Trial vs Control Store Analysis

Quantium Virtual Internship case study | Store 77 deep dive and cross-store summary

Executive Summary

This analysis evaluated whether a store trial produced a meaningful uplift in sales and customer numbers. Trial store 77 was matched with control store 233 because their pre-trial sales and customer trends were similar. The comparison used both trend correlation and standardised magnitude similarity, with equal weight assigned to sales and customer measures.

During the three-month evaluation window, store 77 exceeded the control store's expected sales range in two of three months. The sales test produced t-values of 7.34 in March and 12.48 in April, both well above the 1.895 critical value. Customer numbers were also significantly above the matched control in two of three months. The trial therefore shows a positive and statistically meaningful effect in specific periods, but not a uniformly sustained uplift across every month.

77	233	2 of 3	1.895
Trial store	Matched control	Significant sales months	Critical t-value

Methodology

- Calculated monthly sales, unique customers, transactions per customer, units per transaction and average price per unit for each store.
- Restricted control candidates to stores with complete observations across the pre-trial period.
- Scored candidate controls using correlation and standardised magnitude similarity for sales and customer counts.
- Combined the sales and customer similarity scores with equal weights; store 233 was selected as the best control for store 77.
- Scaled control performance and compared trial results with the control store's 5th-95th percentile confidence range.

Matched Store Summary

Trial store	Matched control	Sales result	Customer result
77	233	Significant uplift in 2 of 3 months	Significant uplift in 2 of 3 months
86	155	Not consistently significant	Significantly higher in all 3 months
88	237	Significant uplift in 2 of 3 months	Significant uplift in 2 of 3 months

The cross-store pattern indicates that the intervention was promising but not identical across locations. Store 86 attracted more customers without a comparably strong sales uplift, which may indicate lower prices, promotional discounting or smaller baskets.

Pre-Trial Sales Alignment

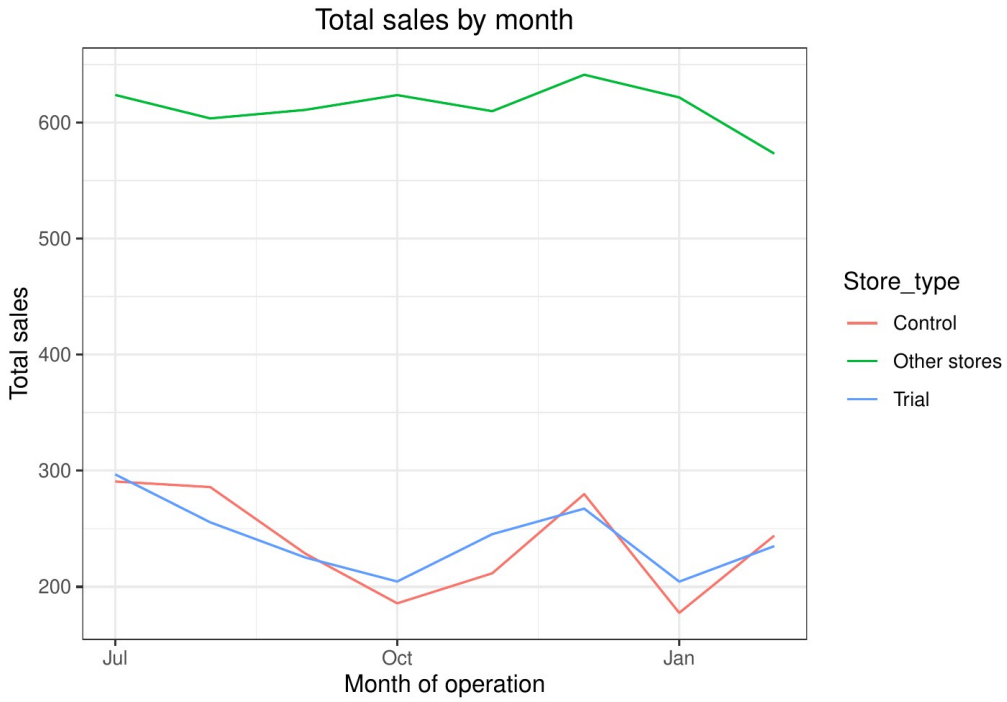

Figure 1. Monthly sales before the trial: store 77, control store 233 and the wider-store benchmark.

Visual interpretation

The trial and control lines move in the same direction across the pre-trial months and remain within a similar sales range of roughly 180-300 units. The broader-store average is much higher, at approximately 575-645, but it is not the comparison target. What matters for causal assessment is that store 77 and store 233 display similar movement before the intervention, reducing the likelihood that later differences are caused by pre-existing trend differences.

