Marketing Mix Modeling Analysis - DT Mart

MGT 6203 Group Project Final Report

Team # 31

Dan Zhang
Wan Wang-Geissler
Beibei Cheng
Chin-Hsien Tsai

Abstract

It is well known in the industry that measuring the effectiveness of various marketing channels is difficult and that makes allocating the marketing budget optimally an arduous task. Nowadays, marketers have powerful statistical tools to help with analysis and decision making, one of which is Marketing Mix Modelling (MMM).

In our project we have conducted an MMM analysis on DT Mart's marketing data. Several regression-based models are built with R. Lag and adstock transformation are applied to the variables to reflect the lag and decay effect of advertising, respectively. As severe multicollinearity in the data is detected, two methods are exploited to tackle this challenge: variable selection and ridge regression. Subset regression models are built and analyzed after variable selection. The thirteen models built using training data are then validated using test data. Prediction accuracies are compared using Mean Squared Prediction Error (MSPE) and Precision Error (PM). The ridge regression model is selected as the best performing model and used to analyze the most and least effective media channels. A media investment optimization is proposed based on the ridge regression's suggestion and would lead to a 50% growth in revenue.

Contents

1. Introduction

1.1 Background

The marketing mix is an essential business theory, which is the set of actions, or tactics, that a company uses to promote its brand or product in the market [1]. Price, Product, Promotion and Place are the 4Ps that make up a typical marketing mix. The 4Ps of the marketing mix influence each other. They set up the business plans for companies, to pursue their marketing objectives in the target market [2].

Marketing mix modeling (MMM) is a statistical modeling that uses historic data, to determine market attributions, evaluate the impact of each marketing channel, and then forecast the impact of future sets of tactics [3].

1.2 Literature Review

To build and apply MMM, Yuan J. (2020) introduced the guide with dataset using Python, R, and Excel. The article describes common data types for MMM, based variables, incremental variables, and others. Trent Y. (2022) built a simple MMM and make prediction using R. This paper introduced the difference between MMM and Multi Touch Attribution (MTA). For the MTA model, last click attribution and first click attribution and time decay were discussed. Then a MMM was built using R. Correlation between variables were checked using the PerformanceAnalytics package. Adstock was defined and applied. Tripathi H. (2021) referred to MMM as the butterfly effect, since a small change in the advertising strategy could make a high impact on the investments. This paper talked about the objective of MMM. Marketing mix elements were discussed. A blog introduced the definitions of marketing mix and marketing mix modeling. It compared the differences between marketing mix modeling and attribution modeling. Technologies such as linear regression model, multiplicative regression models were covered.

2. Project Overview

2.1 Problem Statement and Objective

In this project, DT Mart's marketing and sales data using marketing mix modeling will be studied. DT Mart is an Indian retailer for electronic products including cameras, gaming hardware, and entertainment products. We aim to find out the most and least effective marketing channels, try to optimize the marketing investment allocation, and see how much revenue growth can be achieved by doing so.

2.2 Approach

Our main approach is to combine the marketing and sales data first, split the data into training and test sets. Then we build and train models based on the training data, compare the models' performance using the test data. Finally, we'll optimize the marketing budget allocation as suggested by the best model and analyze the influence on revenue.

2.3 Hypothesis

The main hypothesis is that given the nature of DT Mart's business, and their target clientele, digital marketing channels such as online marketing, affiliates, and SEM should be more effective than traditional marketing channels such as TV and radio.

3. Data Transformation and Exploratory Data Analysis (EDA)

3.1 Data Cleaning and Transformation

We have two data sets, both from Kaggle. These are raw data, so preparing and cleaning data is an important job we should do before we start our model building and analyzing. In this process we use python to help us refine our job. After a preliminary inspection, we make the following assumptions about our data:

- Currency: we assume that the currency is use is Indian Rupee (INR) as the data suggests that the promotions occur on Indian public holidays
- Media investment and revenue amount: The values seen in the media investment data set are very small. A rule of thumb for marketing investment is that it should be 2% - 5% of the revenue. Working backwards, we deduce that the media investment amount must be in the hundreds of thousands (00,000) so that the total media investment of DT Mart is 2.1% of the total revenue. As such, we will transform the "gmv_new" data into hundreds of thousands (gmv_new / 100000) to keep the units consistent

A quick overview of the original data can be seen in Table 3-1.

