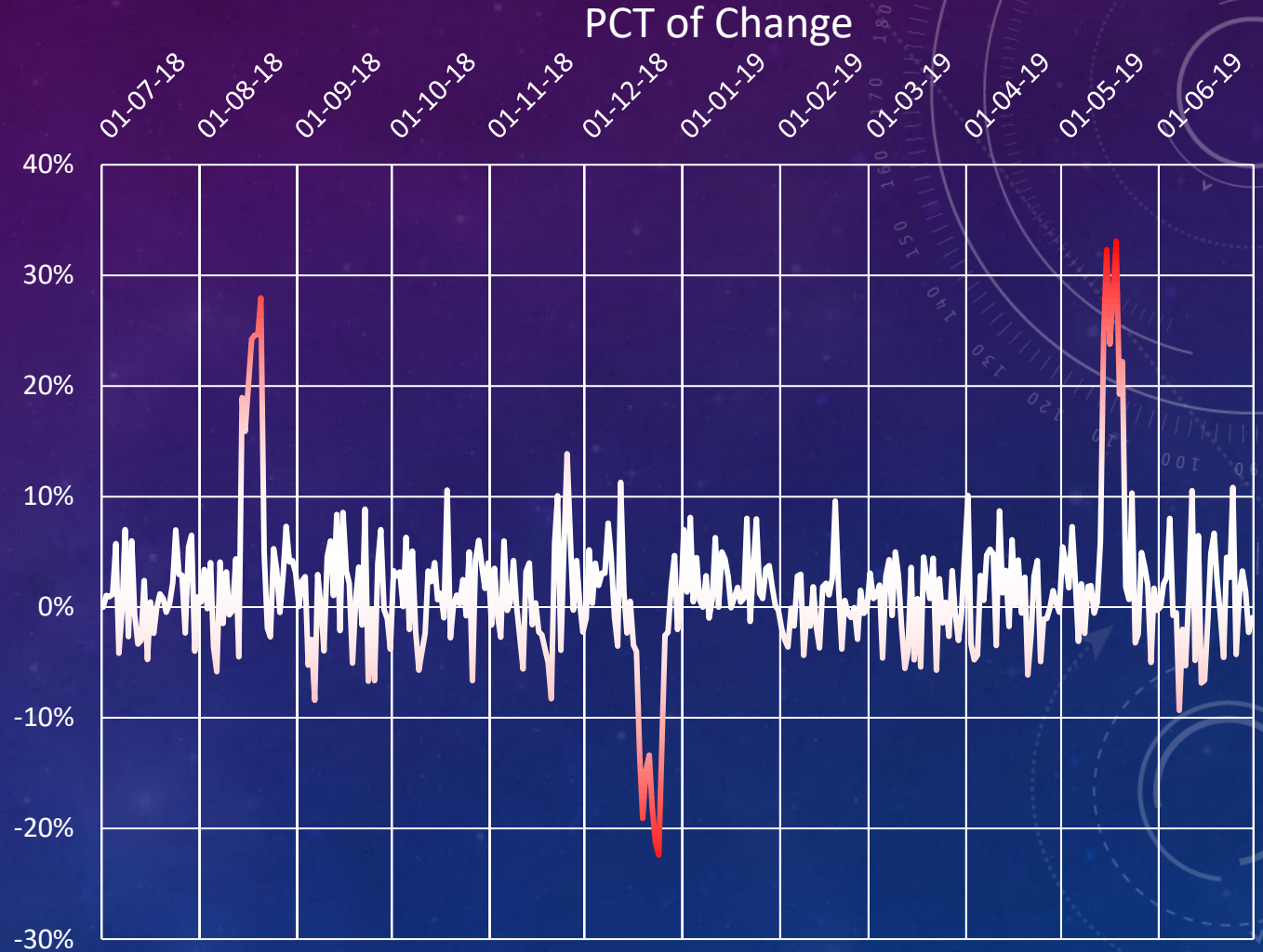


DATA ANALYSIS

SALES ANALYSIS

Key Observations:

- **Significant Spikes:**
 - Around **August 2018**, the percentage change sharply **ris**es to approximately **30%**.
 - Another major spike occurs in **May 2019**, **reaching** around **30%** again.
- **Notable Drops:**
 - A **significant dip** is observed around **December 2018**, with the percentage change dropping to approximately **-20%**.
- **Overall Volatility:**
 - The data exhibits frequent fluctuations throughout the period, indicating volatility in whatever metric is being measured.



