SHOEDAZZLE

MEDIA ANALYSIS & RECOMMENDATION

CAL POLY - 2019

INTRODUCTION

Our industry project is an in-depth analysis of the social media advertising data for the ShoeDazzle retail subscription business. We looked into how ShoeDazzle can optimize its ads in order to most efficiently increase subscriber ("VIP") conversions. The data that we received was Facebook Campaign Data from November 2017 through December 2018. We were provided with all of the dimensions against which the ads are purchased (platform, subchannel, post type, audience, date, dynamic type, bidding window, offer, and spend) along with the performance data (impressions, clicks, VIP vs. lead and view-through vs. click through conversions). The focal points for our analysis were evaluating the impact of different ad dimensions, finding opportunities and impediments in terms of cost-efficiencies, and identifying patterns in seasonality.

This report will cover the data analysis and decision-making process that allowed us to come up with our recommendations. Our team first started our process by cleansing our data to address any anomalies, whether that be outliers or missing values, and ensure accuracy in our findings. We then utilized a variety of analytical methods to find the most important factors for optimizing VIP efficiency. We used cluster analysis, multiple linear regression models, and logistic regressions. We identified from these analyses which variables were statistically significant and which ones had the largest on ad performance, either positive or negative. Lastly, we outlined our findings in the form of recommendations and a quarterly media mix that can be implemented in ShoeDazzle's future advertising campaigns.

II. METHODOLOGY & ANALYSIS

APPROACH

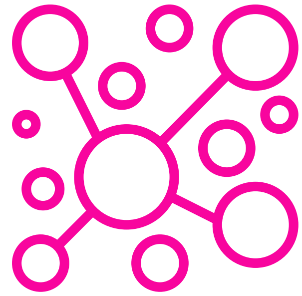
We broke out our overall goal of optimizing VIP conversions into two main objectives: increase VIP efficiency and increase leads. The strategy for optimizing VIPs is minimizing cost-per-VIP acquisition (CPA). We also investigated the leads segment with the same approach and methods as VIPs. It is important to increase leads because they are a highly qualified audience that are more likely to drive conversion. Maximizing this mid-funnel audience will allow for more cost-efficient conversion later on and also help build a larger retargeting pool from which ShoeDazzle can draw from.

DATA CLEANSING

We first began our data cleansing process by looking for any anomalies or outliers in the data and either modified or omitted the items in question. The specific items we omitted were the ads with missing values for Spend, acquisition duplication in October and November, and one day in June that had blanks for Pixel values. We also increased the CPM of all [redacted] ads by [redacted]% to take into account a [redacted]% premium that [redacted] charges to make the ad. Lastly in terms of client-directed changes, we filled in the missing Facebook Objective values as “Conversions,” and the missing Audience values to “Broad.”

Next, we chose to modify the data in a few ways that would help our in our analysis. We first grouped the ads by Ad ID. We also grouped together Quarter 4 from 2017 and Quarter 4 from 2018, so that we could more concisely analyze the seasonality. We calculated and added variables for CPM (cost-per-thousand impressions), CPA (cost-per-VIP acquisition), CPL (cost-per-lead), click rates, and conversion rates for leads and VIPs. Another calculation we added was Day Count (the number of days an Ad ID was served). Although it didn’t result in a significant finding, we added a dimension for Day of Week. Lastly, we renamed the audiences to fit into larger groups. We bucketed them into Retargeting, Broad, DLX, LAL, GEO, Facebook Fan Friend, Facebook Fan Cancelled VIP, and Interest. Drilling down more into Retargeting, we also created variables to separate Lead vs. Site Retargeting.

CLUSTER ANALYSIS



We began our analysis of the data by running hierarchical clustering. We clustered our ads based on conversion rate and click rate. We investigated results for both the Agglomerative and Divisive method before settling on the former. We clustered a random sample of 5,500 Ads from the entire year. We split the dendrogram at seven clusters for the quarterly analysis and nine for the entire year analysis. After having our data in clusters, we used visualization software (Tableau) to look for any general patterns in the data. We then ran regressions within each cluster to identify which variables were statistically significant in contributing to performance. The goal of our cluster analysis was to identify hyper-performing ad profiles that drove high conversion and low CPAs. The clusters show us specific combinations of ad dimensions that worked well together, that are more difficult to identify with regressions.