Table 3-1 Overview of the Original Data

Dataset	Key Variables	Source	Cleaning Process
Media Investment	TV, Digital, Sponsorship, Content Marketing, Online Marketing, Affliates, SEM, Radio, Other	Kaggle - DT MART: Market Mix Modeling	Transform monthly data into weekly data, merge with Sales Revenue data
Sales Revenue (firstfile.csv)	gmv_new, special_sale	Kaggle - DT MART: Market Mix Modeling	Transform daily data into weekly data, merge with Media Investment data

The media investment dataset is relatively simple, displaying the monthly spending of each marketing channel, spanning July 2015 to June 2016. We can easily check that there are no null values, duplicates or obvious faults.

The values in this dataset need to be transformed from monthly to weekly, as our team will be using marketing mix modeling, where it is standard to use weekly data. Furthermore, the sales dataset has daily values, so it makes sense to unify them to a weekly time resolution. An excerpt of the transformed dataset can be seen in Table 3-2:

Table 3-2 An Excerpt of the Transformed Media Investment Data

week_of_year	total_invest	TV	Digital	sponsorship	content_marketing	online_marketing	affiliates	SEM	radio	other
1	3.42	0.04	0.5	1.48	0.0	0.26	0.1	1.0	0.0	0.0
2	3.42	0.04	0.5	1.48	0.0	0.26	0.1	1.0	0.0	0.0
3	3.42	0.04	0.5	1.48	0.0	0.26	0.1	1.0	0.0	0.0
4	3.42	0.04	0.5	1.48	0.0	0.26	0.1	1.0	0.0	0.0
5	3.42	0.04	0.5	1.48	0.0	0.26	0.1	1.0	0.0	0.0
6	1.28	0.0	0.33	0.28	0.0	0.03	0.03	0.63	0.0	0.0
7	1.28	0.0	0.33	0.28	0.0	0.03	0.03	0.63	0.0	0.0
8	1.28	0.0	0.33	0.28	0.0	0.03	0.03	0.63	0.0	0.0
9	1.28	0.0	0.33	0.28	0.0	0.03	0.03	0.63	0.0	0.0
10	19.26	0.78	0.28	12.56	0.12	3.28	1.0	1.24	0.0	0.0
11	19.26	0.78	0.28	12.56	0.12	3.28	1.0	1.24	0.0	0.0
12	19.26	0.78	0.28	12.56	0.12	3.28	1.0	1.24	0.0	0.0
13	19.26	0.78	0.28	12.56	0.12	3.28	1.0	1.24	0.0	0.0
14	19.26	0.78	0.28	12.56	0.12	3.28	1.0	1.24	0.0	0.0
15	42.55	1.53	3.15	21.18	0.85	6.1	1.75	7.98	0.0	0.0

The sales revenue dataset contains detailed information for each individual sale. We first check the dataset for null entries and duplications, then remove the columns that irrelevant to our analysis. After this, we proceed to accumulate the individual sales into weekly data, taking care that the column for gmv_new (revenue) is recognized as numeric.

After doing the above, the first few rows of the weekly sales data are displayed in Table 3-3:

Table 3-3 An Excerpt of the Transformed Sales Data

week_of_year	sales_name	gmv_new
1	No Promotion	623671
2	No Promotion	41002829
3	Eid & Rathayatra ...	12851746
3	No Promotion	34728672
4	No Promotion	45187859
5	No Promotion	31331539
6	No Promotion	20264
7	No Promotion	7434
7	Independence Sale	4278
8	No Promotion	11440
8	Independence Sale	689
9	No Promotion	18500
9	Rakshabandhan Sale	60269

A new boolean variable called special_sale will be used to indicate whether a special promotion has taken place in a given week.

Finally, after the two raw datasets have been cleaned and transformed, we use the week number as key to join the media investment data and sales data. The final data looks like so:

Table 3-4 Cleaned Data

week_of_year	total_invest	TV	Digital	sponsorship	content_marketing	online_marketing	affiliates	SEM	radio	other	gmv_new	special_sale
1	3.42	0.04	0.5	1.48	0	0.26	0.1	1	0	0	6.23671	0
2	3.42	0.04	0.5	1.48	0	0.26	0.1	1	0	0	410.02829	0
3	3.42	0.04	0.5	1.48	0	0.26	0.1	1	0	0	475.80418	1
4	3.42	0.04	0.5	1.48	0	0.26	0.1	1	0	0	451.87859	0
5	3.42	0.04	0.5	1.48	0	0.26	0.1	1	0	0	313.31539	0
6	1.28	0	0.33	0.28	0	0.03	0.03	0.63	0	0	0.20264	0
7	1.28	0	0.33	0.28	0	0.03	0.03	0.63	0	0	0.11712	1
8	1.28	0	0.33	0.28	0	0.03	0.03	0.63	0	0	0.12129	1
9	1.28	0	0.33	0.28	0	0.03	0.03	0.63	0	0	0.78769	1
10	19.26	0.78	0.28	12.56	0.12	3.28	1	1.24	0	0	818.77355	0

3.2 Exploratory Data Analysis

After data cleaning and preparation, exploratory data analysis was carried out using R. A time-series chart of annual Gross Merchandise Value (GMV) over a year was created using ggplot. This is illustrated in Figure 3-1.

