GloBox A/B Test Analysis:
Food and Drink Product Banner's Influence on Revenue
(JANUARY 25, 2023 - February 6, 2023)

BY

AGUZUE, OLUEBUBE GOODNESS

AUGUST 2023

PRESENTED TO

Growth Product & Engineering Team Stakeholders

LED BY

Leila Al-Farsi, Product Manager

Alejandro Gonzalez, User Experience Designer

Mei Kim, Head of Marketing

ABSTRACT

This experiment delves into the analysis of an A/B test conducted for GloBox, an e-commerce company specializing in unique and high-quality products. The experiment aims to extract insights from the A/B test results and provide data-driven recommendations. The first stage involves aggregating user-level data using SQL, followed by statistical analysis utilizing spreadsheets and visualizations in Tableau as well as the inclusion of a power analysis test in addition to other statistical evaluations. This power analysis assesses the experiment's statistical power and sample size adequacy, ensuring that it can effectively detect meaningful differences between the test and control groups. The experiment focuses on GloBox's efforts to enhance revenue by spotlighting its food and drink offerings through a website banner. By implementing an A/B test that presents the banner to a test group while excluding it from a control group, the company aims to determine the banner's impact on user behavior. The analysis centers on conversion rates alongside average amount spent in both groups and the potential to launch the banner for all users.

TABLE OF CONTENT

CHAPTER THREE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

CHAPTER ONE

INTRODUCTION

1.1 Background of the experiment

GloBox, an online marketplace renowned for its boutique fashion and high-end decor products recently experienced substantial growth in the company's food and drink offerings, prompting a strategic focus on raising awareness and boosting revenue within this category.

The Growth team plans an A/B test featuring a prominent banner showcasing food and drink products on the website's header. The test group views the banner, while the control group does not.

1.2 Significance of the experiment

This experiment lies in its potential to drive strategic decisions and revenue growth for GloBox. By conducting an A/B test that highlights food and drink products through a website banner, the company aims to leverage user behavior insights to make informed choices. This experiment aligns with GloBox's evolving focus on diversifying its offerings, particularly in the food and drink category. The experiment's outcomes could shed light on the effectiveness of the banner in increasing user engagement and conversion rates. growth trajectory.

1.3 Experiment Objectives

i. The A/B test is conducted exclusively on the mobile version of the website.

ii. Upon visiting the GloBox main page, users are randomly placed in either the control or test group, with the date of joining recorded.

iii. For the test group, the page displays the banner, while the control group sees no banner.

iv. Users could make purchases on the same day they join the experiment or later. If they make one or more purchases, it is classified as a "conversion."

1.4 Experiment Questions /Answers

Before using SQL query to extract the appropriate dataset needed to perform the A/B Test Statistics, we made sure to account for these questions:

i. Can a user show up more than once? **Yes, there were 139 users with more than 1 purchases for different days.**

ii. What type of join should we use to join the tables? **We used the Left Join**

iii. What are the start and end dates of the experiment? **Jan 25,2023– Feb 6, 2023**

iv. How many total distinct users were in the experiment? **48943**

v. How many users were in the Control and Test groups? **Control - 24343, Test – 24600**

vi. What was the total number of conversions of all users? **2094**

vii. What was the conversion rate of all users? **0.043**

viii. What was the conversion rate for the Control and Test groups? **Control - 955, Test – 1139**

ix. What is the user conversion rate for the Control and Test groups? **Control – 0.046(4.6%), Test – 0.039(3.9%)**

x. What is the average amount spent per user for the Control and Test groups, including users who did not convert? **Control – $3.374, Test – $3.909**

xi. Why does it matter to include users who did not convert when calculating the average amount spent per user? **The experiment could result in more users converting, but spending less when they do. Or, it could result in fewer users converting, but spending more when they do.**

1.5 Experiment Hypothesis

H_0: There is no difference in the conversion rate between the control and treatment group.

H_1: There is a difference in the conversion rate between the control and treatment group.

U_0: There is no difference in the average amount spent between the control and treatment group.

U_1: There is a difference in the average amount spent between the control and treatment group.

CHAPTER TWO

RESULTS AND DISCUSSION

2.1 Data Presentation

The dataset comprises user demographic details, A/B test group assignments, and user purchase activity. It includes information such as user IDs, country codes, gender (M = male, F = female, O = other), test group membership (Control, Test), devices used (I = iOS, A = android), and purchase amounts in USD. This data that was extracted using a SQL query enables analysis of user behavior and A/B test outcomes.

2.2 Data Analysis

A Hypothesis Test and 95% Confidence Interval was conducted to see if there is a difference in the conversion rate between the Control and Test group, and the average amount spent per user between the two groups.

2.3 Descriptive Statistics

Below are the results of the Test Statistics:

Table 2.3.1: Hypothesis Test and 95% Confidence Interval for Difference in Conversion Rate

HYPOTHESIS TEST	
Null Hypothesis	There is no difference in the conversion rate between the control and treatment group i.e Ho : P1- P2 = P0
Alternative Hypothesis	There is a difference in the conversion rate between the control and treatment group i.e Ha : P1 - P2 <> P0
Number of tails	2 sided/tailed test
Alpha Value	0.05
control sample size(n1)	24343
test sample size(n2)	24600
Control sample proportion(p^1)	0.0392
Test sample proportion(p^2)	0.04630081301
pooled sample proportion(p^)	0.04278446356
z test statistics numerator	-0.0071
z test statistics denomenator	0.001829526081
z test statistics	3.86429177
z Critical Value(left tail)	-1.959963986
z Critical Value(right tail)	1.959963986
P-value	0.0001114119853
Test Result	Since p-value ≤ α(0.05) and our test statistic is also greater than our critical value, we reject the null hypothesis that there is no difference in the conversion rate between the control and treatment group in favor of the alternative hypothesis that there is a difference in the conversion rate between both groups
Conclucion:	We can conclude that there is enough sample evidence to support the alternative hypothesis that suggest that indeed there is a difference in conversion rate between the control and treatment group and the results are statistically significant at the chosen 0.05 significance level. We do run a 5% Type 1 Error risk of saying that there is a difference in the conversion rate between both groups, when in fact there is no difference.

95% CONFIDENCE INTERVAL	
Confidence level	0.95
Alpha Value	0.05
SAMPLE STATISTIC	0.0071
Alpha	0.05
z Critical Value(left tail)	-1.959963986
z Critical Value(right tail)	1.959963986
Control sample size A(n1)	24343
Test sample size B(n2)	24600
Control sample proportion A(p^1)	0.0392
Test sample proportion B(p^2)	0.04630081301
SEp^1	0.000001548367901
SE2p^2	0.00000179500194
STANDARD ERROR(SE)	0.001828488403
MARGIN OF ERROR	0.00358377142
LOWER CONFIDENCE INTERVAL	0.0035
UPPER CONFIDENCE INTERVAL	0.0107
CONCLUSION	The 95% confidence interval for the proportion difference between the control and treatment group is 0.0035 to 0.0107

SOURCE: Author's computation using Google Spreadsheet, 2023.

Table 2.3.1 shows the summary of the descriptive statistics of the Hypothesis Test and 95% Confidence Interval to see if there is a difference in the conversion rate between the control and test group.

In the provided hypothesis test, the calculated z-test statistic shows that the sample statistic (the difference in sample proportions) is 3.864 standard deviations away from the hypothesized population parameter under the null hypothesis. The critical values for the two-tailed test at a significance level of 0.05 is -1.96 and 1.96. These values show us the boundaries beyond which our test statistics results are considered statistically significant. The calculated probability of

observing a test statistic as extreme as the one calculated, assuming the null hypothesis is true is 0.000111. Since the p-value is low, it suggests stronger evidence against the null hypothesis.