DATA ANALYSIS

TRIAL AND CONTROL STORE

Control Store Selection.

- **Correlation Analysis:** Analyzed relationships between variables to understand their co-movement.
- **Standardized Magnitude Difference (SMD):** Measured the standardized difference in magnitude between variables.

Trial Impact Analysis.

- **Sales Impact:** Assessed the trial's influence on sales figures.
- **Customer Impact:** Examined how the trial affected customer behavior and metrics.
- **Statistical Significance:** Determined the meaningfulness of trial results statistically.

TRIAL STORE 77



Control Store Selection

Correlation Analysis: Store 233 was selected as the control store for Trial Store 77 due to its high correlation in both total sales (0.97) and number of customers (0.96) during the pre-trial period.

Standardized Magnitude Difference (SMD): The SMD value was low (0.02), indicating that Store 233 was a suitable control store for Trial Store 77.

DATA ANALYSIS TRIAL AND CONTROL STORE

Pre-Trial Period Analysis

- **Sales Comparison:** In the pre-trial period, the total sales of Trial Store 77 and Control Store 233 were closely aligned, with a scaling factor of 1.04 applied to the control store to match the trial store's sales.
- **Customer Comparison:** The number of customers in both stores was also similar, with a scaling factor of 1.02 applied to the control store.

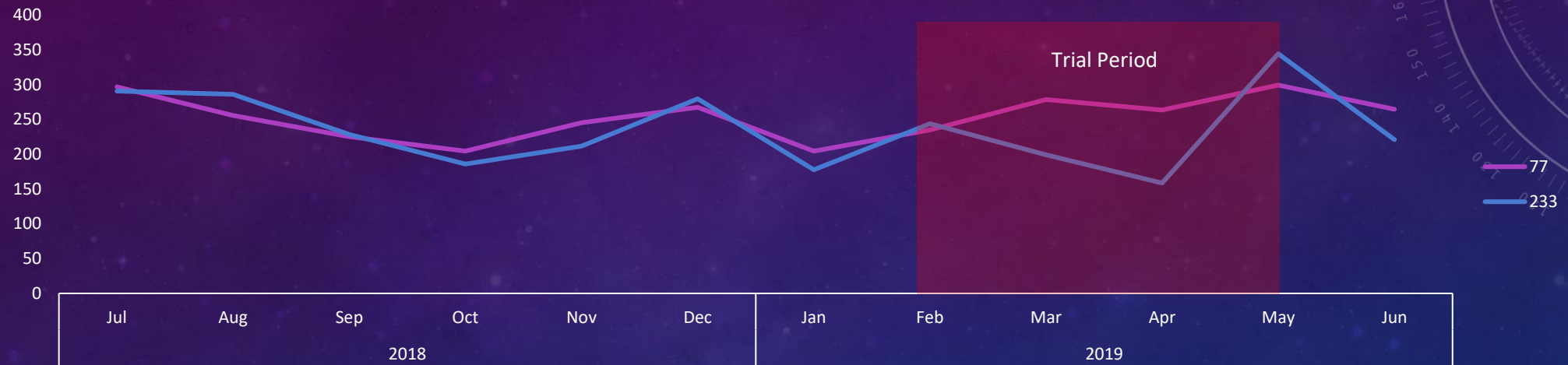
Trial Period Analysis

- **Sales Impact:** During the trial period, Trial Store 77 showed a significant increase in sales compared to the control store. For example:
 - In **March** 2019, Trial Store 77's sales were 26.39% higher than the control store.
 - In **April** 2019, the difference increased to 41.91%.
- **Statistical Significance:** The t-test results confirmed that the differences in sales during the trial period were statistically significant ($p\text{-value} < 0.05$) for all months except February 2019.
- **Customer Impact:** The number of customers in Trial Store 77 also increased significantly during the trial period, with a 22.26% increase in March 2019 compared to the control store.

