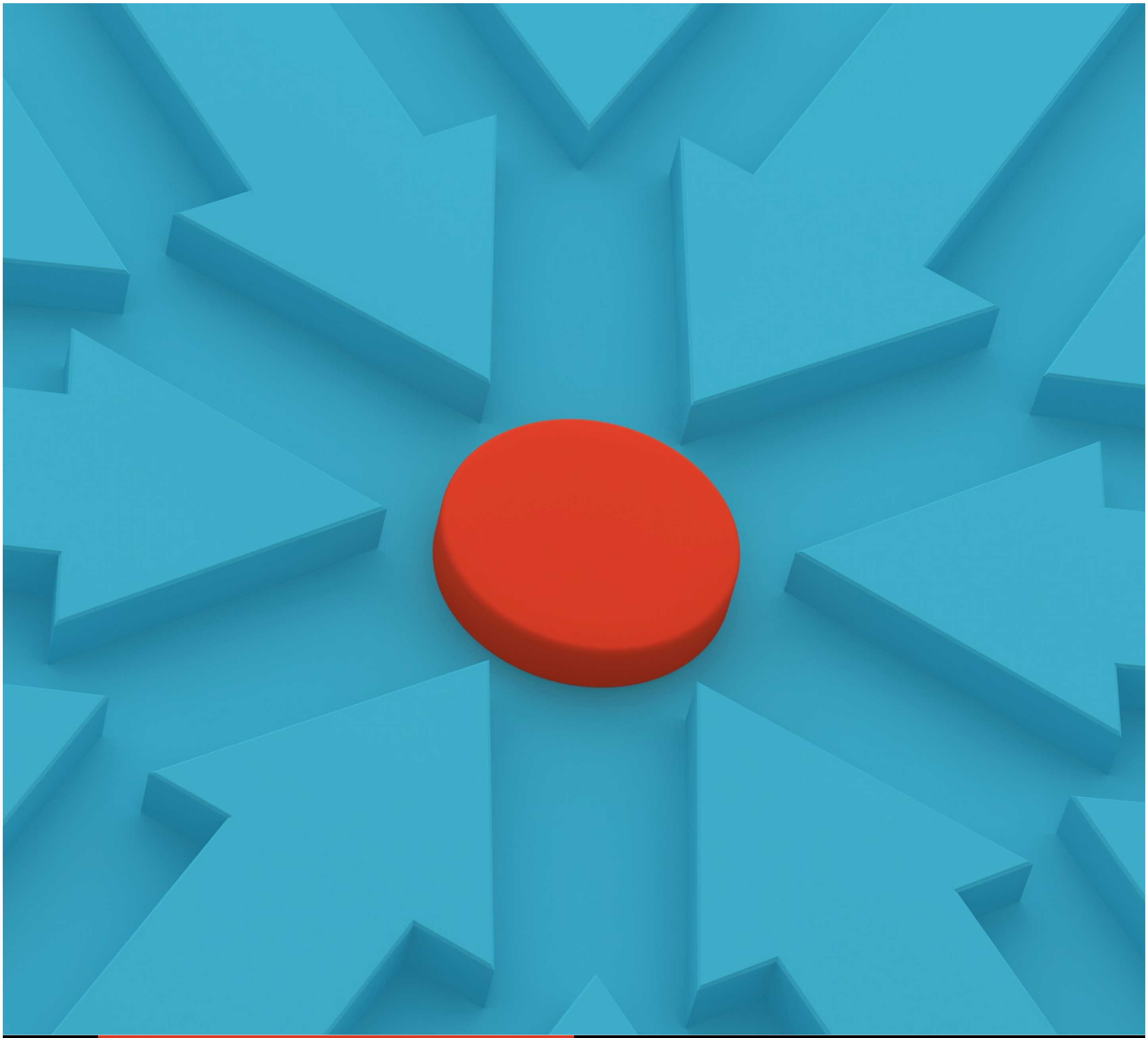


A/B TEST ANALYSIS

GLOBOX FOOD & DRINKS BANNER

Submitted By: Himanshu Saini

MASTERSCHOOL

SUMARY

The thorough analysis of the A/B test outcomes and the associated insights, it is advisable to proceed with the deployment of the banner. The A/B test has revealed a statistically significant variance in conversion rates between the control and treatment groups, suggesting a positive influence of the banner on conversion metrics.

