Digital Marketing Budget Allocation using Predictive/Prescriptive Analytics techniques

Mihir Padsumbiya mxp210085@utdallas.edu
Preet Desai pxd210046@utdallas.edu
Gautam Iyer gautam.iyer2@utdallas.edu
Poorwa Kunwar Poorwa.Kunwar@utdallas.edu

Abstract

This paper proposes an approach that utilizes predictive and prescriptive analysis to optimize digital marketing budget allocation and maximize the overall reach of digital marketing campaigns based on 5 key metrics. The proposed model uses a Zero Inflated Regression model with Decision Trees Classifier and Linear regression (With the assumption of linear relationship between dependent and independent variables). Based on the numerous statistical tests, and linear programming conclusions, maximum budget should be allocated to Programmatic Display Ads.

Introduction

This paper presents a data-driven approach to optimize digital marketing budget allocation and maximize campaign reach across five key metrics. Using a combination of Zero Inflated Regression, Decision trees Classifier, and Linear Regression, the study concludes that Programmatic Display Ads are the most effective channel. By leveraging data-driven insights, companies can make informed decisions about how to allocate their resources, minimizing waste and maximising impact of their marketing campaigns. This work offers valuable insights for companies seeking to make informed decisions about budget allocation and stay competitive in a rapidly changing world.

Solution Formulation

For the purpose of this project, numerous statistical and analytical tests were performed. Firstly, the significance of each variable in the model were checked using difference in means tests- t test/ ANOVA. The results of which are tabulated below. Initial analysis helped omit duplicate observations and null rows in some variables- Ad Device, Ad Group, and Ad Content ID were inspected and corrected for. Further, the data was inspected for below insights:

1. On average, the number of web-visits through Search channels are on par for both Engagement and Purchase Goals.
2. For traffic generated through Search engine, on average, the response increases linearly with increase in spend.
3. For traffic generated through Programmatic Display ads, on average, the response increases exponentially after spending about $30000.
4. For traffic generated through Social media, on average, the response is constant above an expenditure of $2000.
5. Amount of social engagement is most cost efficient, on average, through Social Media Channel.
6. Amount of web visits is most cost efficient, on average, through Search engines.
7. Number of Impressions is most cost efficient, on average, through programmatic ad displays.
8. Programmatic Ad displays are, on average, the costliest form of advertising in terms of generating leads.
9. Audience type 3 has, on average, the most increase in impressions with increase in $ Spend.
10. Audience type 3 has, on average, the most increase in number of clicks with increase in $ Spend.
11. Audience type 1 has, on average, the most increase in web visits with an increase in $ Spend.
12. Audience type 3 has, on average, the most increase in collateral views with increase in $ Spend.
13. Customer response rate stays constant for the entire duration of the campaign, on average, for campaign IDs 1, 4, and 5.
14. Customer response rate increases after an average of 127 days since the start of campaign for campaign IDs 2 and 3.

The conclusions from exploratory data analysis helped clean the data for noise. The next step was to define appropriate metrics to evaluate the success of each campaign goal on any given day. For this a number of factors were taken into consideration including but not limited to the given Audience responses on the ad platform as well as the company websites, along with the Campaign goal code which was necessary to evaluate the meaning of success for any given campaign. The KPI's defined to evaluate the success of a campaign are:

1. Click through rate = CTR = clicks/Impressions
2. Completion rate= VCR= Video Completes/Impressions

3. Social Media Engagement Rate= SMER= Likes+Shares/Impressions
4. Conversion Rate- For Goal Code = Engagement= Web visits/Impressions
5. Conversion Rate- For Goal Code = Awareness = Collateral View/Impressions
6. Conversion Rate- For Goal Code = Intent to Buy = Product View/Impressions
7. Conversion Rate- For Goal Code = Consideration = Forms Complete/Impressions
8. Cost Per Action = Spend/ Sum of all actions on the website
9. Cost Per Click = Spend/ # Clicks

For goal code "Purchase" the rows were omitted as not enough data is available to generate an accurate model on the same. Observations were also omitted where the number of impressions is 0 as there is no data to suggest the reach of audience response.

A number of Linear Regression models on different metrics with interaction terms were trained and evaluated based on R2 score. However, due to the presence of a large number of 0's in the target variable (as a result of 0 actions by the audience on several occasions), the dataset demanded the use of Zero Inflated Regression model. The solution led to training a Zero Inflated regression model with Decision Trees for Binomial classification and then a Linear Regression model on non-zeros. The results of the Regression model were used as an objective function in a Linear Programming optimization model.