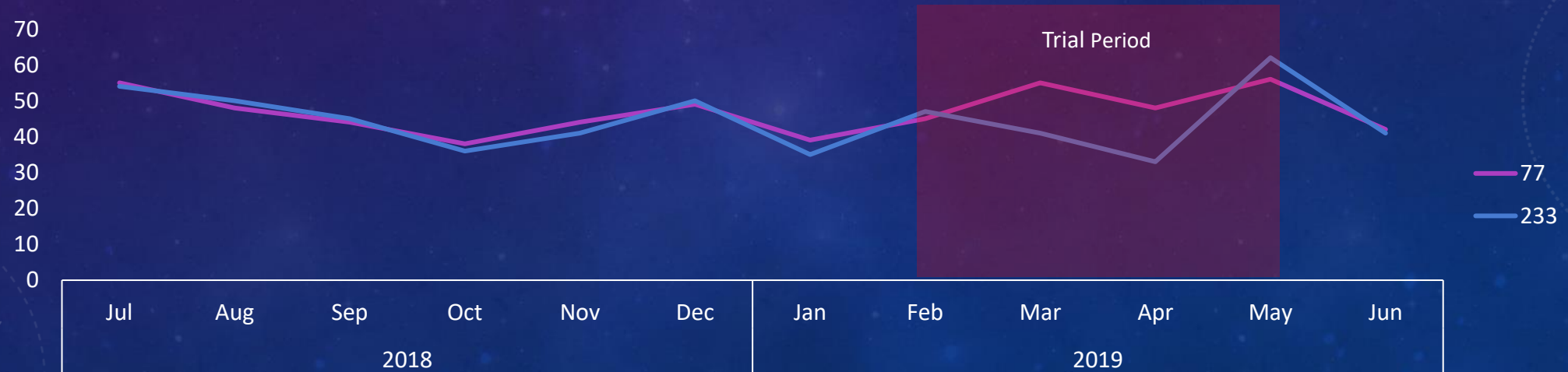
DATA ANALYSIS

TRIAL AND CONTROL STORE

Total Sales



Number Of Customers



TRIAL STORE 86



Control Store Selection

Correlation Analysis: Store 155 was selected as the control store for Trial Store 86 due to its high correlation in total sales (0.87) and number of customers (0.83) during the pre-trial period.

Standardized Magnitude Difference (SMD): The SMD value was low (0.00), indicating that Store 155 was a suitable control store for Trial Store 86.

