Effectiveness Analysis of Advertising Activities

Information, Insight, and Impact

OCTOBER 23, 2021

Table of Contents

Executive Summary:

This report deals with the relationship between advertising and sales performance of products launched by the cosmetics firm four years ago. It is imperative to assess the advertising channels that significantly contribute to sales. The goal is to grasp the metrics that best describe the relationship by finding the model that best fits the data. Variables are selected to obtain a best focal model using square root (sqrt) functions explaining 31% of the variation. A comparison is also carried out between the focal model with intercept and without intercept and it is found that with intercept is a precise representation of diminishing returns. The four variables: Lagged sales, Catalogs Winback, Catalogs NewCust and Portals have strong implications on sales as per the focal model. Portals have the highest elasticity amongst the rest in the model. The firm can accordingly allocate a budget for the advertising spends to augment sales and subsequently maximize profit. The firm should allocate more resources to the portals and lesser to Catalogs_Winback and Catalogs_NewCust based on the elasticities. Log models are run as an extension to understand if the log transformations are better explaining the variation in sales.

Introduction:

In this report we will determine the relationship between advertising and sales performance of products using multiple regression. Advertisers track the relationship between advertising and sales performance in a variety of ways. The dataset encompasses two types of advertisement: Online and Offline. Online advertising consists of digital and internet-based channels like social media, banner ads etc. Offline advertising consists of real-world channels and is also known as the "traditional advertising". The model we have developed will help identify the advertising channels that drive the sales over time using historical sales data. Firms can plan the budgets on promotions and advertisement channels that bring the greatest return on investment.

Problem Formulation:

Any sort of compensated promotion of a sponsor's ideas, products, or services constitutes advertising. To evaluate the effectiveness of the marketing mix activity-advertising spending on

sales, regression models are developed with the most relevant factors to suit the data. As advertising undergoes diminishing returns effect, square root function is applied on the model to capture that. Besides this, log functions are also implemented on the model to compute significant differences in effectiveness of the parameter. Backward Stepwise Regression is implemented to find a reduced model that can best explain the data. We gauge the performance of our regression model using Adjusted R^2 and AIC.(*An Introduction to the Akaike Information Criterion*, 2020)

Data Description:

There are a total of *15* variables in the dataset. Sales are used as the response variable.

	Sales (units)	Catalogs_ExistCust	Catalogs_Winback	Catalogs_NewCust	Mailings	Search	Newsletter	Retargeting	Portals
Min.	3,355	0	0	0	0	38.17	7.06	0.00	2.54
1st Qu.	4,406	328.7	0	0	0	45.38	16.69	0.00	3.39
Median	4,690	598	0	43.63	0	66.11	19.78	0.00	4.71
Mean	4,809	567.6	83.42	272.87	11.42	69.83	20.73	10.85	5.25
3rd Qu.	5,195	625.6	174.15	487.42	19.24	88.19	25.14	18.56	6.87
Max.	6,976	1298.7	438.54	1131.57	84.47	134.87	53.61	49.30	9.30

Figure 1: Descriptive statistics of the variables of interest that are considered along with sales for modelling.

Catalogs_ExistCust, Catalogs_Winback, Catalogs_NewCust, Mailings, Search, Newsletter, Retargeting, Portals are taken as the predictor variables. After analyzing data, we omitted the ADV_Total, ADV_Online, and ADV_Offline variables as they are aggregations of total, online and offline metrics respectively. This is done to eliminate the possibility of multicollinearity between the predictor variables. Multicollinearity can lead to overfitting of the model, creeping in of insignificant variables, redundancy, and biased R^2(Wu, 2021). Banner and social media are also removed as they have the *100%* and *91%* of nulls. There is a strong correlation between Search, Retargeting, and Portals [Search – Retargeting (*83%*), Search – Portals (*88%*), Retargeting – Portals (*70%*)], which can lead to multicollinearity, however we went ahead with them in developing models. Since, AIC would discount them and would lead us to find the best fit. Zeros in selected predictor variables are assumed as the zero-money spent on the respective channels.

Model Development:

We used multiple regression to identify the best fit model for different advertising channels. In multiple regression, we are opting for the dynamic model as the current sales might depend on past sales. However, we restricted the model to only one period of lagged sales.

As a first step, post choosing variables of interest we applied square root functional transformation on them to transform the nonlinear relationships observed to linear relationships. We then included all the selected variables (Catalogs_ExistCust, Catalogs_Winback, Catalogs_NewCust, Mailings, Search, Newsletter, Retargeting, Portals) along with the lagged sales variable to develop our initial model. Post this we followed a stepwise elimination methodology and gradually eliminated and observed the variable's effect on the model by understanding the AIC values.

