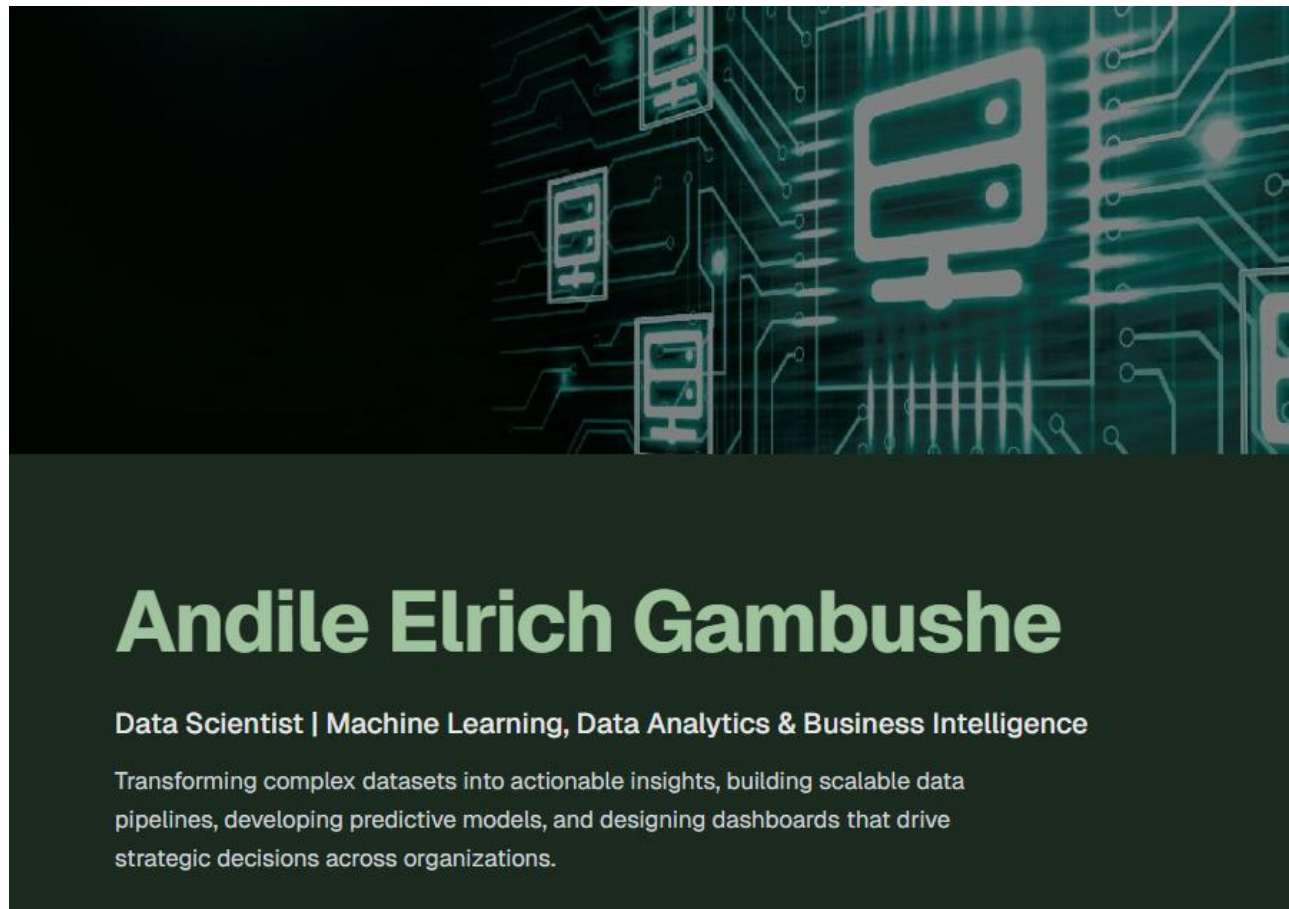


Quantum Retail Layout Trial Data Analysis Report



Andile Elrich Gambushe

Data Scientist | Machine Learning, Data Analytics & Business Intelligence

Transforming complex datasets into actionable insights, building scalable data pipelines, developing predictive models, and designing dashboards that drive strategic decisions across organizations.

