Title
Experimenting with Online Advertisements Using Full Factorial Design

Author
TRINH, MICHELLE NHU-Y

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Experimenting with Online Advertisements Using Full Factorial Design

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Applied Statistics

by

Michelle Nhu-Y Trinh

2022

ABSTRACT OF THE THESIS

Experimenting with Online Advertisements

Using Full Factorial Design

by

Michelle Nhu-Y Trinh

Master of Applied Statistics

University of California, Los Angeles, 2022

Professor Yingnian Wu, Chair

With the rising use of technology, online advertising has become an effective tool for many businesses. This research paper explores the different factors of digital ads and its effect on ad clicks. The full 2^k factorial experimental design provides a better understanding on which factors can optimize CTR (clickthrough rate) and CPC (cost per click). This experiment was designed with new e-commerce entrepreneurs in mind to build a starting knowledge of digital ads based on statistical procedures and results. While there were capital limitations and minimal brand awareness, the experiment still showed actionable results that can drive future business decisions.

The thesis of Michelle Nhu-Y Trinh is approved.

Hongquan Xu

Michael Tsiang

Yingnian Wu, Committee Chair

University of California, Los Angeles

2022

To Newton . . .

who inspired the business venture for this thesis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Purpose of Experiment

Online advertising has rapidly grown in the last decade thanks to the increased dependence on technology and its expanding digital footprint. What started in the 1990s with simple clickable banners above ancient search engines has now grown into a 356 billion dollar industry [2]. It has become one of the essential forms of advertising in today's marketing standards. The COVID-19 pandemic played a big part in continuing its growth; people were forced to remain at home and find other ways to be entertained. To distract themselves from the uncertainty and chaos of the global pandemic, many turned to social media platforms, such as YouTube, Instagram, and TikTok, to still feel that sense of connection with others. According to surveys conducted by HubSpot, social media usage increased by 11% from 3.4 billion in 2019 to 3.78 billion in 2021 [2]. The customer experience has been a huge driver in growth and 2020 only accelerated the importance of buyer-sell relationships [2]. Social media marketing is more personalized and conversational, making it the top approach and priority. With those staggering numbers, the digital ecosystem grows with more user data on people's searches, interests, spending behaviors, and online engagement. Advertisers now have the power to gather and generate new strategies to bring a personalized touch to the customer, all through the use of online advertising.

A quick search on how to make the best digital ads will list out dozens of tips and tricks from the "top" marketing experts. For example, online surveys claim that video ads produce two to three times higher click rate results over static images. Some marketers emphasize the importance of call to action words that advertise discounts or sales. Other marketing blogs suggest it comes down to defining the proper ad keywords that best fit the target audience. With all these different factors in mind, it can be overwhelming for a new small business to figure out where to start. This experiment aims to provide insight on the most significant ad factors and its impacts on CTR (clickthrough rate) and CPC (cost per click).

The objective of this experiment is to apply a statistical approach to a customer-centric e-commerce world. The full 2^k fractional experimental design will help drive actionable business decisions. Results show which factors may have a significant impact on bringing traffic and sales to a new online store. Digital advertising results can be optimized by appropriately setting basic ad factors, such as format, enticing sales, and daily budget.

1.2 Dropshipping E-Commerce Business Model

In order to start this experiment, it is important to review the business model, technology tools, and target audience. Dropshipping is an order fulfillment method that does not require an online store to keep products in stock. Instead, the business orders inventory from the wholesale supplier to ship directly to the customer [3]. While it can be a low-risk business model, there are many other factors of an online store to keep the business afloat. The benefit of dropshipping from a supplier perspective is that the seller produces the marketing material for their products. From a customer perspective, the dropshipping website is designed to provide a curated unique experience and exceptional customer support.

Suppliers and products are found on Aliexpress, a B2C e-commerce marketplace based in China that sells items at wholesale prices without requiring customers to purchase at wholesale quantities. One of the advantages of using Aliexpress is overall lower costs compared to brick-and-mortar retail stores. However, the largest disadvantage of using Aliexpress is the long shipping times of 2 to 3 weeks and lack of quality control. When selecting the products

Figure 1.1: Dropshipping Business Model [1]

for the dropshipping store, supplier ratings were carefully reviewed to ensure the quality of the product and supplier reliability. The dropshipping business is in close contact with the suppliers to assist with logistic efficiency.

An online software, Shopify, was utilized for website hosting and building. Shopify is an e-commerce platform that offers entrepreneurs a quick way to open their online store. It provides the necessary tools to customize the website and manage products, inventory, payments and shipping [4]. Shopify also has reporting features to observe website traffic, sales, and even customer behavior.

In this experiment, pomteahouse.com is used as the dropshipping store front. The small business, Pom Tea House, was founded in 2021 and falls under the drinkware niche. The target audience of Pom Tea House is a younger demographic of females who have a strong interest in whimsical home decor products. While the selected demographic has its limitations, its niche design adds emphasis on the unique user experience with the business. At the start of the experiment, there were more than 50 products available for purchase and an active 20% off all items discount code was made available.

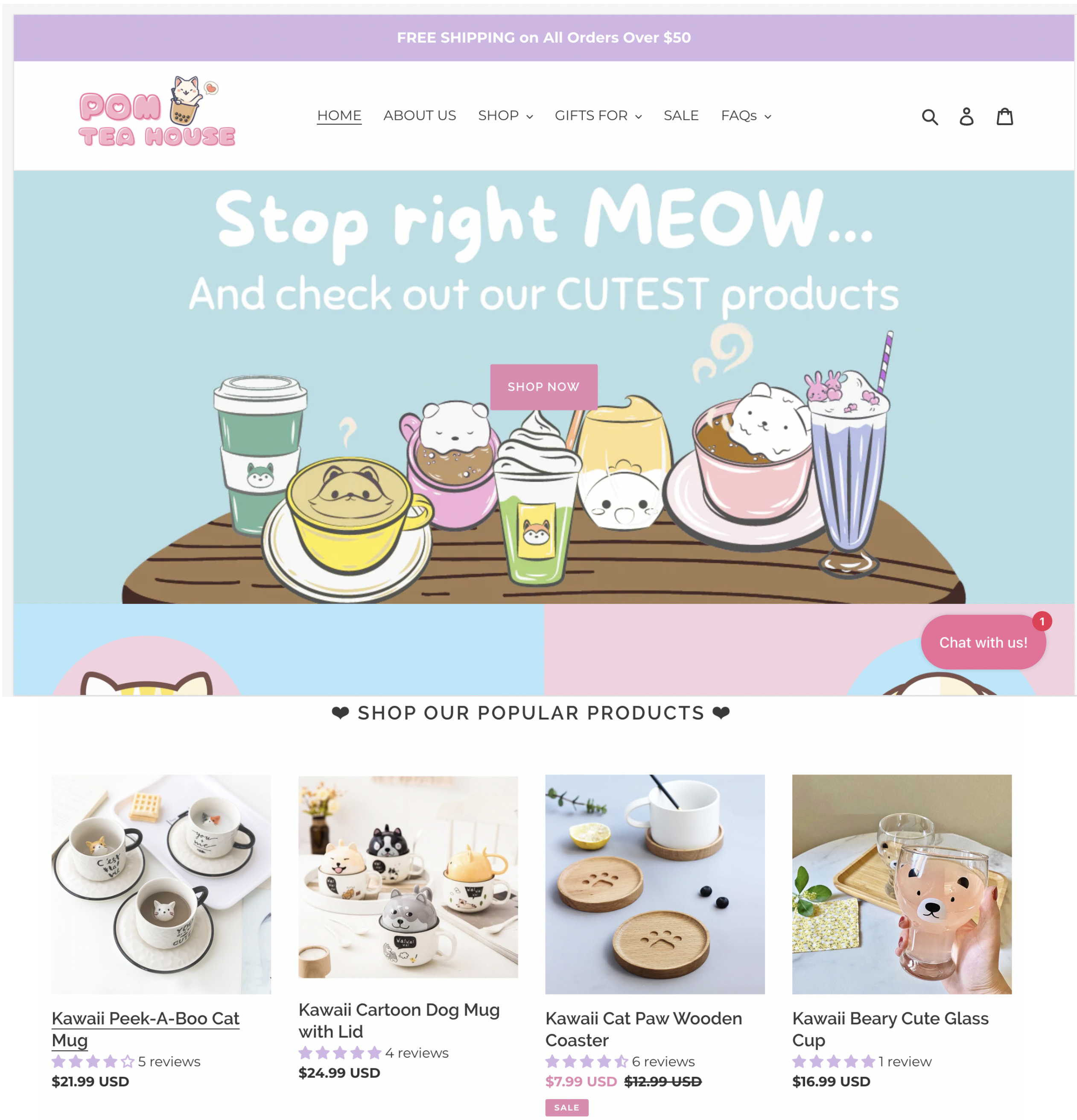

Figure 1.2: Dropshipping Business Website Storefront and Products

1.3 Online Advertising Platforms

This experiment will focus on two of the top advertising platforms, Google Ads and Facebook Ads. Both platforms have the option of creating display ads, which allows us to use picture and/or video formats. While Google leads in the search engine ads, Facebook is dominating the world of display ads [5]. To choose which platform would be the best option, it is important to determine the business goals and how each platform can help achieve them.