REGRESSION MODELS

The most significant portion of our analysis came from running regressions on the dimensions both by quarter and entire year. For VIPs, we ran the dimensions against CPA (cost-per-VIP acquisition). For leads, we ran them against CPL (cost-per-lead). Our main strategy was to look at our initial media mix for each variable on each dimension and see which ones had the most significant impact. We mapped out a chart (Figure 1 in Appendix) of each quarter and the entire year and marked the important variables.

We looked at the potential change in CPA or CPL (coefficients) and the level of significance, which we capped at the alpha=0.05 level. We used log(CPA) and log(CPL) to normalize the distribution in the regressions after seeing that the lead conversion data had a skewed distribution. The end result gave us the percent changes that would occur to the CPA and CPL if certain variables were changed.

REGRESSION MODELS

REFERENCE VARIABLES

The variables that we selected as the reference variables in the regressions were strategically selected. In order to see what would happen with a change from the current mix, we chose to select reference variables (missing variables) that were the variables composing the largest majority in each dimension category. For example, we saw that, for [redacted], the majority of ads in were [redacted] therefore we made this the missing variable in the regression so that we would read [redacted] coefficients as the potential change in CPA if there was a shift away from [redacted]

SEASONAL INDEX

A notable aspect of our regressions was our use of a Seasonal Index for the entire year regressions. In order to see the impact a variable would have, regardless of quarter, we deseasonalized the CPAs. For example, [redacted] is only in Q[redacted], and showed a high potential % change in CPA with its coefficient. It would be inaccurate to apply that % change to the CPA in another quarter, because some quarters have significantly different costs. This way, if we chose to add [redacted] dollars to another quarter other than Q[redacted], we could more accurately predict the CPA in that quarter.

SEASONALITY

Another way we looked at Seasonality in our regressions was by quarter and by month. We added these variables to drill deeper into performance to identify seasonal opportunities for cost-efficiency. We wanted to establish significance for certain quarters and months beyond viewing the trends in Tableau. We did this for both VIP and Lead conversion.