DATA ANALYSIS

TRIAL AND CONTROL STORE

Pre-Trial Period Analysis

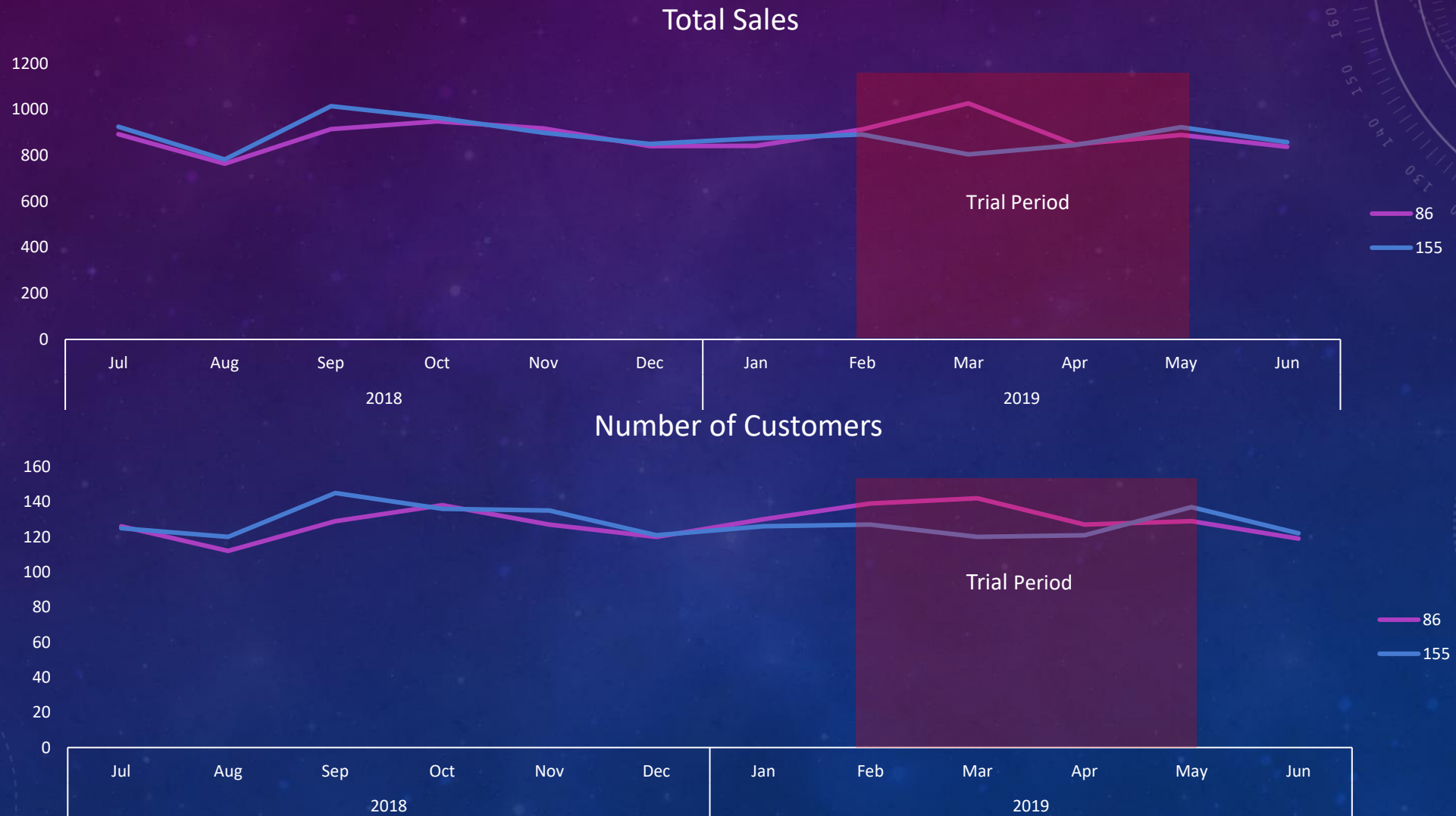
- **Sales Comparison:** In the pre-trial period, the total sales of Trial Store 86 and Control Store 155 were closely aligned, with a scaling factor of 0.97 applied to the control store to match the trial store's sales.
- **Customer Comparison:** The number of customers in both stores was also similar, with a scaling factor of 0.98 applied to the control store.

Trial Period Analysis

- **Sales Impact:** During the trial period, Trial Store 86 showed mixed results compared to the control store:
 - In **March 2019**, Trial Store 86's sales were 21.14% higher than the control store.
 - However, in **April 2019**, the difference was only 3.23%, and in May 2019, it was 1.48%.
- **Statistical Significance:** The t-test results confirmed that the differences in sales during the trial period were statistically significant ($p\text{-value} < 0.05$) for March 2019 but not for April or May 2019.
- **Customer Impact:** The number of customers in Trial Store 86 increased significantly in March 2019 (15.49% higher than the control store), but the differences in April and May were not statistically significant.

DATA ANALYSIS

TRIAL AND CONTROL STORE



TRIAL STORE 88



Control Store Selection

Correlation Analysis: Store 237 was selected as the control store for Trial Store 88 due to its high correlation in total sales (0.48) and number of customers (0.96) during the pre-trial period.

Standardized Magnitude Difference (SMD): The SMD value was low (0.00), indicating that Store 237 was a suitable control store for Trial Store 88.

