Analysis of Promotion's Effects on Sales

Introduction

When individuals know there is a promotion going on, whether that is in a supermarket, retail store or movie theater, they are more likely to be willing to purchase the product since they are getting a discounted price for it. This is because with promotional discounts, consumers stop paying attention to original pricing of products, and instead start focusing on how much they are saving (Graciola et al., 2018). As such, they stockpile so that they don't have to make these purchases in the future when prices are not discounted. This is often the case in supermarkets where individuals buy long shelf-life items. What this means for the firm is that sales may increase during the promotional period. However, what happens when the promotion ends? Oftentimes we believe in a post promotion drop in sales, as consumers delay the timing to their next purchase since they have more than enough to keep them satisfied now. Throughout our report we will be analyzing: how much do promotions change present sales for the Blue Mart supermarket in the United Arab Emirates and how does this affect post promotion sales?

Data Description

The dataset contains transaction-level sales data with product identifiers, quantities, unit prices, total transaction value, and timestamps, allowing aggregation at different time levels. It also includes promotion identifiers with start and end dates, so each transaction can be linked to specific marketing events. Sales channel information is recorded, enabling comparisons across online and offline channels and supporting both descriptive analysis and more advanced demand and promotion modelling.

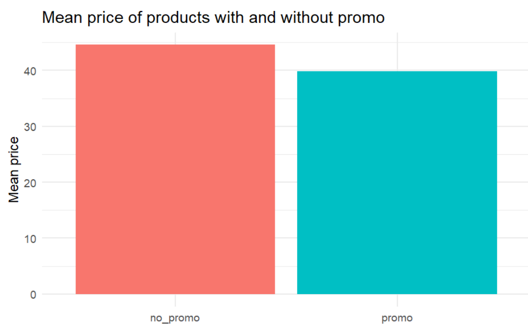
Each record includes: date of sale, store_id, sku_id, customer_id, quantity, unit_price, total_value, channel (e.g. in-store, app), discount_pct, a unique sale_row_id, and, where relevant, promo_id with its active period.

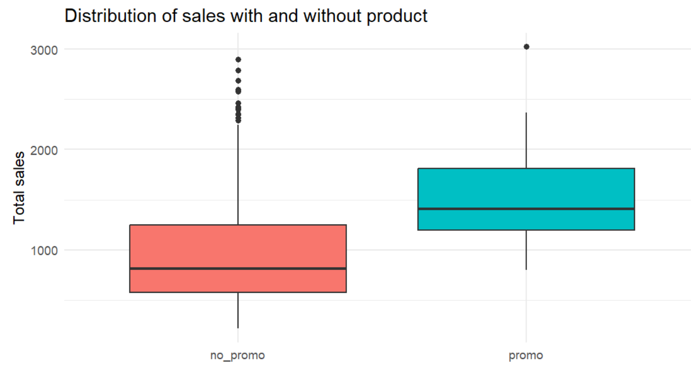
The dataset spans five years and covers around 200 products sold through Amazon.ae, the mobile app, Noon, physical stores, and the website. It comprises 641,843 observations: 274,802 from physical stores, 174,549 from the website, 111,264 from the app, 55,171 from Amazon.ae, and 26,057 from Noon. Of these, 204,368 transactions occur during major promotional events (e.g. Black Friday, Dubai Shopping Festival, Eid Al Adha, Ramadan, Summer Sale, UAE National Day), while 437,475 occur outside promotion periods.

Data Preparation

We construct two datasets to estimate the first and second regressions. The first tests whether sales volumes increase during promotional periods; the second examines how sales deviate from baseline in the days immediately before and after a promotion.

For the first dataset, we aggregate transactions at the daily level and compute: total units sold, number of transactions, number of distinct products, an indicator for whether any promotion is active, average unit price, and average discount. As expected, average prices are lower on promotion days, while sales volumes are higher.