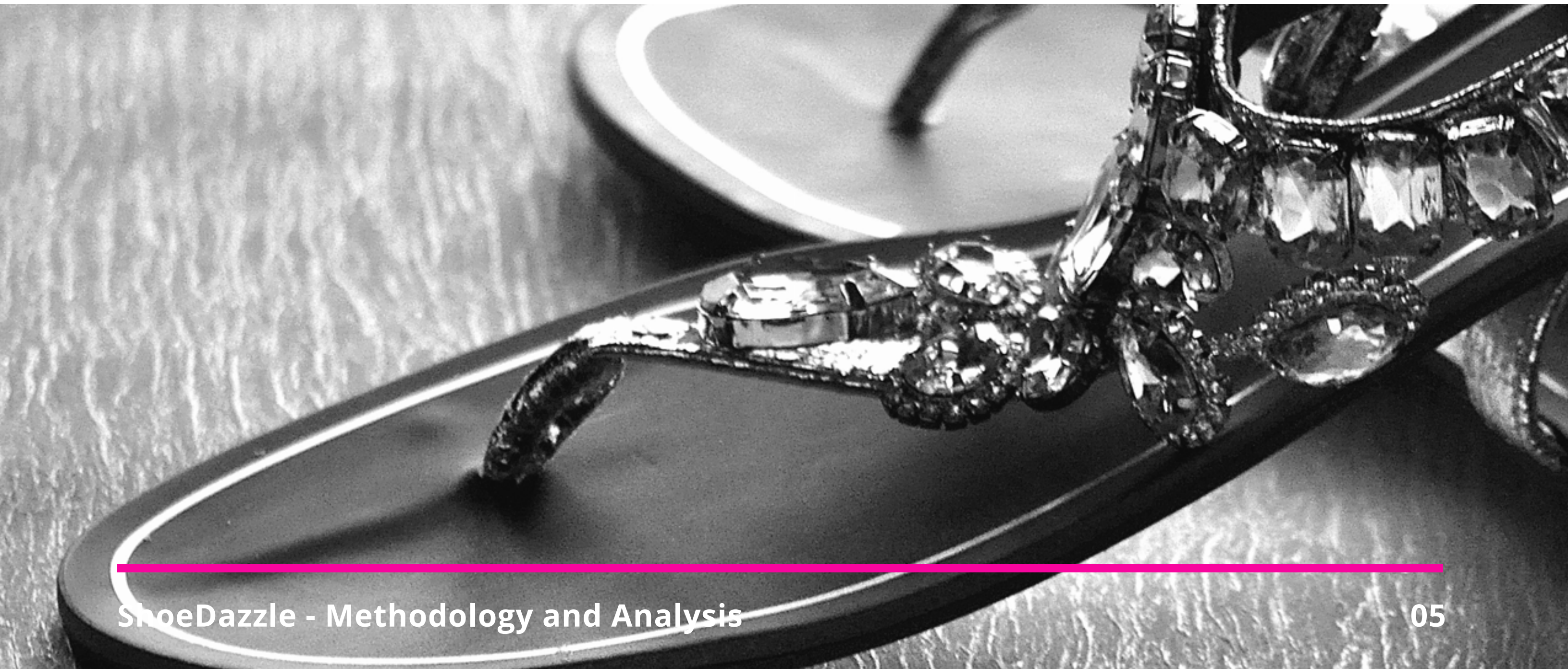
RELEVANCE SCORES

Our research yielded that Facebook Relevance Score above [redacted] positively contribute to the CPA of an advertisement. We created a binary variable for a “Good” score (greater than [redacted]), and utilized a logistic regression to determine variables that contribute to better odds of receiving a positive advertisement score. Advertisements that catered to [redacted] audiences had a significant probability of receiving a relevance score above [redacted], which is to be expected. The [redacted] and [redacted] advertisement also had significantly increased odds of receiving a favorable relevance score.

Otherwise, the most profound influence on this score was had by [redacted] videos. Shoedazzle advertisements currently have an average relevance score of [redacted], while [redacted] advertisements average a score of [redacted]% of [redacted] advertisements received a “Good Score”, likely due to its especially appealing creative effects.

PROJECTING CPAS

Our quarterly CPA projections were found by creating a sample data set that had distributions of four dimensions --Subchannel, Audience, Offer, and Dynamic that corresponded to our recommended media mix. We then used our model to predict the CPAs for each ad in the sample data set. The median for each quarter was adjusted using our seasonality index for the quarterly median.



III. RESULTS & RECOMMENDATIONS

RECOMMENDATION #1

Our first recommendation is to shift allocation be hannels, specifically away from [redacted] and towards [redacted] and [redacted]. [redacted] was the subchannel with the highest allocation of ads served ([redacted]%), and based on our regressions, the CPA would decrease if we increased [redacted] and [redacted] tead. Incremental investment in [redacted] would decrease CPA by [redacted]% vs. [redacted] [redacted]. We recommend increasing [redacted] from [redacted]% to [redacted]%.

[redacted], the 3rd party vendor that ShoeDazzle partners with to create video content, was a standout performer. [redacted] thrived with the lowest overall CPA for the year, even while only flighted in Q[redacted] when CPMs are the highest. Using regressions integrating a Seasonal Index verified that investment in [redacted] would be advantageous overall and not just due to the fact that it ran in a specific quarter.

Our recommended Subchannel mix for the entire year would be [redacted]% [redacted] [redacted], [redacted]% [redacted], [redacted]% [redacted], and [redacted]% [redacted]. We took into account the coefficients/changes in CPA by quarter when deciding the optimal levels, but we recommend a consistent mix year-round, as we do not have sufficient evidence or rationale to support the idea that Subchannel mixes should differ by quarter. Multiple years of data would be needed to draw those conclusions. (Note: Our quarterly differences will, therefore, largely be regarding the volume of spend allocated to each quarter, outlined below, as opposed to the ad mix).