Given that the calculated z-test statistic (3.864) is much larger than the critical values (±1.96) and the p-value (0.000111) is well below the significance level (0.05), we have strong evidence to reject the null hypothesis. This implies that there is a statistically significant difference in the conversion rate between the control and treatment groups, supporting the alternative hypothesis. We do run 5% Types I Error risk of saying that there is a difference in the conversion rate between both groups, when in fact there is no difference.

On the other hand, the confidence interval test shows that our sample statistic varies by 0.00183 from the true population parameter with the result shown by our standard error.

Given the margin of error (MOE) which shows us the range around our sample statistic where the true population parameter is likely to fall, we expect the true difference to be within 0.00358 of our sample statistic. Using the MOE, our confidence interval is 0.0035 to 0.0107 this suggests that we are 95% confident that the true difference in conversion rates between the control and treatment groups lies within this interval. Since this interval does not include zero, it indicates that the difference is statistically significant. This aligns with the earlier hypothesis test results that led us to reject the null hypothesis in favor of the alternative hypothesis.

Table 2.3.2: Hypothesis Test and 95% Confidence Interval for Difference in Average Amount Spent

HYPOTHESIS TEST	
Null Hypothesis	There is no difference in the average amount spent between the control and treatment group i.e Uo : U1- U2 = U0
Alternative Hypothesis	There is a difference in the average amount spent between the control and treatment group i.e Uo : U1- U2 <> U0
Number of tails	2 sided/tailed test
Alpha Value	0.05
Control sample mean A(xbar1)	3.3745
Test sample mean B(xbar2)	3.39
Control standard deviation A(s1)	25.93639056
Test standard deviation B(s2)	25.41382684
Control sample sizeA(n1)	24343
Test sample size B(n2)	24600
Degree of freedom	24342
Test statistics numerator	0.01584179109
Test statistics denomenator	0.2321393004
Test statistics	0.06824260718
T Critical Value(left tail)	-1.959963986
T Critical Value(right tail)	1.959963986
P-value	0.9455925104
Test Result	Since p-value > α(0.05) and our test statistic is also less than our critical value, we fail to reject the null hypothesis that there is no difference in the average amount spent between the control and treatment group
Conclucion:	We can conclude that there isn't enough sample evidence to support the alternative hypothesis that suggest that there is a difference in average amount spent between the control and treatment group and the results are statistically insignificant at the chosen 0.05 significance level. We do run a 5% Type II Error risk of saying that there is no difference in the average amount spemt between both groups, when in fact there truely is a difference.

95% CONFIDENCE INTERVAL	
Confidence level	0.95
Alpha Value	0.05
Control sample mean (xbar1)	3.3745
Test sample mean (xbar2)	3.39
Sample Statistics	0.0158
Control sample sizeA(n1)	24343
Test sample size B(n2)	24600
Alpha Value	0.05
Degree of freedom	24342
T Critical Value(left tail)	-1.960061445
T Critical Value(right tail)	1.960061445
Control standard deviation A(s1)	25.93639056
Test standard deviation B(s2)	25.41382684
STANDARD ERROR(SE)	0.2321393004
MARGIN OF ERROR	0.4550072927
LOWER CONFIDENCE INTERVAL	-0.4392
UPPER CONFIDENCI	0.4708
CONCLUSION	The 95% confidence interval for the Difference in Means between the control and treatment group is -0.4392 to 0.4708

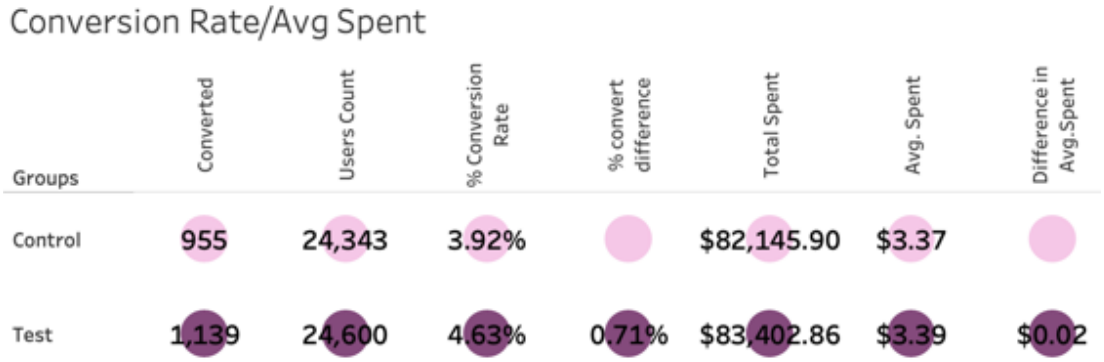
SOURCE: Author's computation using Google Spreadsheet, 2023.

The provided hypothesis test results indicate a lack of strong evidence to show a statistically significant difference in the average amount spent between the control and test groups. The calculated Test statistic is (0.0682) which is much smaller than the critical values (±1.96) and the p-value (0.9456) is much greater than the significance level (0.05). Consequently, we fail to reject the null hypothesis in favor of the alternative hypothesis, indicating a statistically insignificant distinction in average amount spent. This is consistent with the confidence interval test, where given that the confidence interval includes both negative and positive values (from -0.4392 to 0.4708) and considering the t critical values, you don't have strong evidence to conclude a statistically significant difference in average amount spent between the groups. This suggests that the observed difference could reasonably be explained by random variability. It's important to note

a 5% risk of Type II Error of saying that there is no difference in the average amount spent between both groups, when in fact there truly is a difference.

2.4 Tableau Visualization Analysis

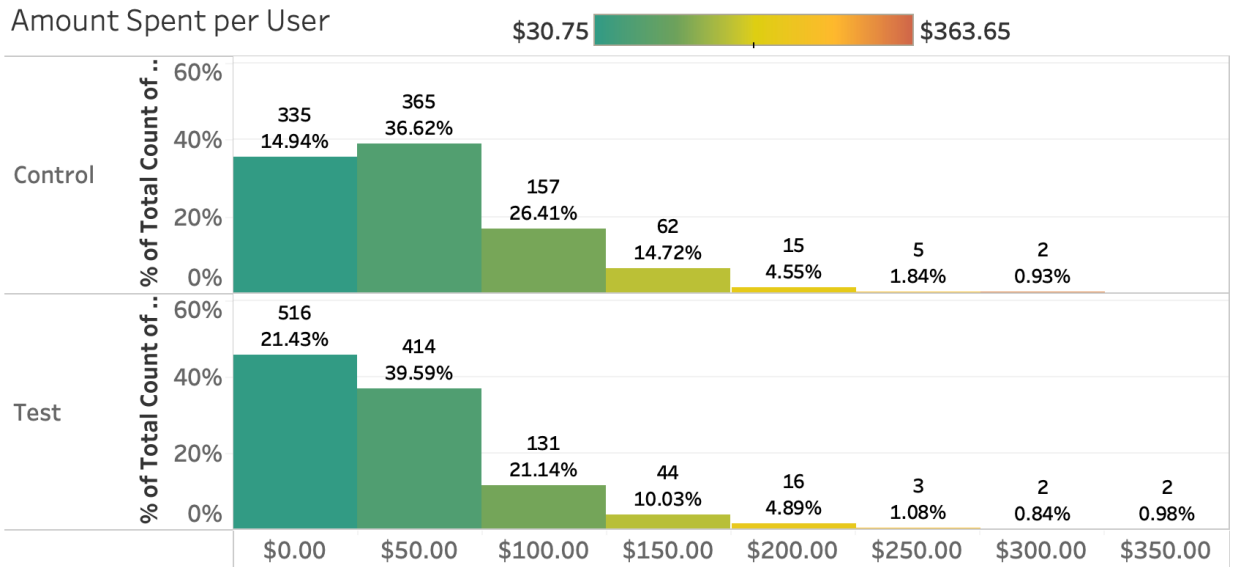
Figure 2.4.1: Viz of Experiment Groups and test metrics (conversion rate and average amount spent)



SOURCE: Author's computation using Tableau Public, 2023.