Introducing the banner is expected to make food and drink products more noticeable, leading to more people making purchases. However, it's important to mention that we haven't seen any clear changes in the average amount customers spend. To truly understand how the banner affects what users spend, it's suggested to keep an eye on the average transaction amounts consistently. Since we don't have definite proof yet, it might be a good idea to think about doing more experiments or increasing the number of people we're studying. This careful approach will help us draw stronger conclusions about any differences in results, allowing us to make informed decisions in future projects.

CONTEXT

Introduction

Welcome to GloBox, a premier destination in the online marketplace for distinctive and premium products sourced globally. As part of their commitment to providing an unparalleled shopping experience, they are currently conducting an A/B test experiment aimed at enhancing their homepage to increase the awareness of food and drink offerings to increase the revenue.

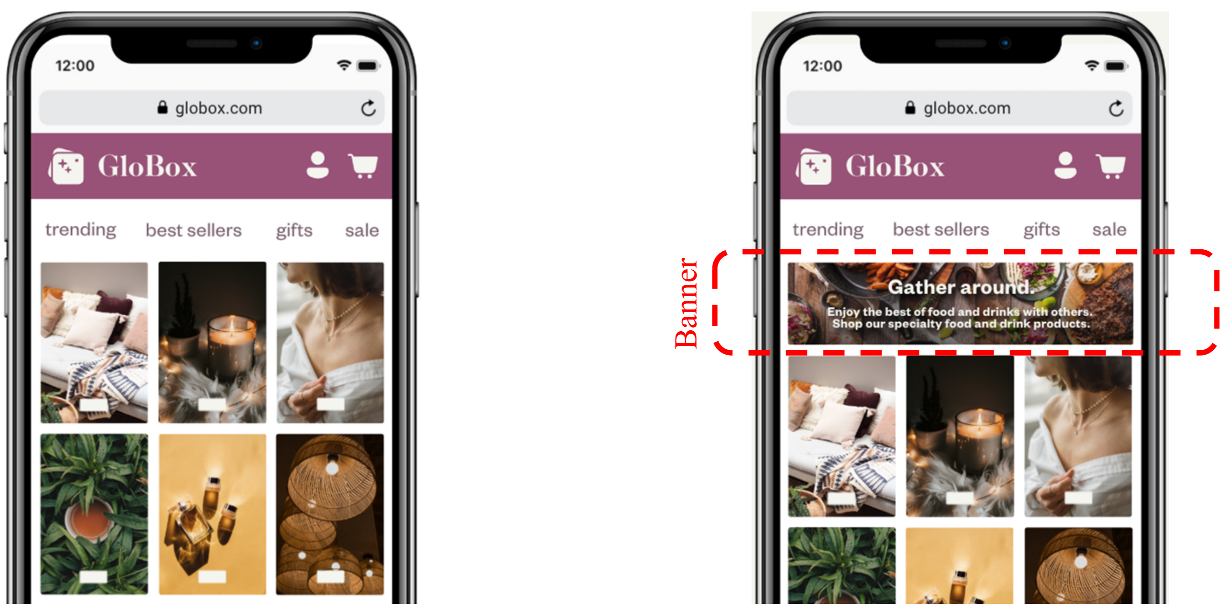
A/B Test

A/B testing is smart strategy businesses use to compare two different versions of a webpage, ad, or product feature and figure out which one works better. It works by randomly putting customers or users into either the A or B version, helping the business see which version is more successful in reaching a specific goal. It's like trying out two options to see which one does the job better.

To conduct the test growth team has decided to highlight key products in food and drinks category as a banner at the top of their website. The control group lands to the existing page whereas the treatment group gets the updated page with banner, as shown in Figure 1.