1.3.1 Facebook Ads

Facebook ads target their 2.89 billion monthly active users [6] on Facebook, Facebook Messenger, and Instagram. Facebook ads are managed through the Meta Business Suite which allows for campaign creation and content planning across the platform. Facebook leverages individual user behavior to bring forward the most relevant and interesting feed to the user. A target audience can be set to be part of the user's feed. This setting can be based off a variety of categories, including gender, age, and even upcoming life events. If the business goal is to grow brand awareness, Facebook's targeting capabilities would be the best fit.

Facebook display ads were set to maximize the volume of page views and obtain the highest number of clicks. A Pom Tea House Facebook page and Instagram account was created to add validity to the business. During the period of the experiment, Instagram posts of the products and other related interests were created daily in order to attract the target audience.

1.3.2 Google Ads

Google ads target users on YouTube and partners within their Search Network. Google reports lower than Facebook's monthly users at 272.4 million [6], but the platform also receives over 5 billion searches every day [7]. Google prioritizes ads based on its relevance to a user's searched keywords. If the business goal is to reach customers with high purchase intent, Google would be the winning option.

Performance Max campaigns are a type of display ads that incorporates automation technology to optimize performance in real time. It is the best option if a business wants to maximize campaign performance and access all of Google's advertising channels [8]. The Google ads were set to maximize conversions, which include page views and shopping cart checkouts. Due to this setting, the ad algorithm was not consistent with the daily budget. Because it was in the early stages of business, the online store was unable to reach the top of Google searches, which could affect the algorithm's output.

CHAPTER 2

Methodology

2.1 Full Factorial Design at Two Levels

It is common for marketing research to perform A/B testing, also known as split testing, which compares two versions of an ad against each other to observe which version performs better. The limitations with this method is that in order to determine which factor(s) are affecting the version, comparisons can only be made one at a time. The "one at a time" method can be very time consuming and costly if multiple factors are to be experimented with. The solution to make this approach more efficient is to integrate a factorial design.

A full factorial design is an experiment that studies the effects of two or more factors. Each factor has only two levels represented by a "low" and "high" defined as -1/+1 respectively. The experiment tests all possible 2^k combinations and observe its effects on the response variable. To display the design in terms of factor levels, a planning matrix should be created to avoid confusion with the experiment factors and levels [9].

The key properties of this design provide many advantages for analysis. One of the most important properties is that the design must have orthogonality. This means that all level combinations of any two factors must appear in the same number of runs. This allows for factor effects to not be obscured by planned changes in other factors [10]. The design should also be reproducible, so that there is a wider inductive basis for the experiment conclusions. The runs can be either replicated or unreplicated, but having replication will help reduce any noise variation. Another procedure that helps reduce those unwanted effects is randomization. This means that the actual run order should be different than the order of runs in the design matrix [9]. The symmetry of a full factorial design allows each response variable and interaction effect to be analyzed. This leads to solid conclusions that are valid over a range of different factors and levels [9]. Since there are only two levels for each factor, we can assume that the response is linear over the range of factor levels. To easily formulate and test a hypothesis about the factorial effects, the averages and variation of the responses should be considered. This can be done by generating the analysis of variance model also known as ANOVA [11].

The general formula for a three-factor analysis of variance model can be written as:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl} \tag{2.1}$$

where μ is the overall mean effect, α_i is the ith level of factor A, β_j is the jth level of factor B, γ_k is the kth level of factor C, $(\alpha\beta)_{ij}$ is the interaction effect between α_i and β_j, $(\alpha\gamma)_{ik}$ is the interaction effect between α_i and γ_k, $(\beta\gamma)_{jk}$ is the interaction effect between β_j and γ_k, $(\alpha\beta\gamma)_{ijk}$ is the interaction effect between α_i, β_j and γ_k, and ϵ_{ijkl} is the random error component [11].

2.1.1 Blocking in a Factorial Design

To reduce the influence of nuisance factors in an experiment, blocking is introduced to the design. A block is a group of homogeneous units, such as days, weeks, batches, or gender [9], that can be incorporated into the design to improve precision in comparing factorial effects. Within the block, the order in which the treatment combinations are ran is randomized. This allows us to assume that the interaction between blocks and treatments are negligible [11]. The experiment becomes more efficient so that block-to-block variation can be accounted for and eliminated.

The effects model of a two-factor design with a block can be defined as [11]

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_k + \epsilon_{ijk} \tag{2.2}$$

where μ is the overall mean effect, α_i is the ith level of factor A, β_j is the jth level of factor B, $(\alpha\beta)_{ij}$ is the interaction effect between α_i and β_j, δ_k is the effect of the kth block, and ϵ_{ijk} is the random error component

2.1.2 Factorial Effects

Since the ANOVA does not illustrate the magnitude or direction of each factor, it is important to also analyze the factorial effect plots [11]. Factorial effects are the main effects and interaction effects of all orders. The main effect is defined as the change in response produced by a change in the level of the factor. The interaction effect is defined in terms of two conditional main effects, both being first-order effects. The conditional main effect measures the main effect of one factor that another is at the same level [9]. When an interaction effect is large, the main effects would appear impractical. Significant interactions often mask the significance of main effects.

The main effects plot shows the averages of all data points at each level of the factors. The main effect is the difference between two averages, as evidenced by the vertical change or regression slope. Interaction plots show the average of each combination of levels of two factors. The effect of one factor depends on the level of the other interacted factor [10]. The interaction is considered synergistic if the two conditional main effects have the same sign (+/+). If they have the opposite signs (+/-), then it is antagonistic. It can be insightful to switch roles of two factors to confirm any valuable information.

Since the design is orthogonal, all factorial effects have the same variance and can be compared directly. The least square estimates are half of the factorial effects. Therefore, the factorial effects can be determined by doubling the corresponding regression estimate calculated by the coded factor levels of the linear model in ANOVA [11].

To justify data analysis of factorial design theory, there are three fundamental principles for factorial effects [9]:

1. Effect Hierarchy Principle
    (a) Lower-order effects are more likely to be important than higher-order effects
    (b) Effects of the same order are equally likely to be important
2. Effect Sparsity Principle
    (a) The number of relatively important effects is small
3. Effect Heredity Principle
    (a) In order for an interaction to be significant, at least one of its parent main effects should be significant.

2.1.3 Regression Modeling

The results of the experiment can be expressed in terms of a regression model. If normality and homogeneity assumptions are met, then the model intercept is the average of all observations and the regression coefficients are one-half of the corresponding factor effect estimates [11]. The multiple R-squared measures the proportion of total variability explained by the model. The adjusted R-squared accounts for the number of factors in the model and measures how well it predicts the new data.

With the regression model, the model residuals are plot to visualize the differences between the observed and fitted values of the response variable [11]. Residual plots are checked to verify the model error assumptions of normality and homogeneity. When experimental errors follow a normal distribution, the calculated factorial effects that have negligible influence on the response should also be normally distributed with mean zero [10].

2.1.4 Advantages of Factorial Design

By designing the experiment as factorial, multiple factors can be analyzed without having to compare each “one-factor-at-a-time”. This method would require more experiment runs for the same precision in effect estimation [9]. For example, the average of two estimates of a main effect is just as precise as when producing that average using a single-factor [11]. Another disadvantage of the "one-factor-at-a-time" approach is that it cannot always estimate interactions [9] and often assumes its effects are negligible. This may lead to inaccurate conclusions that will not support the experiment objective and hypothesis. Therefore, a factorial design is necessary when there is a presence of interactions. It is a much more efficient and cost effective approach especially for those with limited capital for advertising.

CHAPTER 3

Experiment Design and Procedure

3.1 Factorial Design

Prior to the start of this experiment, practice ads were tested on the Facebook and Google Platforms. Google CTR results were generally lower than Facebook CTR by approximately two to three percent. In order to support a small budget, there is only one factor, the daily budget, that is observed in the Google ads. To be able to study some effects between Facebook and Google, the experiment is split into two different designs. The 2^3 design explores the effects of factors between Facebook ads only. The 2^2 design explores the effect of daily budget and the two platforms. Reference Table 3.1 for the factor levels and complete design matrix.

Table 3.1: Complete Design Matrix for Experiment

	Factor Levels			Facebook			Google		
	A	B	C	A	B	C	A	B	C
Ad Run	Format	Discount	Budget	Format	Discount	Budget	Format	Discount	Budget
1	Picture	No	$5	1	-1	1			
2	Picture	Yes	$5	1	1	1			
3	Picture	No	$10	1	-1	-1			
4	Picture	Yes	$10	1	1	-1			
5	Video	No	$5	-1	-1	1			
6	Video	Yes	$5	-1	1	1	-1	1	1
7	Video	No	$10	-1	-1	-1			
8	Video	Yes	$10	-1	1	-1	-1	1	-1

3.1.1 2^3 Factorial Design

The first factorial experiment is a 2^3 design, meaning there are three factors with two levels each. A block is also included to optimize results and reduce unwanted variations.

Table 3.2: Factor and Block Levels for 2^3 Design

Factor	(+) Level	(-) Level
Format	Picture	Video
Discount Included	Yes	No
Daily Budget	\$5	\$10
Block	Day of the Week	

With the three factors and their interactions, there are a total of eight ad treatments on the Facebook platform. The majority of the experiments will be tested on the Facebook platform which includes ad placements on Facebook Messenger and Instagram. There are two replicated sets of data by the day block factor, which results in the ads being ran for a minimum of 14 days. On the first day of experimentation, it is unclear if the ads began running in the morning or later in the day. To account for this, an additional day of results is included for analysis. Experimental order is listed by ad number. Randomization is being relied on by the Facebook display ad algorithm so randomized order is not recorded each day. Since the data is replicated and randomized, this is considered a complete block design.

