A/B Test Analysis Report for Globox Food and Drink Banner Experience

Sumanth Bandaru

2023-12-24

Summary:

The purpose of this report is to provide the findings of an A/B test done to assess the influence of a new banner experience on user behaviour on the Globox mobile website. The banner's purpose is to raise awareness of the food and beverage options. The test separated users into two groups: those who viewed the previous site and those who saw the new banner. The conversion rate and average amount spent per user were the major metrics of interest. The experiment was place between January 26 and February 6.

The study found that the test group had a statistically significant higher conversion rate than the control group. This implies that the banner resulted in a larger proportion of consumers making a purchase. The average amount spent per user, however, did not change much between the two groups.

Based on these findings, it is recommended that the banner be made available to all users. There is also evidence to propose that the launch be delayed in order to increase the sample size, validate the results, and further explore the influence on the average amount spent per user.

The parts that follow give a full description of the test, data analysis, and the reasoning behind these results.

Summary of Findings:

The new food and drink banner had a good influence on user behaviour, according to our A/B test. The test group had a much greater conversion rate than the control group, with a 12.47% increase. Although there was no significant difference in average spending per user between the two groups, the rise in the number of users making a purchase in the test group suggests the possibility of higher income. To validate these findings, we propose rolling out the updated banner to all users and undertaking additional testing with a bigger sample size.

Context:

The goal of the experiment was to see how a new banner emphasising food and drink offers affected user behaviour. The new food and drink banner was launched with the hope of increasing user engagement and encouraging more purchases by emphasising the diversity of items available. If GloBox is effective, this might lead to higher sales and money.Users who visited the site during the testing period were allocated to either the control or test groups at random. The original webpage was seen by the control group, whereas the food and drink banner was seen by the test group. The users were assigned at random to ensure that the two groups were equivalent in terms of the user attributes listed below.

The conversion rate and average amount spent per user were the major metrics of interest for the test. The conversion rate was defined as the percentage of users who made a purchase on the site during their first or subsequent visits. The average amount spent was determined by dividing the total amount spent by the number of users.

Both measures were computed individually for each group. These indicators are significant because an increase in one or both would result in an increase in overall revenue for GloBox.The dataset utilised for the investigation contains information from about 49,000 people who visited the site throughout the testing period. The information gathered included the group they were assigned to, whether or not they made a purchase, the amount they spent, and other user attributes such as gender, device, and country of residence.

Data Overview:

The A/B test had 48,943 total users, including 24,343 in the Control group and 24,600 in the Test group. The Control group converted at a rate of 3.923%, whereas the Test group converted at a rate of 4.630%. The Control group spent $3.37 on average, whereas the Test group spent $3.39. This is a statistically significant difference in conversion rate but not in average spend, as will be proved in the next sections.

In terms of user characteristics, the United States had the most users, followed by Brazil and Mexico. Gender distribution was roughly similar, with slightly more females in the Control group (10,069 vs 10,054) and slightly more men in the Test group (10,235 vs 10,061). There were also 1,669 users of 'Other' gender (Transgender, Non-Binary, a gender, gender fluid, and other gender identities) (808 in Control, 861 in Test). There were additional 6,855 users of unknown gender who were evenly distributed throughout the groups. In each category, there were around 5,900 more Android users than iOS users. In each nation, Android users accounted for around 62% of all smart phones.

Hypothesis Testing for Conversion Rate

Our goal is to determine whether the two groups' conversion rates differ statistically significantly. To finish the computations, Excel was utilised. A significance threshold of 0.05 is being used.

H0: p1 = p2 (The conversion rates in the control and treatment groups are equal)
HA: p1 ≠ p2 (The conversion rates in the control and treatment groups are not equal)

Control Group: 955 Conversions from 24,343 users wih conversion rate of 0.03923099
Test Group: 1139 Conversions from 24,600 users with conversion rate of 0.046300813

Pooled Proportion: (955 + 1139) / (24343 + 24600) = 0.042784464

Standard Error: √(0.042784464 * (1-0.042784464) * (1 / 24343 + 1 / 24600) = 0.001829526

Test Statistic: (0.03923099 - 0.046300813) / 0.001829526 = 3.86429177

P-Value: 2 * (1 – NORM.S.DIST(3.86429177, TRUE)) = 0.000111

Since P is less than α we reject the null hypothesis, there is a statistically significant difference in the conversion rate between the groups.

95% Confidence Interval for Conversion Rate

The 95% confidence interval for the difference in the conversion rate between the treatment and control groups (treatment - control) is approximately [0.0035, 0.0107].

Standard Error Unpooled: √((0.03923099 * (1 – 0.03923099) / 24343) + (0.046300813 * (1 – 0.046300813) / 24600)) = 0.001828488

Difference in Conversion Rates: 0.00706982

Lower Bound: 0.00706982 – 1.96 * 0.001828488 = 0.003485985

Upper Bound: 0.00706982 + 1.96 * 0.001828488 = 0.010654

Hypothesis Testing for Average Spent

We are attempting to determine whether the average expenditure for the two groups differs in a way that is statistically significant. Once more, Excel was utilised to finish the computations. Welch's t-test was employed, with a significance threshold of 0.05 and the assumption of unequal variances.

H0: $\mu1 = \mu2$ (The average amount spent per user in the control and treatment groups is equal)
HA: $\mu1 \neq \mu2$ (The average amount spent per user in the control and treatment groups is not equal)

Control Group average spent is 3.37451752 with a Standard Deviation of 25.93585054
Test Group average spent is 3.390866667 with a Standard Deviation of 25.41358872

Standard Error: $\sqrt{(25.93585054^2 / 24343) + (25.41358872^2 / 24600)} = 0.232135762$

T-Test p-value: 0.943853 (Using T.TEST() function in Excel)

Since this P-value is greater than α we fail to reject the null hypothesis. This means that there is not a statistically significant difference in the average amount spent between the two groups.