Step 1: Model with all selected variables + lagged sales and intercept. The AIC is at *577.39*.

Step 2: We removed 'Mailing' which was not significant, and AIC reduced from *577.39 to 555.46.*

Step 3: We removed 'Search', the next least significant variable: AIC reduced further *to 553.97*.

Step 4: We removed 'Retargeting', next least significant variable: AIC reduced *to 552.54*.

Step 5: We removed 'Newsletter', next least significant variable: AIC reduced *to 552.48*.

Step 6: Finally, we removed 'Catalogs_Winback', the next least significant arriving at AIC of *552.2.*

Post the above process we checked the AIC by further removing and adding different combinations of variables. However, we weren't able to reduce the AIC much further than *552.2* when using square root functional form.The below is the optimized focal model with least AIC after removing Catalogs_ExistCust, Mailings, Search, Newsletter and Retargeting variables.

$$Sales = \mathbf{2019.33} + \mathbf{0.2169} * \sqrt{\boldsymbol{stm1(lagged\ sales)}} + \mathbf{50.63} * \sqrt{\mathbf{Catalogs\_Winback}} - \mathbf{24.89} * \sqrt{\mathbf{Catalogs\_NewCust}} + \mathbf{788.73} * \sqrt{\mathbf{Portals}}$$

We also developed models without an intercept and lagged sales, but the above model stood out with the least AIC. In appendix section, we will explore different functional forms and synergies.

Results:

Sqrt with Intercept	Coefficients	Std. Error	t-value	p-value	Elasticity
(Intercept)	2019.33	639.84	3.16	0.00	
Lagged variable(stm1)	0.22	0.11	2.02	0.05	
Catalogs_Winback	50.63	22.11	2.29	0.03	0.06
Catalogs_NewCust	-24.89	13.24	-1.88	0.07	-0.05
Portals	788.73	275.94	2.86	0.01	0.24
AIC	552.2	**Adjusted R2**	30.96%		

Figure 2: Statistics for the focal model

Above are the outputs we obtained from the focal model. We interpreted from the output that the model explains approximately *~31%* of the variation in the sales. And, only *3* variables Catalogs_Winback,Catalogs_NewCust, Portals along with the lagged sales variable contributed to the variation with reasonable significance at *10%* significance level. We deduce from the elasticities that only advertising spend on Portals is having a significant impact. For every *1%* increase in 'Portals' spend we interpret that sales increase by *0.24%*. The other two significant variables (Catalogs_Winback and Catalogs_NewCust) elasticities are 0.06% and -0.05% respectively. Catalogs_Winback is having very little impact and Catalogs_NewCust spend is indeed presenting a negative relationship with sales (for every *1*% increase in Catalogs_NewCust spend we interpret that sales decrease by *0.05%*). However, we still need to examine the negative coefficient for Catalogs_NewCust in detail. Since, the negative coefficient might be arising due to a variety of reasons like insufficient data, multicollinearity etc. However, based on the focal model, we suggest that the firm should allocate more resources to the portals and spend comparatively less resources to Catalogs_Winback and no resources to Catalogs_NewCust.

Recommendations and Managerial Implications:

The focal model which has been created to estimate the effectiveness of marketing mix suggests two offline advertisement variables 'Catalogs_Winback', 'Catalogs_NewCust' and one online advertisement variable 'Portals' have significant impact on the sales. However, further exploring log transformation models and models with synergies. We developed a better model that explains close to ~70% variation in the sales when compared to *~30.96%* by the focal model. The square

root model with synergies and intercept, which consists individual variables 'Catalogs_ExistCust', 'Catalogs Winback', 'Catalogs NewCust', 'Search', 'Newsletter', 'Retargeting', 'Portals' and various synergies[1] , is the best and recommended model to management because it has least AIC among square root models and higher Adj. R^2 (*~70%*) when compared with other functional form transformed models. Hence, it better explains the variation in sales when compared to the focal model. The management should adopt to the new best fit model and focus on the synergies and new variables identified in the model to improve their advertising spend.