Figure 2.4.1 displays user conversion data, including counts, percentages, total spending, and averages in both experiment groups. The Test group has more users and conversions, leading to a higher conversion rate compared to the Control group. A 0.71% difference in conversion rates is observed. However, analyzing average spending reveals only a $0.02 difference between groups, which isn't significant to assert higher spending in the Test group. The conclusions are in line with prior hypothesis tests, supporting conversion rate differences but lacking evidence for significant spending differences.

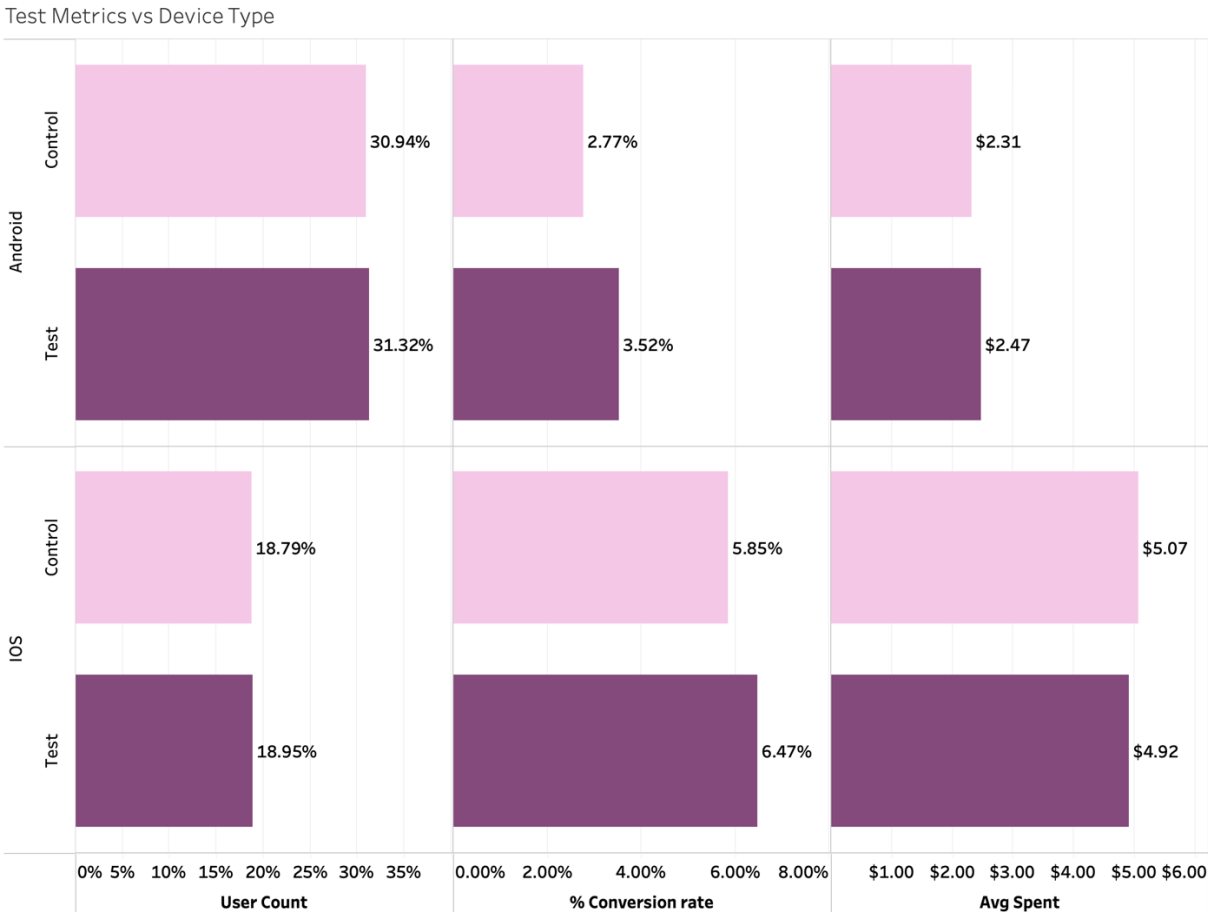
Figure 2.4.2: Viz of Experiment Groups and Amount Spent per User



SOURCE: Author's computation using Tableau Public, 2023.

Figure 2.4.2 illustrates spending patterns in both groups, highlighting that the $50-$99 range contributed the most to total spending: $26,452.91 in Control and $29,309.52 in Test. The Control group saw most spending user count (365) in the $50-$99 range, while the Test group's highest user count (516) was in the 0-$49 range. Significant spending in the Control group mainly came from $50-$99, $100-$149 and $150-$199 ranges, totaling $56,166.88. Comparing the total spending for $150-$199 and 0-$49 ranges with a $156.85 difference isn't considerable due to a less user count and a higher users spending in the former. Focusing on top 2 spending ranges for the Control group users spending between $50-$99 and $100-$149 (522 users, $45,530.93) slightly surpassed Test group users in it's top spending range 0-$49 and $50-$99 (930 users, $44,995.63). While external factors like date and experiment duration may have an influence on the outcomes, there's no substantial average spending difference between both groups.