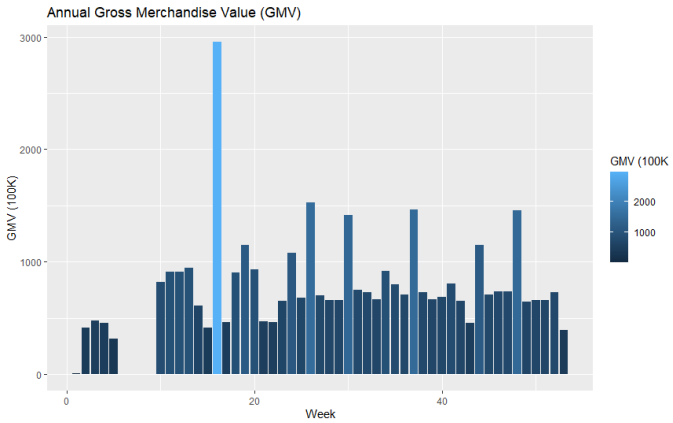

Figure 3-1 Annual Gross Merchandise Value

Figure 3-2 Correlation Between Variables

The GMV of 52 weeks in 100K are shown above. There are several values higher than the average that could be related to special holiday sales. The correlation between variables were checked using PerformanceAnalytics package. The results are shown in Figure 3-2.

All channels have a positive relationship with GMV. Significance ranges from 0.15 to 0.54. Online marketing and affiliates have a strong relationship with GMV. However, some collinearities between different channels are very high, since the factors affect each other. Multicollinearity can

make the model unstable because a lot of variances in the coefficient estimates were created. The relationship between the total investment and GMV were generated using ggplot, as illustrated in Figure 3-3.

With the total invest increases, the GMV also increases. The special sales were also analyzed. Sales without promotions were set as 0, and promotions were set as 1. The distributions were shown between. The chart in Figure 3-4. shows that the special sales have the effect on GMV.

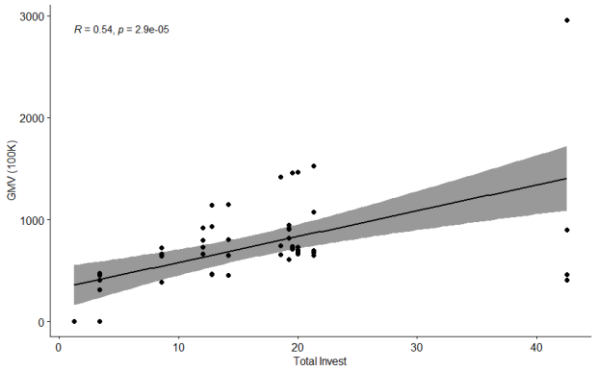

Figure 3-3 GMV vs. Total Investment

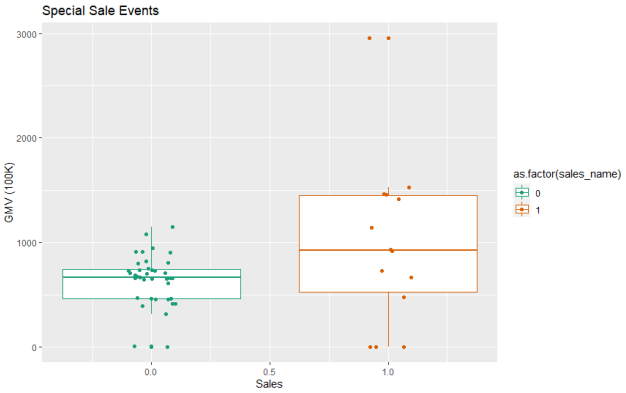

Figure 3-4 Special Sale Events

4. Modeling

4.1 The Baseline Multiple Linear Regression Model

Multiple linear regression is the most common model used in MMM. As the correlation matrix suggests that most of the explanatory variables have linear relationships with the response variable, it is worth trying multiple linear regression (model_base) as the first step. The regression formula is as follows:

lm(formula = gmv_new ~ TV + Digital + sponsorship + content_marketing + online_marketing + affiliates + SEM + radio + other + special_sale, data = datatrain)

According to the model summary, the overall regression is statistically significant. However, there are only three statistically significant variables, including sponsorship, content_marketing, and special_sale, at a 95% confidence interval. The adjusted R^2 value is 0.5484, which suggests that 54.84% of the variance in gmv_new could be explained by the media investments and special sales.