Conclusion:

In this report we examine the relationship between the advertising spending and the sales of the cosmetics firm. The methodological complexity inherent in analyzing the advertising spending-sales relationship is overcome in stages. The initial focal model with the square root function considered with intercept results in an AIC of *552.2* which is comparatively less and hence chosen. The proportional variance measure estimated through Adjusted R^2 is *30.96%*. We achieved the best regression fit through this. Further, the sqrt model with synergies and intercept fared relatively better than the focal model as the variance is well explained because the operating variables and synergies influence has been accounted for, in this scenario unlike in the focal model. Since, the sales response to advertising is complex and dynamic, regression estimates can be inconclusive as there are limitations about functions considered which provide results that may vary over each run. There are many possible extended analyses like trying to resolve multicollinearity, use different functional forms that fits better for each individual variable, identifying and developing synergies to test using domain knowledge etc. These would help in further refinement of the model and explain more variation among the sales. We are not very sure if the analysis is causal as there might be many confounding variables present which are affecting the sales. Having more data and domain knowledge would help in determining the causality.

[1] Refer to 5th extension model in appendix

Bibliography:

An introduction to the Akaike information criterion. (2020, March 26). Scribbr. https://www.scribbr.com/statistics/akaike-information-criterion/

Wu, S. (2021, June 5). *Multi-Collinearity in Regression*. Medium. https://towardsdatascience.com/multi-collinearity-in-regression-fe7a2c1467ea

Appendix:

Additional models:

Adding to the focal model, models with natural log and synergy are tested. Indeed, the natural log model with intercept and synergy and square root model with intercept and synergy explains the variation in sales better than the focal model when their Adj R^2 and AIC are compared.

There exists a synergy between and within offline and online advertisements. The model which should be used by management for allocation of advertising spend is the sqrt model with intercept and synergy. It not only explains better variation of sales but also considers various synergies which might happen in real scenarios. However, in this model we have many synergies out of which few might be just noise. Few individual variables and synergies are not significant. However, when considered as a model they are significant. Hence, it is better to refine the model with more domain understanding and larger dataset. But, for now with the current data 'Catalogs_Winback','Retargeting' and 'Portals' individually and 'Mailings', 'Retargeting', 'Search', 'Newsletter' when in synergy with others are explaining variation in sales.

Models	Best Model	AIC	Adjusted R2	Comments
1.Sqrt with Intercept	Sales = 2019.33+0.2169*sqrt(stm1(lagged sales))+50.63*sqrt((Catalogs_Winback))-24.8909*sqrt((Catalogs_NewCust))+788.728*sqrt((Portals))	552.2	30.96%	Chosen as a focal model as it is parsimonious and has less AIC
2.Sqrt Model without Intercept	Sales = 0.1772*sqrt(stm1(lagged sales))+51.51*sqrt(Catalogs_Winback)-24.99*sqrt(Catalogs_NewCust)+278.546*sqrt(Search)-169.12*sqrt(Retargeting)+896.504*sqrt(Portals)	553.59	98.05%	More AIC compared to focal model
3.Ln Model with Intercept	Sales = 2560.15-210.76*ln(Catalogs_ExistCust)+206.82*ln((Catalogs_Winback))-135.74*ln((Catalogs_NewCust))+428.84*ln((Newsletter))+1234.64*ln((Portals))	539.76	49.68%	
4.Ln Model without Intercept	Sales =-197.05*ln(Catalogs_ExistCust) +215.64*ln((Catalogs_Winback))-142.86*ln((Catalogs_NewCust))+809.45*ln((Search))+362.98*ln((Newsletter))-126.23*ln((Retargeting))+925.36*ln((Portals))	539.68	98.62%	
5.Sqrt Model with Intercept and Synergy	Sales = 6.83e+04- 8.75e+02 *sqrt(Catalogs_ExistCust)+ 1.40e+03 *sqrt(Catalogs_Winback)-8.04e+02 *sqrt(Catalogs_NewCust)- 1.90e+03 *sqrt(Mailings)- 1.34e+04 *sqrt(Search)-6.53e+03 *sqrt(Newsletter)+ 5.11e+03 *sqrt(Retargeting)+ 9.00e+03 *sqrt(Portals)- 1.92e+02 *sqrt(`Catalogs_ExistCust:Catalogs_Winback`)+ 1.36e+02 *sqrt(`Catalogs_ExistCust:Catalogs_NewCust`)+ 2.61e+01 *sqrt(`Catalogs_ExistCust:Mailings`)+ 4.84e+02 *sqrt(`Catalogs_ExistCust:Search`)- 1.05e+02 *sqrt(`Catalogs_ExistCust:Newsletter`)- 3.18e+01 *sqrt(`Catalogs_ExistCust:Retargeting`)- 1.18e+03 *sqrt(`Catalogs_ExistCust:Portals`)- 1.74e+01 *sqrt(`Catalogs_Winback:Catalogs_NewCust`)- 3.59e+02 *sqrt(`Catalogs_Winback:Mailings`)- 9.00e+01 *sqrt(`Catalogs_Winback:Search`)+ 5.64e+02 *sqrt(`Catalogs_Winback:Newsletter`)+ 9.82e+01 *sqrt(`Catalogs_Winback:Retargeting`)+ 8.76e+02 *sqrt(`Catalogs_Winback:Portals`)+ 2.70e+02 *sqrt(`Catalogs_NewCust:Mailings`)- 2.45e+02 *sqrt(`Catalogs_NewCust:Newsletter`)- 4.32e+01 *sqrt(`Catalogs_NewCust:Retargeting`)- 6.88e+02 *sqrt(`Catalogs_NewCust:Portals`)- 4.44e+02 *sqrt(`Mailings:Search`)+ 5.95e+02 *sqrt(`Mailings:Newsletter`)+ 2.45e+02 *sqrt(`Mailings:Retargeting`)+ 7.95e+02 *sqrt(`Mailings:Portals`)+ 2.83e+02 *sqrt(`Search:Newsletter`)+ 4.41e+02 *sqrt(`Search:Portals`)- 1.12e+03 *sqrt(`Newsletter:Retargeting`)+ 3.92e+03 *sqrt(`Newsletter:Portals`)- 1.78e-01 *sqrt(Stm1)	507.75	69.65%	The best model with the least AIC among square root models and has a higher Adjusted R square compared to best LN model