3.1.2 2^2 Factorial Design

To be able to compare ads between platforms, the observed factors are platform and budget. The format and discount factors are still be part of the ads, but remain static for this 2^2 design. Therefore, the only ad treatments for this design include Ad #6 and #8, where both are in video format with an included discount. The daily budget for each treatment is \$5 for Ad #6 and \$10 for Ad #8. Similarly to the previous design, the ads are ran on both Facebook and Google for at least 14 days, resulting in two replicated sets of data by the day block factor. Since the ads are assumed to be randomized by the platform, this is also a complete block design.

Table 3.3: Factor and Block Levels for 2^2 Design

Factor	(+) Level	(-) Level
Platform	Facebook	Google
Daily Budget	\$5	\$10
Block	Day of the Week	

3.2 Data Collection Procedure

To collect the data, multiple reports from Shopify, Meta Business Suite, and Google Ads were generated and exported. The collected data included the number of impressions, number of clicks, cost per day, click through rate, cost per click, and conversions. The data variables were extracted from each reporting platform and consolidated into one data set. Due to the different ad settings for conversions on Facebook and Google, number of conversions is not included in the analysis. Because of the small number of factors of the experiment, data collection was manageable with little cleaning. Impressions are defined as the number of views of the ad. Clicks are counted if the viewer clicks on the ad. Daily cost is the amount of the daily budget spent for that ad and day. During the second week of experimenting, Google's Performance Max Display Ads adjusted the daily cost from \$0 on some days to \$20 which caused concern in data variation.

3.3 Response Variables

The response variables for both experiments are CTR (clickthrough rate) and CPC (cost per click). CTR is the number of clicks the ad received divided by the number of impressions. For optimal performance results, CTR should be high; this can indicate the ad's interest and engagement with viewers. CPC can help gauge the quality and costs of keywords used in the ad settings. A lower CPC means more clicks for a lower unit price. Depending on the ad campaign goals, this can be of benefit, especially if the volume of clicks is to be maximized instead of using more expensive keywords. Ad settings on Facebook and Google were set to maximize click-rate conversions.

3.4 Hypothesis

For the 2^3 factorial design, it is hypothesized that factors with the highest CTR and lowest CPC would be in video format, display the 20% discount, and require the higher daily budget of $10. These assumptions were made based on previous subject research. For the 2^2 factorial design, it was already observed prior to this experiment that Google Ads has a lower CTR than Facebook Ads. Therefore, we assume Facebook will perform significantly better than Google. However, we also hypothesize that the higher daily budget will improve the Google ads compared to the lower budget.

3.5 Creating Ad Content

A total of four ads were created prior to setting up the budget in the platforms. Other possible ad variables, such as colors, fonts, and featured products were controlled and kept the same on all ads to minimize confounding effects on the response.

- Picture with no discount noted (Figure 3.1a)
- Picture format with a discount added (Figure 3.1b)
- Video slideshow ad with no discount or sale shown in the beginning or ending slides
- Video slideshow ad with discount shown in the beginning and ending slides

(a) Picture Ad with No Discount

(b) Picture Ad with Discount

Figure 3.1: Ad Designs for Experiment

Video ads were designed similarly to the picture format with the exception of animated graphics. The video was also structured like a slideshow, which featured additional products and complimentary color backdrops. Since video ads shorter than 15 seconds tend to have higher engagement rate compared to longer ones [2], the duration of the video ads was 12 to 15 seconds, in order to maximize clicks.

CHAPTER 4

Data Analysis and Results

4.1 Exploratory Analysis

The experiment ran eight ads on Facebook and two ads on Google for 15 days, totaling 150 ad runs. As previously noted in Chapter 3, a fifteenth day of results was included in the data set to ensure a completed second set of replicates. After combining the platforms, there were 291,881 impressions and 7,957 clicks. When looking at the raw number of clicks each factor received (Figure 4.1), results appear to support the hypothesis that video format, displayed discount, and higher budget lead to better performance. However, the number of clicks do not yet account for the actual number of impressions, like CTR does, so this first observation can be misleading.

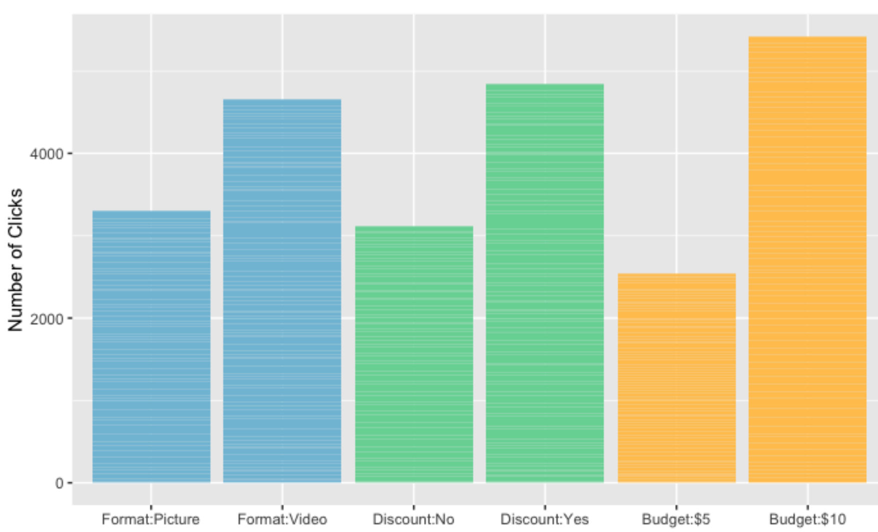

Figure 4.1: Bar Plot of Clicks by Factor

Table 4.1 shows the calculated averages of the CTR and CPC response variables by factor level, ad treatment, and day of the week. Since the Google ads only varied in the budget factor, there are no averages for the picture format and displayed discount ads. Overall, Facebook CTR is higher than Google CTR while the CPC values are similar in range. This gives us some evidence that our hypothesis is true regarding the better CTR-performing platform. Since the CPC averages for both platforms are similar, additional analysis is required to understand its effect.

Table 4.1: Summary of Average Responses

	Facebook		**Google**	
	Avg CTR	Avg CPC	Avg CTR	Avg CPC
Factor: Level				
Factor: Picture	4.511	0.143		
Factor: Video	5.075	0.152	1.978	0.157
Discount: Yes	4.690	0.145	1.978	0.157
Discount: No	4.897	0.150		
Budget: $5	4.788	0.156	1.169	0.140
Budget: $10	4.799	0.139	2.787	0.173
Ad Treatment				
#1 Pic/No/$5	4.475	0.153		
#2 Pic/Yes/$5	4.513	0.149		
#3 Pic/No/$10	4.557	0.147		
#4 Pic/Yes/$10	4.501	0.122		
#5 Vid/No/$5	5.367	0.151		
#6 Vid/Yes/$5	4.796	0.170	1.169	0.140
#7 Vid/No/$10	5.189	0.146		
#8 Vid/Yes/$10	4.949	0.140	2.787	0.173
Day of the Week				
Sunday	5.237	0.119	1.595	0.133
Monday	4.461	0.135	1.555	0.133
Tuesday	4.489	0.147	4.563	0.040
Wednesday	4.557	0.149	2.320	0.170
Thursday	4.355	0.161	0.930	0.185
Friday	4.802	0.168	1.692	0.200
Saturday	5.648	0.143	1.335	0.215

To further visualize the data and its variation, box plots were generated by ad treatment and day of the week. With these box plots, the overall data distribution and outliers become evident. The box plots by ad treatment (Figure 4.2) illustrate how Facebook CTR averages are overall higher than Google's which fits the hypothesis. The highest CTR were the ads in video format that had no discount. The budget difference from Ad #5 and Ad #7 shows that the lower $5 daily budget produced better CTR results.

While the Facebook and Google CPC medians appear similar by ad treatment, the box plots show the Google CPC range is much wider than Facebook CPC. This large difference could be due to Google's new Performance Max algorithm, which fluctuated the daily budget throughout the second week of experimenting. The box plots also detect more outliers in Facebook CPC compared to Google's. These outliers may have a strong influence in estimating the regression coefficients and result in a poorly fitted model. Therefore, Facebook CPC may be overall more precise, but the three outliers would have more impact on the model's fit than the wide Google CPC range.

The box plots by day of the week (Figure 4.3) show an increase in Facebook CTR and decrease in Facebook CPC on "fri", "sat", and "sun". This suggests that social media users are more likely to click on ads, making it more cost effective, over the weekend. The ranges of the Google CTR and CPC box plots vary greatly by day with multiple outliers. In fact, outliers are detected for all four box plots by day. This could result in misleading conclusions, regarding the best day of the week for ads. Therefore, trying to fit a model to the CPC may not be reliable or impactful.