Figure 2.4.3: Viz of Experiment Groups and Test Metrics vs Device type

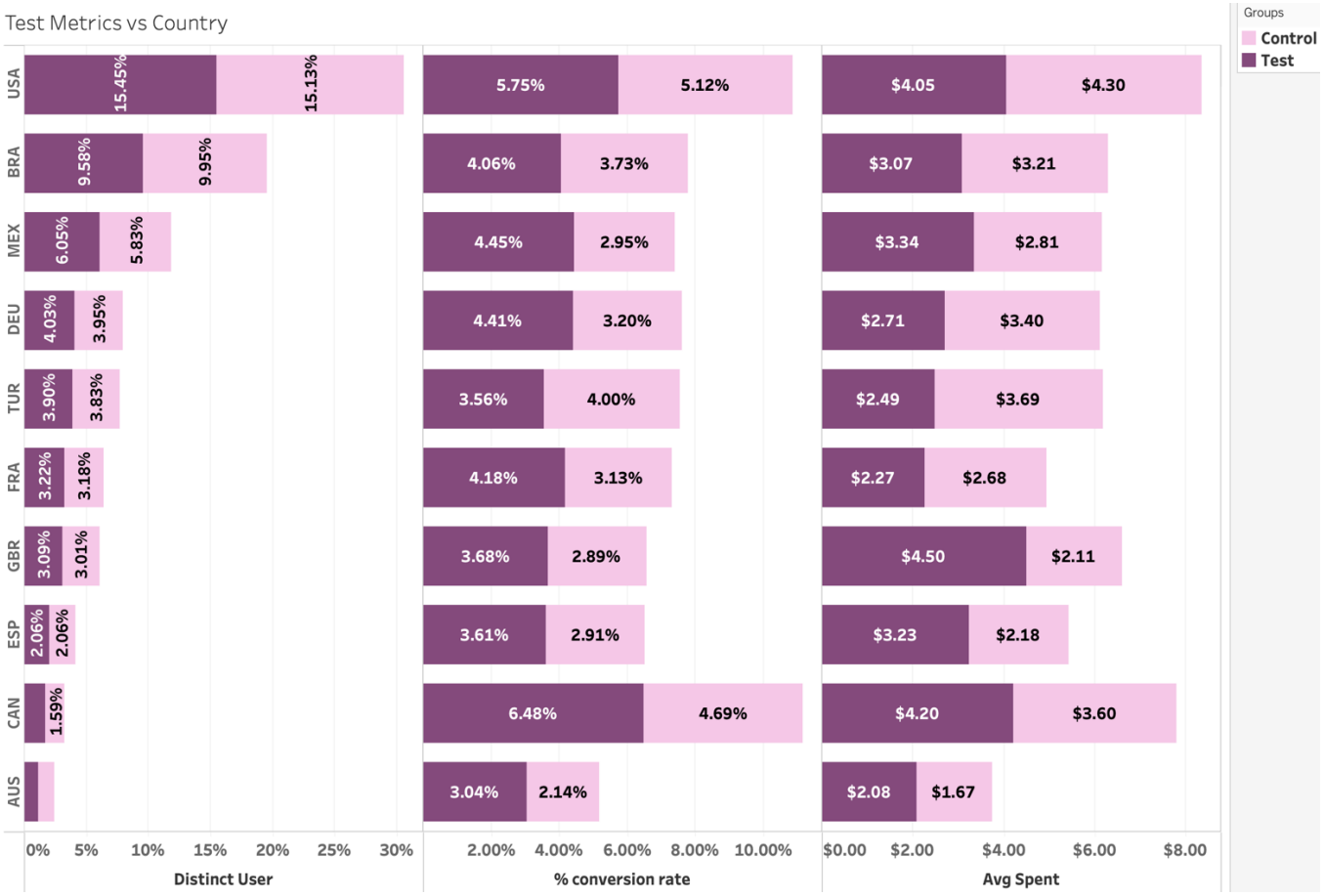


SOURCE: Author's computation using Tableau Public, 2023.

Figure 2.4.3 reveals that both experiment groups have more users using Android over IOS, with Control and Test groups having 15,054 and 15,235 Android users, and 9,142 and 9,218 IOS users, respectively. However, despite Android having more users, IOS exhibits higher conversion rates and average spending among converters in both groups. For Android, the Test group's 0.75% higher conversion rate difference indicates more conversions compared to the Control group unlike the average amount spent between both groups with only an average difference of $0.16, there isn't much of a significant difference between the average spent for both groups.

While IOS users show a 0.62% conversion rate difference, the $0.15 average spending gap isn't significant. The data suggests IOS users convert more and spend more on average. Both IOS and Android have statistically significant conversion rate differences in Test groups versus Control. Yet, no substantial spending variance exists between the device user groups. In essence, the GloBox homepage predominant converts are majorly IOS users.

Figure 2.4.4: Viz of Experiment Groups and Test Metrics vs Country



SOURCE: Author's computation using Tableau Public, 2023.

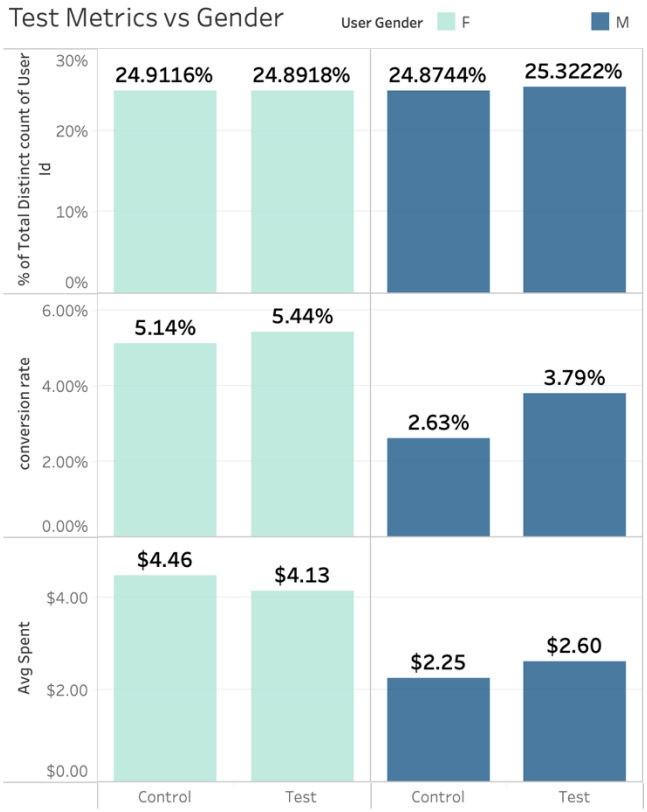
Figure 2.4.4 presents an insightful exploration of the relationship between user countries, test groups, user counts, % conversion rates, and average amount spent. Although USA has the highest count of users for the Test and Control group, Canada (CAN) had the highest conversion

rate for the Test group and USA for the Control group. While it may look like United Kingdom (GBR) has the highest average amount spent for the Test group and USA for the Control group, when comparing the difference in the average spent for GBR Test group with that of USA or CAN and the difference in the average spent for USA Control group with that of CAN, Turkey (TUR), Germany (DEU) etc., the difference may not be so significant. Australia (AUS) is seen to have the lower user count, conversion rate and average amount spent compared to other countries. When considering the conversion rate for each country, a higher rate % means that more users converted based on each country's average user count compared to the other countries. United States (USA) Test group exhibits a higher conversion rate and average amount spent, indicating positive test impacts. Despite slightly higher conversions, spending remains comparable to the Control group. For Brazil (BRA), both Test and Control groups show improved conversion rates, suggesting effective test changes. Spending differences between the groups are minor. Mexico (MEX) with a slightly better conversion rate in the Test group implies positive test impacts, with marginal spending differences. Germany (DEU) Test group demonstrates a higher conversion rate, indicating successful test changes. Spending differences between groups are negligible. While both groups in Turkey (TUR) have respectable conversion rates, the Control group's spending is higher, highlighting the varied impact of test changes on spending. France (FRA) enhanced conversion rates in both groups are observed, with spending remaining relatively stable. In United Kingdom (GBR), the Test group's higher conversion rate and significantly increased average amount spent suggest successful test changes. Spain (ESP) with a slightly higher conversion rate in the Test group indicates positive test effects, with increased spending. Canada (CAN) Test group significantly outperforms the Control group in conversion rate and average amount spent, signaling substantial benefits from test changes.

Lastly, Australia (AUS) Test group's higher conversion rate suggests favorable test impacts on conversions, with a modest increase in spending.

Analyzing the dataset in tandem with the prior hypothesis test outcomes, we observe that the test changes have consistently influenced the conversion rates across various countries. However, the impact on average amount spent varies by country. Some countries exhibit significant differences in spending, while others show minor variations. Given the insights gained, it's recommended to tailor marketing and promotional strategies based on country-specific trends. For countries where the test changes have led to significant increases in both conversions and spending (e.g., Canada, UK), investing in further optimization of these strategies could yield substantial benefits. For countries where spending hasn't been significantly impacted (e.g., Germany, France), exploring additional enhancements to encourage higher spending might be worthwhile. Continuous monitoring of user behavior, particularly in relation to the test changes, is advised to refine strategies and maximize user engagement and revenue generation.

Figure 2.4.5: Viz of Experiment Groups and Test Metrics vs Gender



SOURCE: Author's computation using Tableau Public, 2023.