Figure 3: Extended models summary

6.Sqrt Model without Intercept and Synergy	Sales = - 123.4 *sqrt(Catalogs_ExistCust+ 244.6 *sqrt(Catalogs_Winback)- 340.7 *sqrt(Catalogs_NewCust)- 1105.1 *sqrt(Newsletter)+ 7375.9 *sqrt(Portals)- 46.6 *sqrt(`Catalogs_ExistCust:Catalogs_Winback`)+ 38.3 *sqrt(`Catalogs_ExistCust:Catalogs_NewCust`)+ 37.1 *sqrt(`Catalogs_ExistCust:Search`)- 105.1 *sqrt(`Catalogs_ExistCust:Portals`)- 89.1 *sqrt(`Catalogs_Winback:Mailings`)+ 133.7 *sqrt(`Catalogs_Winback:Newsletter`)+ 16.7 *sqrt(`Catalogs_Winback:Retargeting`)+ 157.3 *sqrt(`Catalogs_Winback:Portals`)+ 72.7 *sqrt(`Catalogs_NewCust:Mailings`)- 43.3 *sqrt(`Catalogs_NewCust:Search`)- 62.4 *sqrt(`Catalogs_NewCust:Newsletter`)- 12.4 *sqrt(`Catalogs_NewCust:Retargeting`)- 73.7 *sqrt(`Mailings:Search`)+ 103.5 *sqrt(`Mailings:Newsletter`)+ 41.8 *sqrt(`Mailings:Retargeting`)+ 169.7 *sqrt(`Search:Newsletter`)+ 45.9 *sqrt(`Search:Retargeting`)- 623.2 *sqrt(`Search:Portals`)- 93.9 *sqrt(`Newsletter:Retargeting`)	518.13	99.29%	
7.Ln Model with Intercept and Synergy	Sales = 5.53e+04- 5.66e+03 *ln(Catalogs_ExistCust)- 1.05e+04 *ln(Catalogs_Winback)- 3.81e+04 *ln(Catalogs_NewCust)- 3.01e+03 *ln(Mailings)- 1.15e+05 *ln(Newsletter)- 5.87e+03 *ln(Retargeting)- 9.06e+04 *ln(Portals)+ 2.42e+04 *ln(`Catalogs_ExistCust:Catalogs_Winback`)- 2.43e+04 *ln(`Catalogs_ExistCust:Catalogs_NewCust`)- 8.28e+02 *ln(`Catalogs_ExistCust:Mailings`)- 5.94e+03 *ln(`Catalogs_ExistCust:Search`)- 7.74e+03 *ln(`Catalogs_ExistCust:Newsletter`)- 1.46e+03 *ln(`Catalogs_ExistCust:Retargeting`)+ 2.13e+04 *ln(`Catalogs_ExistCust:Portals`)+ 2.04e+04 *ln(`Catalogs_Winback:Catalogs_NewCust`)+ 1.68e+03 *ln(`Catalogs_Winback:Mailings`)- 7.80e+04 *ln(`Catalogs_Winback:Search`)- 7.31e+02 *ln(`Catalogs_Winback:Retargeting`)+ 4.30e+04 *ln(`Catalogs_Winback:Portals`)- 1.48e+03 *ln(`Catalogs_NewCust:Mailings`)+ 7.93e+04 *ln(`Catalogs_NewCust:Search`)- 1.08e+03 *ln(`Catalogs_NewCust:Newsletter`)- 3.68e+04 *ln(`Catalogs_NewCust:Portals`)+ 1.99e+03 *ln(`Mailings:Search`)+ 3.18e+03 *ln(`Mailings:Newsletter`)+ 5.68e+02 *ln(`Mailings:Retargeting`)- 3.89e+03 *ln(`Mailings:Portals`)+ 3.76e+04 *ln(`Search:Newsletter`)- 2.69e+03 *ln(`Search:Retargeting`)- 2.95e+04 *ln(`Search:Portals`)+ 2.71e+03 *ln(`Newsletter:Retargeting`)+ 7.91e+04 *ln(`Newsletter:Portals`)+ 7.63e+03 *ln(`Retargeting:Portals`)- 3.44e-01 *ln(Stm1)	507.75	61.61%	Best among the LN models with least AIC
8.Ln Model without Intercept and Synergy	Sales = - 18589.5*ln(Catalogs_Winback)- 47313.9 *ln(Catalogs_NewCust)+ 70184.2 *ln(Search)-15723.0 *ln(Portals)+ 29165.5 *ln(`Catalogs_ExistCust:Catalogs_Winback`)- 28404.1 *ln(`Catalogs_ExistCust:Catalogs_NewCust`)- 197.2 *ln(`Catalogs_ExistCust:Portals`)+ 32011.1 *ln(`Catalogs_Winback:Catalogs_NewCust`)+ 1416.6 *ln(`Catalogs_Winback:Mailings`)- 113989.5 *ln(`Catalogs_Winback:Search`)+ 8927.4 *ln(`Catalogs_Winback:Newsletter`)+ 60840.6 *ln(`Catalogs_Winback:Portals`)- 1294.1 *ln(`Catalogs_NewCust:Mailings`)+ 112526.7 *ln(`Catalogs_NewCust:Search`)- 8378.1 *ln(`Catalogs_NewCust:Newsletter`)- 71.2 *ln(`Catalogs_NewCust:Retargeting`)- 59315.9 *ln(`Catalogs_NewCust:Portals`)- 40061.0 *ln(`Search:Newsletter`)- 27766.6 *ln(`Search:Portals`)+ 40800.7 *ln(`Newsletter:Portals`)	511.21	99.40%	