Four of the variable coefficients are negative, which means that increasing the investment in the relevant media channels, the revenue would decrease. In reality, having negative coefficients does not make sense as marketing spending should never cause the revenue to decrease. In our model, one explanation for the negative coefficients is that the corresponding investment is a very small percentage of the total investment. Another explanation would be the presence of multicollinearity in the data.

VIF value larger than 5 suggests multicollinearity. The VIF values of this model are mostly in hundreds or thousands, which suggests severe multicollinearity.

In marketing data, it is common to encounter multicollinearity because everything usually happens at the same time during marketing campaigns. For example, online marketing could be supplemented by SEM or affiliates. However, multicollinearity in our model generates high variance and weakens the statistical power of it, which is undesirable.

4.2 Lag and Adstock Transformed Models

Since advertising does not usually take effect immediately, we need to consider the "lag effect", illustrated in Appendix. It is also common that advertising or marketing events can have a lasting impact, also known as "decay", which we model by introducing the adstock transformation.

The formula of measuring adstock is as follows:

$$A_t = T_t + \lambda A_{t-1} \qquad t = 1,2,\dots,n \qquad \textit{(Joy, 2006)}$$

Where A_t is the Adstock at time t, T_t is the value of the advertising variable at time t, λ is the adstock rate, and n is the maximum number of weeks of decay.

To find the best value for these lag and adstock hyper-parameters, we iterated over a range of values and selected the ones which return the highest adjusted R^2 value for our model. The best combinations of the lag (in weeks) and adstock rate are displayed in Table 4-1. During research, we also found the "industry standard" lag and adstock rates for several marketing channels in this table, and our empirically determined values were very close to them.

Table 4-1 Lag and Adstock Hyper-parameters

Variables	Lag Weeks	Adstock Rate
TV	5	0.7
Digital	0	0.5
Sponsorship	7	0.25
Content_marketing	0	0.2
Online_marketing	6	0.2
Affiliates	0	0.5
SEM	2	0.25
Radio	0	0.2
Other	6	0.1

We used these lag and adstock hyper-parameters to transform the predictors and built two new multiple linear regression models, one with only lag transformed predictors and one with both lag and adstock transformed variables.

The overall regression is statistically significant for both models. The lag transformed model (model_lag) has improved in statistical significance as all the variables are significant and the adjusted R^2 increased to 0.6588, but multicollinearity still exists.

The lag and adstock transformed model (model_lag_ad) has two statistically insignificant variables and an adjusted R^2 value of 0.6197. It has done slightly worse in statistical significance, and multicollinearity is still a big problem, which leads us to 2 options: using variable selection to delete correlated variables and build subset models, or using ridge regression, a model tuning method to analyze data suffering from multicollinearity.

4.3 Variable Selection and Subset Models

Several methods have been used to select subsets of predictors, including best subset regression, stepwise, LASSO, and Elastic Net. The summary of the results is displayed in Table 4-2, including the best subset regression results based on Adjusted R^2 value, CP value and BIC value, and the results of stepwise, LASSO, and Elastic Net model selection procedures.

Table 4-2 Subsets of Predictors Selected by Each Variable Selection Method

	Variables Selected									
Methods	TV	Digital	Sponsorships	Content Marketing	Online Marketing	Affiliates	SEM	Radio	Other	Special Sale
Best Subset Regression (Adjusted R-squared)		✓	✓	✓	✓	✓	✓	✓	✓	✓
Best Subset Regression (CP value)			✓	✓	✓			✓		✓
Best Subset Regression (BIC value)			✓			✓				✓
Stepwise Regression		✓	✓	✓	✓	✓	✓	✓	✓	✓
LASSO			✓	✓	✓	✓		✓		✓
Elastic Net			✓	✓	✓	✓		✓		✓

Both best subset regression and stepwise regression selected the same nine variables: Digital, Sponsorships, content_marketing, online_marketing, Affiliates, SEM, Radio, Other, and special_sale. Both LASSO and Elastic Net selected the following six variables: Sponsorships, content_marketing, online_marketing, Affiliates, Radio, and special_sale . Best subset regression also selected five variables and three variables based on the CP value and BIC value respectively. Hence, we decided to build subset models with the nine, six, and five variables selected. We hoped to find the biggest subset model with multicollinearity dealt with so that we could analyze as many media channels as we could. We also added lag and adstock transformation to the subset models and generated six more subset models. The summary of the subset models is displayed in Table 4-3.

