Measurement When Data Isn't Clean

Written by
Rhonda Carraway Petty
Marketing Insights Data Scientist
rpetty@mathnetix.com

The impact of a marketing campaign is measured from quantifiable metrics that are compared before and after a promotion launches. Many times, however, unforeseen circumstances influence these metrics making it difficult to isolate the true impact of a promotion.

This paper is a case study of sorts. It presents three real-life examples where post-promotion data was 'imperfect' because of an external happening that was not anticipated. For each example, the situational backdrop is presented. Then, the method used to perform hygiene on the data time-series prior to measuring performance is illustrated.

Because these examples are real-life case studies, figures have been altered in order to share the story without disclosing protected information. The text is written in the first person to emphasize the subjective nature of the hygiene rules. The reader should use these illustrations as a guide for dealing with similar scenarios, recognizing that campaign measurement requires a great attention to detail.

CASE STUDY #1: Widespread Winter Storms

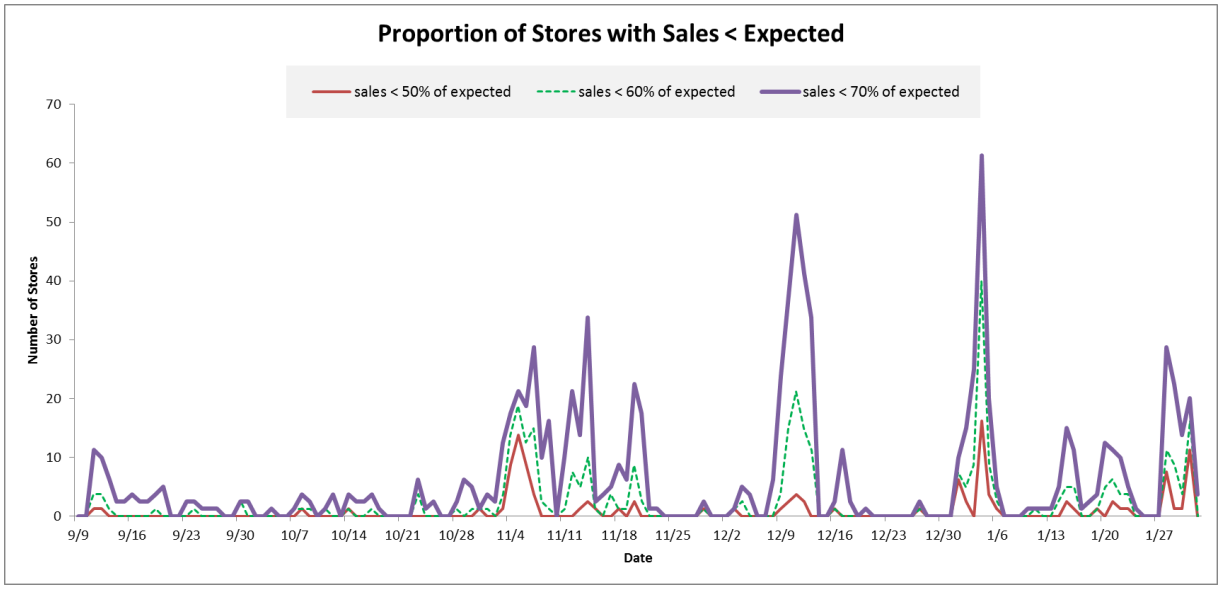
Situational Backdrop: A retailer decided to pilot a new marketing program in a few test markets before a national roll-out. At the conclusion of the pilot test period, a deep-dive analysis of the program's impact on sales and several other performance metrics was conducted. Impact was reported on an overall basis---over all markets in the pilot---and at a very detailed level to include a break-down by market, store and customer segment.

The Issue: After carefully matching each Test store with a lookalike Control store, historical sales and other metrics were summarized by store date across the analysis time period. These summaries fed impact calculations which, in turn, yielded impact reports for executive management. But, in the process of studying results, several Test stores in the Northeast had counter-intuitive figures. Instead of these Test stores reporting a positive impact on sales, negative results were calculated. This suggested the Control stores witnessed more favorable results than the Test stores exposed to the marketing program. No report would be delivered to senior management before a microscopic review of the underlying data patterns. These counter-intuitive results had to be explained.

The Depiction: Because the counter-intuitive results were concentrated among stores in the Northeast, the recent plethora of winter snow storms was the suspected culprit. To prove this suspicion in a way lending an easy visual, I counted the number of stores with actual sales below expected sales, plotting the results across time. Expected sales were estimated by looking at the store's average sales on the same weekday in the same time period, but omitting known holiday dates.

