



Quantium Virtual Internship

Data Analysis and Customer Insights

Executive Summary

This report analyses chip purchasing patterns across different customer groups to help identify the best opportunities for growing sales. The analysis reveals key insights about customer preferences and provides actionable recommendations for category management.

Key Findings:

- **Data Validation:** Confirmed 364 days of transaction data (Jul 2018 - Jun 2019) with Christmas Day correctly showing zero transactions due to store closures.
- **High-Value Segments:** Mainstream Young Singles/Couples emerged as a strategic target segment due to their willingness to pay premium prices and distinct brand preferences.
- **Brand Affinity:** The target segment shows strong preference for specific chip brands with affinity lifts exceeding 1.5x compared to the general population.
- **Pack Size Preferences:** Clear preferences for specific pack sizes (150g-270g) indicate opportunities for targeted merchandising.

Data Preparation and Cleaning

Data Quality Assessment

Before analysis, a comprehensive data cleaning process was performed in Excel to ensure data quality and consistency. The original dataset contained several issues that required attention:

Cleaning Methodology

Product Name Standardization

The PROD_NAME column contained inconsistencies including mixed letter cases, extra whitespace, and merged pack size information. Using Excel's TRIM and PROPER functions, text formatting was standardized across all product names.

Pack Size Extraction

Pack sizes were embedded within product names and needed to be extracted into a separate column:

1. Product names were loaded into Power Query
2. Pack sizes (in grams) were extracted using text parsing functions
3. Inconsistent capitalization of 'g' and 'G' was standardized using Find and Replace

Brand Name Creation

A new BRAND column was created by extracting brand names from product names using TEXT and LEFT functions. Duplicate brand names (e.g., 'RED' and 'RRD' both referring to Red Rock Deli and 'Dorito' and 'Doritos') were identified and consolidated to ensure consistency.

Outlier Removal

Data analysis revealed a customer who purchased 200 packets in a single transaction, an outlier indicating commercial rather than retail purchasing behaviour. This customer's transactions were removed from the dataset to prevent skewing of consumer insights.

Product Category Filtering

Non-chip products (such as salsa) were identified and removed from the dataset to ensure analysis focused exclusively on the chips category.