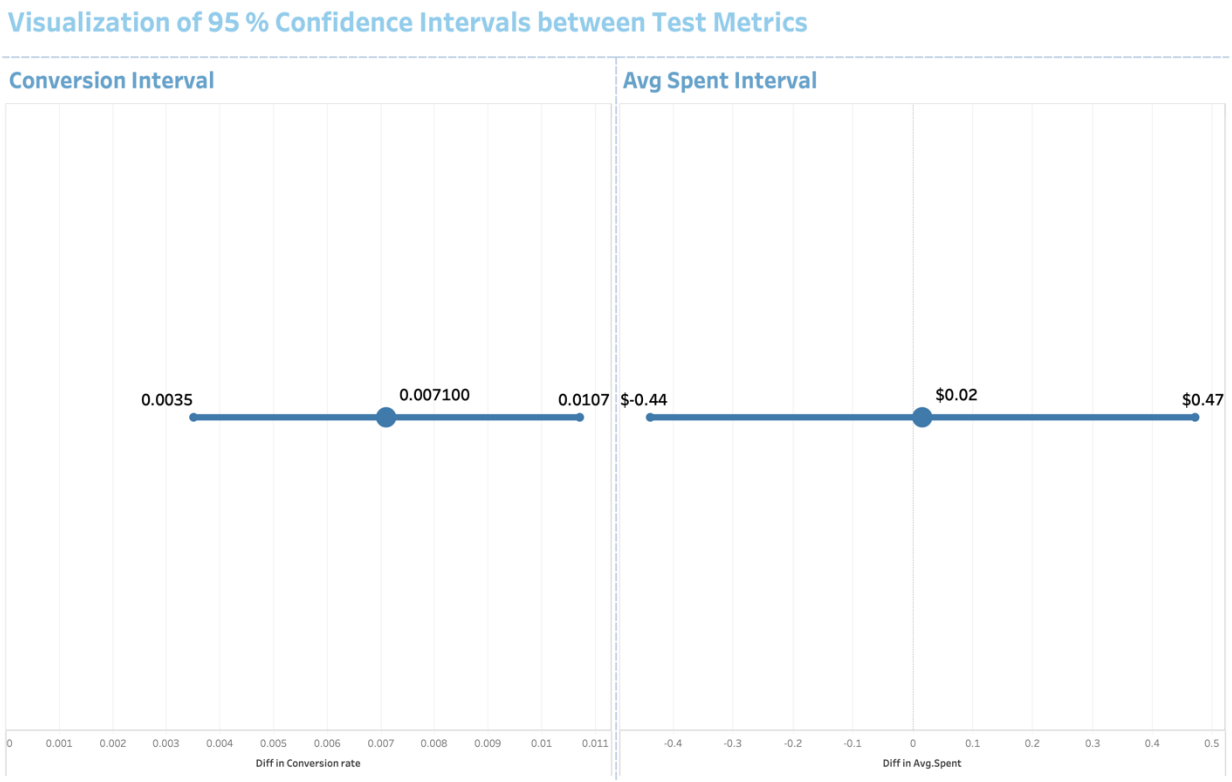
Figure 2.4.5 explores the relationship between user gender and the test groups, focusing on their user counts, average conversion rates, and average amount spent.

The analysis reveals gender-based differences in conversion rates and average spending. Test Group consistently demonstrates higher conversion rates compared to Control Group for both genders, indicating that the changes have positively impacted conversion rates. For average amount spent, there is a variation across the genders and experiment groups, but these differences are not substantial. Based on the findings, it's recommended to focus on refining the strategies that contributed to the improved conversion rates in the Test Group. However, considering that the average amount spent doesn't vary significantly between the groups and genders, it might be beneficial to further investigate factors beyond gender, such as user preferences or purchasing

behavior, that could influence spending. Continued monitoring of user behavior, especially in relation to gender-specific trends, is advised to tailor marketing approaches and maximize both conversion rates and average spending. It's important to understand that while there are positive shifts in conversion rates, these changes might not always translate directly into increased spending, indicating the need for more refined strategies tailored to different user segments.

2.5 Visualize the Confidence Intervals

Figure 2.5.1: Visualization of 95 % Confidence Intervals between Test Metrics



SOURCE: Author's computation using Tableau Public, 2023.

Figure 2.5.1 presents two insights with corresponding 95% confidence intervals. The proportion difference between control and treatment group ranges from 0.0035 to 0.0107, indicating 95% confidence in this range. The sample statistic of 0.0071 falls within, suggesting proportions lie between 0.0035 and 0.0107. Similarly, the interval for difference in means between groups is -$0.4392 to $0.4708, indicating 95% confidence. The sample statistic of 0.0158 aligns, implying

means could range from a decrease of $0.4392 to an increase of $0.4708. These intervals provide certainty on metric values. The analysis via Tableau yields valuable insights, with sample statistics within intervals, bolstering credibility. These intervals serve as essential benchmarks, aiding data interpretation.

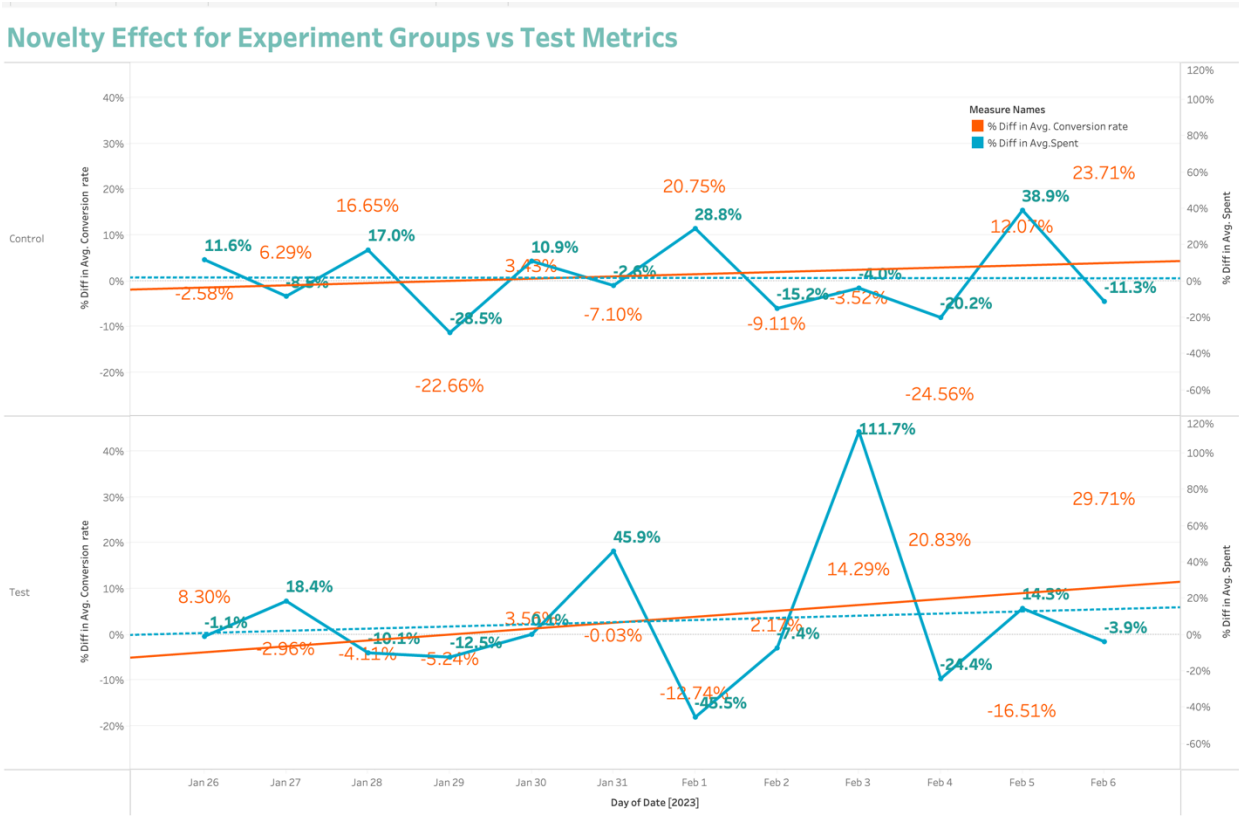
2.6 Novelty Effects

Since users might behave differently when the Test group is new, we decided to inspect the difference in the key metrics between the groups over time. We want to notice if the effectiveness of the banner is short-lived which explains the tendency of individuals to show increased interest, engagement, or behavioral changes in response to new or novel stimuli. In this case, the introduction of the banner showcasing food and drink products on the website is the novel element. GloBox has been predominantly associated with boutique fashion and high-end decor items, so the sudden prominence of the food and drink category represents a departure from the familiar offerings. The novelty effect motivates users to engage more with the website due to the intrigue generated by this new addition. Users are likely to explore the highlighted products, spending more time on the website and potentially making purchases they might not have considered before. Novelty effect tends to be more pronounced in the short term. Over time, as users become accustomed to the change, the initial excitement may diminish, potentially leading to a decline in engagement related to the food and drink category.