DATA ANALYSIS

TRIAL AND CONTROL STORE

Pre-Trial Period Analysis

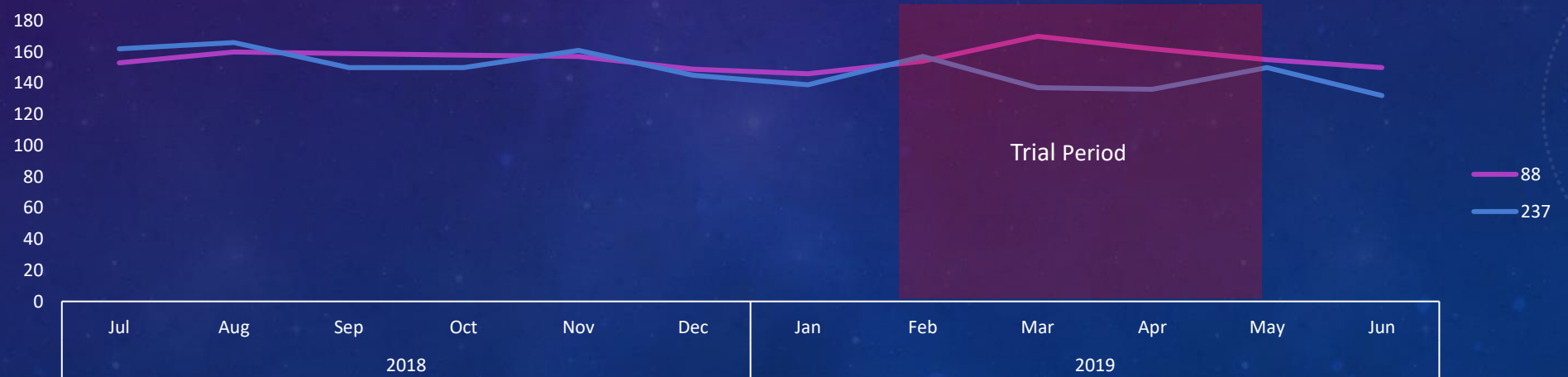
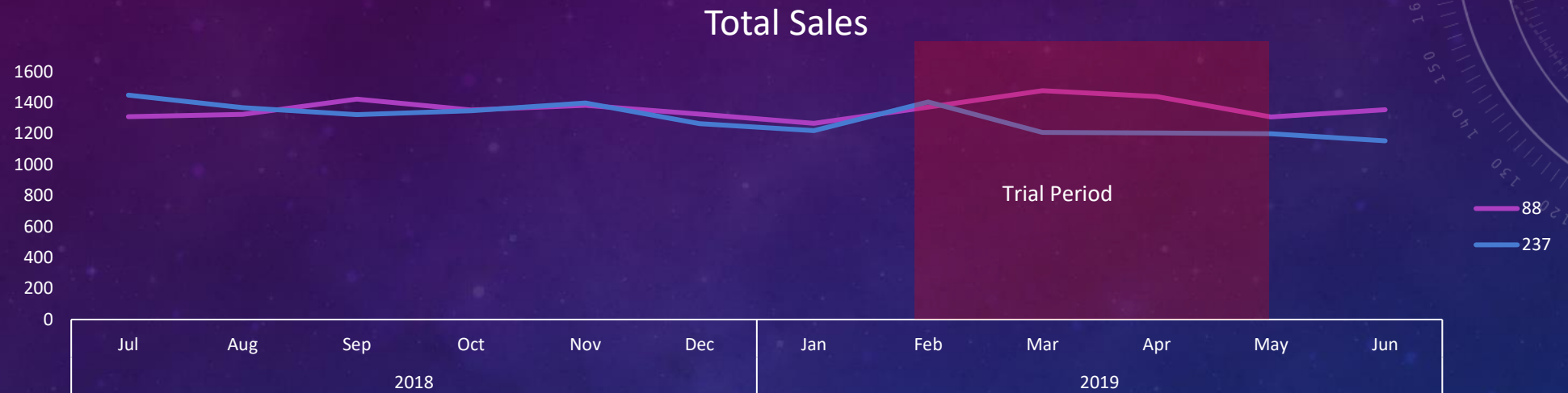
- **Sales Comparison:** In the pre-trial period, the total sales of Trial Store 88 and Control Store 237 were closely aligned, with a scaling factor of 0.99 applied to the control store to match the trial store's sales.
- **Customer Comparison:** The number of customers in both stores was also similar, with a scaling factor of 1.00 applied to the control store.

