

A/B Testing: A Data-Driven Approach to Optimize Marketing Campaigns

Akhmad Taufik Ismail

April 2, 2023

Introduction and Background

Marketing campaigns play a pivotal role in promoting products and establishing a connection with potential customers. Nonetheless, the success of these campaigns is not uniform across the board, as it can be impacted by several factors, including message type, content, design, and audience preferences. Hence, it is imperative for marketing firms to employ data-driven techniques to test and optimize their campaigns to achieve their desired objectives.

The process of A/B testing, which is also commonly referred to as split testing or bucket testing, is a viable means of achieving this objective. Essentially, A/B testing is a method utilized to assess the effectiveness of various elements of a marketing campaign, including ads, landing pages, and other associated components. In order to carry out an A/B test, a specific aspect of the campaign is modified, and both versions are executed simultaneously, with performance data being collected. By analyzing the test results, it is possible to identify the superior version and subsequently implement any necessary changes [2].

In the present circumstance, the marketing approach of the company is restricted to Public Service Announcements (PSAs), which are informational messages intended to enlighten the public about a social concern. The organization's novel approach is to utilize ads, which are persuasive messages designed to influence the public into purchasing a product or service. Nevertheless, the company is uncertain about the efficacy of ads in enhancing user conversions. Consequently, the company has been requested to conduct A/B testing to compare the impact of PSA and advertisements.

The challenge lies in determining the optimal marketing campaign by selecting the most suitable message type (PSA or ads) tailored to the target audience. The objective is to boost the count of users who convert.

This experiment will employ A/B testing methods, which are a technique for comparing two versions of a marketing element and determining which performs better. A/B testing can assist marketers in determining what resonates best with their target audiences and improving campaign effectiveness [1].

This report will summarize the results of an A/B test conducted by a marketing firm to compare two different message types: public service announcements (PSA) and advertisements. We will go over the test's methodology, data analysis, and conclusions.

Setting Up Problem

Experiments Goal

In order to increase revenue, the company seeks to determine the most effective marketing approach by conducting an A/B test to comprehensively evaluate the impact of two distinct types of marketing messages on user behavior. Presently, the company utilizes Public Service Announcements (PSA) to educate users about its products and services, but it wants to explore the potential of ads as an alternative method to attract and retain customers. The A/B test will involve randomly assigning users to either a PSA or an ad

group, and then measuring their conversion rates. In this context, conversion rate pertains to the percentage of users who take a desired action, such as subscribing to a newsletter, purchasing a product, or downloading an application. The experiment aims to test the hypothesis that ads will produce a higher conversion rate than PSA, which will provide valuable insights into the optimal marketing strategy for the company to increase revenue.

Metrics

A driver metric can be defined in this experiment as the key metric that the company seeks to improve or increase in order to achieve its primary goal of increasing revenue. In this case, the conversion rate could be the driver metric because it is directly related to revenue generation. The company's goal is to increase conversion rates by implementing the best marketing strategy, which will be determined through A/B testing of PSA and ads.

A guardrail metric, on the other hand, is one that the company will monitor to ensure that the marketing strategy has no unintended negative consequences for other important aspects of the business. A guardrail metric for this experiment could be the bounce rate, which is the percentage of users who leave the website without taking any action. If the company notices a significant increase in bounce rate after implementing the new marketing strategy, this could indicate that the ads are not resonating with the target audience, and the company may need to adjust its approach to avoid any negative impact on user behavior.

Variants

1. **Control Group:** The control group will be exposed to PSA, which is the current marketing strategy used by the company. The control group will be used as a baseline to compare the performance of the test group.
2. **Treatment Group:** The treatment group will be exposed to ads, which is the new marketing strategy that the company is considering. The test group will be used to determine the effectiveness of the new marketing strategy.

Hypothesis

- H_0 : The conversion rate of the control group (PSA) is equal to or lower than the conversion rate of the treatment group (ads).
- H_1 : The conversion rate of the treatment group (ads) is higher than the conversion rate of the control group (PSA).

In simpler terms, the null hypothesis (H_0) for this experiment is that there is no difference in conversion rates between the control group (PSA) and the treatment group (ads). The alternative hypothesis (H_1) is that the conversion rate of the treatment group (ads) is higher than the conversion rate of the control group (PSA).