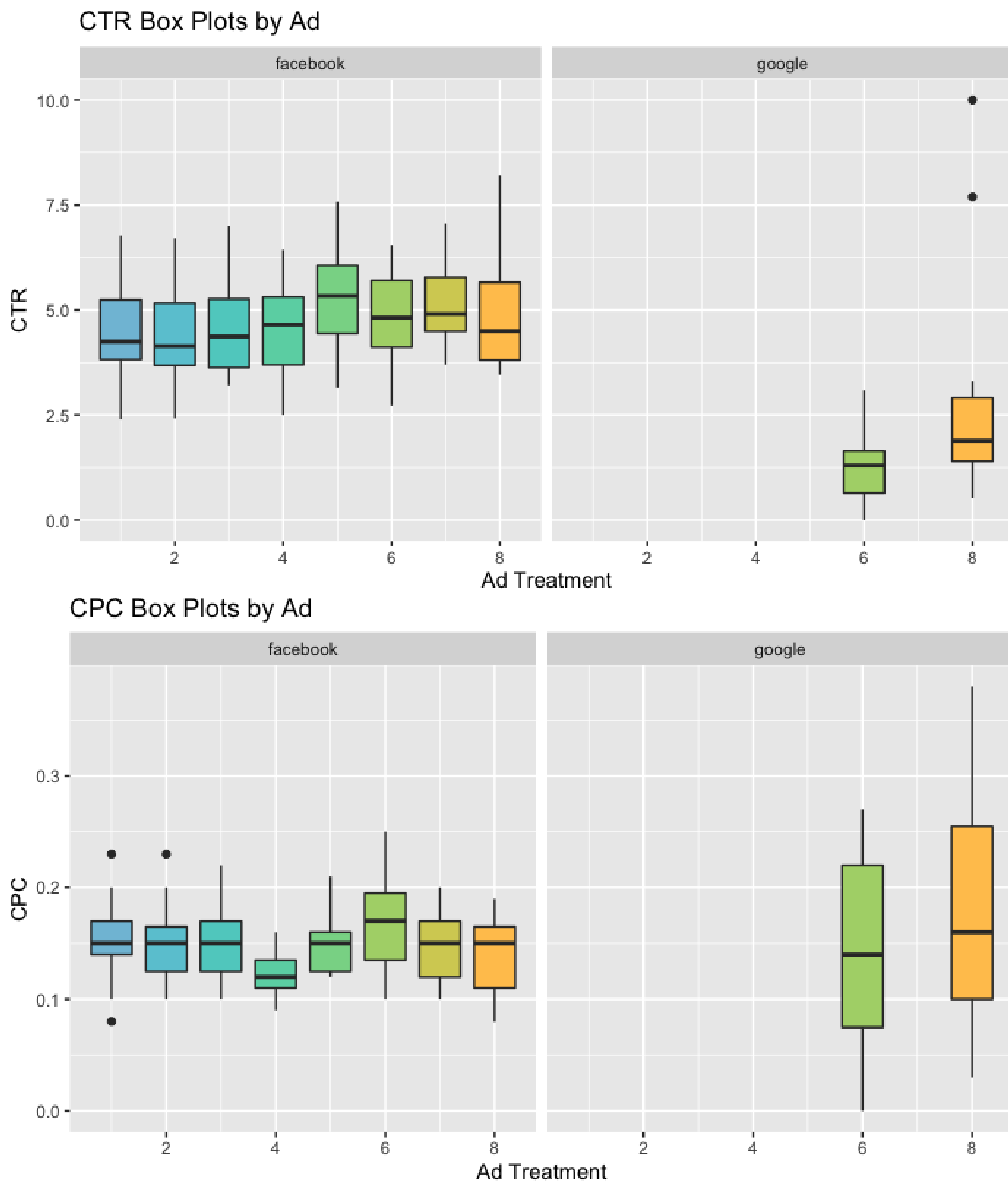

Figure 4.2: Box Plots of Response Variables by Ad Treatment

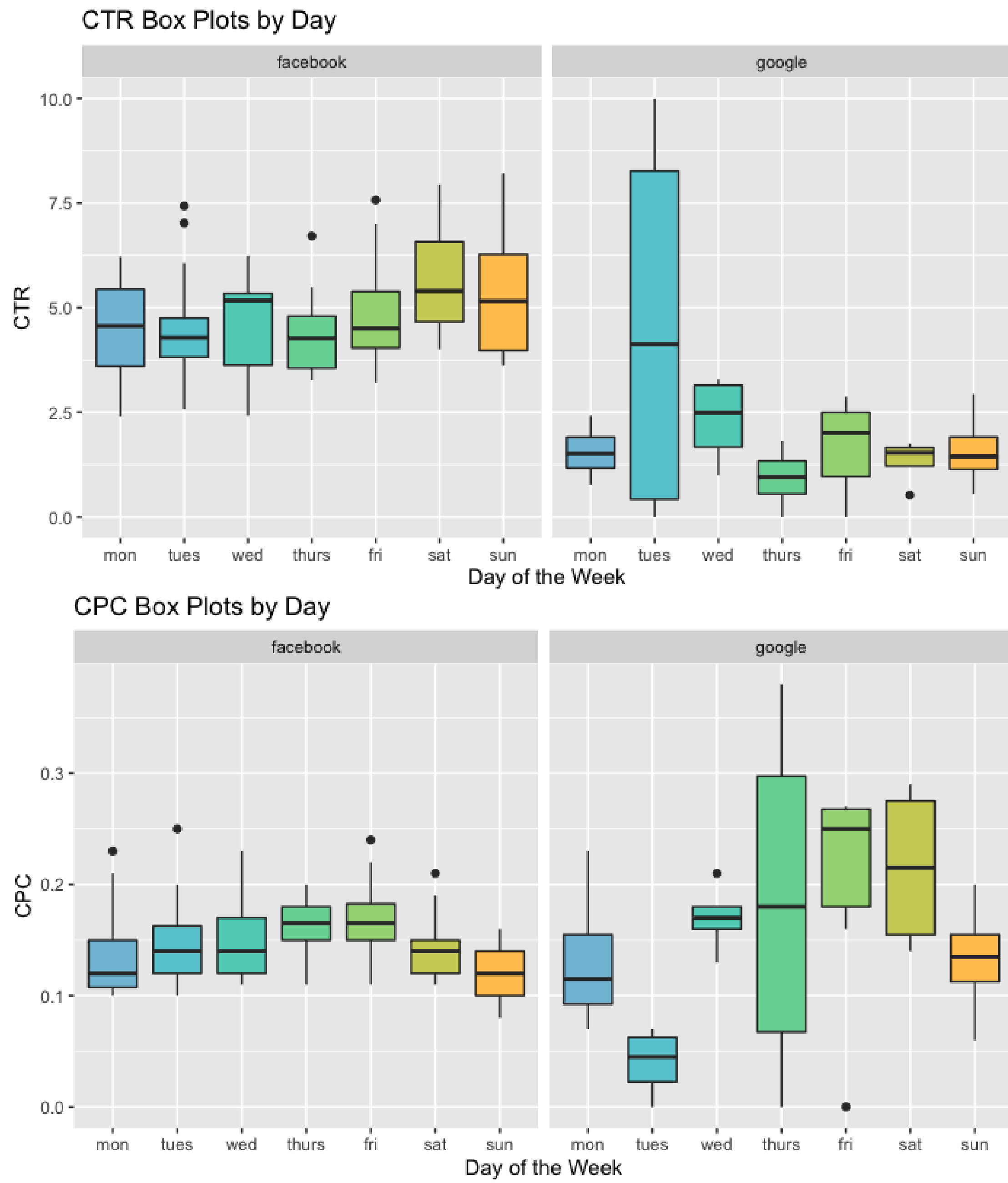

Figure 4.3: Box Plots of Response Variables by Day of the Week

Two-sample t-tests were performed by factor for each platform. The t-test results by factor are summarized in Table 4.2. A p-value less than 0.05 is considered statistically significant, which means there is strong evidence in favor of the hypothesis. For Facebook ads, the CTR format and CPC budget were significant. It follows that, for Facebook, more consideration should be taken with the format and daily budget when creating the ads; however, applying a discount does not have the same weight of impact. For Google, only the CTR budget was significant, so having enough marketing funds would be important when considering Google ads. When combining both, the CTR discount is significant, while the CPC p-values did not have any significance. Therefore, the significance of the discount factor supports the hypothesis of its effect on CTR between platforms, but that cannot be concluded with the CPC response.

Table 4.2: Summary of t.Tests by Factor

	Combined		Facebook		Google	
Factor	CTR p-value	CPC p-value	CTR p-value	CPC p-value	CTR p-value	CPC p-value
Format	0.0867	0.1927	0.0118	0.1686		
Discount	4.58e-05	0.9611	0.3604	0.5174		
Budget	0.2682	0.4219	0.9614	0.0068	0.0379	0.3627

4.2 2^3 Design Results - Comparing Ad Factors

The main effect plots, shown in Figure 4.4, are reviewed in order to note any possible significant factors. By examining the magnitude and direction of each main effect against the average CTR and CPC, the efficacy with which a factor impacts the response variables is revealed. For both CTR and CPC, video format averages higher than picture format, which supports the hypothesis on the CTR response. The budget factor makes little change on CTR but has a larger magnitude on CPC, implying possible significant effect. The slope of the discount main effect is similar for both CTR and CPC. The interaction plots, shown in

Figure 4.5, shed light on the relationship between each factor and its interaction effect on the response variables. For both CTR and CPC, the interaction between discount and budget, as well as the interaction between format and discount, may be significant, but would need additional analysis to confirm.