Group A: Control Group
Present Landing Page

Group B: Treatment Group
Experimental Banner Page



(a) (b)

Figure 1. Page layout for (a) Control and (b) Treatment Group.

The experiment takes place on the mobile website, where all users visiting the main GloBox page are randomly assigned to either the control or treatment group. The date of their website visit is considered their joining date. Subsequently, users who make one or more purchases either on the same day of joining or later subsequent days are classified as a "conversion". Additionally, the analysis includes the examination of the average user spending on products. Furthermore, a detailed analysis has been conducted to capture user behavior based on gender, the device they used, and the country to which they belong.

METHODOLOGY

The following steps are followed in this study:

1. Organizing data and cleaning: SQL
2. Statistical analysis: Spreadsheets
3. Visualization: Tableau

Data Extraction & Cleaning

GloBox stores its data in a relational database, therefore data has been extracted and processed by using Structured Query Language (SQL). The data set contains three tables named as `users`, `groups`, and `activity`. Figure 2 illustrate the table and its field with datatypes.

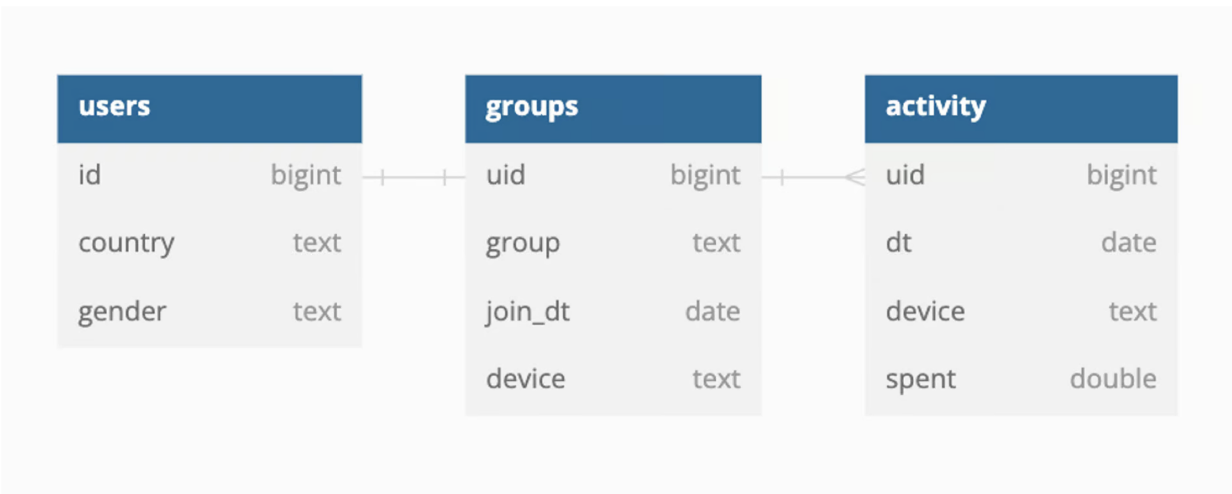


Figure 2: Database diagram.

- `user` table: It contains the demographic information of all the users.
    - id: the user ID.
    - country: ISO 3166 alpha-3 code.
    - gender: the user's gender.
- `grouops` table: it contains the user's A/B test group assignment.
    - `uid`: the user ID.
    - `group`: the user's test group.
    - `join_dt`: the date user joined the test.
    - `device`: the device user used to visit the page.

- `activity` table: It contains user purchase activity, containing 1 row per day that a user made a purchase.
    - `uid`: the user ID.
    - `dt`: date of purchase activity.
    - `device`: the device types of the user purchased on (I = iOS, A = Android).
    - `spent`: the purchase amount in USD.

The statistical parameters are then calculated using the SQL queries on these tables the conversion rate (CR) and the average amount spent. The formula for the CR is given below.

$$CR = \frac{Total\ number\ of\ users\ converted\ in\ the\ group}{Total\ number\ of\ users\ in\ the\ group} \times 100\ \%$$

Statistical Analysis

The statistical analysis has been done by using the spreadsheet for both the parameters conversion rate and the average amount spent.