95% Confidence Interval for Conversion Rate:

The 95% confidence interval for the difference in the average spent between the treatment and control groups (treatment - control) is approximately [-0.439, 0.471].

Lower Bound:
(3.390866667 – 3.37451752) - T.INV (0.975, (24343 + 24600)-2) * 0.232135762 = -0.43863984

Upper Bound:
(3.390866667 – 3.37451752) + T.INV (0.975, (24343 + 24600)-2) * 0.232135762 = 0.471338

Further Analysis

Novelty Test

After observing a significant difference in the conversion rates between the control and test groups, we conducted a Novelty Effect Test. The novelty effect refers to the tendency of an individual to respond more positively to a new experience, object, or piece of information than to a familiar one. In the context of an A/B test, this could mean that users might change their behavior just because something is new, not necessarily better. Therefore, it's important to check if the observed effect was due to the novelty of the change.

Some differences were visible when the average amount spent by all users in both groups was visualised. But there was no distinct pattern that suggested a sizable novelty impact. There was not a persistent enough difference between the Control and Treatment groups to indicate conduct influenced by novelty.

Analysing the conversion rates over time for every user in both groups also revealed some slight variations. Nevertheless, there was no evidence that the conversion rates were impacted by the novelty effect. There was insufficient variation between the Control and Treatment groups to suggest that the reaction was motivated by novelty.

I looked at the average amount spent by converted users in both groups and found minor variances but no trend that was persistent enough to point to a novelty impact. There was inconsistent lack of significance in the differences observed between the Control and Treatment groups.

The parallel trends in both the conversion rate and site visits suggest that there was no Novelty Effect in this test. The observed increase in conversion rate in the test group is likely due to the new banner, rather than just a response to something new.

Power Analysis

A crucial statistical technique for figuring out the sample size required for an experiment to reach a particular degree of statistical power is power analysis. The likelihood of accurately rejecting a null hypothesis when it is incorrect and identifying a genuine effect is known as statistical power. Stated differently, power analysis aids in determining the probability of discovering a noteworthy impact, should one exist within the population.

Before starting an experiment or A/B test, it is crucial to do a power analysis to make sure the sample size is sufficient to identify significant effects. An experiment may not have the power to identify true differences if the sample size is too small, which might provide inconsistent or ambiguous results. Conversely, if a lower sample size offers enough power to make meaningful conclusions, then a larger sample size may be superfluous and resource-intensive.

We may maximise the sample size, raise the likelihood of finding significant effects, and improve the dependability of the experiment's results by carrying out a power analysis. This crucial stage helps researchers and data analysts to plan reliable trials, make well-informed decisions, and maximise the results of their data-driven projects. Let's examine the main conclusions in more detail:

Comparing Conversion Rates: A total sample size of 60,600 was suggested by the power analysis, which would have been adequate to detect the intended minimum detectable influence of 10% on conversion rates. The trustworthiness of the data was further reinforced by the much bigger observed impact size of 18% for the conversion rate.

Comparison of Average expenditure: Based on the power analysis, a significantly higher sample size of 182,164 would be needed for the average expenditure analysis in order to detect a tiny effect size of 0.5% with the appropriate level of statistical power. With limited resources and time, it would not be feasible to get such a huge sample size.

The findings of the power analysis highlight the significance of concentrating on methods to raise user engagement and conversion rates. In this regard, we may reliably draw conclusions and make data-driven decisions because the conversion rate study has a large enough sample size.

It might not be as important to pay close attention to this parameter, though, considering the tiny impact size and difficulties in acquiring a considerably bigger sample size for average spending research. Rather, the focus should be on increasing income through the optimisation of user engagement and conversion rates.

Conclusions and Recommendations

It is confidently advised to deploy the food and drink category banner based on the analysis of the A/B test results and further insights supplied. The test findings showed a substantial difference in conversion rates between the treatment and control groups, demonstrating the banner's beneficial effect on promoting conversions.

Although there was a noticeable increase in the conversion rate, there was a trend in the average amount spent per user, which may indicate that users are becoming more likely to purchase less expensive goods. In light of this, it appears like a wise strategic choice to give user acquisition priority over quick income increases. Increased brand engagement, a bigger client base, and the possible long-term advantages of repeat business and customer loyalty might result from drawing in more users with the banner

One benefit of this suggestion is that it requires very little money to launch the banner, which lowers the financial risk involved in the choice. Even while it won't have a major immediate impact on income per user, the cheap cost makes it a good choice, especially in light of its possible long-term effects.

Maintaining careful tabs on the banner's performance and carrying out additional research on user spending patterns are crucial for finding a balance and optimising income creation. We are able to obtain a better understanding of the data's influence on various client categories and adjust our marketing strategy by dividing it into segments according to order value, product preferences, and demographics.

Additionally, our analysis revealed a higher conversion rate among female users compared to male and other gender users. Understanding the factors behind this difference could provide valuable insights for improving targeting strategies and further increasing conversion rates. Future research could focus on exploring these gender differences in more detail.

As a result, I advise introducing the food and drink category banner and doing a thorough study of user purchasing patterns at the same time. We will be able to maximise tactics, make well-informed decisions, and take use of the banner's potential to propel GloBox's development and success thanks to this data-driven strategy.