The red line of Figure 1 depicts the proportion of stores across time that witnessed actual sales less than half of the expected sales level. The blue and purple lines follow suit but with a different tolerance. These different lines parallel each other. They peak on dates when sales are far below average for a relatively large proportion of stores---and these unusual dates do indeed align with the timing of known winter storms.

Figure 1: Proportion of Stores with Actual Sales Less than Expected Sales



The Resolution: The measurement dilemma caused by the localized winter storms would have been less of an issue had each Test store been matched with a lookalike Control store from the same geographic market—both impacted by the storm. This, however, was not an option. In order to prevent confusion in the Test market, Control stores were sampled from other geographic areas----aligned with their Test counterparts on pre-period behavior, type of store and age of store---but not aligned on geography. So, when sales in the Test stores were diminished because of the bad weather---sales in the Control stores were not---making the marketing program look unfavorable.

In order to correct for the weather anomaly, a data hygiene rule was devised. This rule (actually, two rules) made assumptions about *how* the weather impacted sales---either displacing sales altogether or shifting sales to another date.

Rule #1. If a store-date combination experienced actual sales less than half of the expected sales, it was presumed impacted by a large weather system. Under this scenario, we assumed most sales on that date were lost and would not be recovered. To bring sales more in line with a 'without storm' scenario, sales were artificially increased but only to 50% of the business-as-usual expectation. In this way, sales were increased while remaining conservative with the adjustment.

If actual sales < 50% of expected, then sales default to 50% of expected

Rule #2. If a store-date combination experienced decreased sales, but not to an extreme, it was presumed impacted by a smaller scale weather system. Under this scenario, we assumed most sales on that date were eventually recovered. To bring sales more in line with a 'without storm' scenario, sales were artificially increased to 70% of the business-as-usual expectation.

If actual sales < 70% of expected, then sales default to 70% of expected

After implementing these hygiene rules---which adjusted less than 2% of the store-date combinations—impact was recalculated. This time, the impact of the program in the Northeastern test markets was much more favorable, in line with intuitive expectation, and similar in magnitude to the results of the other geographic test markets.

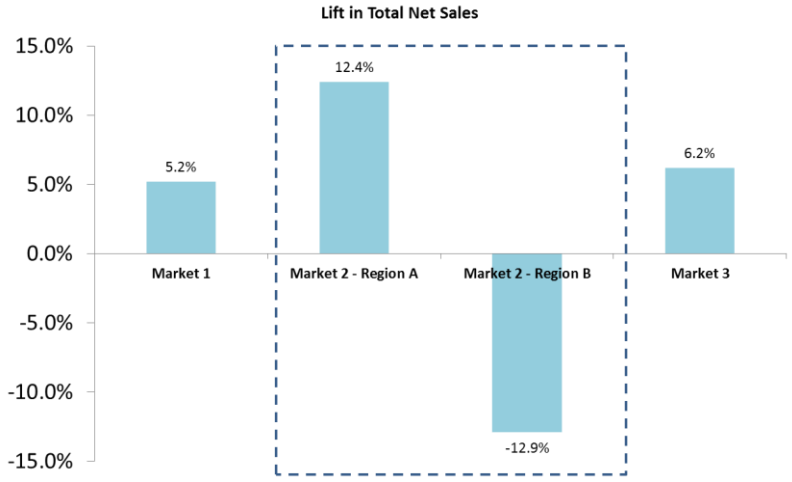
CASE STUDY #2: New Store Cannibalization

Situational Backdrop: A national retailer launched a market-wide test to study the impact of a tiered promotional offer by customer segment. The promotion was advertised in select Test stores and available to all customers who shopped those stores during the promotional time period. I was tasked with reporting the impact of the promotion by customer segment within each store of the Test market.

The Issue: After the customary data-preparation tasks, I calculated the incremental impact on sales and noticed that one of the three Test markets exhibited a negative impact on total sales. I was able to narrow the negative performance to a handful of stores in a smaller geographic region of the larger Test market. There had to be something unusual going on in this smaller geographic subset of the Test market.

The Depiction: I brought this discovery to the attention of middle management as illustrated in Figure 2. Region A of Market 2 depicts a positive impact in sales while Region B of Market 2 depicts a negative impact in sales. What would cause Region B to be negatively impacted by the marketing promotion? Could cannibalization of sales by a new store in Region B be behind the mischief?