Trial Period Analysis

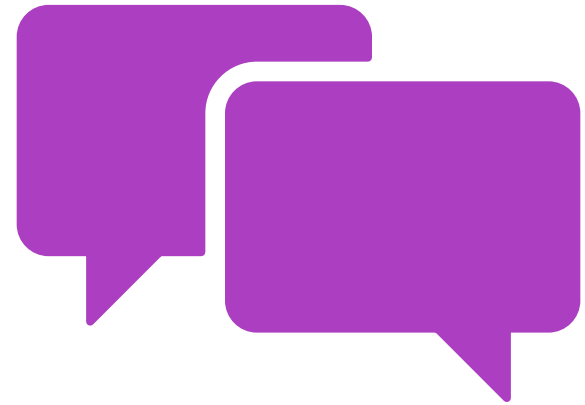
- **Sales Impact:** During the trial period, Trial Store 88 showed a significant increase in sales compared to the control store
 - In March 2019, Trial Store 88's sales were 20.47% higher than the control store.
 - In April 2019, the difference was 13.22%, and in May 2019, it was 9.65%.
- **Statistical Significance:** The t-test results confirmed that the differences in sales during the trial period were statistically significant ($p\text{-value} < 0.05$) for all months except February 2019.
- **Customer Impact:** The number of customers in Trial Store 88 increased significantly in March 2019 (13.19% higher than the control store), but the differences in April and May were not statistically significant.

DATA ANALYSIS

TRIAL AND CONTROL STORE



DISCUSSION





DISCUSSION RECOMMENDATIONS AND BUSINESS INSIGHTS

Positive Impact of the Trial

- **Trial Stores 77 and 88:** The trial had a significant positive impact on both sales and customer numbers in these stores. This indicates that the strategies implemented during the trial were effective in driving customer engagement and increasing revenue.
- **Recommendation:** Consider rolling out the trial strategy to other stores with similar customer demographics and purchasing behaviors. The success in these stores suggests that the strategy is scalable and can be applied to other locations.

Inconsistent Results in Trial Store 86

- **Trial Store 86:** While there was a significant increase in sales and customer numbers in March 2019, the effect diminished in the following months. This inconsistency suggests potential issues with the trial implementation in this store.
- **Recommendation:** Investigate the implementation process in Trial Store 86 to identify any deviations from the trial strategy. Factors such as staff training, in-store promotions, or customer engagement tactics may have differed from the other trial stores.
- **Insight:** The inconsistent results in Trial Store 86 highlight the importance of consistent execution across all trial locations. Any deviations in strategy implementation can lead to mixed results, even in stores with similar customer profiles.



DISCUSSION RECOMMENDATIONS AND BUSINESS INSIGHTS

Customer Segmentation and Targeting

- **Lifestage and Membership Analysis:** The analysis revealed that "Mainstream" memberships are most popular among Young Singles/Couples, while "Budget" memberships are more common among Older Families. This segmentation provides valuable insights into customer preferences and purchasing behavior.
- **Recommendation:** Tailor marketing campaigns and promotions to specific customer segments. For example:
 - Offer premium products and loyalty rewards to "Mainstream" customers (Young Singles/Couples) to encourage repeat purchases.
 - Focus on value-for-money promotions for "Budget" customers (Older Families) to drive sales in this segment.
- **Insight:** Understanding customer segmentation allows for more targeted and effective marketing strategies, which can lead to higher customer retention and increased sales.



DISCUSSION RECOMMENDATIONS AND BUSINESS INSIGHTS

Product and Brand Insights

- **Top-Selling Products:** Products like "Kettle Mozzarella Basil & Pesto 175g" and "Smiths Crinkle Cut Chips Chicken 170g" were among the top sellers. Additionally, the most popular pack size was 175g.
 - **Recommendation:** Increase stock levels of top-selling products and consider introducing new flavors or variants in the most popular pack sizes (e.g., 175g). This can help capitalize on existing customer preferences and drive additional sales.
 - **Insight:** Focusing on high-performing products and brands can optimize inventory management and maximize revenue.