Job simulation focused on Data Analytics and Commercial Insights

29/11/2025

Contents

Introduction.....	3
Data Exploration	3
Methods.....	4
Modelling and Results.....	4
Conclusion & Recommendations	6

Introduction

The purpose of this analysis was to understand customer purchase behaviour and store transaction patterns for a retail chain, with the ultimate goal of providing actionable insights for strategic decision-making. Using transaction data alongside customer purchase behaviour data, we aimed to identify sales trends, uncover inconsistencies, and evaluate key performance metrics across stores. Furthermore, the project involved defining metrics to select appropriate control stores for trial experiments, enabling comparison between trial stores and their respective controls. By examining both historical and trial-period data, we intended to assess the effectiveness of interventions and support evidence-based recommendations to optimize sales performance.

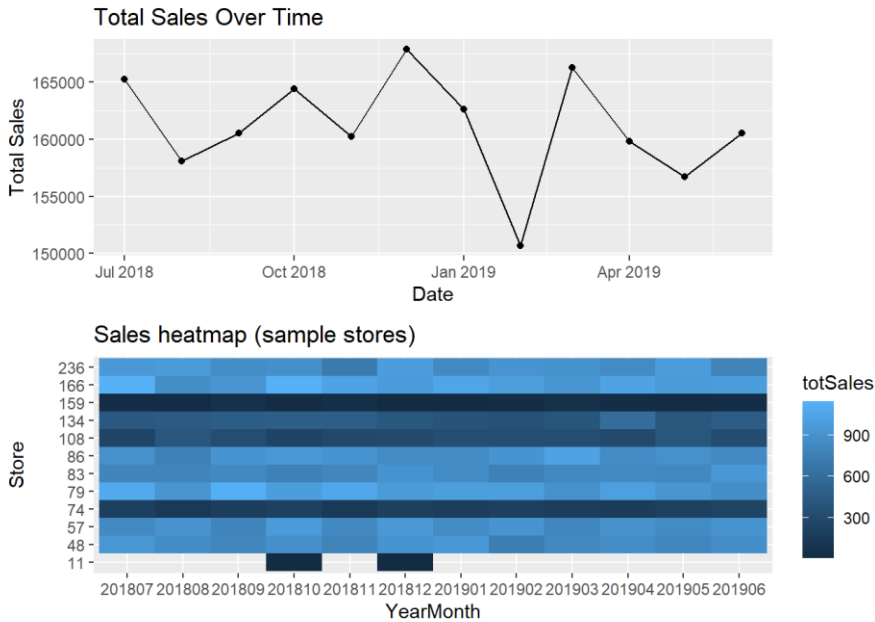
Data Exploration

Initial exploration of the datasets revealed a comprehensive record of transactions and customer activities spanning 40 stores over the period from 2018-01-01 to 2019-12-31. Data cleaning steps included standardizing column names, converting date formats, removing invalid loyalty card numbers and ensuring numeric fields were properly formatted. Product names were processed to extract pack sizes and brands, while yearmonth allowed for time-based aggregation.

Aggregating data at the store-month level produced key metrics such as total sales (totSales), number of transactions (nTransactions), number of unique customers (nCustomers), average price per unit (avgPricePerUnit), transactions per customer, and quantity per transaction. There were no missing values, esp. in tot_sales, indicating clean transaction data.

Visualizations:

- **Total Sales Over Time** (gg1) shows an overall upward trend with seasonal peaks in December and dips in mid-year.
- **Sales Heatmap** (gg2) highlights variability across stores, with some stores consistently performing above the median and others below.



Methods

To select suitable control stores for trial experiments, we applied a combination of correlation and magnitude similarity metrics. Correlation measured how closely a trial store's metrics (total sales and customer count) matched other stores over the pre-trial period, while magnitude similarity captured differences in sales levels. A weighted driver score ($0.5 * \text{correlation} + 0.5 * \text{magnitude}$) was used to rank potential control stores.