CPA EFFECT:

The suggested new mix of Subchannels, based on recommendation #1 would drive CPA down by the following amounts per quarter:

[redacted]%

QUARTER 1

$[redacted] vs. $[redacted]

[redacted]%

QUARTER 2

$[redacted] vs. $[redacted]

[redacted]%

QUARTER 3

$[redacted] vs. $[redacted]

[redacted]%

QUARTER 4

$[redacted] vs. $[redacted]

RECOMMENDATION #2

Our second recommendation is to shift dollars between months to leverage seasonality patterns. To acquire VIPs, we found through our Tableau time series charts that there were opportunities for higher cost-efficiency. The heaviest spending was in [redacted] and [redacted] ([redacted]% and [redacted]% of spend respectively). The highest conversion rate is in [redacted], along with the lowest CPA. After our initial observations in Tableau, we created new variables in R to run regressions against specific months. We confirmed that there was statistically significant evidence in making shifts in spend between months to decrease CPA.

We found in the regression that it would be more cost-efficient to shift dollars out of Q[redacted] and into Q[redacted], Q[redacted], and Q[redacted] in that priority order. Their respective marginal effects on CPA are: -[redacted]%, -[redacted]%, and -[redacted]% respectively. Drilling down to the month, we found that it was optimal to shift dollars from [redacted] and [redacted] to [redacted], [redacted], [redacted], and [redacted]. These months all have large marginal effect %s on CPA compared to [redacted] and [redacted] (as demonstrated in Figure 3 in the Appendix).

CPA Effect:

The suggested new flighting allocations, which drop [redacted] by [redacted]% spend, [redacted] by [redacted]%, increase [redacted] ([redacted]% each), [redacted] ([redacted]%), [redacted] ([redacted]%), and [redacted] ([redacted]%) would drive CPA down to $[redacted] from $[redacted] (a decrease of [redacted]%).

Recommendation #3

Our third recommendation is to optimize specific ad profiles that we discovered in our cluster analysis. From our cluster analysis we found two high-performing cluster profiles that we recommend increasing (seen in Figures 7 & 8 in Appendix). The first one is a combination of: [redacted], [redacted], [redacted], [redacted], and [redacted] or [redacted]. The clusters that had this type of ad dimension combination drove a CPA of roughly $[redacted]-[redacted]. Since retargeting reaches a limit based on the size of the lead pool, we made sure to look at a cluster type that didn't include retargeting. The second cluster profile we found is: [redacted], [redacted], [redacted], [redacted], and [redacted], which saw roughly a CPA of $[redacted].

Recommendation #4

Our fourth recommendation is to increase leads, as leads are more valuable to target in order to optimize conversion. We found that [redacted]% of leads convert to VIPs, which is much higher than the average conversion rate of [redacted]%. When we analyzed the ad dimensions against cost-per-lead vs. cost-per-VIP acquisition, the main differences we found were that leads responded better for the following ad profile vs. VIPs: [redacted], [redacted], and [redacted]. We also found that there was a difference in seasonality for leads vs. VIPs. Based on our regressions, it would decrease CPL by allocating more dollars to [redacted], [redacted], and [redacted].

OVERALL PROJECTED CPAS BY QUARTER:

Per the explanation in the Analysis section, we used a sample of of our data with a similar media mix to that which we recommended to project quarterly CPAs.

QUARTER 1

$[redacted]

vs. $[redacted]

QUARTER 2

$[redacted]

vs. $[redacted]

QUARTER 3

$[redacted]

QUARTER 4

$[redacted]

vs. $[redacted]

YEARLY TOTAL

[redacted]%

[redacted]ase

IV. Appendix

MEDIA MIX ONE SHEET

Overall our analysis focused on the top four recommendations, however we also created a full media mix recommendation one-sheet (Figure 2 in Appendix) outlining recommended allocations for all dimension types with significant potential for optimization.

NOTES

We wanted to note that we looked into Day of the Week, Black Friday/Cyber Monday seasonality effects, Frequency/Duration of ads, and Interactions and did not find either statistically significant evidence or a large enough impact to make a recommendation on them. We also excluded Platform in some regressions as we had to run them with Platform or Subchannel separately due to multicollinearity.