For the second regression, we introduce time-window indicators around promotions. To avoid overlap between events and retain sufficient data, we classify each day into four phases: baseline (non-promotion days), pre-promotion (five days before a promotion), during-promotion, and post-promotion (five days after it ends). This structure allows us to track how sales evolve around promotions rather than only comparing promotion and non-promotion days.

Modeling & Estimation

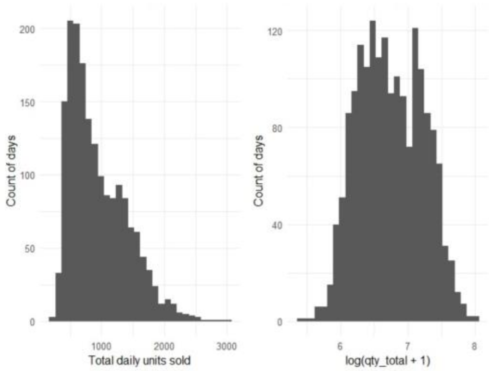
Regression 1 : To quantify the effect of voucher promotions on sales, we followed a step-by-step modeling strategy, gradually increasing the complexity of the regression specification based on both theory and diagnostics starting from:

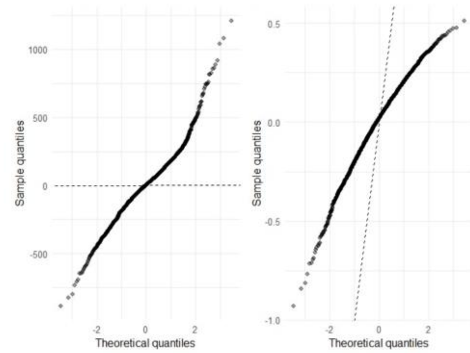
daily_sales = beta_0 + beta_1 promo_active + error,

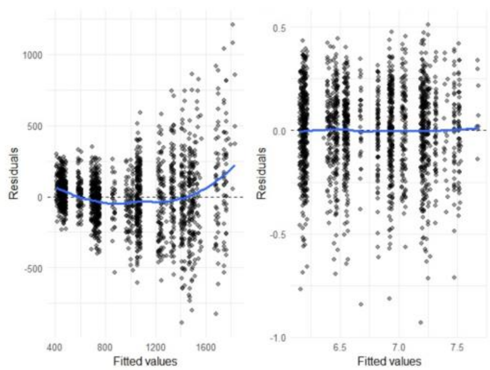
To account for seasonality, we extended the model by adding sets of dummy variables for day-of-week and month:

- beta_2 * dow + beta_3 * month

where “dow” is a vector of day-of-week indicators and month is a vector of month indicators (with one category omitted in each case as the reference group). When estimating these linear models using ordinary least squares (OLS), we observed two common issues with sales data such as highly right-skewed sales and heteroskedasticity.







1) After transforming the dependent variable to log(qty_total+1), the distribution becomes much more symmetric and closer to a bell shape. 2) The linear model exhibited large deviations from normality, especially in the upper tail, driven by extreme sales spikes. 3) Residual-versus-fitted plots confirm this improvement. In the linear model, residual variance increases with fitted values, indicating heteroskedasticity and violating OLS assumptions.

A log transformation is used because: 1) **Scale and variance stabilization:** Taking logs compresses large values more than small ones, which typically reduces heteroskedasticity and makes residuals more homoscedastic. 2) **Approximate percentage interpretation:** In a log-linear model, coefficients on dummy variables can be interpreted (after a simple transformation) as approximate percentage changes, which is intuitive in a marketing context. 3) **Diminishing marginal effects:** A log specification naturally captures the idea that the marginal effect of promotions may be smaller at very high sales levels than at low levels.