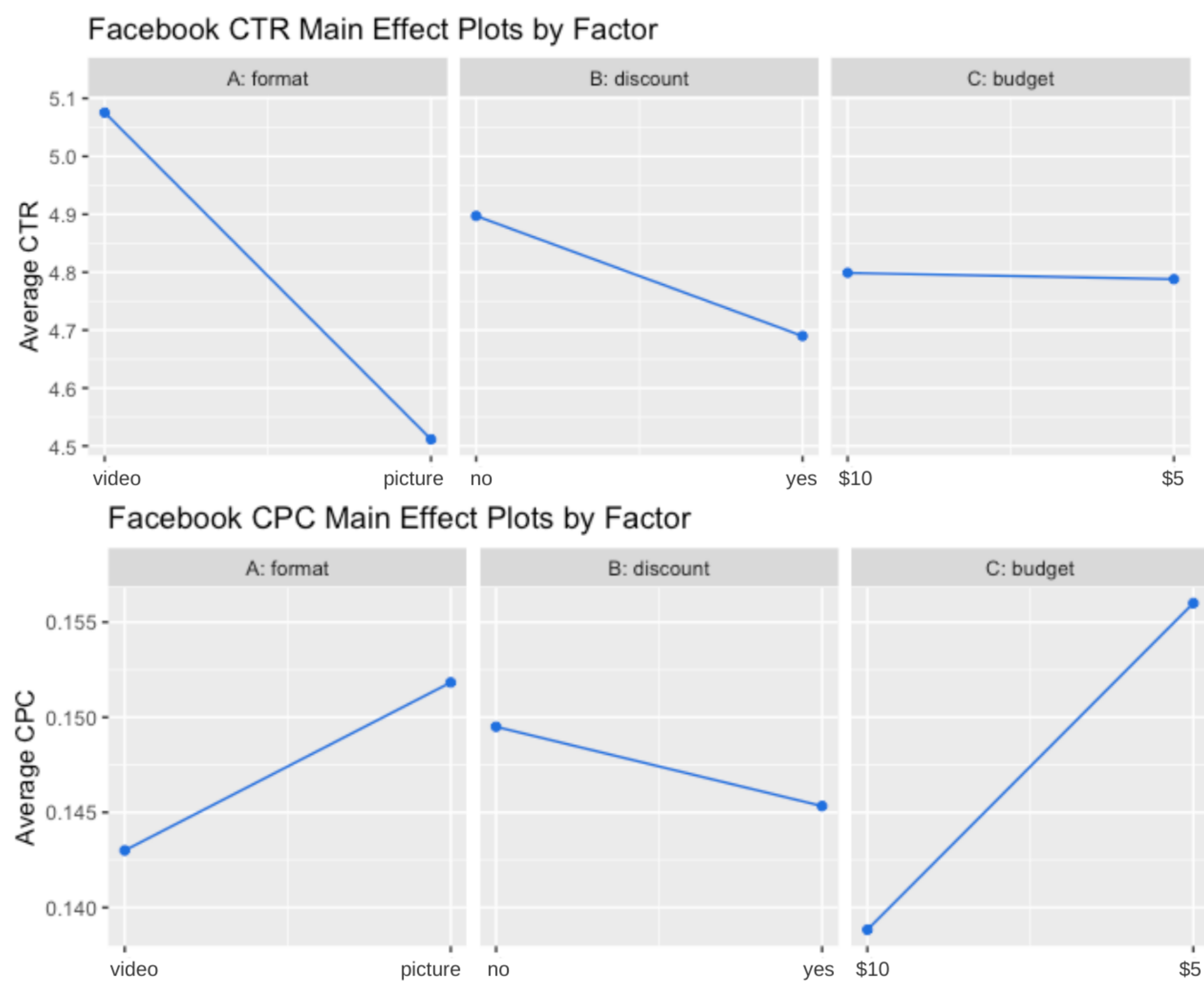

Figure 4.4: Main Effects Plots of Facebook Response Variables

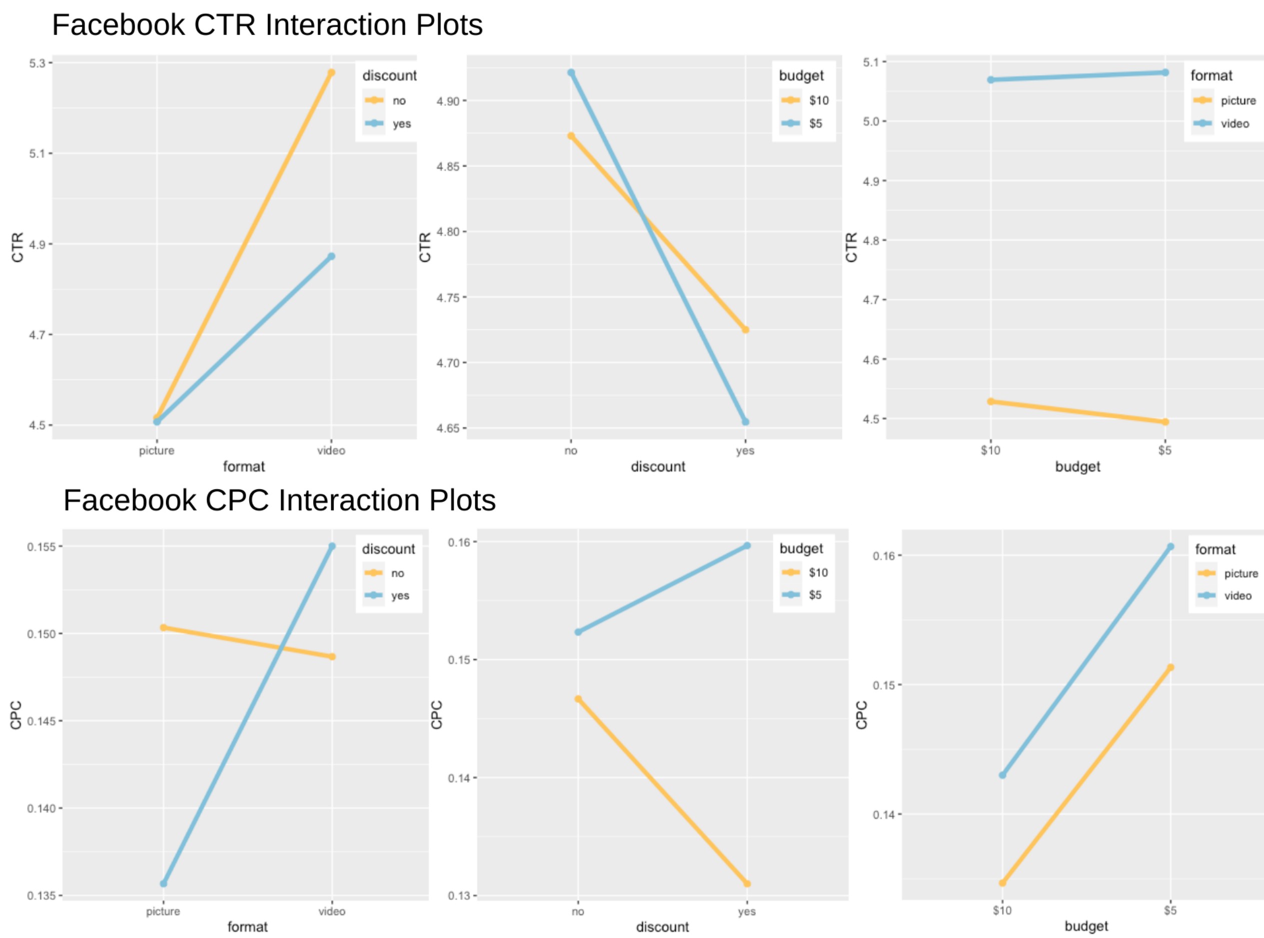

Figure 4.5: Interaction Plots Facebook Response Variables

4.2.1 CTR Model

The main effect and interaction plots show that there may be significant effects on CTR with the format main effect as well as the interaction between discount and budget. To better interpret these initial findings, a regression model is developed for Facebook CTR.

For this first model, the main effects, interaction effects, and blocking factor are all included to determine which factorial effects influence CTR. This model will be referred to as “model1A”. The ANOVA table for model1A is shown in Table 4.3. The last right column displays the p-value for all of the main and interaction effects of the model. An effect with

Table 4.3: ANOVA Facebook CTR Full Model1A

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	9.5429	9.5429	6.8690	0.0101
discount	1	1.2896	1.2896	0.9283	0.3375
budget	1	0.0036	0.0036	0.0026	0.9593
day	6	22.0606	3.6768	2.6466	0.0197
format:discount	1	1.1801	1.1801	0.8494	0.3588
format:budget	1	0.0163	0.0163	0.0118	0.9139
discount:budget	1	0.1056	0.1056	0.0760	0.7833
format:discount:budget	1	0.3392	0.3392	0.2442	0.6222
Residuals	106	147.2614	1.3893		

a p-value of less than 0.05 is considered statistically significant. From the p-value results, it is confirmed that format and day are significant. The full model1A is also significant, which supports the hypothesis. The adjusted R-squared shows how strongly the model explains the variation within the data. Since the results are low with an adjusted R-squared of 0.091, additional steps are taken to improve the fit of the model.