The results of the test will provide insights into whether the initial interest translates into sustained benefits over time.

2.7 Check for Novelty Effects

Figure 2.7.1: Novelty Effect for Experiment Groups vs Test Metrics



SOURCE: Author's computation using Tableau Public, 2023.

Figure 2.6.1 offers valuable insights into the potential novelty effect over time as a result of the A/B test conducted by GloBox. The analysis of difference in average total spending and difference in conversion rates over a series of dates can help us understand whether the introduction of the banner highlighting food and drink products had a significant and sustained impact or if it was influenced by the novelty effect. The visualization captures the spending behavior and conversion rates for the Control group and the Test group over a span of time, shedding light on the potential novelty effect. From January 25 - February 6, the difference in average spent fluctuates over time for both groups. Looking at the trend line, the Control group has a more stable line overall of a less than 1% increase in the percentage difference in the average spent while it started with an

initial negative of close to 2% decrease in the percentage difference in the conversion rate and went up to a positive of 10% increase by the end date of the experiment. The Test group on the other hand started good with an initial of close to 0 percentage difference in the average spent at the start of the experiment and went close to a 20% increase in the percentage difference in the average spent and has an initial negative of close to 8% decrease in the percentage difference in the conversion rate generally and went up a positive over 30% increase in the difference in the conversion rate.

The visualization also captures the percentage difference in average spent and average conversion rates over time for both groups. The Control group initially exhibits a higher percentage difference in average spending compared to the Test group. However, this trend changes on January 31 and February 3 when the Test group experiences a surge in spending, leading to a higher positive percentage difference of 45.9% and 111.7% respectively but experienced a decline of a 24.4% difference the next day Feb 4 and kept fluctuating at a lower percentage afterwards. The Control group consistently shows lower percentage differences in average conversion rates compared to the Test group. On February 4 and February 6, the Test group demonstrates a notably higher positive percentage difference in conversion rates of 20.8% and 29.71% respectively, indicating a potential novelty effect or other factors influencing user behavior.

The fluctuations in percentage differences reflect changes in user behavior due to the novelty effect. The Test group shows higher positive percentage differences in both average spending and average conversion rates, suggesting that the introduction of the banner has influenced users to explore the food and drink category more actively.

The observed differences in percentage metrics between the Control group and the Test group highlight the presence of a novelty effect. This effect is indicated by the changing patterns in user

behavior over time, particularly the notable increase in percentage differences around the introduction of the banner. Further analysis is advised to understand the sustainability and long-term impact of these changes and to refine strategies accordingly. In brief, the percentage differences in average spending and average conversion rates provide a succinct picture of how the novelty effect is influencing user behavior over time in both groups.

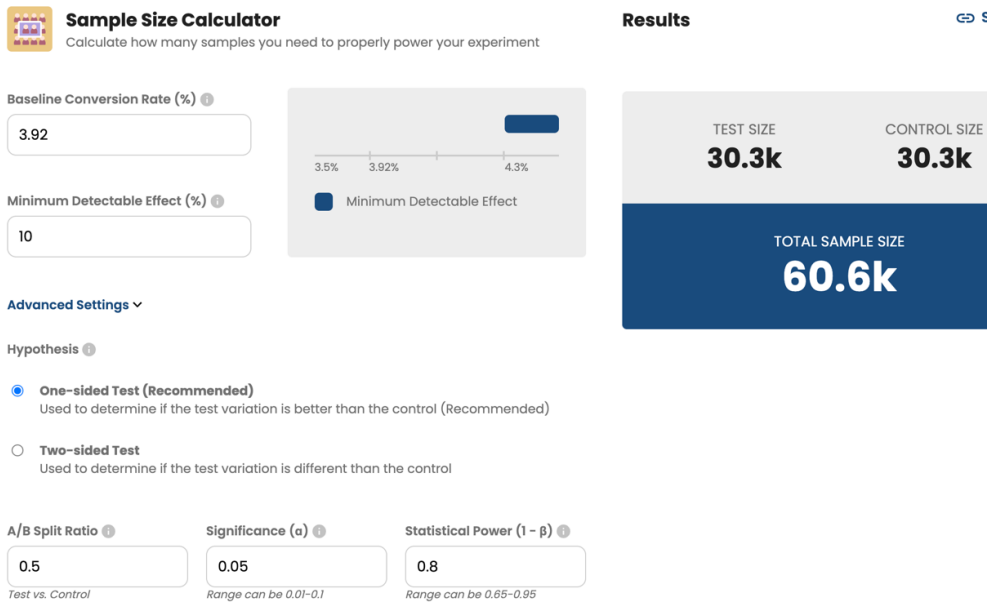
2.8 Power Analysis

A power analysis in this experiment is essential to assess the practical significance of differences in conversion rates and average spending between the control and test groups. Practical significance determines whether these differences are meaningful in a real-world context for GloBox. It helps answer questions like whether the increase in conversion rate justifies implementing the banner for all users and if the slight difference in average spending is meaningful for revenue growth.

The Minimum Detectable Effect (MDE) is a crucial concept in a power analysis as it defines the smallest effect size the experiment can reliably detect. It guides the determination of the required sample size for adequate statistical power, which is the probability of detecting a true effect if it exists. Smaller sample sizes reduce the experiment's power, making it less likely to detect real differences between the control and test groups, even if such differences exist in the population.

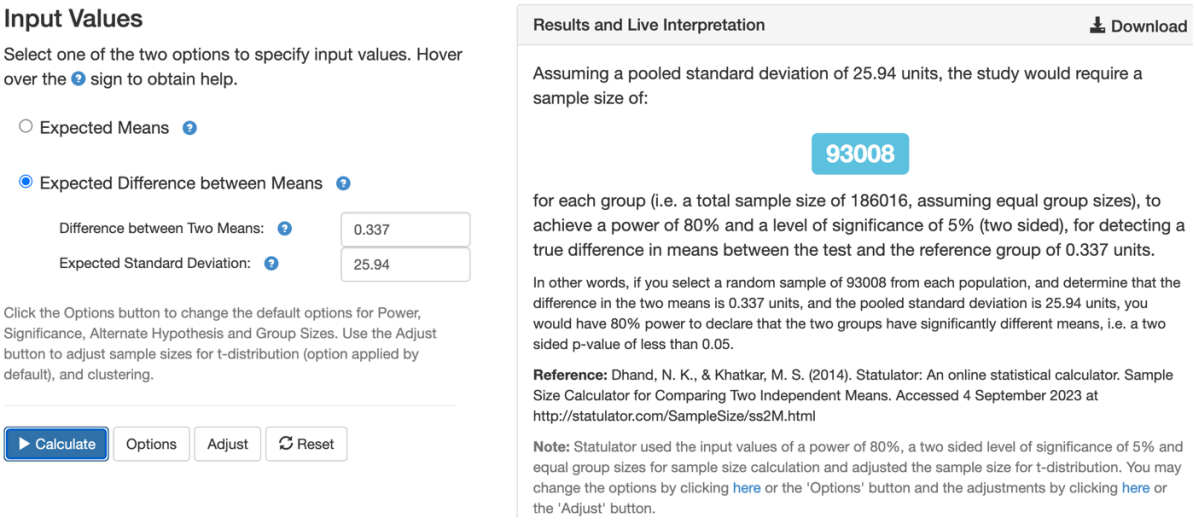
2.9 Sample Size Calculator

Figure 2.9.1: Conversion Rate Sample Size Estimation



SOURCE: Author's computation using Statsig, 2023.