- Conversion Rate Analysis: The hypothesis has been made with null hypothesis that there is no significant difference between the conversion rate of both the groups (Control and Treatment group). Whereas the alternative hypothesis says that there is a significant change in the conversion rate. The mathematical expressions are,

  $$\boldsymbol{H_0: p_1 - p_2 = 0}$$

  $$\boldsymbol{H_1: p_1 - p_2 \neq 0}$$

  where, p_1 and p_2 are the conversion rate in control and treatment group respectively.
  To perform the analysis the z-test has been used,

  $$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{P * (1 - P) * (\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{(\hat{p}_1 - \hat{p}_2)}{SE}$$

- Average Spending Analysis: Similarly, the average amount spent has been analyzed. The mathematical expressions of hypothesis are given,

  $$\boldsymbol{H_0: \mu_1 - \mu_2 = 0}$$

  $$\boldsymbol{H_1: \mu_1 - \mu_2 \neq 0}$$

  where, μ_1 and μ_2 are the average spending of control and treatment group respectively. Two-sided t-test has performed in this case.

  $$\boldsymbol{T = \frac{(\overline{x}_1 - \overline{x}_1) - \mu_0}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}}$$

Visualization

To get the insights of the data the Tableau has been used to visualize the data. All the statistical parameters for example, conversion rate, average amount spent per user, average spending distribution, relationship with respect to gender, device, and country are all visualized. These graphs helped to find the main characteristics of a dataset, gain insights, and uncover patterns or anomalies in the data.

RESULTS

GloBox Dataset

The GloBox dataset extraction process utilized the tables outlined in the Methodology/Data Extraction & Cleaning section employing SQL techniques. In the initial data extraction phase, field values were cleaned using the `COALESCE()` function. Specifically, fields from `user.country`, `user.gender`, `group.device`, and `group.group` were considered, and any `NULL` values encountered were systematically replaced with the designated label `"Unknown".`

To calculate the conversion rate, a new field named `"usr_conversion"` was generated with the following logic: `if a user made a purchase, they were assigned a value of 1; otherwise, a value of 0 was assigned`. This new field, thus, contains binary values (`0` or `1`) representing the occurrence of a user making a purchase. The `"spent"` field has been reformatted using the `CAST()` function, ensuring precision up to three decimal places. This adjustment enhances the accuracy and clarity of the expenditure data. Subsequently, the final organized and cleaned dataset has been extracted and utilized for further data analysis. For more information see Appendix.

Data Analysis

Inferential Statistics

1. **Conversion Rate:** The conversion rate, defined as the ratio of users converted to the total number of users in each group as illustrated in Figure 3, was analyzed using a proportion two-tailed z-test in this instance, given the comparison of two samples. The data was organized using a pivot table in a spreadsheet. Hypotheses were formulated, with the Null hypothesis (H_0) positing "`no significant difference in the conversion rate between the groups`," and the alternative hypothesis (H_1) stating "`a significant difference in the conversion rate between the groups`." A significance level of 0.05 was chosen, resulting in a calculated p-value of 0.0001. The 95% confidence interval was determined to be 0.0035 to 0.0107. The calculated parameters for the test are presented below.

Calculation			
Statistics Parameters	**Notation**	**Control Group (A)**	**Treatment Group (B)**
Sample Size	n1, n2	24343	24600
Number of Conversions	x1, x2	955	1139
Conversion Rate	$\hat{p}_1, \hat{p}_2$	0.03923099	0.046300813
Standard Error (SE)	SE1, SE2	0.001244334	0.001339777
Pooled Proportion	P	0.042784464	
Difference in Conversion Rate	$\hat{p}_1 - \hat{p}_2$	0.007069823	
Total SE (Pooled Proportion)	SE	0.001829526	
Degree of Freedom	df	24342	
z-Test	z	3.86429177	
z Critical Value for 95% CI	z*	1.96	
p-value	p_val	0.000112	
Margin Error	ME	0.003585871	
95% Confidence Interval with α = 0.05	CI	Lower Bound	Upper Bound
		0.003483951	0.010655694