Table 4-3 Summary of Subset Models

Subset Models	Formula	Adjusted R-squared	Number of Insignificant Variables	Multicollinearity Dealt With
model_9	lm(gmv_new ~ Digital + sponsorship + content_marketing + online_marketing + affiliates + SEM + radio + other + special_sale, data = datatrain)	0.561	2	X
model_6	online_marketing + affiliates + radio + special_sale, data = datatrain)	0.519	4	X
model_5	lm(gmv_new ~ sponsorship + content_marketing + online_marketing + radio + special_sale, data = datatrain)	0.532	1	✓
model_9_lag	lm(gmv_new ~ Digital + sponsorship + content_marketing + online_marketing + affiliates + SEM + radio + other + special_sale, data = datatrain_lag)	0.551	6	X
model_6_lag	online_marketing + affiliates + radio + special_sale, data = datatrain_lag)	0.524	3	✓
model_5_lag	lm(gmv_new ~ sponsorship + content_marketing + online_marketing + radio + special_sale, data = datatrain_lag)	0.524	1	✓
model_9_lag_ad	lm(gmv_new ~ ads_digital + ads_sp + ads_cm + ads_om + ads_af + ads_sem + ads_radio + ads_other + special_sale, data = datatrain_lag)	0.605	3	X
model_6_lag_ad	lm(gmv_new ~ ads_sp + ads_cm + ads_om + ads_af + ads_radio+ special_sale, data = datatrain_lag)	0.515	3	✓
model_5_lag_ad	lm(gmv_new ~ ads_sp + ads_cm + ads_om + ads_radio+ special_sale, data = datatrain_lag)	0.525	2	✓

We define “Multicollinearity Dealt With” as all the VIF values are smaller than or equal to five. “Number of Insignificant Variables” means the number of variables in the specific model which are not statistically significance at a 95% confidence interval. The overall regression in each model is statistically significant while some variables are not. The five-variable subset model shows good statistical significance and has multicollinearity dealt with. The six-variable subset model with lag transformation also has multicollinearity dealt with but has a little bit more insignificant variables according to the model coefficients.

4.4 Ridge Regression Model

Ridge regression is a regularization method for analyzing data suffering from multicollinearity. When there is multicollinearity among the explanatory variables, the coefficient of one variable

depends on the others. By adding a penalty term λ, the coefficients of collinear variables would shrink, except for the significant variables.

To find the optimal amount of penalty, λ, we performed ten-fold cross validation using the cv.glmnet function in R to search for the λ that gives the minimum MSE (Mean-Squared Error). Figure 4-3 shows the MSE values for different λ values. The minimum MSE is achieved when λ = 27.89. The impact of different λs in the estimated coefficients is displayed in Figure 4-4. The coefficients would be shrunk towards zero when λ is very large.