Figure 2.9.2: Average Spent Sample Size Estimation



SOURCE: Author's computation using Statulator, 2023.

In Figure 2.9.1 and Figure 2.9.2, we used the control group as our baseline and a relative percent change of 10% for our Minimum Detectable Effect (MDE) to be able to estimate if we had enough sample size for our test to be sufficiently sensitive. The result indicates that the sample sizes used in our experiment are smaller than what is recommended for achieving a desired level of statistical power. For conversion rates, a sample size of 30.3k for each group is recommended, while for average spending, it's 93k for each group. However, our actual sample sizes are considerably smaller, with 24,343 in the control group and 24,600 in the test group. With an MDE of 10%, we want our experiment to be able to detect differences of at least 10% between the control and test groups. However, with smaller sample sizes, the experiment may not have the power to detect such differences accurately.

To improve the statistical power of our experiment and increase the ability to detect meaningful differences, we might consider increasing our sample sizes to be closer to what the power analysis suggests. This would involve expanding our control and test groups to match the recommended sample sizes for conversions and means.

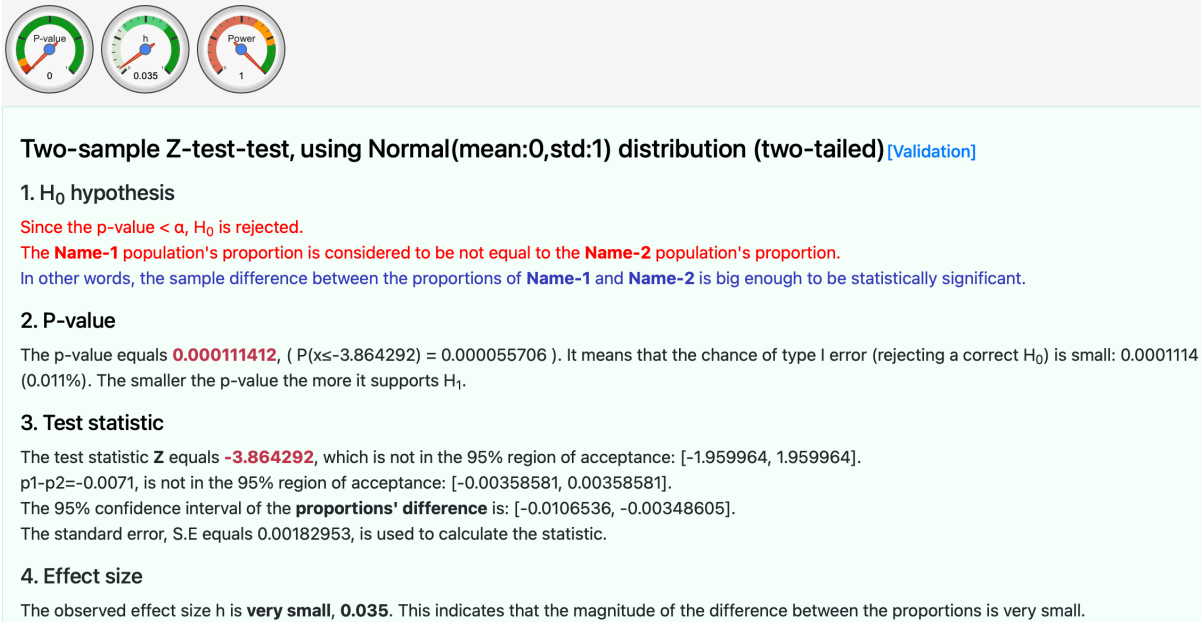
2.10 Effect Size

Effect size helps with interpreting the practical significance of results by providing a quantitative measure of the magnitude or strength of an observed effect. While statistical significance tells you whether an effect is likely to be real or not based on your sample data, effect size tells you how substantial or meaningful that effect is in practical terms.

Effect size quantifies the size or strength of an observed relationship or difference between groups. It answers questions like, "How much of an impact does this intervention have?" or "How strongly are these variables related?" By providing a numerical value, effect size allows you to gauge the extent of the effect.

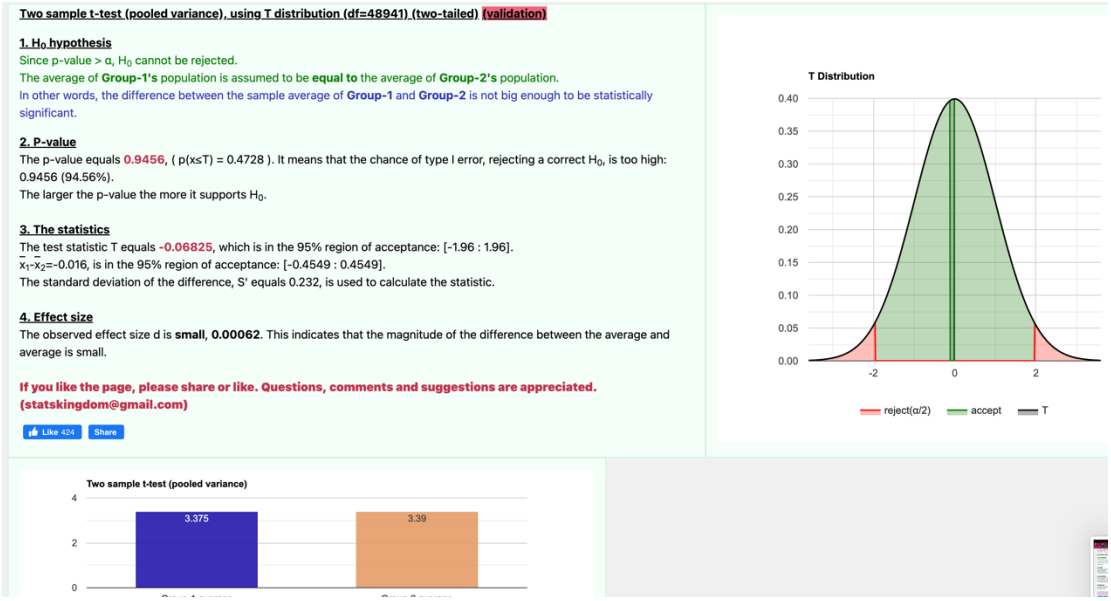
2.11 Effect Size Calculator

Figure 2.11.1 Conversion Rate Effect Size Estimation



SOURCE: Author's computation using Statistics Kingdom, 2023.

