Analytics & Advertising: What to do and what not to do



**GROUP 6**

10/30/2019

BAX 401 Business Analytics

# Zunfeng Huang

# Mitesh Jain

# John Elmar Loretizo

# Mayank Mani

# Shalini Mishra

Prakriti Rastogi

**Introduction**

Marketing has always been about connecting to customers at the right time in the right place.

Marketing mix refers to a set of tactics/strategies that a company uses for a successful product offering. When properly understood and applied, it’s a key factor driving a product's success.

A cosmetic firm launched a product 4 years ago and has been utilizing multiple media for advertising. For Offline Advertising, catalogs and mailings were used. Mailings included flyers, postcards, and letters. For effective targeting, different catalogs were sent to different customer segments. Existing customers were sent catalogs different to customers who had not made any purchase in the previous 6 months. Whereas entirely different catalogs were sent to potential new customers. Their online marketing activities included Banner ads, Search ads, Social Media ads and Newsletter ads. The firm has also invested in “Retargeting ads” and ad portal advertising. As a pixel-based technology, retargeting ads anonymously follow the audience throughout their online journey and later targets them with personalized product offers. The firm wants to measure the efficacy of advertising activities on product sales.

**Problem Formulation -- Shalini**

**Uploaded as a separate file, equations getting distorted in google docs**

**Data Description -- Mayank**

We have to carefully select variables from our dataset to avoid a Garbage In Garbage Out model. We removed certain variables from our dataset based on the following criteria :

* Variables like Catalogs\_WinBack, Catalogs\_NewCust, Mailings, Banner, Social Media & Retargeting can’t be relied on for modeling as they have >40% data missing
* Overall spend metrics like ADV\_Total, Adv\_offline and Adv\_online were removed as they are a linear combination of other variables included in the data. Incorporating them into the model would result in multicollinearity.

Note, we have duplicates in Search and Portals for 5 and 10 months respectively. But we still have data points >= (2n+20=32)with reliable data.

Thus, we are working with remaining 4 variables:

* Catalog\_ExistCust : The amount spent on sending the catalogs to the existing customers.
* Search: The amount spent on search Ads.
* Portals: In online advertising, consumers can find what they need at various other locations online via an ad portal
* Newsletter: Amount spent on newsletter ads in a month. The variable seems fairly clean data with odd duplicates which are less than three.
* Lag\_Sales : This variable takes into account the effect of sales with a lag of 1 month. So, the sales of last month will have an effect on this month’s sales.

Descriptive statistics of the dependent as well as the independent variables:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sales | CatExistCust | Search | Newsletter | Portals | Sales\_lag |
| Min. :3355 | Min. : 0.0 | Min. : 38.17 | Min. : 7.057 | Min. :2.544 | Min. :3355 |
| 1st Qu.:4406 | 1st Qu.: 328.7 | 1st Qu.: 45.38 | 1st Qu.:16.691 | 1st Qu.:3.393 | 1st Qu.:4169 |
| Median :4690 | Median : 598.0 | Median : 66.11 | Median :19.779 | Median: 4.707 | Median :4631 |
| Mean :4809 | Mean : 567.6 | Mean : 69.83 | Mean :20.734 | Mean :5.246 | Mean :4716 |
| 3rd Qu.:5195 | 3rd Qu.: 625.6 | 3rd Qu.: 88.19 | 3rd Qu.:25.139 | 3rd Qu.:6.867 | 3rd Qu.:5165 |
| Max. :6976 | Max. :1298.7 | Max. :134.87 | Max. :53.609 | Max. :9.303 | Max. :6467 |

**Model Development -- John H (Finished, docs uploaded to the google drive)**

**Points --**

Focal Model: Sales (Y)= lag\_sales + Diminishing Returns( of all the variables we decided to go for)

**Which variables we are taking?**

**Eq of our focal model**

**Explain the sign used for each variable in the eq?**

**variable definition, effectiveness, significance of each variable**

**Intercept being too high: two possibilities we can think of:**

1. **We are not taking into account the fixed component of advertising spend**

**2.Due to lack of independent variables, we are not able to explain the sales contributed by different advertising channels. There might be other channels which are more effective but since their data is not reliable, we can’t use them in our model.**

**5. Results -John E and Prakriti Rastogi**

**Finalize on one model**

**Explain the summary stats of that model**

**Why does it make sense given what we have?**

**Compute elasticity**

In order to establish the relationship between various advertising channels and the sales for each month, a linear regression model was created. In total, there are 5 variables in the final model. These are the (1) lag of sales, the carryover effect of previous month’s sales, (2) search advertising medium, (3) shopping catalog for existing customers. The last two variables are derived from variables (2), and (3) which was transformed by taking the square root of their values. The transformation was done because advertising exhibits a non-constant and decreasing marginal return on sales (diminishing returns [4]).