Design Experiments

Randomization Unit

In order to guarantee that the experiment generates precise and significant results, it is imperative to perform the A/B test at the user level. This implies that each individual user will be randomly allocated to either the control group, which will be exposed to the PSA message, or the treatment group, which will receive the ad message. By employing this randomization process, it ensures that there is no partiality in the allocation of users to the distinct groups, and any observed differences in the conversion rates between the two groups can be credited to the message type, and not to any other extraneous factors.

Target of Randomization Unit

In order to achieve the objective of determining the impact of message type on user behavior, targeting the randomization unit at the individual user level is the most appropriate approach for this experiment. This is because user behavior is influenced by their individual preferences and interests. By randomly assigning each individual user to either the control group or the treatment group, the experiment can be conducted in a way that is consistent with the company's goal of identifying the optimal marketing strategy.

Sample Size

The sample size for the experiment must be calculated to ensure that the results are statistically significant. The following formula can be used to calculate sample size [3]:

$$n = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2}$$

- n : Sample size
- σ : Standard deviation
- $z_{1-\alpha/2}$: Critical value for $\alpha/2$
- $z_{1-\beta}$: Critical value for β
- δ : Minimum detectable effect

Because the standard deviation of the conversion rate is unknown, it can be estimated using the Bernoulli distribution approach, which is given by:

$$\sigma = \sqrt{\hat{p}(1 - \hat{p})}$$

- \hat{p} : Estimated conversion rate / baseline of conversion rate.

Therefore, to calculate the sample size for the experiment, an estimate of the baseline conversion rate is required, which can be obtained from historical data or industry benchmarks. Once the baseline conversion rate is estimated, the minimum detectable effect and the desired levels of statistical significance (i.e., α and β) can be determined. Using these values, the sample size can be calculated using the formula provided above. This ensures that the experiment is conducted with a sufficient sample size to yield statistically significant results.

Significant Level (α)

In order to ensure that the results of the experiment are reliable and meaningful, it is essential to set a significance level that is appropriate for the study. The significance level, denoted by α , is the probability of obtaining a result that is as extreme or more extreme than the one observed in the experiment, assuming the null hypothesis is true. In this experiment, a significance level of 5% has been chosen, which means that there is a 5% chance of obtaining a result as extreme or more extreme than the one observed, even if the null hypothesis is true.

The choice of a significance level of 5% is common in A/B testing because it strikes a balance between detecting meaningful differences between the control and treatment groups, while minimizing the likelihood of false positives. This value is widely used in industry and has been shown to be effective in a variety of experimental settings. To ensure that the results of the experiment are statistically significant, the p-value, which represents the probability of obtaining a result as extreme or more extreme than the observed result, must be less than or equal to α .

Power Level ($1 - \beta$)

Standard Deviation (σ)

Differences Between Control and Treatment Groups (δ)

Experiment Duration

To ensure that the results of the experiment are reliable and accurate, the duration of the experiment has been set to two weeks. The reason for this duration is because the experiment is conducted on a website or app, and it is observed that the experiences seasonality in the number of visitors. Therefore, to avoid any bias in the results due to variations in the number of visitors, the experiment is being conducted during the peak season. In this way, the experiment can be conducted under conditions that are more representative of the typical user experience, and the results obtained can be relied upon to make data-driven decisions [4]. The two-week duration provides sufficient time to collect data from a large number of users in both the control and treatment groups, while minimizing any potential external factors that could influence the results.

Analyzing and Interpreting the Data

Conclusion and Recommendation

References

- [1] Tom Baldwin. *How A/B Testing Can Improve Your Marketing Campaigns*. www.linkedin.com, Mar. 2023. URL: <https://www.linkedin.com/pulse/how-ab-testing-can-improve-your-marketing-campaigns-tom-baldwin> (visited on 04/02/2023).
- [2] *How to Use A/B Testing to Maximize Marketing Campaigns*. Lotame, Mar. 2019. URL: <https://www.lotame.com/how-to-use-a-b-testing-to-maximize-marketing-campaign-performance/> (visited on 04/02/2023).
- [3] Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press, 2020. DOI: 10.1017/9781108653985.
- [4] Khalid Saleh. *How Long Should You Run an A/B Test for and How to Calculate...* Invespcro, Nov. 2017. URL: <https://www.invespcro.com/blog/how-long-should-you-run-an-ab-test-for/> (visited on 04/02/2023).