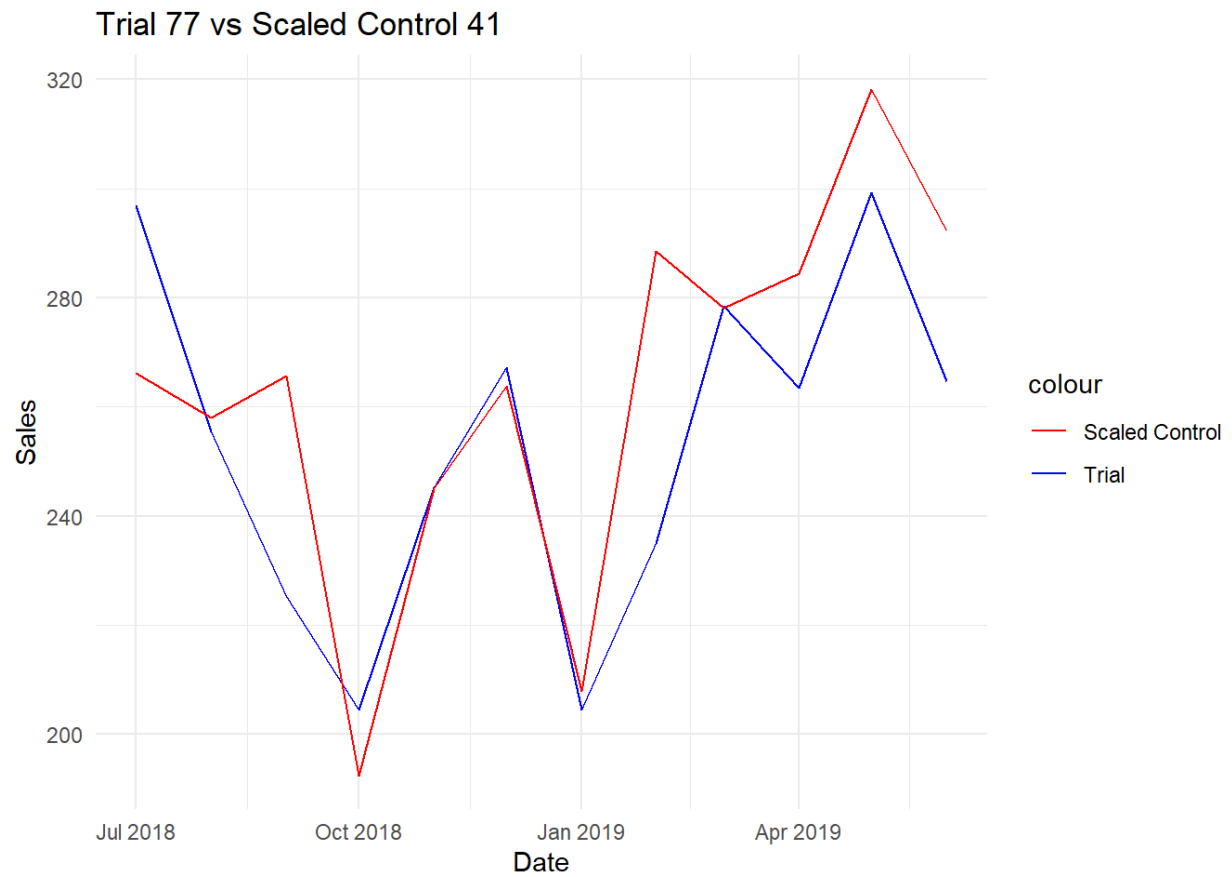
Three trial stores (77, 86, 88) were analyzed. Only stores with at least 8 pre-trial months were considered eligible, resulting in 36 eligible stores. For trial analysis, pre-trial sales were used to compute scaling factors so that control store sales aligned with the trial store baseline. Absolute percentage differences (abs_pct_diff) were calculated for each month, and significant deviations were identified using a t-like threshold ($t_{\text{crit}} = 2.31$ for trial store 77, $df = 11$).