Because some days in our data have zero sales, we used log(qty_total + 1) as the dependent variable rather than log(qty_total), which would be undefined at zero.

log(qty_total + 1) = b_0 + b_1 promo_active + b_2 * dow + b_3 * month + error

We estimated this model using OLS. Compared with the earlier linear specifications, the log-linear model achieved the best overall fit, with **$R^2 \approx 0.81$**, and residuals that are much closer to being symmetric and homoscedastic on the log scale. In addition, the coefficient on promo_active in this model can be interpreted as the approximate percentage change in daily units sold when a voucher promotion is active, holding day-of-week and month fixed. In our case, this corresponds to an average **+134% increase in daily units sold** on promotion days relative to non-promotion days (i.e., sales more than double on promo days).

Regression 2

log(qty_total + 1) = b_0 + b_1 * phase + b_2 * dow + b_3 * month + error

The objective is to determine whether promotional activity induces systematic shifts in purchasing behavior such as early demand anticipation or a post-promotion drop, effects that cannot be captured when contrasting only promotion versus non-promotion days.

To do this, we had to reconstruct a refined dataset in which every day is assigned to one of four mutually exclusive phases (see Data Preparation) and iterate with different window-sizes to get the most reliable result.

- The time window must be long enough to detect meaningful pre/post deviations but short enough to avoid overlaps between consecutive promotions, which would contaminate phase classification and bias estimates.

The modeling pipeline remains consistent with the approach adopted earlier (see Regression 1).

We used log(qty_total + 1) as the dependent variable and estimate a log-linear specification where the phase indicators enter as the main explanatory variables, alongside controls for day-of-week and month to account for predictable seasonality patterns.

In sum, the dynamic-effects specification shows that voucher promotions generate a large temporary boost in sales without introducing distortions in the days surrounding the campaign. Promotional activity appears to generate genuine incremental volume concentrated within the promotion window, rather than merely shifting purchases across adjacent days.

Results & interpretation

Results

The two regression achieves an **$R^2 \approx 0.8$** in both cases, indicating strong explanatory power. The **during-promotion** phase is **highly significant**, with a large positive effect. **Pre-promotion** and **post-promotion** phases show **no statistically significant deviation** from baseline. Day-of-week and month controls behave as expected, with several coefficients strongly significant.

1st and 2nd regression results:

Term	Estimate	Std. Error	p value	Effect %
(Intercept)	6.6628	0.0282	0.00e+00	78171.1
promo_activePromo	0.8500	0.0190	3.64e-291	134.0
dowMon	-0.2833	0.0186	1.98e-49	-24.7
dowTue	-0.2804	0.0186	1.58e-48	-24.5
dowWed	-0.2711	0.0186	1.21e-45	-23.7
dowThu	-0.2573	0.0186	1.63e-41	-22.7
dowFri	-0.2591	0.0185	4.12e-42	-22.8
dowSat	0.0043	0.0186	8.15e-01	0.4
monthFeb	-0.0396	0.0243	1.02e-01	-3.9
monthMar	-0.1990	0.0263	6.67e-14	-18.0
monthApr	-0.1973	0.0253	1.05e-14	-17.9
monthMay	0.1631	0.0290	2.10e-08	17.7
monthJun	0.1575	0.0299	1.57e-07	17.1
monthJul	-0.3059	0.0237	1.50e-36	-26.4
monthAug	-0.3097	0.0237	2.24e-37	-26.6
monthSep	-0.2191	0.0305	1.06e-12	-19.7
monthOct	-0.1848	0.0304	1.44e-09	-16.9
monthNov	-0.1960	0.0307	2.14e-10	-17.8
monthDec	0.0140	0.0304	6.45e-01	1.4

