Retail Transaction and Customer Segmentation Analysis

Quantium Virtual Internship case study | Updated portfolio report

Executive Summary

This analysis examined retail transaction behaviour and customer segments to identify the groups driving sales and translate the findings into commercial recommendations. The cleaned dataset contained 264,834 transaction rows after two abnormal 200-unit purchases were removed. Transaction activity was stable for most of the year, rose sharply above 900 daily transactions immediately before Christmas, fell to zero on Christmas Day because stores were closed, and then returned to its normal range.

The segment visuals show that revenue concentration is driven by a mix of customer volume and spending intensity. Budget Older Families generated 8.7% of sales from 6.4% of customers, while Mainstream Young Singles/Couples formed the largest customer group at 11.1% but produced only 8.2% of sales. A separate price comparison found that Mainstream young and mid-age singles/couples paid 4.05 per unit on average, compared with 3.69 for comparable Budget and Premium customers.

264,836	264,834	8.7%	+9.7%
Raw transaction rows	Rows after cleaning	Largest sales segment	Price difference

Business Objective

The objective was to understand transaction trends, compare customer population share with sales contribution, and identify segments suitable for targeted merchandising, pricing and promotional decisions.

Data Preparation and Analytical Method

- Converted the transaction date field from its spreadsheet serial format into a usable calendar date.
- Validated product records and removed two transactions in which one customer purchased 200 packs at a time; this behaviour was treated as commercial rather than ordinary retail purchasing.
- Retained 264,834 rows for analysis and combined transaction measures with lifestage and affluence classifications.
- Compared total sales, customer share, average units and average price per unit across Budget, Mainstream and Premium segments.
- Used a Welch two-sample t-test to test whether observed differences in unit price were statistically meaningful.

Transaction Trends Over Time

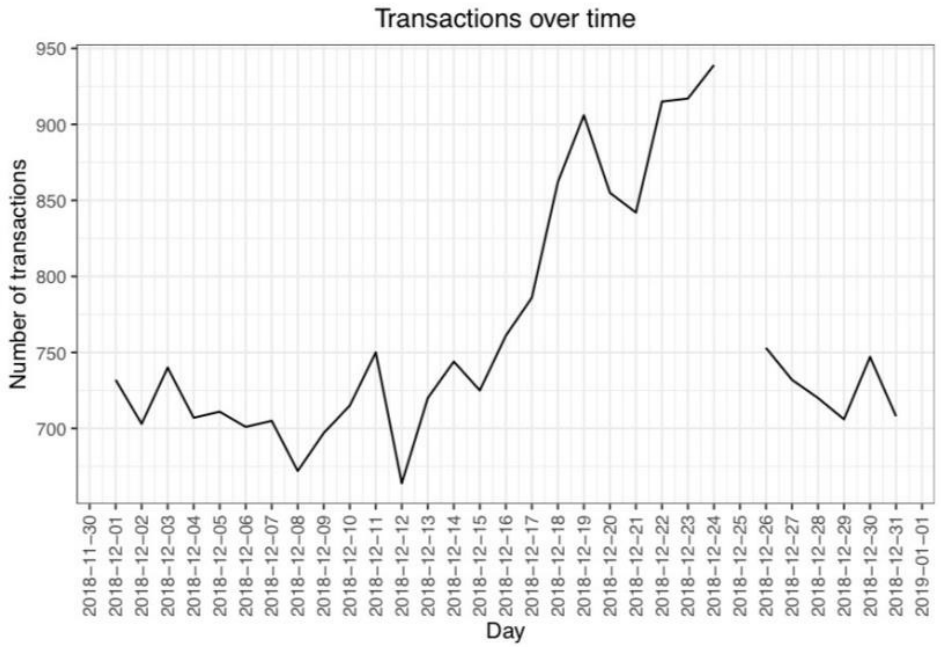

Figure 1. Daily transaction volume across the observation period.

Visual interpretation

Most days cluster around roughly 670-790 transactions, indicating a stable baseline rather than sustained growth. The sharp rise to more than 900 transactions occurs immediately before Christmas and is followed by a gap on Christmas Day. Because activity returns to its previous range afterwards, the spike is best interpreted as seasonal demand caused by holiday shopping rather than a permanent shift in customer behaviour.

Commercial implication: promotional displays and stock availability should be increased in the lead-up to Christmas, but demand forecasts should not treat the short spike as a new long-term baseline.

Customer Segmentation and Sales Contribution

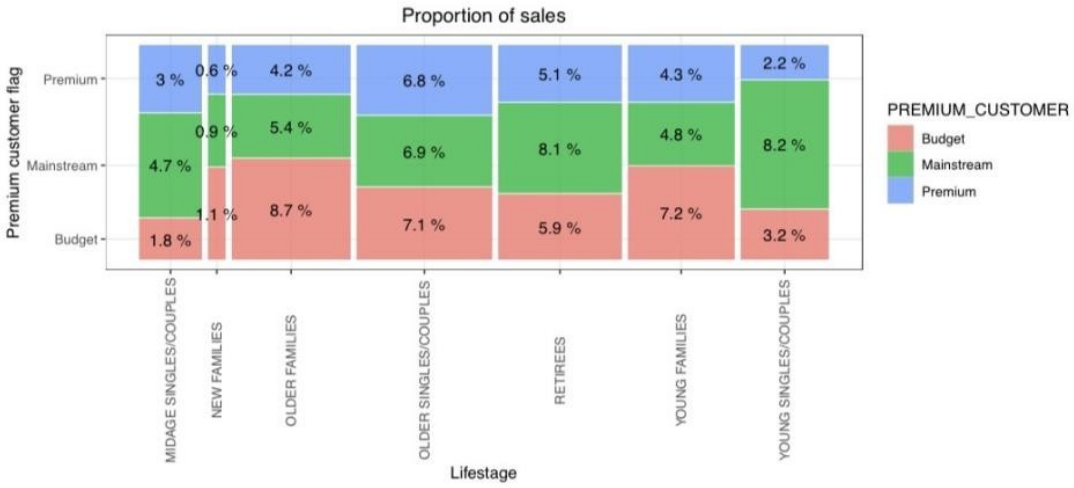

Figure 2. Proportion of total sales by lifestage and customer segment.