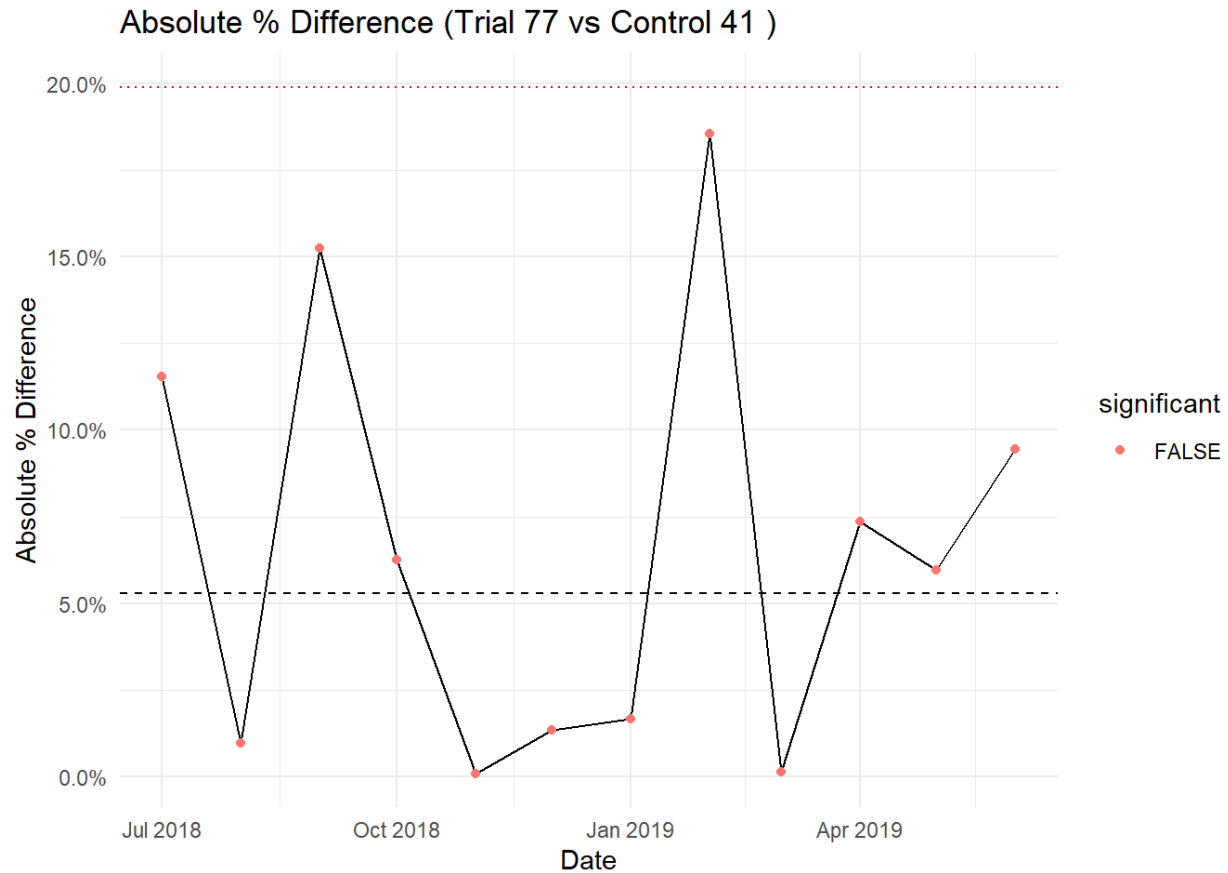
Modelling and Results

The analysis identified trial stores and their most comparable control stores. For example, trial store 77 had store 41 as its top control candidate. Pre-trial baseline metrics showed a mean absolute percentage difference of $\mu = 0.03$ (3%) and standard deviation $\sigma = 0.015$ (1.5%), with a scaling factor of 1.12 applied to the control store's sales.

Time series comparisons between trial and scaled control stores highlighted months with significant deviations. For trial store 77, March 2019 and April 2019 showed absolute percentage differences exceeding the threshold, indicating meaningful increases in sales relative to control.

Correlation analysis confirmed that total sales strongly aligned with customer count ($r = 0.85$) and number of transactions ($r = 0.88$), suggesting these as primary sales drivers.





Conclusion & Recommendations

The findings indicate that the combination of correlation and magnitude-based metrics is effective for selecting control stores comparable to trial stores. Trial stores exhibited significant deviations from controls in specific months, suggesting that interventions had measurable effects on sales performance.

Based on these insights, it is recommended that management continue monitoring trial store outcomes against scaled controls to evaluate future initiatives. Stores showing consistent underperformance may benefit from targeted marketing or operational adjustments, while top-performing stores could serve as models for best practices. Overall, the analysis highlights the importance of data-driven decision-making, leveraging historical sales trends, customer behaviour metrics, and robust control store selection to optimize retail performance.