To improve the CTR model, model1A is simplified by removing the insignificant factors. By excluding the insignificant factors, there are less considerations for the model to make. Only significant factors, which include format and day, are used to fit a new "model1B" and its ANOVA results are generated in Table 4.4. Model1B is also significant with a p-value less than 0.05 and results in an increased adjusted R-squared of 0.1222. Although there is some improvement in the adjusted R-squared, this suggests that only 12.22% of the variation is explained by the model. There may be a better model to fit the Facebook CTR data than model1B. The residual plots, though, will confirm if the model fit meets the linearity and constant variance assumptions or not.

Figure 4.6 compares the residual versus predicted values and normal Q-Q plots of the first full model1A and reduced model1B. The Q-Q plot for both models appear similar and generally remain in the gray region. The residual versus predicted plot of model1A shows more points clustered towards the small predicted values, whereas the reduced model1B

Table 4.4: ANOVA Facebook CTR Reduced Model1B

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	9.5429	9.5429	7.1161	0.0088
day	6	22.0606	3.6768	2.7417	0.0160
Residuals	112	150.1959	1.3410		

shows a better even spread of the residuals along the x-axis. The Q-Q plot for both models appear similar and generally remain in the gray region. Both Q-Q plots show skewness at opposite ends, so it is unclear if the normality assumption is met.

In order to be certain about the models meeting the assumptions of normal distributions, diagnostic tests should be performed. The Breusch-Pagan test checks if the residuals have constant variance about the true model. If the resulting p-value is greater than 0.05, then the null hypothesis, that the model meets the constant variance assumption, is accepted. The Shapiro-Wilk test checks if the data has been sampled by a normal distribution or not. If the resulting p-value is greater than 0.05, then the null hypothesis, that there is normality within the model data, can be accepted. After performing these tests on model1B, it is confirmed that the homoscedasticity assumption passes but the model does not meet normality.

Next, the model is transformed using the square root method to help improve the normality of the data. The CTR response variable is square rooted and a new "model1C" is fitted with the same predictors as model1B. From the ANOVA results (Table 4.5), the model is still significant and has the same adjusted R-squared as model1B. When performing the previous two diagnostic tests, the null hypothesis is accepted for both, so model1C meets homoscedasticity and normality. Therefore, model1C is the better fit to predict format effects on Facebook CTR.

Table 4.5: ANOVA Facebook CTR Transformed Model1C

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	0.5015	0.5015	7.2163	0.0083
day	6	1.1338	0.1890	2.7193	0.0167
Residuals	112	7.7828	0.0695		

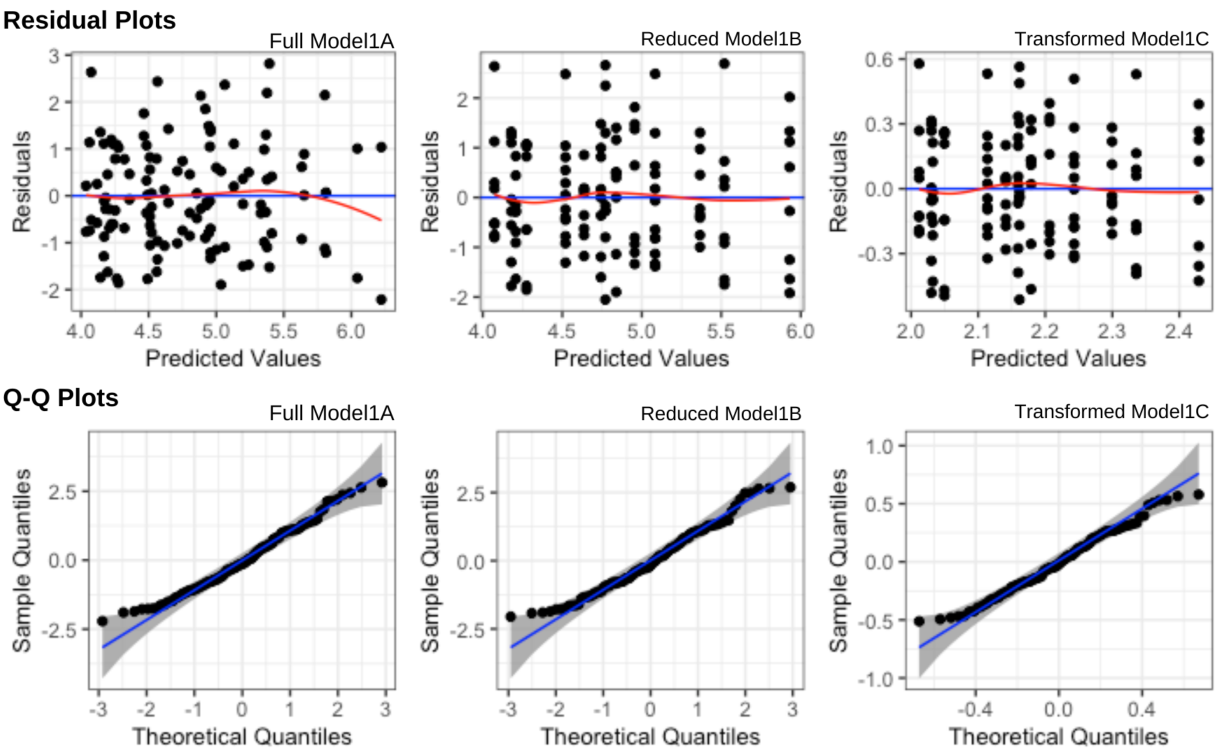

Figure 4.6: Residual Plots Comparison of Facebook CTR Models

4.2.2 CPC Model

Since the CPC main effect plots show all three factors having some impact on the average CPC and the interaction plot shows an intersection between format and discount, a new linear model is fit to the Facebook CPC data. Fitting a model will help better explain the relationship between CPC, the response variable, and the ad factors, the predictors.

Similarly to the previous modeling steps, the main effects, interaction effects, and blocking factor are included to determine which factorial effects influence CPC, which will be referred to as "model2A". The ANOVA results for model2A are shown in Table 4.6.

Table 4.6: ANOVA Facebook CPC Full Model2A

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	0.0023	0.0023	2.5019	0.1167
discount	1	0.0005	0.0005	0.5567	0.4573
budget	1	0.0088	0.0088	9.4492	0.0027
day	6	0.0281	0.0047	5.0088	0.0001
format:discount	1	0.0033	0.0033	3.5351	0.0628
format:budget	1	8.00e-06	7.50e-06	0.0080	0.9288
discount:budget	1	0.0039	0.0039	4.2405	0.0419
format:discount:budget	1	2.10e-05	2.08e-05	0.0223	0.8817
Residuals	106	0.0992	0.0009		

From the p-value results, it is confirmed that budget, day, and discount*budget effects are significant at the 5% level. The format*discount interaction has a close p-value of 0.063, so for modeling purposes, this interaction will be considered significant as well. The p-value for model2A is significant, so the model adds support to the hypothesis. The R-squared is 0.3221 and adjusted R-squared of model2A is 0.239, which indicates that the model does not strongly explain the variation within the data. There is certainly room for improvement with the fit of this initial full model.

For the new "model2B", the insignificant factors from model2A are eliminated to simplify the CPC model. Significant factors, budget, day, discount*budget, and format*discount, are included in the new model. Although main effects, format and discount, were not significant, the interactions between the two factors are. Therefore, both will remain in the new reduced model. Table 4.7 shows the ANOVA results of the new linear model. Model2B is also significant with a p-value less than 0.05 and results in an increased adjusted R-squared of 0.2528, which shows improvement.

To determine next steps, residual plots are used to assess if the model follows a normal distribution. From the residual plot comparison in Figure 4.8, the model2B residuals appear

Table 4.7: ANOVA Facebook CPC Reduced Model2B

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	0.0023	0.0023	2.5484	0.1133
discount	1	0.0005	0.0005	0.5670	0.4531
budget	1	0.0088	0.0088	9.6247	0.0025
day	6	0.0281	0.0047	5.1018	0.0001
format:discount	1	0.0033	0.0033	3.6008	0.0604
discount:budget	1	0.0040	0.0040	4.3193	0.0401
Residuals	108	0.0992	0.0009		

to be spread out more with larger fitted values. This impacts the normal distribution's assumption of constant variance. The Normal Q-Q plot also shows the points at which the highest and lowest quantiles begin to stray from the line, which can indicate there are more outliers than expected in the data. Since the residual plots show that model2B does not meet the assumptions of normal distribution, performing a box-cox transformation may help further improve the model. This type of model transformation helps correct the non-linearity of the data so that the distribution will be closer to normality.

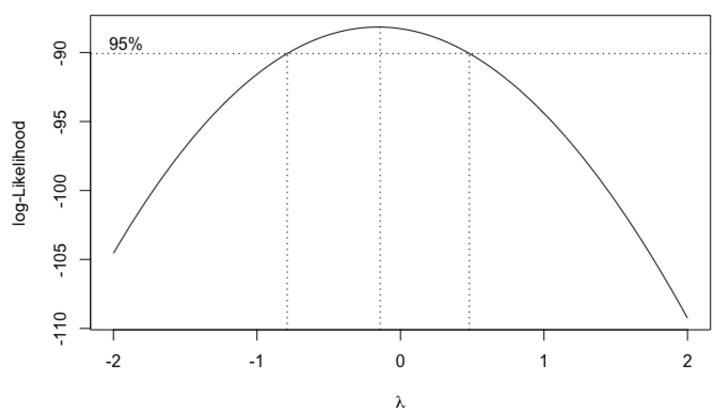

Figure 4.7: Lambda Output of Box-Cox Transformation

The box-cox function finds a transformation parameter, lambda, that can be utilized in the model to adjust its shape. The function calculates lambda to be optimal at -0.141 and is plot against the log-likelihood as shown in Figure 4.7. Since the plot shows that lambda zero is also within the 95% confidence interval of the box-cox transformation plot, a log transformation is also acceptable. By definition, as lambda approaches zero, then the box-cox equation is negligible and essentially becomes the logarithm function.