Term	Estimate	Std. Error	p value	Effect %
(Intercept)	6.6597	0.0408	0.00e+00	77933.1
phaseduring	0.8495	0.0328	2.65e-113	133.8
phasepost	0.0157	0.0365	6.67e-01	1.6
dowMon	-0.2895	0.0255	3.23e-28	-25.1
dowTue	-0.2704	0.0257	1.13e-24	-23.7
dowWed	-0.2741	0.0255	1.65e-25	-24.0
dowThu	-0.2351	0.0255	1.95e-19	-20.9
dowFri	-0.2448	0.0256	8.35e-21	-21.7
dowSat	-0.0068	0.0256	7.90e-01	-0.7
monthFeb	-0.0401	0.0252	1.12e-01	-3.9
monthMar	-0.2021	0.0298	1.98e-11	-18.3
monthApr	-0.1813	0.0279	1.32e-10	-16.6
monthMay	0.1323	0.0451	3.45e-03	14.1
monthJun	0.1706	0.0422	5.67e-05	18.6
monthJul	-0.3060	0.0246	4.95e-33	-26.4
monthAug	-0.3099	0.0246	8.85e-34	-26.6
monthSep	-0.2194	0.0560	9.56e-05	-19.7
monthNov	-0.2118	0.0452	3.17e-06	-19.1
monthDec	-0.0250	0.0401	5.33e-01	-2.5

Interpretation

In our data analysis we concluded that sales have increased by **134%** for Blue Mart supermarkets, with no post promotion dip being visible. The reasoning for the massive increase in sales of more than double what is usually bought on regular days, is due to consumers buying more for current consumption and stockpiling for future consumption too (Hendel & Nevo, 2003). Knowing that items such as toilet paper, water and toothbrushes are for sale, it seems better to buy more than necessary now, to save money in the future. These items do not have a perishable date and so stockpiling seems to be cheaper in the long run. While an increase in sales seems like a logical solution, why is it that we do not observe a dip in sales after the promotional period. There are three reasons which can describe this behavior. First, our data is store-level, rather than consumer level. We can not see how individual households react to promotions, and instead we are capturing what happens throughout the store where a mix of consumer segments shop. If we were to control for household heterogeneity, we would be able to spot a small dip in sales (Hendel & Nevo, 2003), however in our analysis the effects are diluted because we are considering the general population and not specific households. Understanding there are differences in buying habits, such as some consumers stockpiling a lot, some don't stockpile at all and some don't even buy the products, can allow us to analyze in which segments stockpiling occurs, rather than having it obstructed by the mix of data currently presented. The second reason is because long-lasting effects of feature and display, may cloud the post-promotion dip if they are not taken into consideration (Hendel & Nevo, 2003). Retailers tend to extend non-price advertisements, such as store flyers, far beyond promotional periods to attract more purchases from consumers. As well, display tactics are used such as storing specific goods in front aisles to make the product easier to see and encourage impulse buying. Both of these techniques ease the post promotion dip as they encourage continuous purchases. Lastly, in a big supermarket such as Blue Mart, promotions are usually not store-wide but are placed on specific items. Blue Mart offers a wide range of products, such as groceries, electronics and health. While there may be a post promotion dip in one product, it will not be seen when taking into consideration store-wide data.

Validation & Robustness

Validation. The validation we performed is intentionally narrow: it demonstrates that the model predicts daily volumes reasonably well for the same assortment, calendar structure, and promotion mechanics observed in the historical data, but it does not certify performance under regime changes. We did not incorporate exogenous drivers such as discount depth, list-price updates, advertising, competitor actions, weather, stockouts, or calendar anomalies beyond basic day-of-week and month effects, so any future shift in these factors could degrade accuracy even if in-sample fit looks fine. We did not calibrate probabilistic forecasts or conformal intervals, so uncertainty bands remain model-based rather than coverage-guaranteed. Label error in promotion dates, silent re-definitions of what counts as "during," and covariate drift are untested failure modes. The scope therefore ends at like-for-like backtesting on the same data-generating environment; it does not extend to new categories, new stores, different cadence or intensity of promotions, long-horizon forecasting, or shock scenarios.