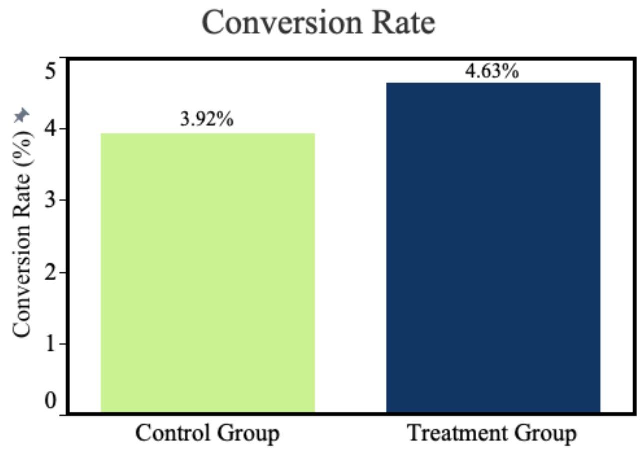


Figure 3: Conversion Rate between the groups.

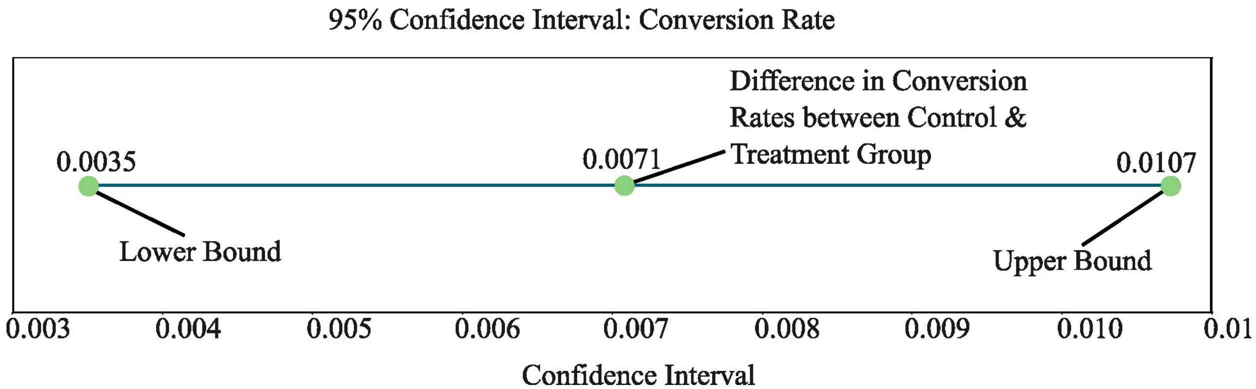


Figure 4: Illustration of confidence interval for conversion rate.

The statistical analysis results (`p-value = 0.0001 < 0.05`) lead to the **rejection of the null hypothesis**, indicating a significant difference in conversion rates between the two groups. The 95% confidence interval for the difference in conversion rates (0.0035 to 0.0107) as shown in Figure 4, for both the control and treatment groups suggests that repeating the experiment multiple times would likely yield differences within this range in 95% of cases. This analysis implies that launching the banner may positively impact user conversion rates between the groups. However, for a comprehensive understanding of the long-term impact and potential variations, additional studies may be warranted.

2. **Average Amount Spent:** For the average spending analysis (in Figure 5), a `two-tailed t-test` was employed. The hypotheses were as follows: the Null hypothesis (`H0`) suggested "`no significant difference in the average spent between the groups`," while the alternative hypothesis (`H1`) proposed "`a significant difference in the average spent between the groups.`" Using a significance level of 0.05, the calculated p-value was 0.944. The 95% confidence interval as shown in Figure 6, was established as -0.44 to 0.47. The parameters for the test are detailed below.