After fitting the new transformed "model2C", there is improvement in the residual plots as shown in Figure 4.8. The residuals appear more spread out and lack a clear pattern. The Q-Q plot shows the transformed data points closer to the straight blue line, compared to the previous Q-Q plot of model2B. The residual plots improved and appear to better fit the normality and constant variance assumptions. From the ANOVA Table 4.8 results, the p-value remains significant, the R-squared value increased to 0.3452, and the adjusted R-squared value increased to 0.2785.

Since the transformed model2C is significant, has improvement in its fit from the original, and the residual plots support normality, this third model2C for Facebook CPC is better than the first full model2A and reduced model2B.

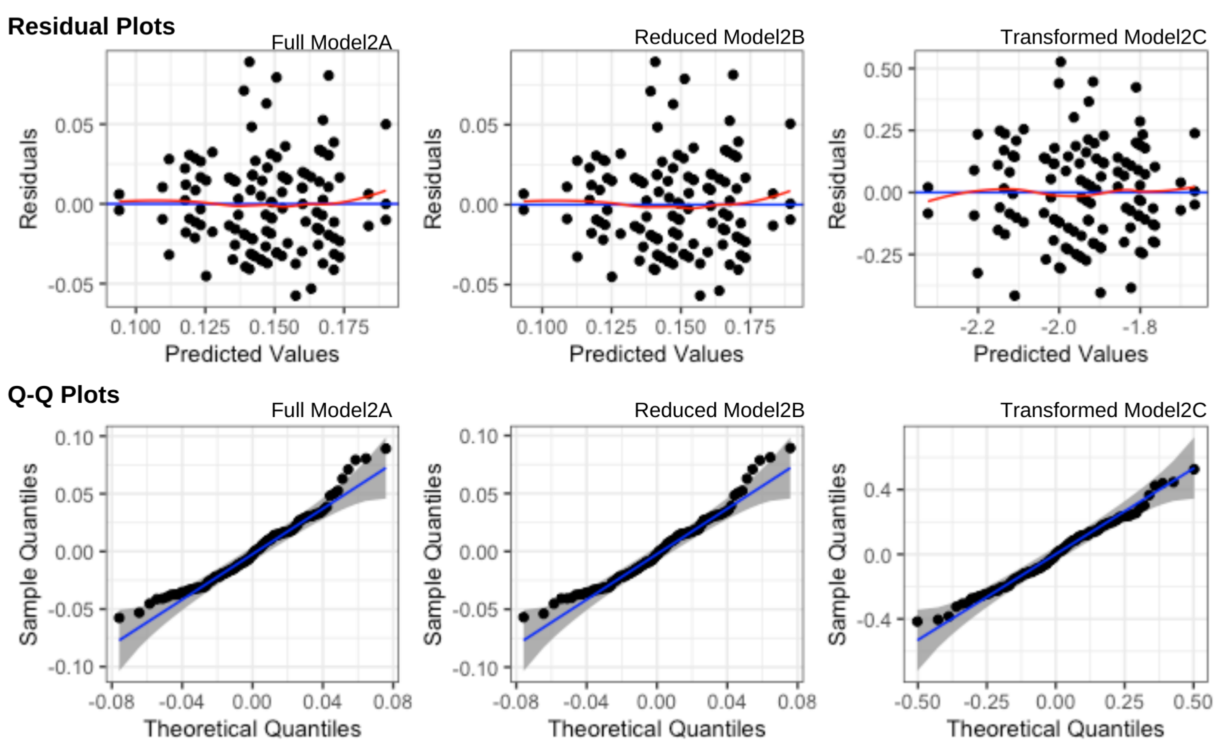

Figure 4.8: Residual Plots Comparison of Facebook CPC Models

Table 4.8: ANOVA Facebook CPC Transformed Model2C

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
format	1	0.1040	0.1040	2.5785	0.1112
discount	1	0.0414	0.0414	1.0273	0.3130
budget	1	0.3839	0.3839	9.5211	0.0026
day	6	1.4698	0.2450	6.0758	0.0000
format:discount	1	0.1208	0.1208	2.9968	0.0863
discount:budget	1	0.1753	0.1753	4.3468	0.0394
Residuals	108	4.3545	0.0403		

4.3 2^2 Design Results - Comparing Ad Platforms

In order to compare Facebook and Google, the platform variable is now treated as a factor. The second and last factor of this 2^2 design is the daily budget. To maximize performance and support the hypothesis, video format and displayed discount were chosen as the static factor levels. Therefore, only data from Ad #6 and #8 are observed in this analysis, resulting in a much smaller sample size compared to the previous 2^3 design.

Since there are only two factors (factor and budget) involved for the 2^2 design, observing the interaction plots will suffice for initial insights. Platform and budget behave inversely when comparing CTR to CPC. The interaction plots (Figure 4.9) show a synergistic relationship for CTR but an antagonistic one for CPC. However, formal tests need to be performed to determine whether the interactions are significant. Increase in budget does not affect Facebook CTR compared to Google CTR. However, the larger magnitude in CPC interaction plot could be explained by the Google Performance Max learning algorithm. Since the daily budget spent for the Google ads varied day-to-day during the second week of the experiment, the data contains many outliers. In order to predict a reliable response without this known nuisance variation, a model will be fit to CTR instead of CPC.

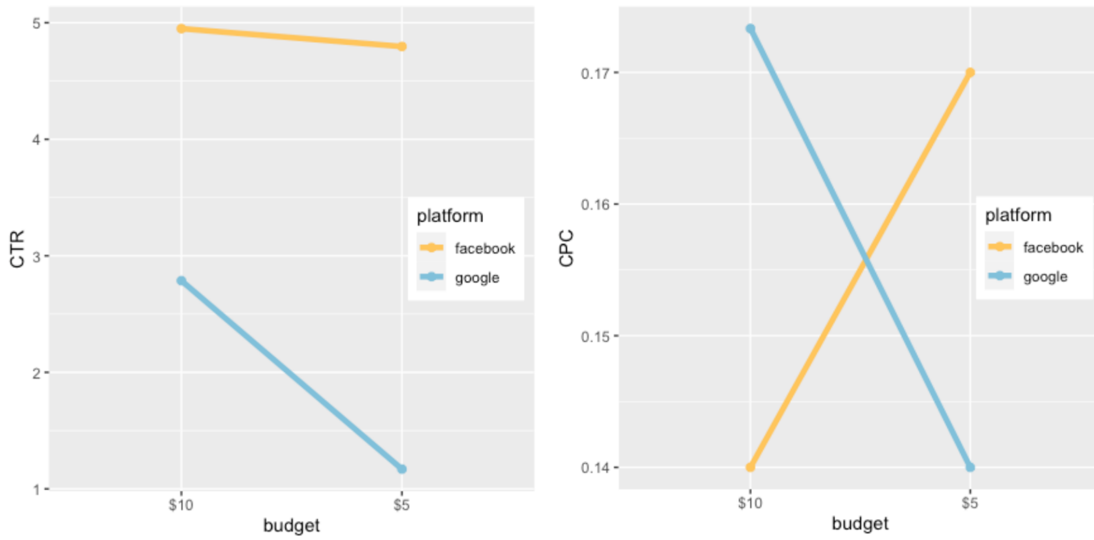

Figure 4.9: Interaction Plots Between Platform and Budget

Table 4.9: ANOVA Results of CTR Full Model3A

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
platform	1	125.6864	125.6864	44.2225	2.12e-08
budget	1	11.7572	11.7572	4.1368	0.0473
day	6	17.8175	2.9696	1.0448	0.4080
platform:budget	1	8.0374	8.0374	2.8279	0.0989
Residuals	50	142.1069	2.8421		

4.3.1 CTR Model

Since there are two sets of replicates for this design as well, regression is initiated by fitting the full model for CTR. The ANOVA results in Table 4.9 confirm that both main effects, platform and budget, are significant. At an alpha level of 10%, the interaction between platform and budget is considered statistically significant as well.