DISCUSSION RECOMMENDATIONS AND BUSINESS INSIGHTS

Trial Store Performance Metrics

- **Sales and Customer Metrics:** The trial stores showed significant improvements in both sales and customer numbers, particularly in March and April 2019.
- **Recommendation:** Use the trial period metrics (e.g., sales growth, customer increase) as benchmarks for future store performance evaluations. This can help identify high-performing stores and areas for improvement.
- **Insight:** The trial period provided valuable data on the potential impact of strategic changes, which can be used to inform future business decisions.

Scaling the Trial Strategy

- **Scalability of the Trial:** The success of the trial in Stores 77 and 88 suggests that the strategy is scalable and can be applied to other stores with similar characteristics.
- **Recommendation:** Conduct a phased rollout of the trial strategy to other stores, starting with those that have similar customer demographics and purchasing patterns. Monitor the results closely to ensure consistency and effectiveness.
- **Insight:** A phased approach allows for controlled implementation and provides opportunities to make adjustments based on real-time feedback and performance data.



DISCUSSION RECOMMENDATIONS AND BUSINESS INSIGHTS

Statistical Significance and Data-Driven Decisions

- **Statistical Analysis:** The t-test results confirmed that the differences in sales and customer numbers during the trial period were statistically significant for most months.
- **Recommendation:** Continue to use statistical analysis to evaluate the impact of future initiatives. This ensures that business decisions are data-driven and based on reliable evidence.
- **Insight:** Statistical significance provides confidence in the results and helps validate the effectiveness of business strategies.

Customer Experience and Engagement

- **Customer Behavior:** The trial stores saw an increase in customer numbers, indicating that the trial strategy may have improved customer engagement and satisfaction.
- **Recommendation:** Focus on enhancing the customer experience through personalized promotions, loyalty programs, and in-store events. This can help build long-term customer relationships and drive repeat business.
- **Insight:** A positive customer experience is key to driving customer loyalty and increasing sales.

DISCUSSION

RECOMMENDATIONS AND BUSINESS INSIGHTS

Monitoring and Continuous Improvement

- **Ongoing Monitoring:** The trial period provided valuable insights, but continuous monitoring is essential to ensure sustained success.
- **Recommendation:** Implement a system for ongoing monitoring of key performance indicators (KPIs) such as sales, customer numbers, and product performance. Use this data to make informed decisions and continuously improve strategies.
- **Insight:** Continuous improvement is essential for maintaining competitive advantage and adapting to changing customer preferences.

CONCLUSION



CONCLUSION



THE TRIAL STRATEGY WAS SUCCESSFUL IN DRIVING SALES AND CUSTOMER ENGAGEMENT IN TWO OUT OF THREE TRIAL STORES (77 AND 88).



INCONSISTENT RESULTS IN TRIAL STORE 86 SUGGEST THE NEED FOR FURTHER INVESTIGATION INTO THE IMPLEMENTATION PROCESS.



CUSTOMER SEGMENTATION AND TARGETED MARKETING CAN OPTIMIZE SALES AND CUSTOMER RETENTION.



FOCUSING ON TOP-SELLING PRODUCTS AND POPULAR PACK SIZES CAN MAXIMIZE REVENUE.