Calculations			
Statistics Parameters	**Notation**	**Control Group (A)**	**Treatment Group (B)**
Sample Size	n1, n2	24343	24600
Average spending	$\bar{x}_1,\ \bar{x}_2$	3.374518506	3.39086687
Sample Standard Deviation	s1, s2	25.93639031	25.41410952
Difference in average spending	$\bar{x}_1 - \bar{x}_2$	0.016348364	
Standard Error	SE	0.232140557	
Degree of Freedom	df	24342	
T-Statisctics	t	-0.070424418	
$p - value\ \approx 2\ \times P(T < -0.07042)$	p_{val}	0.943856435	
p-value using T.TEST Function	p_{val}	0.943856145	
Confidence Level	$1 - \alpha(=0.05)$	0.95	
Critical Value for t	$t_{critical}$	± 1.960	
Margin Error	ME	0.454995492	
95% Confidence Interval with α = 0.05	CI	Lower Bound	Upper Bound
		-0.438647129	0.471343856

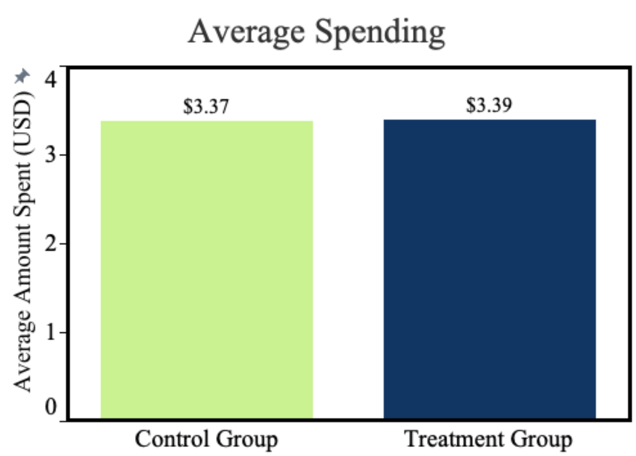


Figure 5: Average spending between the group.

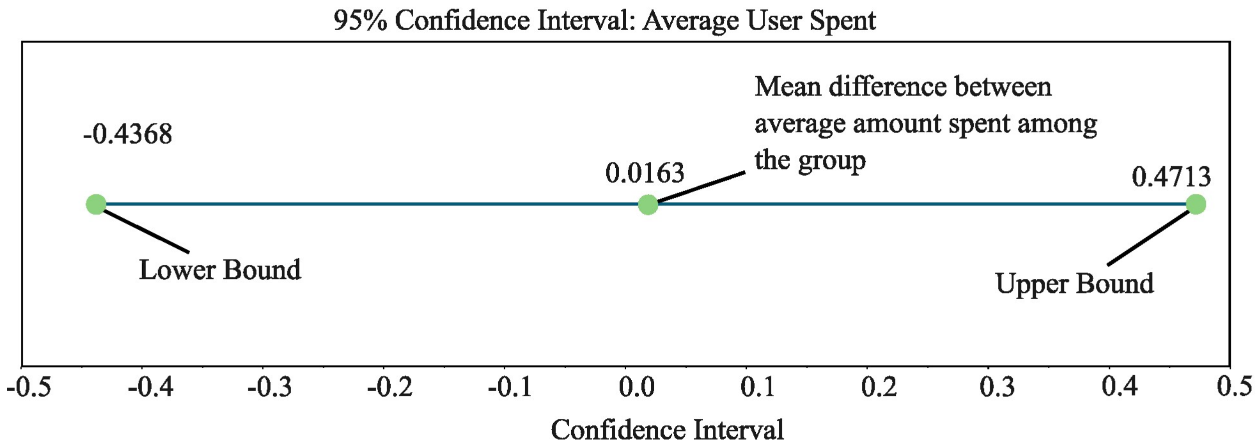


Figure 6: Illustration of confidence interval for average user spent.

It's noteworthy that the high `p-value` and the confidence interval spanning zero indicate a **lack of evidence to reject the null hypothesis**, suggesting no significant difference in average spending between the two groups. Further investigation may be required to explore any nuances or potential factors influencing spending behavior.