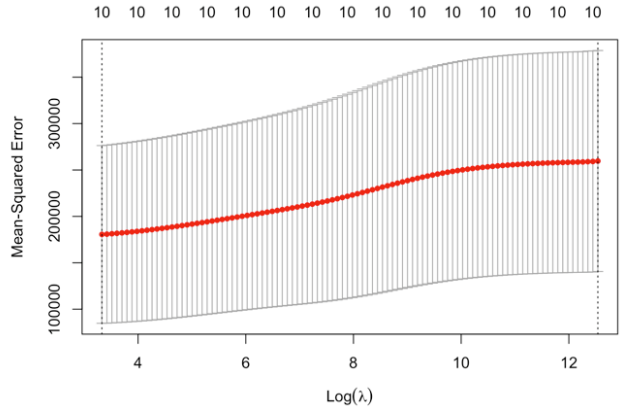

Figure 4-3 Log(λ) against MSE

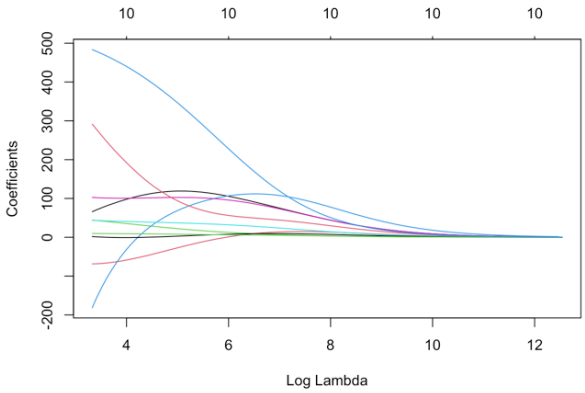

Figure 4-4 Log(λ) Against Scaled Coefficients

With λ = 27.89, ridge regression was performed using the function glmnet in R. The ridge regression (model_ridge) coefficient estimates are displayed in Figure 4-5, in comparison with the baseline multiple linear regression coefficients. The coefficients have shrunk a lot compared with the baseline model, which indicates that multicollinearity has been dealt with.

```
##                    model_base     ridge
## (Intercept)         -1193.8271  -43.4454
## TV                    401.5364   65.3841
## Digital              5439.6274  -68.7930
## sponsorship           115.0745   44.3046
## content_marketing  -13698.5494 -180.4339
## online_marketing     4037.1430   43.6812
## affiliates         -11036.5046  103.1729
## SEM                  -1459.2110    1.8647
## radio                13336.8821  288.6247
## other                 -954.5247   10.0537
## special_sale           661.4936  483.4719
```

Figure 4-5 Coefficients of the Ridge Regression Model and the Baseline Multiple Linear Regression Model

4.5 Model Comparison

Based on the training data, we have built 13 models. In order to compare these models' performance, we made predictions on the test data using these models and analyzed the prediction accuracy of each model. MSPE and PM were calculated and displayed together with adjusted R^2 in Table 4-4.

Table 4-4 Adjusted R^2, MSPE, and PM Values of Each Model

	Model	Adjusted R-Squared	MSPE	PM
Full Models	model_base	0.548	2932659.75	36.359
	model_lag	0.659	2258804.06	28.005
	model_lag_ad	0.620	232897.44	2.887
	model_ridge	0.439	86232.04	1.069
Subset Models	model_9	0.561	2294132.98	28.443
	model_6	0.519	103598.40	1.284
	model_5	0.532	92979.20	1.153
	model_9_lag	0.551	270505.10	3.354
	model_6_lag	0.524	87620.11	1.086
	model_5_lag	0.524	89556.69	1.110
	model_9_lag_ad	0.605	223024.83	2.765
	model_6_lag_ad	0.515	154677.59	1.918
	model_5_lag_ad	0.525	135058.48	1.674

This table shows that the baseline model has the highest prediction error, which affirmed our job in finding the best model and improving the model performance in the following steps. Among the models with all the variables, the ridge model performs the best while among the subset models, the model with 6 variables and with lag transformation performs the best.

Overall, model_ridge is the best performing model, although its adjusted R^2 is slightly lower. We would prefer a full model so that we could analyze all the media investment channels. Therefore, we would pick the ridge model for further analysis.

4.6 Ridge Regression Model Interpretation

With the coefficients estimated by the ridge regression, a formula of the response variable and predictors was generated:

*Gmv_new = -43.445 +65.384*TV – 68.793*Digital + 44.305*Sponsorship – 180.434*Content_marketing + 43.681*Online_marketing + 103.173*Affiliates + 1.865*SEM + 288.625*Radio + 10.054*Other + 483.472*special_sale*

The negative intercept value indicates that without media investment, DT Mart's revenue would decrease. The coefficients suggest that with 1INR increase in the specific media investment channel, the revenue will increase or decrease the amount of INR indicated by the relevant coefficients, holding all other variables constant. When there is a special sale, the revenue is predicted to increase 48.3M INR, holding all other variables constant. However, at this stage we focus our comparisons and analysis on the nine media investment channels.

According to the coefficient values, among the nine media investment channels, affiliates, TV, sponsorship and Online Marketing are the main drivers on revenue, while Digital and content marketing are the least effective as they have negative coefficients. Radio has a very high coefficient but as radio's spending is small and sporadic, it is unlikely to be a main driver of revenue.

4.7 Marketing Investment Optimization and Revenue Forecast

Based on the findings in 4.6, we adjusted the investment allocation of each marketing channel. The percentage of each media channel's investment is displayed in Figure 4-6.