Statistical tests

Sr. No	Test conducted to check the significance at 95% confidence interval.	Results
1	Check for the significant difference in spending between different AD channels.	Social>Programmatic Display>Search.
2	ANOVA test to check for significant differences in spend between different Audience Type.	Audience Type3> Audience Type2> Audience Type4> Audience Type1> Audience Type5
3	ANOVA test to check for significant difference in clicks between different AD channels.	Social>Programmatic Display>Search.
4	ANOVA test to check for significant differences in clicks between different Audience Type.	Type3>Type4>Type2>Type1>Type5.
5	ANOVA test to check for significant differences in impressions between different AD channels.	Social>Programmatic Display>Search.
6	ANOVA test to check for significant differences in impressions between different Audience Type.	Type3>Type4>Type2>Type5>Type1
7	ANOVA test to check for significant difference in video completes between different AD channels	Social>Programmatic Display>Search
8	ANOVA test to check for significant differences in video completes between different Audience Type.	Type4>Type2>Type3>Type5=Type1
9	ANOVA test to check for significant difference in social likes completes between different AD channels	Social>Programmatic Display=Search
10	ANOVA test to check for significant difference in social likes between different Audience Type	Type3>Type2>Type4=Type5=Type1
11	ANOVA test to check for significant difference in social shares between different AD channels	Social>Programmatic Display=Search
12	ANOVA test to check for significant difference in social shares between different Audience Type	Type3>Type2>Type4=Type5=Type1
13	ANOVA test to check for significant difference in web visits between different AD channels	Search>Programmatic Display>Social
14	ANOVA test to check for significant difference in web visits between different Audience Type	Type1>Type3>Type5>Type4=Type2
15	ANOVA test to check for significant difference in collateral views between different AD channels	Search>Programmatic Display>Social
16	ANOVA test to check for significant difference in collateral between different Audience Type	Type1>Type3>Type5>Type4=Type2
17	ANOVA test to check for significant difference in product views between different AD channels	Search>Programmatic Display=Social
18	ANOVA test to check for significant difference in product views between different Audience Type	Type1>Type5=Type3>Type4=Type2
19	ANOVA test to check for significant difference in form completes	Search>Programmatic Display=Social

	between different AD channels	
20	ANOVA test to check for significant difference in form completes between different Audience Type	Type1>Type5=Type3=Type4=Type2

By conducting ANOVA tests to check for significant differences in spend, clicks, impressions, video completes, social likes, social shares, web visits, collateral views, product views, and form completes between different AD channels and audience types, the business can identify which channels and audience types are performing better and allocate the budget accordingly to maximize ROI. The results of these tests can also help the business in making data-driven decisions to improve the digital marketing strategy.

Solution Implementation

This paper proposes the use of the results of a Zero Inflated Regression predictive model as the objective function to a Linear Programming optimization problem for optimal marketing budget allocation. The Zero Inflated regression model uses Decision Tree as binomial classifier (for large number of 0's in the target variable) along with a Linear Regression model. The linear equation generated along with the coefficients of the model are used as a maximization problem in Solver (MS Excel).

Results

We have allocated the budget based on both audience type and channel wise, while ensuring the constraints have been satisfied. The total budget of 1 million has been split as follows:

Allocation per channel wise:
Social: $140,000
Search: $360,000
Programmatic: $500,000
Allocation per audience type:
Audience 1: $334,962
Audience 2: $234,800 (Social - $114,800, Programmatic - $120,000)
Audience 3: $185,200 (Social - $25,200, Programmatic - $160,000)
Audience 4: $160,000
Audience 5: $85,038 (Search - $25,038, Programmatic - $60,000

Our team has allocated the highest budget to audience 1, as they represent the most important target audience for your brand. We have also allocated a substantial budget to audience 2 and 3, as they have shown potential for significant growth and engagement.

Additionally, we have allocated a significant portion of the budget to programmatic advertising, as it has shown to be an effective channel for targeting specific audiences and generating a high return on investment. Social media advertising and search engine advertising have also been given ample budget to ensure that we are reaching our target audience through multiple channels.

We believe that this allocation will provide the greatest return on investment and maximize the effectiveness of our marketing efforts. We are confident that our plan will exceed your expectations and help your brand achieve its marketing goals.

Conclusions

The proposed model assumes a Linear relationship between the dependent and independent variables which in practicality is not mandatorily true. Furthermore, the proposed model ignores customer segmentation, and synergy effect between different variables which might come into effect. The EDA suggested that the campaigns gain traction after a period of 150 days after the start of campaigns for campaign 2 and 3. The time factor has been ignored and a linear constant relationship has been assumed between time and the effect of campaigns for all audience types through all channels.

Next steps, to improve the practical use of the model, a time series model is proposed as for any marketing campaign, the response has a threshold and a decay rate. Furthermore, the variables could possess a non-linear log relationship which must be tested for and an appropriate non-linear regression model should be trained. The model can be improved further by formulating the model to include Sales data.

References

https://www.lytics.com/blog/15-important-digital-marketing-analytics-metrics/
https://www.winclap.com/how-machine-learning-is-helping-us-to-optimize-cross-channel-budget-allocation-decisions/
https://pubsonline.informs.org/doi/10.1287/ijoc.2020.1023
https://dl.acm.org/doi/10.1145/2983323.2983347
https://www.latentview.com/marketing-mix-modeling/
https://doordash.engineering/2020/07/31/optimizing-marketing-spend-with-ml/
https://tryolabs.com/blog/price-optimization-machine-learning