[4] - The law of diminishing returns states that the relationship between marginal input and marginal output will be more than proportional at the initial stage, but after a certain point, with all other inputs held constant, the marginal product of each unit of input will drop as the amount of the input increases (Heibroner and Thurow 1981; Picard 1989; Samuelson and Nordhaus 1989).

The relationship can be defined by the equation below:

(we will put this equation in word)

After analyzing the related coefficients of variables used in the final model, the square root transformed search variable yielded the highest coefficient at 1223.68 which meant that for every unit increase in budget for search increases the mean of sales by 1223.68 units. This is followed by catalogs for existing customers which increases the mean of sales by 4.76 for every unit increase in its budget. However, from the same model, we can see that search, square root transformed catalog for existing customer and the lag of sales have a negative effect on the mean of sales.

Moreover, analyzing the elasticity of the variables (1) search = -0.99, (2) catalogs for existing customers = 0.08, (3) square root transformed of search = 20.99, (4) square root transformed of catalogs for existing customers = -3.32

After analyzing the related coefficients of variables used in the final model, square root transformed of search budget yielded the highest effect in terms of elasticity and \_\_\_ is

The negative coefficient of diminishing returns implies that every additional dollar spent on such channels would result in an additional return of less than a dollar.

**Executive Summary**

Marketing has always been about connecting consumers to products. It’s centered on the idea of making them aware, and enticing them to eventually buy products. This report quantifies the effects of marketing mediums in the sales of a cosmetic firm. In the last four years, the cosmetic company has been using multiple platforms to advertise their product. The advertising mediums can be generally divided into two major groups - online and offline. Basic exploration of data revealed that only four advertising mediums have been deemed of quality in terms of inclusion in the model. From these the following results have been discovered:

(1)

(2)

(3)

In addition, the effects of diminishing returns of these advertising mediums are also assessed. A final recommended model was developed with the following key points:

(1)

(2)

(3)

From this, we developed allocation strategies for the company based on elasticities of the variables in the recommended model. A reduction in the search advertising budget and an increase in the catalog advertising for existing customers is highly recommended. Moreover, in the course of the development of the model, the team highlights the importance of using high quality data in order to truly assess the effects of the advertising mediums to sales.

**Recommendation ---Shalini & one more person**

**Conclusion**

Implementing infrastructure that captures advertising data from different platforms is only the first step towards developing Advertising Analytics 2.0. Knowing what data to capture and how to use it to the firm’s advantage is a completely different feat at which many have failed. Sales of a company doesn’t depend only on their advertising channels, they also hinge upon the company’s goodwill in the market and their price points, none of which was captured in this case.

**References-**

**Extensions**

-What we discovered:

* Created more than 1000+ models to try all possible combinations of all valid variables. This includes both their log transform and square root transform variables along with diminishing returns of the variables.
* The variables cannot predict sales accurately.
* Both log and square root transformation did not yield a good model with adjusted r squared values falling between 0.5 - 0.6 This explains less than 50% of the standard deviation of the data.
* We need to adjust for seasonality that is more than the 1 period lag indicated.

In order to gauge the robustness of the chosen model above, more than 1000+ other models have been generated to check on the best possible combination of advertising mediums that best explains sales. Few of the most notable iterations include natural log transformation of the variables as well as checking for synergistic effects. The following are some general insights about the model:

* The variables are not able to explain at least 80% of the variation in the sales values.
* In addition, none of the models yielded a very low AIC value as compared to others to stand out as a good fit.

***Log Transformation for Effects***

When checking for variables and their log transformation, the model that considers the lag of sales, search, catalogs for existing customer, log transformation of portals, search, and catalogs for existing customer. The total r-squared for this model