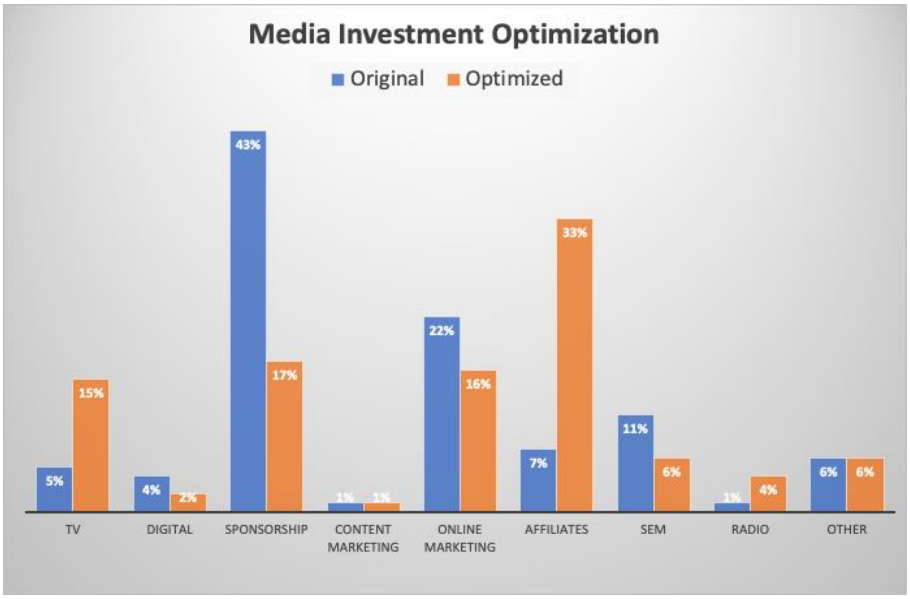

Figure 4-6 Media Investment Optimization Proposal

In this specific example, we increased the budget on TV and affiliates, and decreased the percentage of sponsorship, online marketing and SEM. Then we calculated the revenue using the formula generated by the ridge model in Section 4.6. Our model predicts that doing so the revenue would increase from 3884.73M INR to 5831.56M INR, a 50% growth.

5. Conclusion

5.1 Summary

With the datasets of DT Mart, we have conducted a thorough marketing mix modeling analysis. After training and validating the 13 models, we chose the ridge regression model as the best performing model for our data. Marketing budget allocation could be optimized and the influence on driving the revenue growth is significant.

There are a few benefits of the model. It could help marketers understand the most and least effective marketing channels and enable marketing budget allocation optimizing. It could also facilitate superior sales revenue forecasting.

There are also limitations. With only one year's data, after transformation into weekly data, we only have 53 observations and 10 variables. There is a risk of model overfitting. The data quality is not ideal. And there was severe multicollinearity issue. They all weaken the statistical power of the model. Besides, this model is more valuable in analyzing short term marketing effect on the revenue. Some variables may not be a strong driver of revenue in short term but could contribute significantly to the brand equity.

5.2 Future Work

Although the project is concluded here, there are several future improvements that can be made:

- The data quality could be improved. More data could be collected on a weekly basis at better quality. We could re-fit the model when more data becomes available.
- With at least 2 year's data, we would be able to include trend and seasonality in the model and would be able to predict revenue forecast for future.
- If attribution data could be provided, we could calculate the contribution of each media channel on revenue and perform an ROI analysis. Valuable insights could be gained for better decision making.

Reference

[1] *What is 'Marketing Mix'*, The Economic Times, Oct 06, 2022, https://economictimes.indiatimes.com/definition/marketing-mix;

[2] Kotler, P., Marketing Management, (Millennium Edition), Custom Edition for University of Phoenix, Prentice Hall, 2001, p. 9;

[3] *A Complete Guide to Building a Marketing Mix Model in 2021*, Terence Shin, Towards Data Science, Jul 24, 2021, https://towardsdatascience.com/an-complete-guide-to-building-a-marketing-mix-model-in-2021-fb53975be754

[4] *Marketing Mix Model Guide With Dataset Using Python, R, and Excel*, Jacky Yuan, Analytics Vidhya, Dec 21, 2020, https://medium.com/analytics-vidhya/marketing-mix-model-guide-with-dataset-using-python-r-and-excel-4e319be47b4

[5] *Advertising adstock theory*, Marko Mikes, Aug 27, 2016, Medium, https://medium.com/@Marko_Mi/advertising-adstock-theory-85cc9e26ea0e

[6] *Using R to Build a Simple Marketing Mix Model (MMM) and Make Predictions*, Trent Y., Towards Data Science, https://towardsdatascience.com/building-a-marketing-mix-model-in-r-3a7004d21239

[7] *Marketing Mix Modelling (MMM) — a potential solution*, Hitesh Tripathi, 2021, https://tripathitesh.medium.com/marketing-mix-modelling-mmm-a-potential-solution-48ba3a248de9