Pre-Trial Customer Alignment

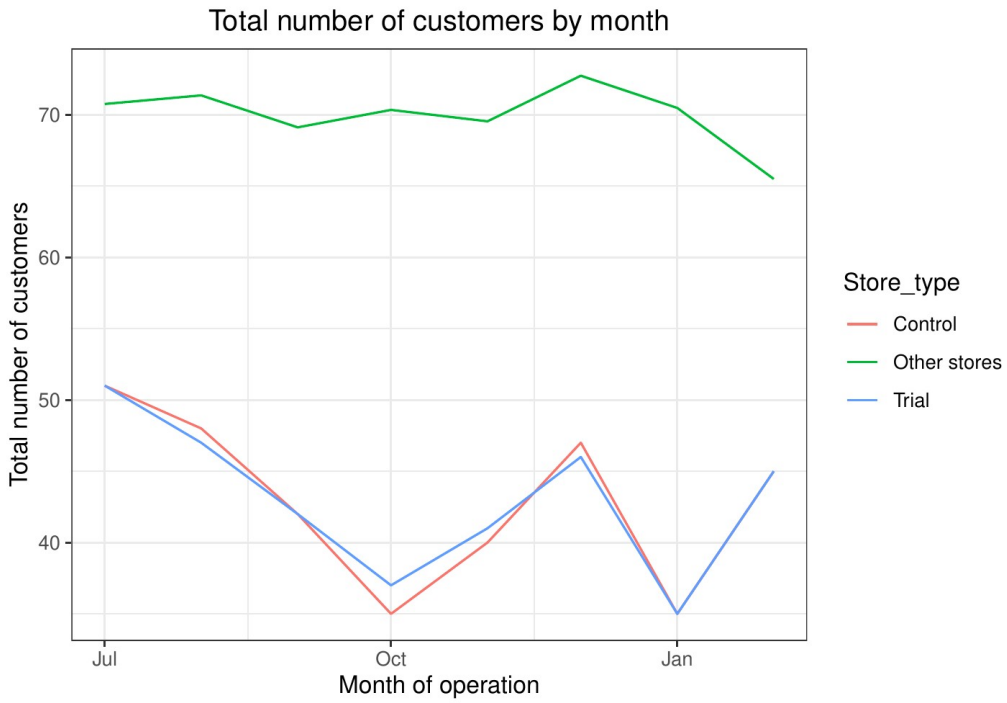

Figure 2. Monthly customer counts before the trial for store 77 and control store 233.

Visual interpretation

Customer counts for the trial and control stores closely overlap, ranging from approximately 35 to 51 customers per month. Both stores decline into October, recover towards December and fall again in January. This close alignment strengthens the control match because it shows similarity in both sales and traffic, rather than in revenue alone.

Sales Uplift During the Trial

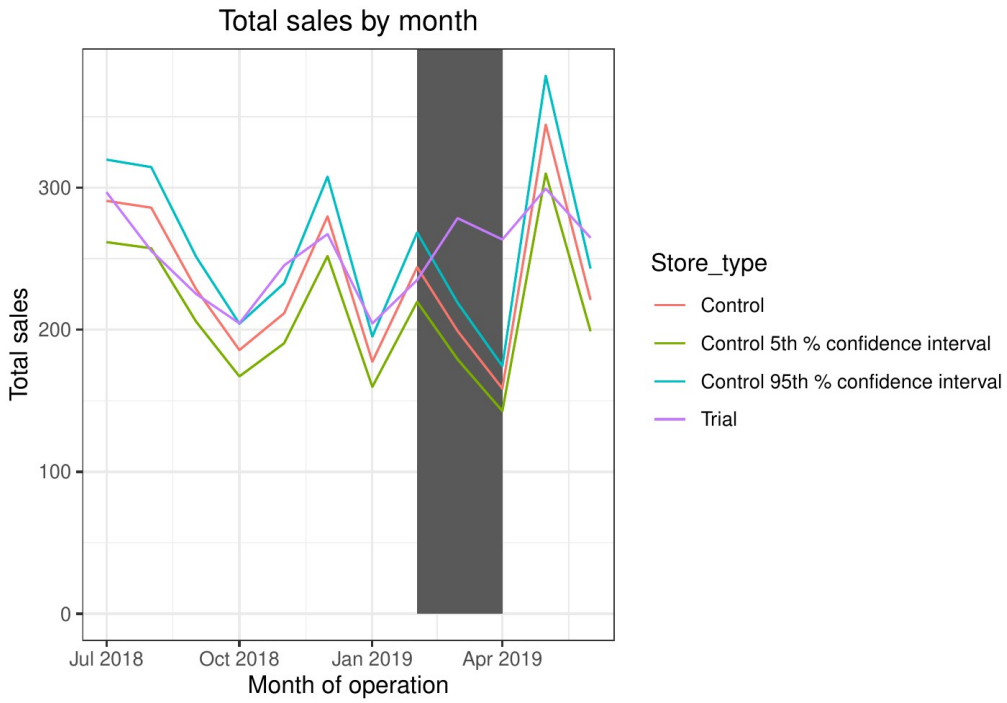

Figure 3. Store 77 sales compared with the scaled control and its 5th-95th percentile confidence range.