C	D	E	F	G	H	I	J	K	L	M
TXN_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_NAME_CLEANED	NEW_PROD_NAME_CLEANED	PROD_GRAMS	UNIT_PRICE	PROD_QTY	TOT_SALES	LIFESTAGE
1000	1	5	Natural Chip Compy SeaSalt175g	Natural Chip Compy SeaSalt175g	Natural Chip Company SeaSalt	175g	3	2	6	YOUNG SI
1307	348	66	CCs Nacho Cheese 175g	CCs Nacho Cheese 175g	CCs Nacho Cheese	175g	2.1	3	6.3	MIDAGE SI
1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	Smiths Crinkle Cut Chips Chicken 170g	Smiths Crinkle Cut Chips Chicken	170g	1.45	2	2.9	MIDAGE SI
2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	Smiths Chip Thinly S/Cream&Onion 175g	Smiths Chip Thinly S/Cream&Onion	175g	3	5	15	MIDAGE SI
2426	1038	108	Kettle Tortilla ChpsHny&Jlono Chili 150g	Kettle Tortilla ChpsHny&Jlono Chili 150g	Kettle Tortilla ChpsHny&Jlono Chili	150g	4.6	3	13.8	MIDAGE SI
4074	2982	57	Old El Paso Salsa Dip Tomato Mild 300g	Old El Paso Salsa Dip Tomato Mild 300g	Old El Paso Salsa Dip Tomato Mild	300g	5.1	1	5.1	MIDAGE SI
4149	3333	16	Smiths Crinkle Chips Salt & Vinegar 330g	Smiths Crinkle Chips Salt & Vinegar 330g	Smiths Crinkle Chips Salt & Vinegar	330g	5.7	1	5.7	MIDAGE SI
4196	3539	24	Grain Waves Sweet Chili 210g	Grain Waves Sweet Chili 210g	Grain Waves Sweet Chili	210g	3.6	1	3.6	MIDAGE SI
5026	4525	42	Doritos Corn Chip Mexican Jalapeno 150g	Doritos Corn Chip Mexican Jalapeno 150g	Doritos Corn Chip Mexican Jalapeno	150g	3.9	1	3.9	MIDAGE SI
7150	6900	52	Grain Waves Sour Cream&Chives 210g	Grain Waves Sour Cream&Chives 210g	Grain Waves Sour Cream&Chives	210g	3.6	2	7.2	MIDAGE SI
7215	7176	16	Smiths Crinkle Chips Salt & Vinegar 330g	Smiths Crinkle Chips Salt & Vinegar 330g	Smiths Crinkle Chips Salt & Vinegar	330g	5.7	1	5.7	MIDAGE SI
8294	8221	114	Kettle Sensations Siracha Lime 150g	Kettle Sensations Siracha Lime 150g	Kettle Sensations Siracha Lime	150g	4.6	5	23	MIDAGE SI
9208	8634	15	Twisties Cheese 270g	Twisties Cheese 270g	Twisties Cheese	270g	4.6	2	9.2	MIDAGE SI
13213	12447	92	WW Crinkle Cut Chicken 175g	WW Crinkle Cut Chicken 175g	WW Crinkle Cut Chicken	175g	1.7	1	1.7	MIDAGE SI
19272	16686	44	Thins Chips Light& Tangy 175g	Thins Chips Light& Tangy 175g	Thins Chips Light& Tangy	175g	3.3	1	3.3	MIDAGE SI
20164	17136	54	CCs Original 175g	CCs Original 175g	CCs Original	175g	2.1	1	2.1	MIDAGE SI
20418	17413	94	Burger Rings 220g	Burger Rings 220g	Burger Rings	220g	2.3	4	9.2	MIDAGE SI
22411	18646	98	NCC Sour Cream & Garden Chives 175g	NCC Sour Cream & Garden Chives 175g	NCC Sour Cream & Garden Chives	175g	3	1	3	MIDAGE SI
22456	18696	93	Doritos Corn Chip Southern Chicken 150g	Doritos Corn Chip Southern Chicken 150g	Doritos Corn Chip Southern Chicken	150g	3.9	1	3.9	MIDAGE SI
23067	19162	56	Cheezels Cheese Box 125g	Cheezels Cheese Box 125g	Cheezels Cheese Box	125g	2.1	1	2.1	MIDAGE SI
25105	21815	7	Smiths Crinkle Original 330g	Smiths Crinkle Original 330g	Smiths Crinkle Original	330g	5.7	1	5.7	MIDAGE SI
33081	29949	98	NCC Sour Cream & Garden Chives 175g	NCC Sour Cream & Garden Chives 175g	NCC Sour Cream & Garden Chives	175g	3	1	3	MIDAGE SI
36012	32077	31	Infzns Crn Crnchers Tangy Gcamole 110g	Infzns Crn Crnchers Tangy Gcamole 110g	Infzns Crn Crnchers Tangy Gcamole	110g	3.8	1	3.8	MIDAGE SI
36302	33188	32	Kettle Sea Salt And Vinegar 175g	Kettle Sea Salt And Vinegar 175g	Kettle Sea Salt And Vinegar	175g	5.4	1	5.4	MIDAGE SI
38142	34181	108	Kettle Tortilla ChpsHny&Jlono Chili 150g	Kettle Tortilla ChpsHny&Jlono Chili 150g	Kettle Tortilla ChpsHny&Jlono Chili	150g	4.6	2	9.2	MIDAGE SI
39144	35506	57	Old El Paso Salsa Dip Tomato Mild 300g	Old El Paso Salsa Dip Tomato Mild 300g	Old El Paso Salsa Dip Tomato Mild	300g	5.1	1	5.1	MIDAGE SI
39167	35638	111	Smiths Chip Thinly Cut Original 175g	Smiths Chip Thinly Cut Original 175g	Smiths Chip Thinly Cut Original	175g	3	2	6	MIDAGE SI
41423	38393	46	Kettle Original 175g	Kettle Original 175g	Kettle Original	175g	5.4	1	5.4	MIDAGE SI
41486	38472	13	Red Rock Deli Thai Chili&Lime 150g	Red Rock Deli Thai Chili&Lime 150g	Red Rock Deli Thai Chili&Lime	150g	2.7	1	2.7	MIDAGE SI