[8] *A Deep Insight into Marketing Mix Modeling - Definition, Methodology, Comparison, Techniques*, https://blog.avada.io/resources/marketing-mix-modeling.html

[9] *The Science Behind the Art of Marketing-Mix Modeling*, Derek L., Nakul P., and Sadashiv G., GAMMA—Part of BCG X, 2021, https://medium.com/bcggamma/the-science-behind-the-art-of-marketing-mix-modeling-ba721d6dab33

[10] *Advertising Adstock with Maximum Period Decay, Gabriel Mohanna*, 2014 https://analyticsartist.wordpress.com/2014/02/26/advertising-adstock-with-maximum-period-decay/

[11] *Understanding Advertising Adstock Transformations,* Joseph Joy, 2006, https://mpra.ub.uni-muenchen.de/7683/4/MPRA_paper_7683.pdf

Appendix

I. Lag and Adstock

Figure I-1 Lag and Decay Effect (Yuan, 2020)

Figure I-2 Advertising Adstock (Mohanna, 2014)

II. VIF Values of Each Model

```
vif(model_base)
```

```
##              TV         Digital      sponsorship content_marketing
##      190.456827     2591.578949       36.417592       549.054431
##  online_marketing      affiliates             SEM           radio
##     7611.817602     6651.689444    2679.609246     1862.247123
##           other    special_sale
##     1945.306391        1.417151
```

```
vif(model_lag)
```

```
##              TV         Digital      sponsorship content_marketing
##       9.868966       96.789955       13.853125      157.785005
##  online_marketing      affiliates             SEM           radio
##       9.006391       11.719613        1.596753      216.056470
##           other    special_sale
##     155.093548        1.443317
```

```
vif(model_lag_ad)
```

```
##       ads_tv  ads_digital       ads_sp       ads_cm       ads_om       ads_af
##    14.490322    33.589700     9.277622    39.953757    16.243674    10.289963
##      ads_sem    ads_radio   ads_other special_sale
##     5.801244    90.459067    69.106832     1.282134
```

```
vif(model_9)
```

```
##           Digital       sponsorship content_marketing  online_marketing
##       1066.472143         26.632958        536.857089       3783.281879
##        affiliates               SEM             radio             other
##       2863.977311        620.982754        304.541898        196.468957
##      special_sale
##          1.402502
```

```
vif(model_6)
```

```
##       sponsorship content_marketing  online_marketing        affiliates
##          5.506919          7.727372        450.976190        383.926866
##             radio      special_sale
##          1.858690          1.240833
```

```
vif(model_5)
```

```
##       sponsorship content_marketing  online_marketing             radio
##          4.380754          3.561388          1.957405          1.835561
##      special_sale
##          1.239981
```

```
vif(model_9_lag)
```

```
##           Digital       sponsorship content_marketing  online_marketing
##         51.868069         11.866599         85.581675          4.236638
##        affiliates               SEM             radio             other
##          6.307113          1.519215        123.589568         83.648537
##      special_sale
##          1.441451
```

```
vif(model_6_lag)
```

```
##       sponsorship content_marketing  online_marketing        affiliates
##          4.373781          3.652673          2.738882          4.087267
##             radio      special_sale
##          1.838191          1.239921
```

```
vif(model_5_lag)
```

```
##       sponsorship content_marketing  online_marketing             radio
##          4.094032          3.436531          1.116647          1.602981
##      special_sale
##          1.232694
```

```
vif(model_9_lag_ad)
```

```
##  ads_digital       ads_sp       ads_cm       ads_om       ads_af      ads_sem
##    33.321697     9.239091    38.702394     6.098383     9.471711     5.787329
##    ads_radio    ads_other special_sale
##    77.791139    55.389518     1.260230
```

```
vif(model_6_lag_ad)
```

```
##       ads_sp       ads_cm       ads_om       ads_af    ads_radio special_sale
##     4.528237     3.790636     3.840905     5.431496     1.811970     1.233278
```

```
vif(model_5_lag_ad)
```

```
##       ads_sp       ads_cm       ads_om    ads_radio special_sale
##     4.181382     3.583875     1.141813     1.597129     1.210529
```

III. Team Information

Dan Zhang
dzhang435

Full-time OMSA student with 6 years' financial experience

Wan Wang-Geissler
wwanggeissler3

Sr. researcher within R&D of a chemical company

Beibei Cheng
bcheng71

Full-time OMSA student with 10 years' marketing experience in tech

Chin-Hsien Tsai
ctsai98

Full-time OMSA student with biology and computer science background