Average Amount Spent Distribution

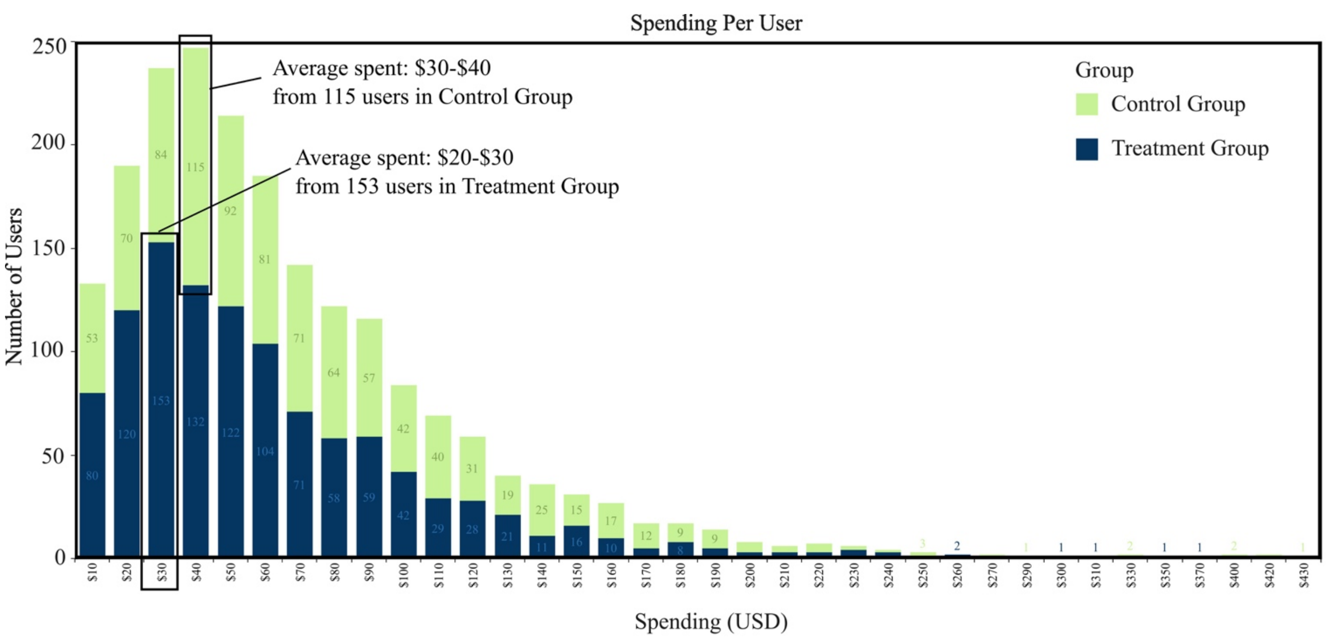


Figure 7: Average spent distribution plot.

The distribution graph of average spending, depicted in Figure 7, illustrates that within the control group, 115 users have an average spending range of $30-$40. In contrast, the treatment group shows 153 users with an average spending range of $20-$30. This graphical representation suggests that there is not a substantial difference in spending behavior between the control and treatment groups, aligning with the findings of the statistical analysis. The visual evidence reinforces the conclusion that the average spending patterns in both groups are comparable.