The residual plots (Figure 4.10) show there is some clustering of residuals near the baseline, but it is not clear if there is a pattern. The Q-Q plot looks overall acceptable with the exception of a few data points deviating from the fitted line in the upper right quadrant. When testing for constant variance and normality, the p-value fails to reject the null hy-

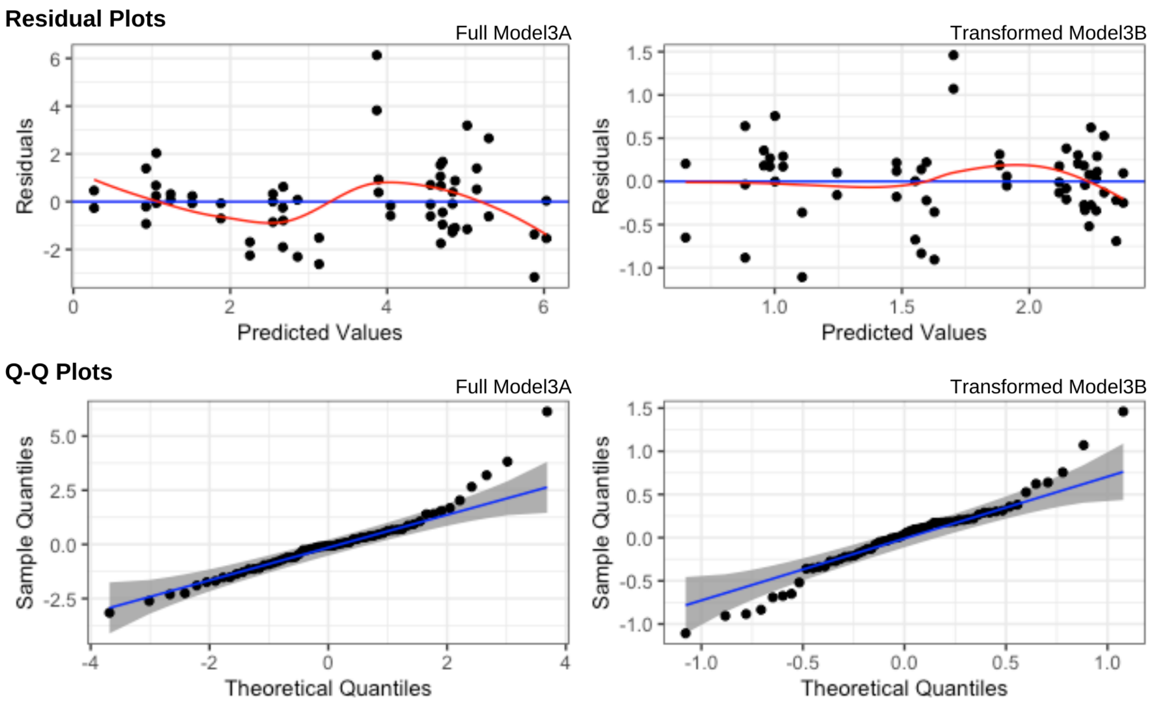

Figure 4.10: Residual Plots Comparison of Platform CTR Models

pothesis, so model3A validates the assumptions of normal distribution. In attempts to find a better model, multiple transformations were completed without success in finding a model that fulfills the normality distribution assumptions. However, it was discovered that a square root transformation improves the model fit by 10% and maintains significance (Table 4.10). Therefore, transformed model3B is the best choice to predict CTR response from platform and budget.

The test results and summary values for each model can be referenced in Table 4.11.

Table 4.10: ANOVA Results of CTR Transformed Model3B

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
platform	1	13.5352	13.5352	55.6570	1.15e-09
budget	1	1.4507	1.4507	5.9653	0.0182
day	6	1.0498	0.1750	0.7195	0.6358
platform:budget	1	1.2046	1.2046	4.9534	0.0306
Residuals	50	12.1595	0.2432		

Table 4.11: Summary Results of Model Diagnostic Tests

Description	**Model Name**	**ANOVA** p-value	**Adjusted** R-Squared	**Breusch-Pagan** p-value	**Shapiro-Wilk** p-value
Facebook CTR					
Full	model1A	0.0365	0.0906	0.8232	0.0966
Reduced	model1B	0.0027	0.1222	0.7126	0.0306
Sqrt Transformation	model1C	0.0027	0.1220	0.8056	0.2274
Facebook CPC					
Full	model2A	4.08e-05	0.2390	0.3761	0.0045
Reduced	model2B	8.43e-06	0.2528	0.2398	0.0044
Box-Cox Transformation	model2C	1.72e-06	0.2785	0.2810	0.3913
Platform + Budget CTR					
Full	model3A	5.64e-06	0.4509	0.0139	0.0018
Sqrt Transformation	model3B	4.00e-07	0.5120	0.0139	0.0525

CHAPTER 5

Conclusion

5.1 Summary

The linear models for CTR and CPC were all found to be significant, which indicates the studied factors of platform, format, discount, and budget do have substantial impact on ad performance. The hypothesis was supported, with video format being more effective than static pictures for higher CTR. Facebook is also the superior platform when it comes to CTR compared to Google. Nevertheless, the experiment's Google CTR is high compared to the retail industry average of 0.21% [12]. Regardless of that small victory, social media is an essential part of branding and marketing, so the results advise to run digital ads on the Facebook platform. Interactions between format*discount and discount*budget have a significant effect on Facebook CPC. Daily budget alone does not have much impact on Facebook metrics, so if a small business is limited in advertising budgeting, CTR and CPC can still reliably perform despite the lower budget.

5.2 Challenges

While there were useful insights provided from the results, there were also many challenges throughout the experiment.

5.2.1 Platform Settings

As previously mentioned, Performance Max display ads were set up on the Google platform. During the second week of the experiment, the learning algorithm adjusted the daily budget from $0 some days and $20 other days, making the budget factor not as consistent as originally designed. Although a block factor was included to account for day to day variation, there were still large outliers in the Google data possibly due to the large difference between certain days.

5.2.2 Small Sample Size in Number of Sales

During the time frame of the experiment, a total of 13 orders were fully converted and generated revenue of $700.09. Ad #2 and Ad #8 were listed as some of those referrals that led to conversion. Due to the small sample size of sales and lack of marketing campaign referral, sales analysis could not be completed. Tracking sales through the campaign was one of the initial intents of this experiment. However, campaign tracking abilities of both platforms were found too unreliable for analysis. Sales from campaigns can be tracked by attaching tags in the ad clicks. However, what often happens is that a customer may initially see an ad on the Instagram platform. Then instead of clicking the ad, they go directly to the profile page, where a direct website link is located. Though the customer may have been acquired through that ad impression, it is not tracked since an ad click was not made. The sale would then be considered a direct referral instead of campaign referral.

5.2.3 Prior Online Presence

Despite daily posting of social media content, online presence was still weak and needed much improvement. Organic traffic could have been improved with more search engine optimized text on the website. Prior to the experiments, the business struggled with moving up on Google search ranks, which negatively impacts the public confidence on the website. Since users tend to gravitate towards familiarity when interacting on the internet, the lack of online presence affects the likelihood of a user clicking on the business ad or website.

5.3 Future Considerations

5.3.1 Adjusting Target Audience and Keywords

Now that the dropshipping business understands the starting factors that make an effective ad, the experimentation continues with testing certain featured products, adjusting target audience settings, testing ad placements, landing pages, and testing out new keywords. Certain keywords cost more than others, which impacts the ad CPC. It would be helpful for a business to know which keywords are most cost effective to ensure the maximum return on investment.

5.3.2 Experimenting on Additional Platforms

Other platforms are currently generating a strong user base and changing the digital ad industry. Tiktok is the current rising star of social media, where users can watch and create short video clips on their cell phone. This platform has been extremely successful in marketing products due to their global reach of 740 million users [13]. Since video format is a foundational hallmark of Tiktok, experimenting would focus on factors other than ad format such as hashtags, featured products, and audio content. Influencer marketing could also be compared between these content-creating platforms, like Tiktok and Instagram. This study recommends further research that compares the return on investment between TikTok and Facebook ad platforms. This knowledge could change the marketing industry and set a new precedent for digital ad standards.

REFERENCES

[1] L. Holton, "Is dropshipping legal? 9 things to consider," Available at https://myva360.com/blog/is-dropshipping-legal (2022/03).

[2] C. P. Megan Moller and M. Bochicchio, "Not another state of marketing report 2021," Tech. Rep., 2021.

[3] C. Ferreira, "What is dropshipping?" Available at https://www.shopify.com/blog/what-is-dropshipping (2022/01/13).

[4] R. Voidonicolas, "What is shopify?" Available at https://www.shopify.com/blog/what-is-shopify (2022/03).

[5] Statista Research Department, "U.s. digital advertising industry - statistics facts," Available at https://www.statista.com/topics/1176/online-advertising/#dossierKeyfigures (2022/05/04).

[6] J.Clement, "Google, amazon, facebook, apple, and microsoft (gafam) - statistics facts," Available at https://www.statista.com/topics/4213/google-apple-facebook-amazon-and-microsoft-gafam/ (2021/02/04).

[7] M. Bretous, "Facebook ads vs. google ads: Which is better for your brand?" Available at https://blog.hubspot.com/marketing/facebook-ads-vs-google-ads (2021/04/06).

[8] E. Indacochea, "Performance max campaigns launch to all advertisers," Available at https://blog.google/products/ads-commerce/performance-max/ (2021/11/02).

[9] C. F. J. Wu and M. S. Hamada, *Experiments Planning, Analysis and Optimization*, 2nd ed. Hoboken NJ: John Wiley and Sons, Inc., 2009.

[10] J. Lawson, *Design and Analysis of Experiments with R*, 1st ed. Boca Raton FL: Taylor and Francis Group, LLC, 2015.

[11] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. Hoboken NJ: John Wiley and Sons, Inc., 2013.

[12] M. Gandhi, "State of the enterprise seo industry," Available at https://www.seoclarity.net/blog/state-of-the-industry-2020 (2019/12/11).

[13] L. Ceci, "Tiktok- statistics facts," Available at https://www.statista.com/topics/6077/tiktok/ (2022/08/04).