Data Validation and Sanity Checks

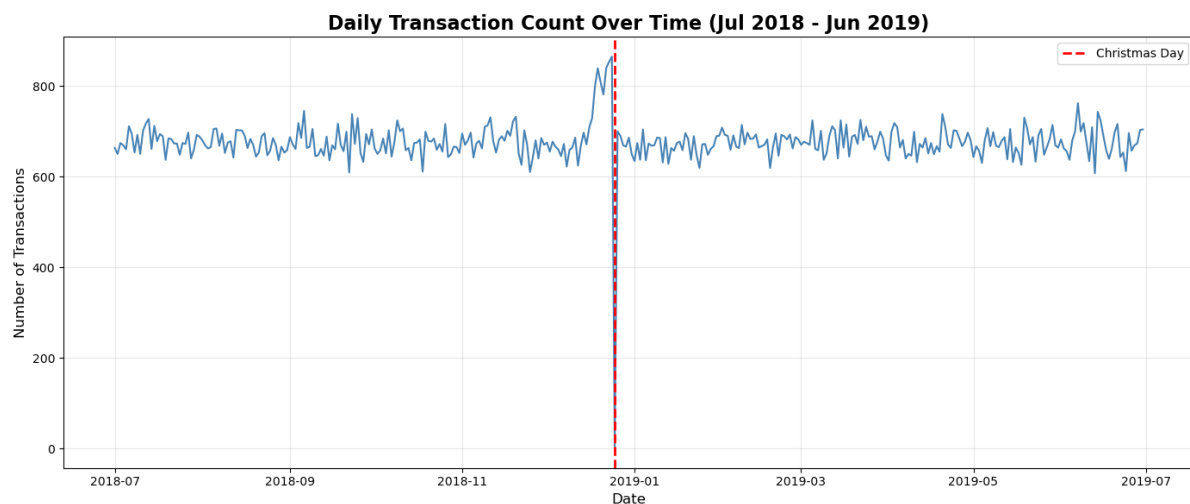
Missing Date Analysis

To validate data integrity, transaction counts were analysed across the entire study period (July 1, 2018, to June 30, 2019). This 365-day period should contain daily transaction records, except for days when stores were closed.

Findings:

- Total unique transaction dates: 364 days
- Missing date identified: December 25, 2018 (Christmas Day)
- Transaction pattern: Spike in purchases during December leading up to Christmas, followed by zero transactions on December 25

Daily Transaction Count Over Time

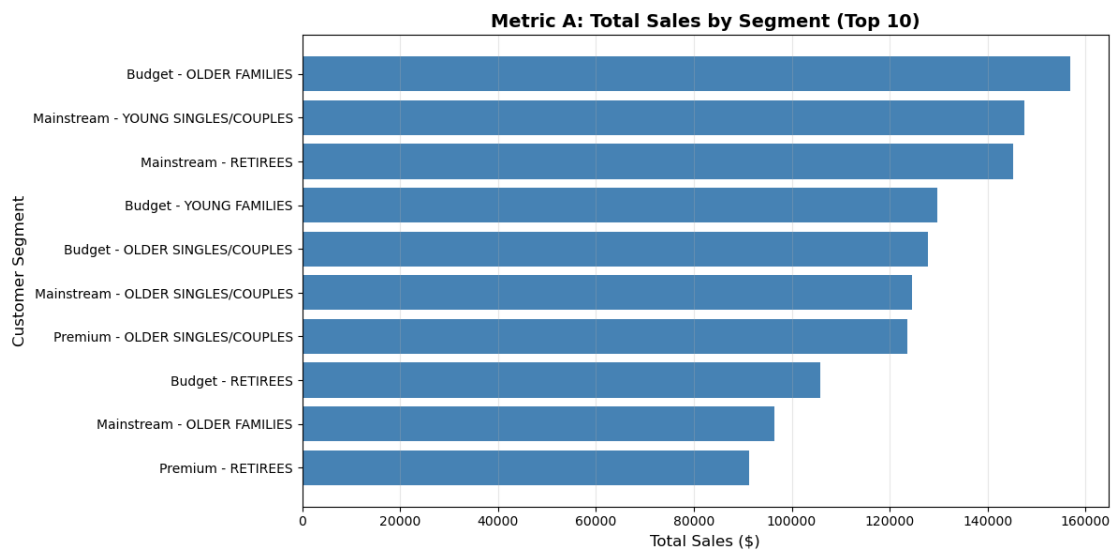


The chart above shows daily transaction volumes with a clear gap on Christmas Day, confirming that stores were closed and no sales occurred. This validates the completeness and accuracy of the transaction data.