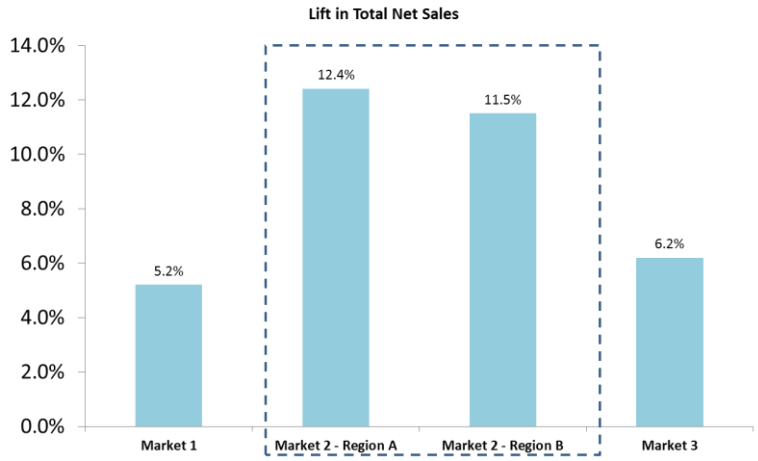
Figure 2: Percentage Lift in Total Sales



The Resolution: Because the focal point of the analysis was a deep-dive review of impact by individual Test store, sales were credited to the store where the transaction took place. So, if a customer transacted at the new store not part of the Test, that customer responded although her sales were not credited to the promotion---making the impact of the promotion in the market with the new store look unfavorable.

Yes, cannibalization was impacting the figures. When existing customers shopped at the new store, their sales were credited to the new store instead of to their primary store – a store in the Test group. To resolve the issue, sales were re-summarized by customer and credited to their primary store instead of to the transacted store. This change brought results in line with expectation as illustrated in Figure 3. Now, both regions exhibited a positive impact.

Figure 3: Percentage Lift in Total Sales after Correcting for Cannibalization



CASE STUDY #3: Insufficient Control Markets Compounded by Winter Storm

Situational Backdrop: A large national retailer with multiple locations in every major US market launched a short-lived promotion across all markets except for two. The goal of the promotion was to drive incremental revenue via more visits and higher spend-per-visit.

The Issue: After the promotion was complete, the retailer wanted to measure the impact of the offer on sales. But, since the promotion was active everywhere except for two markets, there was no representative Control group. Also, a large band of snow hit in the middle of the promotional period which compounded the retailer's difficulty in producing a reliable estimate of the promotion's impact. In other words, both terms on the impact equation were tainted:

Impact = Actual Sales – Expected Sales

- **Expected Sales.** The Control Group was not representative of the Test---thus, it could not be used to estimate a 'no promotion' scenario.

- **Actual Sales.** Actual sales attributable to the promotion was not known because the winter storm impacted behavior.

The Resolution: The resolution addressed both parts of the impact formula:

Expected Sales. Given there was not a sufficient Control Group to use as a proxy for what would have happened without the promotion, I used traditional forecasting methods to estimate what would have happened 'without the promotion and without the snow storm' for each Test Market :

Expected Sales = Expected Transactions * Expected Sales/Transaction

- **Expected Transactions** was estimated by taking a weighted average of a Time-Series forecast and a Regression forecast fed with historical data before the promotional time-period. The time-series extrapolated historical trends to future dates. Regression estimated future transactions as a function of other known variables. The historical accuracy of each of these independent methods was used to establish weights for combining their respective forecasts into a single forecast.

- **Expected Sales-per-Transaction** was set equal to actual productivity before the promotion started.

Actual Sales. Because the winter storm impacted observed sales during the promotional period, its impact had to be removed in order to estimate what actual sales would have been had the storm not occurred. The impact attributable to the winter storm was estimated by observing historical trends before the promotional period for warm markets and cold markets as two mutually-exclusive groups. Trend in the warm markets was used like a Control group---used to estimate what sales would have been in the cold markets had the storm not occurred.

These adjustments, in tandem, produced an estimate of the promotion's impact that was intuitively believable by senior management---and aligned with internal estimates produced by the company's financial analyst. The measurement difficulty also highlighted the need for creating a Control group at the outset of each marketing experiment, which the company adopted as routine protocol.

Rhonda Carraway Petty
rpetty@mathnetix.com
601-384-2488