CONCLUSION



**THE TRIAL STRATEGY IS SCALABLE
AND CAN BE ROLLED OUT TO
OTHER STORES WITH SIMILAR
CHARACTERISTICS.**



**DATA-DRIVEN DECISIONS AND
STATISTICAL ANALYSIS ARE
ESSENTIAL FOR EVALUATING THE
IMPACT OF BUSINESS STRATEGIES.**

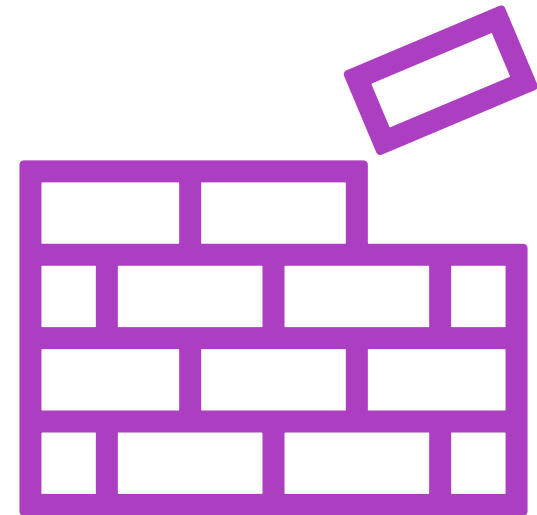


**ENHANCING THE CUSTOMER
EXPERIENCE IS KEY TO DRIVING
LONG-TERM LOYALTY AND REPEAT
BUSINESS.**



**CONTINUOUS MONITORING AND
IMPROVEMENT ARE NECESSARY
TO SUSTAIN SUCCESS AND ADAPT
TO CHANGING MARKET
CONDITIONS.**

APPENDIX



APPENDIX

A1. Data Sources and Description

- **Customer Data:** The customer dataset includes information on loyalty card numbers, lifestage, and membership type (Mainstream, Budget, Premium). It contains 72,637 entries with no missing values or duplicates.
- **Transaction Data:** The transaction dataset includes details such as date, store number, loyalty card number, product number, product name, product quantity, and total sales. It contains 264,836 entries, with one duplicate removed during preprocessing.

APPENDIX

A2. Preprocessing Steps

- **Handling Missing Values:** No null values were found in either dataset.
- **Removing Duplicates:** One duplicate transaction was identified and removed.
- **Date Format Correction:** The date column in the transaction dataset was converted to a datetime format for time-series analysis.
- **Outlier Removal:** Transactions with unusually high quantities (e.g., 200 units) were removed as they were likely not representative of typical retail behavior.

APPENDIX

A3. Customer Segmentation Analysis

- **Lifestage Distribution:**
 - Retirees: 20.38%
 - Older Singles/Couples: 20.11%
 - Young Singles/Couples: 19.88%
 - Older Families: 13.46%
 - Young Families: 12.64%
 - Midage Singles/Couples: 10.02%
 - New Families: 3.51%
- **Membership Distribution:**
 - Mainstream: 40.26%
 - Budget: 33.69%
 - Premium: 26.05%

APPENDIX

A4. Top-Selling Products

The top 5
best-
selling
products
by total
sales are:

Kettle Mozzarella Basil & Pesto 175g: \$33,043.50

Smiths Crinkle Cut Chips Chicken 170g: \$32,960.00

CCs Nacho Cheese 175g: \$29,961.90

Smiths Chip Thinly S/Cream&Onion 175g: \$29,740.50

Kettle Tortilla ChpsHny&Jlpno Chili 150g: \$29,664.00

APPENDIX

A5. Trial and Control Store Metrics

- **Trial Store 77 vs Control Store 233:**
 - Pre-Trial Scaling Factor (Sales): 1.04
 - Pre-Trial Scaling Factor (Customers): 1.02
 - Trial Period Sales Increase: Up to 41.91% in April 2019
 - Trial Period Customer Increase: Up to 22.26% in March 2019
- **Trial Store 86 vs Control Store 155:**
 - Pre-Trial Scaling Factor (Sales): 0.97
 - Pre-Trial Scaling Factor (Customers): 0.98
 - Trial Period Sales Increase: Up to 21.14% in March 2019
 - Trial Period Customer Increase: Up to 15.49% in March 2019
- **Trial Store 88 vs Control Store 237:**
 - Pre-Trial Scaling Factor (Sales): 0.99
 - Pre-Trial Scaling Factor (Customers): 1.00
 - Trial Period Sales Increase: Up to 20.47% in March 2019
 - Trial Period Customer Increase: Up to 13.19% in March 2019

APPENDIX

A6. Statistical Analysis Results

- **Trial Store 77:**
 - March 2019: t-value = 5.43 (Significant)
 - April 2019: t-value = 8.63 (Significant)
 - May 2019: t-value = 3.95 (Significant)
- **Trial Store 86:**
 - March 2019: t-value = 9.76 (Significant)
 - April 2019: t-value = 1.49 (Not Significant)
 - May 2019: t-value = 0.68 (Not Significant)
- **Trial Store 88:**
 - March 2019: t-value = 4.10 (Significant)
 - April 2019: t-value = 2.64 (Significant)
 - May 2019: t-value = 1.93 (Significant)

AHMED ELATWY

THANK YOU