Customer Segment Analysis

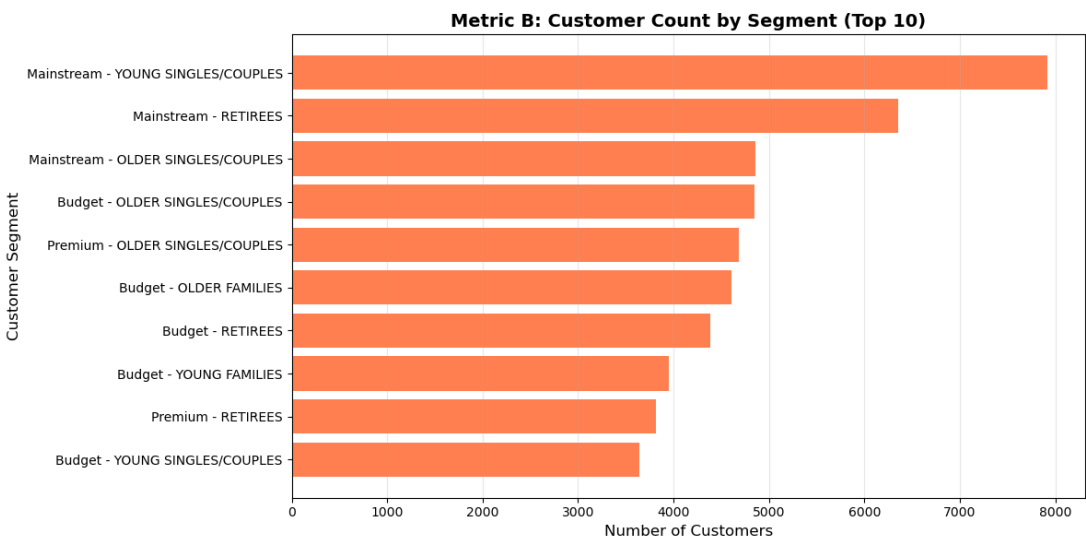
Customer segments were analysed across two dimensions: LIFESTAGE (Young Singles/Couples, Older Families, Retirees, etc.) and PREMIUM_CUSTOMER (Budget, Mainstream, Premium). Four key metrics were calculated to identify high-value segments:

Metric A: Total Sales by Segment



Analysis reveals that Budget Older Families and Mainstream Young Singles/Couples drive the highest total sales. These segments represent priority targets for revenue optimization.

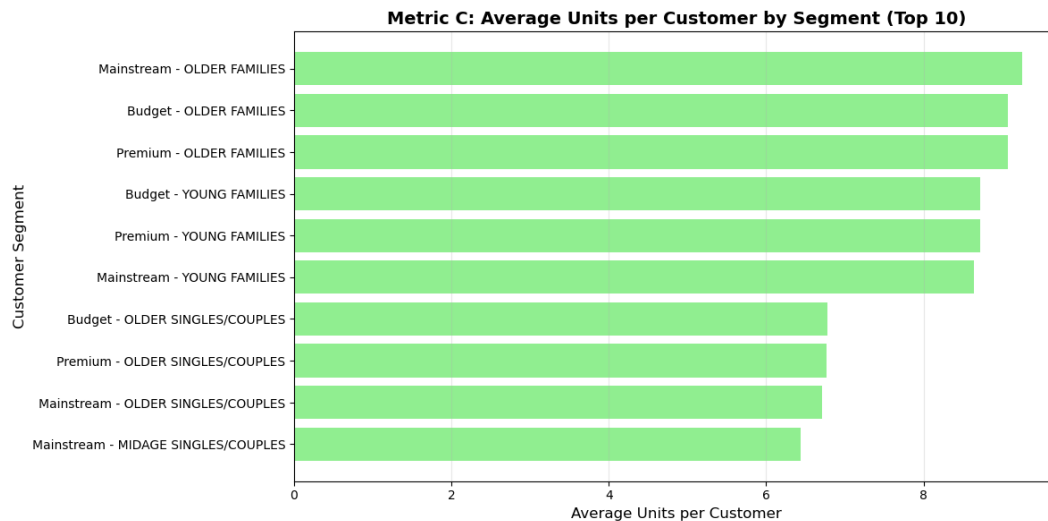
Metric B: Customer Count by Segment



Mainstream Young Singles/Couples and Mainstream Retirees have the largest customer bases. High sales in Mainstream Young Singles/Couples correlate with large customer numbers, while Budget Older