Visual interpretation

The shaded region marks the evaluation window. In the first evaluated month, the trial line remains within the expected control range, so the difference is not statistically distinguishable from normal variation. In March and April, the trial line rises above the upper confidence boundary. The corresponding t-values are 7.34 and 12.48, compared with a critical value of 1.895, providing strong evidence of a real sales uplift in those two months.

Statistical Assessment of Sales

Month	t-value	Critical value	Interpretation
February 2019	1.18	1.895	Not statistically significant
March 2019	7.34	1.895	Statistically significant uplift
April 2019	12.48	1.895	Statistically significant uplift

Because only two of the three months exceed the threshold, the correct conclusion is partial statistical significance rather than a claim that the trial created a uniformly significant uplift throughout the full period.

Customer Uplift During the Trial

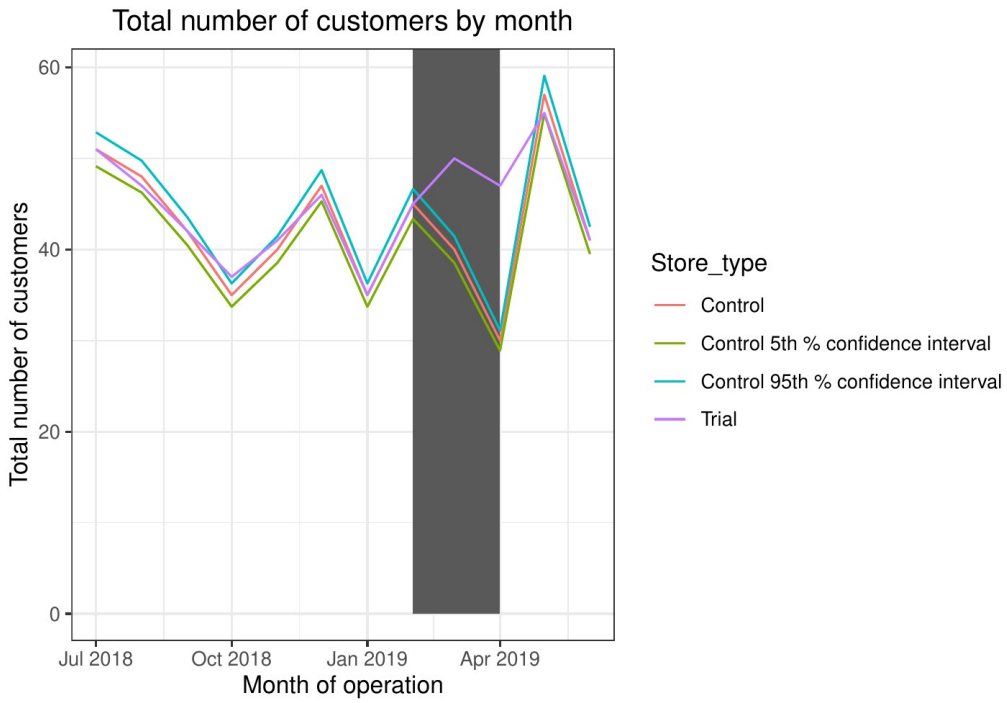

Figure 4. Store 77 customer counts compared with the scaled control confidence range.

Visual interpretation

The trial store exceeds the upper customer-count confidence boundary in two of the three evaluated months. The visual therefore supports the same directional conclusion as the sales analysis: the intervention increased engagement during specific months, but the uplift was not consistently outside the expected range for the entire evaluation window.

Business Interpretation

The combination of higher customer counts and higher sales suggests that the store intervention attracted additional shoppers and converted at least part of that traffic into revenue. However, the store-level variation shows why a full rollout would be premature. Store 86, for example, recorded significantly higher customer numbers in all three months without consistently significant sales, implying that traffic uplift and revenue uplift should be monitored separately.

Recommendations

1. Extend the test period to increase statistical power and reduce the effect of short-term volatility.
2. Replicate the intervention across additional stores matched on pre-trial sales and customer trends.
3. Monitor sales, unique customers, basket size, transactions per customer and average price concurrently.
4. Investigate store 86 for discounting or pricing changes that may explain higher traffic without equivalent sales growth.

5. Use a phased rollout with continued control-store monitoring rather than an immediate national launch.

Limitations

The analysis is based on a short trial window and a limited number of stores. Seasonal effects, local promotions and operational differences may influence results. Confidence intervals support inference, but they do not remove every source of bias; a longer and broader experiment would provide stronger evidence of sustained causal impact.

Source: Quantium Virtual Internship experimentation and uplift-testing case data. Sales amounts are reported in the source dataset's original units.