Visual interpretation

The largest individual sales contributions come from Budget Older Families (8.7%), Mainstream Young Singles/Couples (8.2%) and Mainstream Retirees (8.1%). The width of each lifestage block reflects the relative size of that lifestage, while the coloured sections show how sales split across Budget, Mainstream and Premium customers. A large sales block can therefore reflect either a large customer base, higher spending per customer, or both.

Customer Distribution and Revenue Efficiency

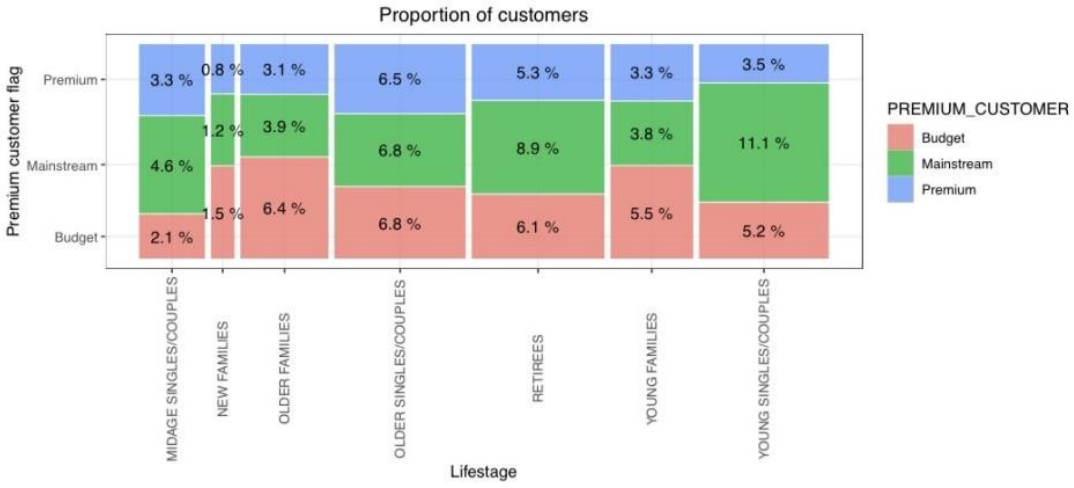

Figure 3. Proportion of customers by lifestage and customer segment.

Visual interpretation

Mainstream Young Singles/Couples are the largest customer segment at 11.1%, followed by Mainstream Retirees at 8.9%. However, customer share alone does not reveal value. Comparing Figures 2 and 3 shows that Budget Older Families and Premium Older Families generate substantially more sales than their population share would suggest, whereas Mainstream Young Singles/Couples are numerous but contribute a smaller share of sales relative to their size.

Quantified Segment Comparison

Segment	Customer share	Sales share	Sales-to-customer index	Interpretation
Budget Older Families	6.4%	8.7%	1.36	36% revenue over-index
Premium Older Families	3.1%	4.2%	1.35	35% revenue over-index
Budget Young Families	5.5%	7.2%	1.31	31% revenue over-index
Mainstream Young Singles/Couples	11.1%	8.2%	0.74	Large base; lower sales per customer
Mainstream Retirees	8.9%	8.1%	0.91	High volume; near-average value

An index above 1.00 means that a segment contributes a larger share of sales than its share of customers. This makes Budget Older Families, Premium Older Families and Budget Young

Families stronger candidates for retention and basket-building initiatives than the customer-count chart alone would suggest.

Price and Product Affinity Findings

Measure	Target group	Comparison group	Result
Average price per unit	Mainstream young/mid-age singles and couples	Equivalent Budget and Premium customers	4.05 vs 3.69 (+9.7%)
Welch t-test	Same groups	One-sided difference test	t = 40.61; p < 2.2e-16
Tyrrells affinity	Mainstream Young Singles/Couples	All other customers	23.6% more likely to purchase
270 g pack affinity	Mainstream Young Singles/Couples	All other customers	Approximately 27% more likely

The price result indicates that the target Mainstream groups are willing to pay more per unit, even though their overall sales contribution is partly driven by population size. The brand and pack-size affinities suggest that merchandising near high-traffic areas could increase impulse purchases, but the effect should be validated through controlled testing.

Recommendations

1. Prioritise Budget Older Families and Premium Older Families for retention offers because both segments materially over-index on sales relative to customer share.
2. Use pre-Christmas displays and stock planning to capture the short seasonal increase without inflating forecasts for the remainder of the year.
3. For Mainstream Young Singles/Couples, focus on basket expansion, product visibility and premium pack mix rather than customer acquisition alone.
4. Test Tyrrells, Twisties and larger-pack placement in discretionary spaces frequented by younger Mainstream shoppers.
5. Track margin, repeat purchase and customer lifetime value before treating revenue share as a complete measure of segment attractiveness.

Limitations and Next Steps

The analysis does not include product-level margin, customer lifetime value or longitudinal retention. The next iteration should add basket analysis, repeat-purchase behaviour and controlled experiments to measure whether recommended merchandising changes produce sustained incremental value.

Source: Quantium Virtual Internship retail analytics case data. Amounts are reported in the source dataset's original units.