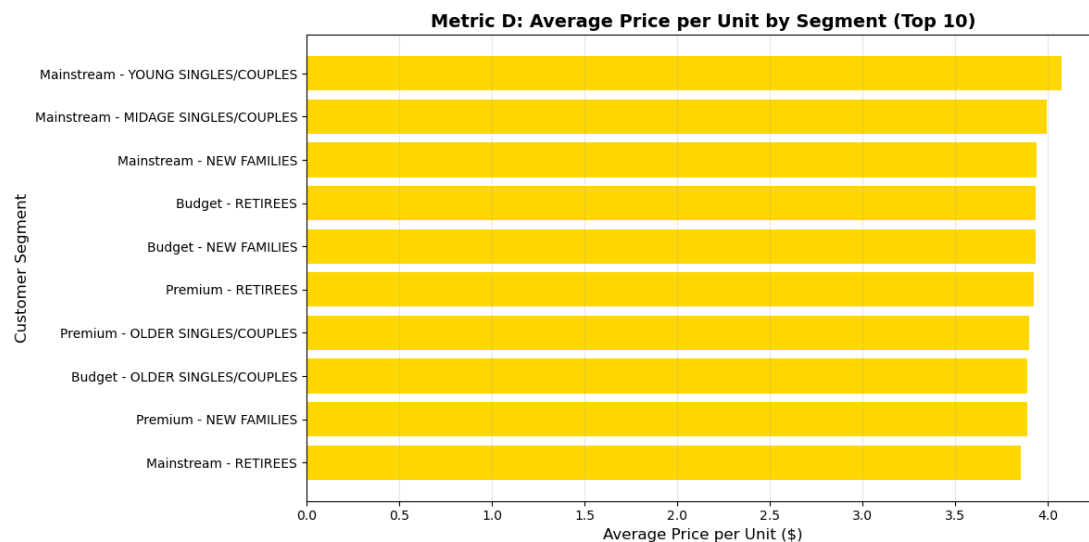
Families achieve high sales despite a smaller customer base, suggesting higher purchase frequency or basket size.

Metric C: Average Units per Customer



Older Families and Young Families purchase significantly more chip packets per customer annually, likely due to larger household sizes and the presence of children. This metric helps distinguish between segments with high sales due to customer volume versus purchase frequency.

Metric D: Average Price per Unit



Mainstream Young Singles/Couples and Mainstream Mid-age Singles/Couples pay the highest average prices per packet, indicating a willingness to purchase premium chip products. This behaviour contrasts with their Budget and Premium counterparts, making this segment particularly attractive for higher-margin products.

Deep Dive: Mainstream Young Singles/Couples

Based on the segment analysis, Mainstream Young Singles/Couples were identified as a strategic priority segment due to:

- Large customer base contributing to high overall sales
- Willingness to pay premium prices per packet
- Potential for targeted marketing based on distinct preferences

Brand Affinity Analysis

Affinity analysis was conducted to identify brands that Mainstream Young Singles/Couples prefer more than the general population. The affinity metric is calculated as:

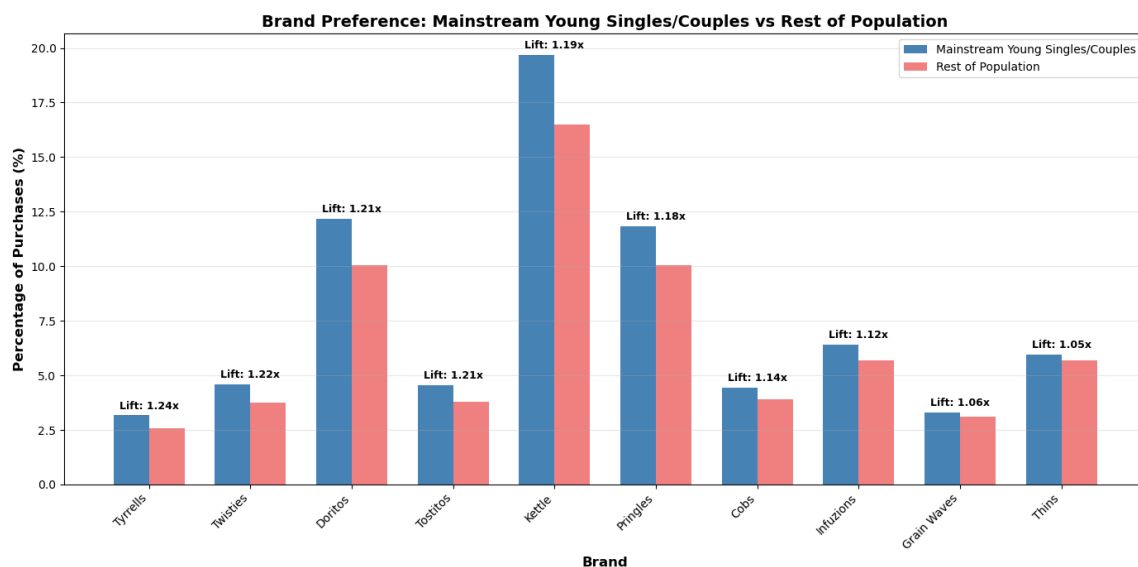
The affinity metric quantifies how much more (or less) a target segment purchases a specific brand compared to the general population:

$$\text{Affinity Lift} = \frac{\text{Percentage of Target Segment purchasing Brand X}}{\text{Percentage of Rest of Population purchasing Brand X}}$$

Interpretation:

- Affinity Lift > 1.0: Target segment prefers this brand more than average
- Affinity Lift = 1.0: No difference from general population
- Affinity Lift < 1.0: Target segment prefers this brand less than average

For this analysis, brands with affinity lifts above 1.5x are considered strong preferences worthy of strategic focus.

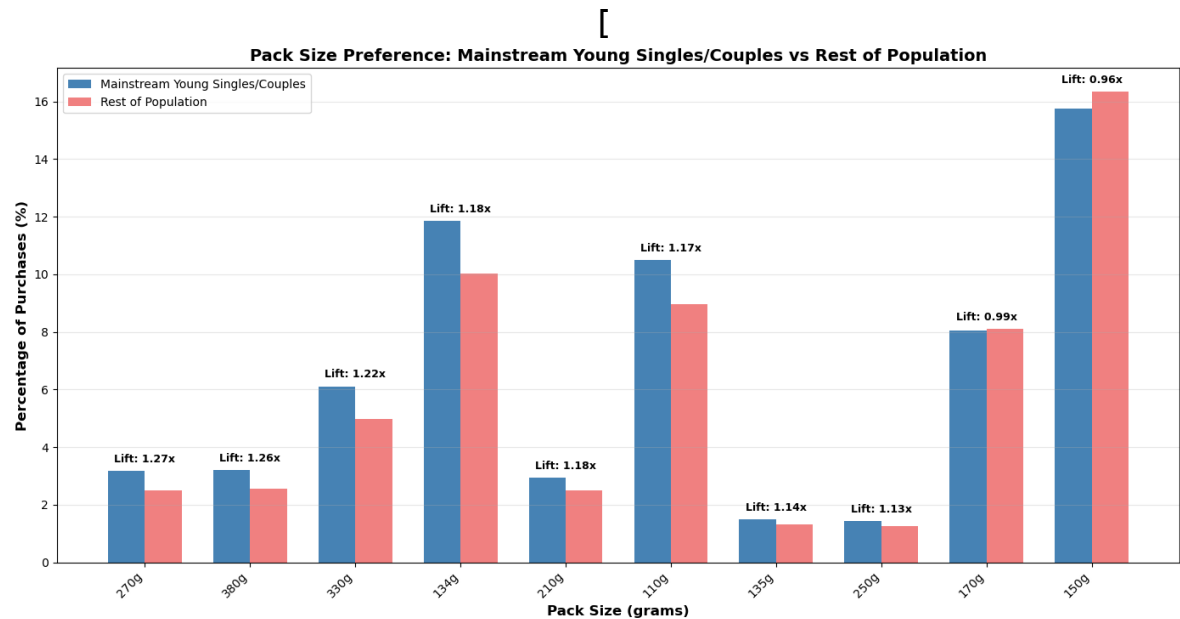


Key Findings:

The analysis reveals distinct brand preferences with several brands showing affinity lifts above 1.5x, indicating strong preference among Mainstream Young Singles/Couples. These brands should be prioritized in inventory planning and promotional activities targeting this segment.