Robustness. Our robustness checks probe specification choices but stop short of full causal identification or market-wide interaction modeling. We treat promotion as exogenous and binary, without accounting for selection into being promoted, discount depth, or simultaneous price changes, which means the estimated effect may partially capture correlated managerial decisions. We do not model cross-product substitution, cannibalization, or channel/store heterogeneity, nor do we include hierarchical random effects or cluster standard errors by campaign, so correlated shocks within promotional spells could make standard errors too optimistic. Anticipation and stockpiling are only examined over short windows and longer carryover is not tested; moving holidays, school breaks, and weather shocks are not explicitly controlled; and measurement issues like stockouts or data censoring are not adjusted for. Under these omissions, things can go wrong in predictable ways: the “during” coefficient can be biased upward by endogenous targeting, diluted by unobserved supply constraints, or destabilized by structural breaks that our seasonal controls don’t capture.

Managerial Implications

Pricing

Understanding that sales increase by 134% during promotional periods, while no dip is observed afterwards, tells Blue Mart that their current pricing for their promotional strategy is effective. Knowing that promotions have a positive effect on short-term sales and no negative effect on long term sales, allows the management team to be confident that their pricing strategy is feasible and profitable. No changes are required on their end to enhance current and present sales, as continuous long-term sales shows they have obtained strong customer loyalty.

Targeting

In our data it was observed that there is no post promotion dip. As previously explained, this could result from heterogeneity of customers. Consumers behave differently, meaning that while some stockpile, others do not. It is the Blue Mart’s responsibility to acknowledge their wide customer base and account for the fact that their customers have different characteristics. As such, they must target to a wide audience, whether that be families, singles, or elderly individuals. Excluding one audience may lead to declines in sales and loyalty. In our data, it is seen there is no decrease in long term sales which means customers have high loyalty to this supermarket. This shows Blue Mart that they are targeting all consumer segments effectively.

Promotions/Advertising

Blue Mart focuses on having holiday sales such as for Ramadan and Eid, but they also account for seasonal promotions such as summer sales. Based on our results, promotions are increasing sales by more than double without harming long term sales. This means that current promotional techniques of having holiday/seasonal clearances should remain. There is no need to alter promotional behavior of the firm. As well, advertising has been a success at Blue Mart. As discussed, the reasoning for the post promotion dip being invisible may be due to strong display advertising long after the promotional period ends.

Limitations and future work

One main limitation is that this dataset considers buying behavior of individuals in the UAE. Trends of consumer buying patterns in the UAE can't necessarily be used as a baseline for companies globally, as different cultures have diverse purchasing behavior. What consumers look for in a good is not the same in the UAE as it is in Italy for example. In the UAE it was noted that product quality is more important than lower prices (Alnahhal et al., 2024). As such, consumers tend to be more loyal to a brand, influencing the inexistence of a post promotion dip. While this analysis can be applied to supermarkets in the UAE, it can not be applied to those in Italy. For the future, it would be crucial to gain an understanding of household level data, rather than just store-level. By obtaining household level data, we could segment customers by family size or income, and then examine stockpiling behavior for these specific segments. Currently, the behavior of all consumers is averaged into one number, flattening the potential dips we would see after a promotion. By examining each household, or segments of households, we can understand which consumers stockpile and create a dip in sales after the promotion ends, and then create tactics of overcoming these effects.

References

Alnahhal, M., Aldhuhoori, E., Al-Omari, M., & Mosab Tabash. (2024). *The impact of pricing on consumer buying behavior in the UAE*. ResearchGate. https://www.researchgate.net/publication/377663750_The_impact_of_pricing_on_consumer_buying_behavior_in_the_UAE

Graciola , A. P., De Toni, D., de Lima, V. Z., & Milan, G. S. (2018). *Does price sensitivity and price level influence store price image and repurchase intention in retail markets?* ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0969698917305775

Hendel, I., & Nevo, A. (2003). *The Post-Promotion Dip Puzzle: What do the data have to say? - quantitative marketing and Economics*. SpringerLink. https://link.springer.com/article/10.1023/B:QMEC.0000004844.32036.5a