Figure 4: Extended models summary(cont.)

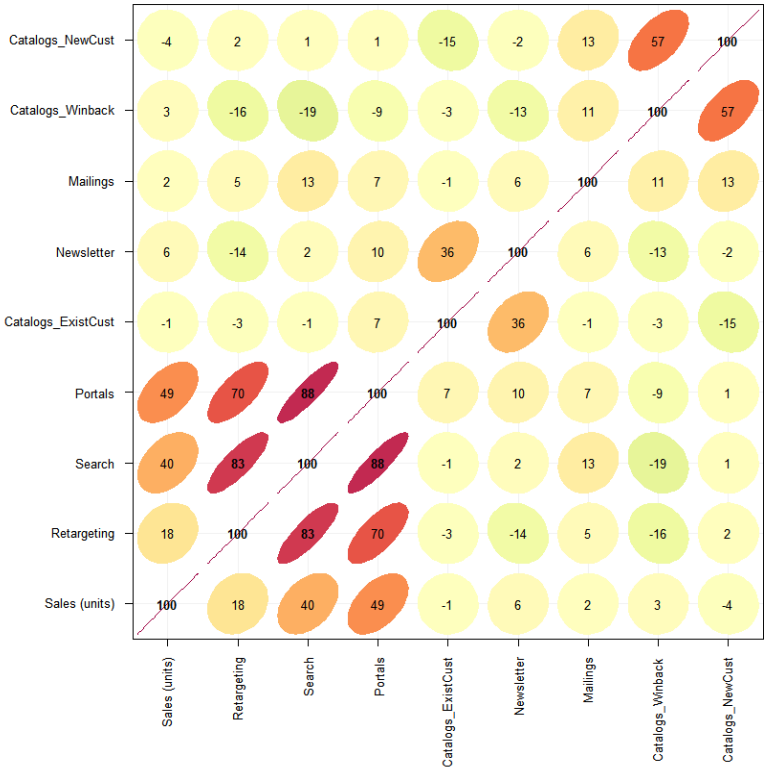

Figure 5: Correlation between different advertising channels.