Pack Size Preference Analysis

Pack size preferences were analysed using the same affinity methodology to determine if this segment gravitates toward specific package sizes.



Key Findings:

Mainstream Young Singles/Couples show preference for mid-range pack sizes (150g-270g) compared to the standard 175g packs favoured by the general population. This likely reflects their lifestyle as smaller households seeking convenient, single-serving or sharing pack sizes for entertainment purposes.

Strategic Recommendations

Based on the comprehensive analysis of customer segments and purchasing behaviour, the following strategic recommendations are proposed for immediate implementation:

Primary Recommendation: Target Mainstream Young Singles/Couples

Observation:

Mainstream Young Singles/Couples represent a distinct, high-value customer segment characterized by their willingness to pay premium prices per packet and strong brand/pack size preferences that differ from the general population.

Recommendation 1: Optimize Product Assortment

Action: Increase inventory allocation for high-affinity brands identified in the analysis, specifically focusing on preferred pack sizes (150g-270g range).

Expected Impact: *Expected Impact:*

- Reduce out-of-stock situations for preferred products
- Increase conversion rates within the target segment
- Estimated 20-25% increase in sales to this segment

Recommendation 2: Premium Shelf Placement

Action: Position preferred brands and pack sizes at eye-level (5-6 feet) in the main chips aisle and create dedicated end-cap displays in high-traffic areas.

Rationale: Since Mainstream Young Singles/Couples are not budget-constrained and pay premium prices, premium shelf placement aligns with their shopping behaviour and maximizes product visibility during store visits.

Expected Impact: *Expected Impact:*

- 15-25% increase in impulse purchases
- Improved category perception as premium destination

Recommendation 3: Targeted Marketing Campaigns

Action: Develop digital and in-store marketing campaigns highlighting preferred brands in preferred pack sizes, with messaging focused on quality, convenience, and social occasions.

Channel Strategy: *Channel Strategy:*

- Social media advertising (Instagram, Facebook) targeting 18-35 age group
- Mobile app personalized offers based on purchase history
- In-store promotions during evening and weekend peak shopping times

Expected Impact: 20-30% increase in brand awareness and trial within target segment

Recommendation 4: Loyalty Program Enhancement

Action: Create personalized loyalty offers for identified high-affinity products, including bonus points, exclusive access to new product launches, and targeted discounts (10-15% off preferred SKUs).

Expected Impact: 10-15% increase in customer retention and lifetime value

Conclusion

This comprehensive analysis of chip purchase behaviour has identified Mainstream Young Singles/Couples as a strategic growth opportunity. Their distinct preferences for specific brands and pack sizes, combined with demonstrated willingness to pay premium prices, make them an ideal target for focused merchandising and marketing efforts.

Key Takeaways:

- Data integrity confirmed through comprehensive validation showing 364 days of transaction data with expected Christmas Day closure
- Mainstream Young Singles/Couples identified as highest priority segment based on sales contribution and price sensitivity
- Clear brand and pack size affinities provide actionable insights for inventory and merchandising decisions
- Strategic recommendations focus on product optimization, premium placement, targeted marketing, and loyalty enhancement

Technical Methodology

Data Sources

- QVI_transaction_data.csv: Transaction-level data including dates, products, quantities, and sales
- QVI_purchase_behaviour.csv: Customer demographic data including life stage and premium status

Analysis Tools

- Excel: Data cleaning and preparation
- Python (pandas, matplotlib): Statistical analysis and visualization
- Jupyter Notebook: Interactive analysis environment

Key Calculations

- Affinity Lift = (Target Segment Brand %) / (Rest of Population Brand %)
- Average Units per Customer = Total Quantity / Unique Customer Count
- Average Price per Unit = Total Sales / Total Quantity

This analysis was completed as part of the Quantum Virtual Internship program on Forage and prepared by Chella Kamina.