Figure 2.11.2 Average Spent Sample Size Estimation



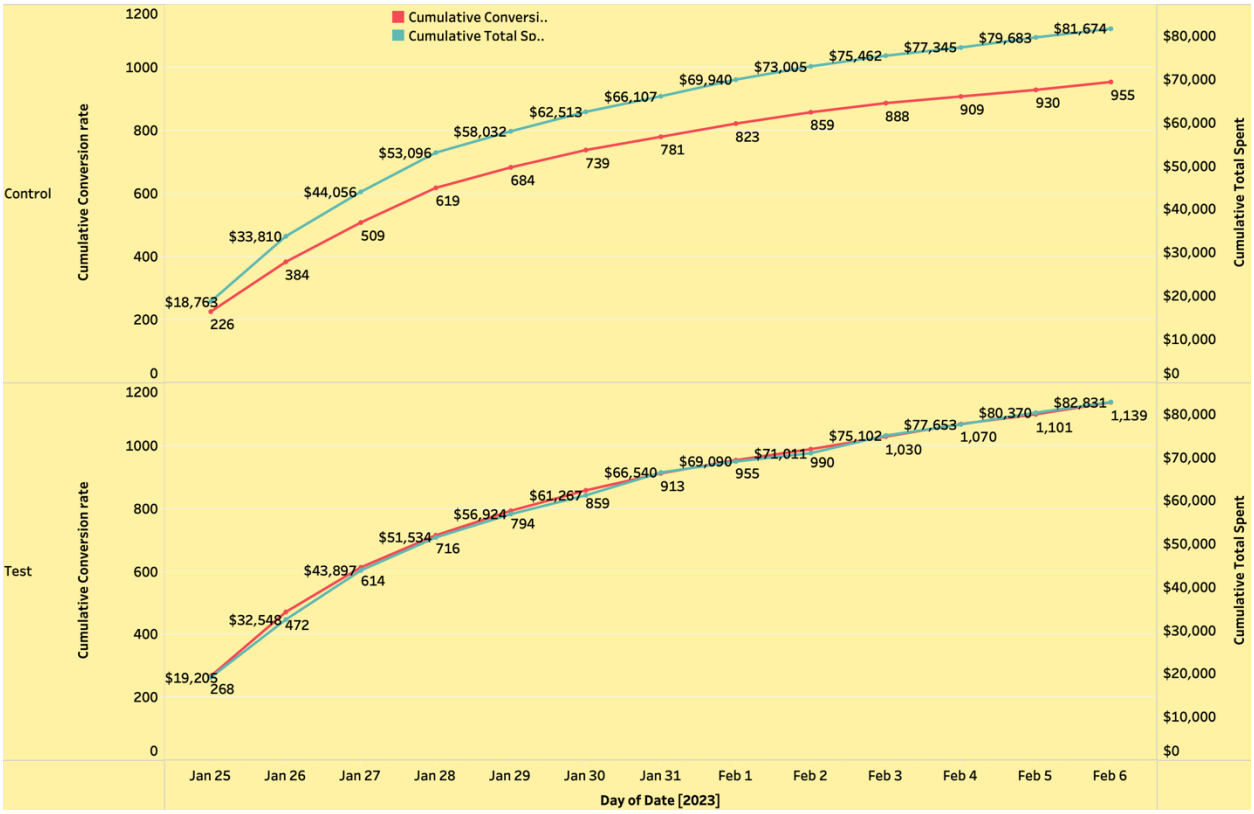
SOURCE: Author's computation using Statistics Kingdom, 2023.

In Figure 2.11.1 and Figure 2.11.2, the observed effect size for conversion rate is moderate. While value of 0.035 suggests that there is a measurable and meaningful difference in conversion rates between the control and test groups. This difference is more substantial compared to the effect size for average spending. The effect size for average spending is very small, indicating that there is a minimal practical difference in average spending between the control and test groups. In other words, the effect of the treatment (or change) on average spending is almost negligible.

In summary, the power analysis suggests that the experiment is well-powered to detect differences in conversion rates due to the moderate effect size. With such a small effect size, it is practically meaningless to draw conclusions about the impact of the test group on average spending.

2.12 Join Curve

Figure 2.12.1 Cumulative join curve for Conversion Rate and Average Spent



SOURCE: Author's computation using Tableau, 2023.

Figure 2.12.1 is a line graph showing the curvature of the cumulative join curve of the Control and Test group and their test metrics. The cumulative frequency graphs generated for the two experimental groups displayed a comparable level of increase over time. This suggests that, based on the chosen metrics, there may not be a statistically significant difference in the performance or behavior of the two groups. The similarity in cumulative frequency graphs implies that the observed data may not align with the assumptions made in the power analysis. The power analysis typically assumes that there will be significant differences between groups if the recommended sample sizes are used. However, our experiment did not yield substantial differences.

The findings from the cumulative frequency graphs indicate that the Globox banner experiment may not have achieved the desired level of statistical power to detect significant differences between the groups. This could be due to factors such as the actual effect size being smaller than expected or the sample sizes used in the experiment being insufficient. Reevaluate the sample sizes used in the experiment to ensure they align with the expected effect size and desired power level. Adjusting the sample sizes may improve the experiment's ability to detect meaningful differences if they exist.

CHAPTER THREE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

3.1 Summary of the Experiment

The study focused on an A/B test conducted by GloBox, an e-commerce company that sought to enhance its revenue by featuring a banner showcasing food and drink products on its website. The goal was to understand the impact of this banner on user behavior, conversion rates, and average spending. Data analysis involved aggregating user-level information, performing hypothesis tests, and creating visualizations.

3.2 Implications of Findings

The findings of the study revealed several key insights. The A/B test results indicated that there was a statistically significant difference in conversion rates between the Control group (no banner) and the Test group (with banner). This suggests that the banner had a positive impact on user engagement and conversion rates. However, the study did not find a statistically significant difference in average spending between the two groups, indicating that the banner's impact on spending was not significant.

The analysis of the novelty effect highlighted that there was a noticeable increase in both average total spending and conversion rates shortly after the introduction of the banner. This suggests the presence of a novelty effect, where users initially engage more due to the new feature. However, further investigation is needed to determine if this effect is sustained over time.

While the cumulative frequency graphs indicate similarity in the performance of the two groups, further examination of the experiment's design, sample sizes, and choice of metrics is warranted to refine our understanding of the Globox banner's impact. Adjustments may be necessary to achieve more conclusive results.

3.3 Conclusion and Recommendations

The experiment suggests that the banner showcasing food and drink products on the website had a positive impact on conversion rates, but its effect on average spending was not significant. The presence of a novelty effect indicates that users responded positively to the introduction of the banner, resulting in increased engagement and conversions. Based on the findings, these recommendations are proposed:

i. Given the potential novelty effect, it's recommended to continue monitoring user behavior over an extended period. This will help determine whether the initial spikes in engagement and conversions are sustained or if there's a decline after the novelty wears off.

ii. The analysis of device types and user countries suggests that different user segments respond differently to the banner. It's recommended to tailor marketing strategies based on these insights to maximize engagement and conversions.

iii. Since the banner positively impacted conversion rates, focusing on refining strategies that contributed to this increase is recommended. Strategies that encourage users to convert could be further optimized to drive sustained results.

iv. Considering the varied impact of the banner on different user segments, exploring personalized recommendations for users based on their preferences and past behaviors could enhance overall user experience and drive conversions.

v. To leverage the novelty effect, the company could consider implementing time-limited promotions, exclusive offers, or limited-edition products to maintain user engagement and interest.

To conclude, the A/B test results and the analysis of the novelty effect provide valuable insights for GloBox's decision-making process. It's common to see an initial peak in user engagement and

behavior as users explore the new feature, followed by a decline as the novelty wears off. This pattern can be attributed to the initial interest generated by the introduction of the banner. It's recommended to continue monitoring user behavior over an extended period. This will help distinguish between the temporary surge due to curiosity and any sustained changes in user engagement. Analyzing data beyond the provided timeframe could reveal whether the observed pattern persists or if there's a shift toward more stable behavior in terms of spending and conversions.

APPENDIX

1. SQL Query:

Query to get user-level analysis dataset GloBox Dataset Query

Query for checking novelty effect GloBox Novelty Effect Query

2. GloBox Dataset

View the GloBox Dataset ,Hypothesis Test and Confidence Interval Test GloBox Dataset

3. Tableau Visualization

Visualization on Experiment Groups vs Test metrics in relation to Users Gender, Device type and Country Control/Treatment Group Visual Analysis

Confidence Interval Visualization CI Viz

Novelty Effect Visualization Novelty Effect Viz

Join Curve Cumulative Join Curve

4. Power Analysis

Sample Size Calculator

Conversion Rate

Average Spent