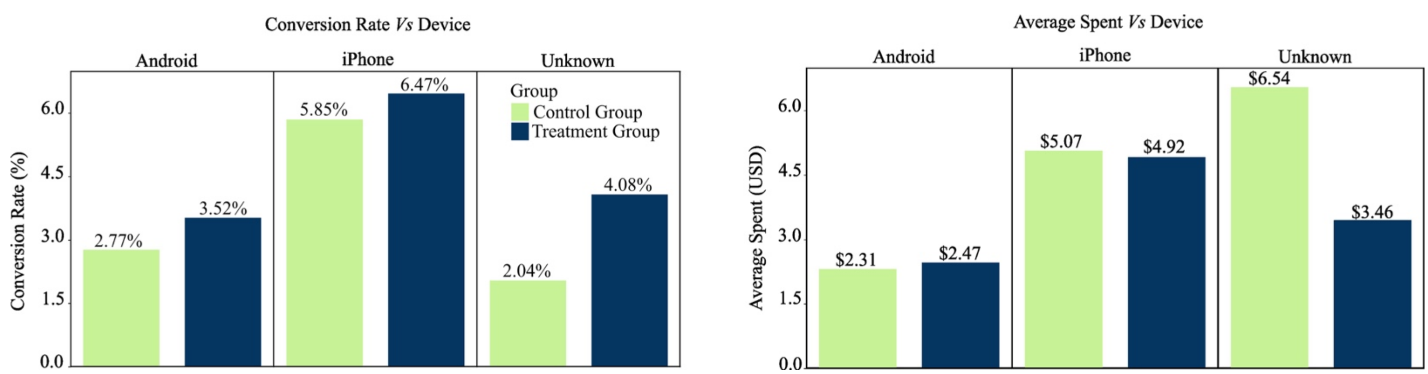
Comparison of Statistical Parameters

To gain deeper insights, additional analysis was conducted by projecting statistical parameters concerning device usage, gender, and regional factors.

In Figure 8, statistical parameters are presented based on the devices users used to visit the page, which are categorized into three groups: Android users, iPhone users, and unknown devices. The unknown category is excluded from the analysis to ensure an unbiased assessment, as it encompasses various devices.

Observations indicate that iPhone users dominate the conversion rate, accounting for 6.47%. However, when analyzing the relative increment in conversion rate concerning the control group, it becomes evident that Android users exhibit a higher overall conversion rate at 27%, surpassing the 10.6% rate observed among iPhone users. The relative increment is calculated using the provided formula, revealing valuable insights into the varying impacts of device types on conversion rates.

$$CR_{relative} = \frac{CR_{Control\ Group} - CR_{Treatment\ Group}}{CR_{Control\ Group}} \times 100\%$$



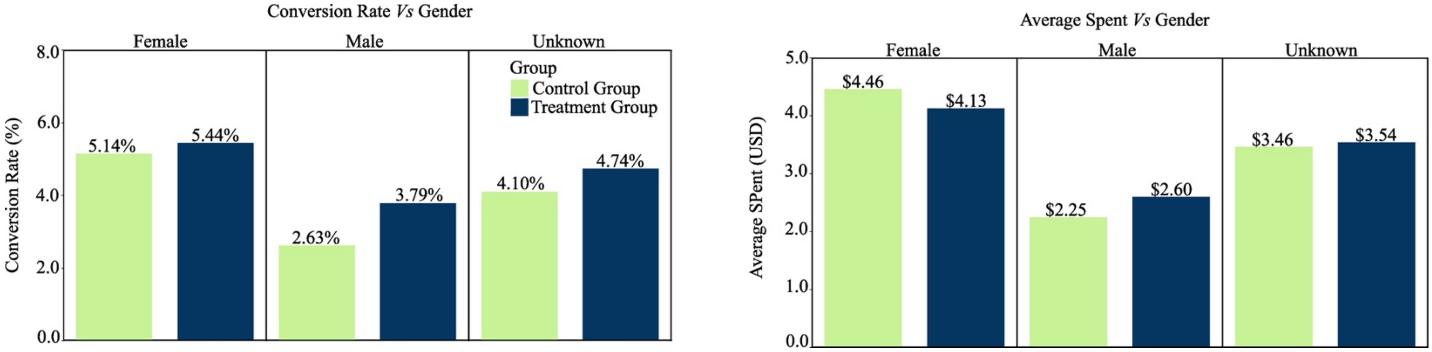
(a) (b)

Figure 8: Illustration of statistical parameters with user device. (a) Conversion Rate, (b) Average Spent.

Furthermore, The average spending analysis also shows the relative change of android users (0.7%) were slighter higher than the iPhone users (0.3%).

Likewise, an analysis was conducted with respect to gender (in Figure 9), revealing a relative improvement in conversion rates of 5.83% for females and 44.10% for males. This indicates a significant outperformance by males, exceeding females by 38.30%. In terms of average spending, males once again outperformed females, showcasing a 15.55% increase, while females exhibited a decrease of -7.4%.

These findings underscore gender-based distinctions in both conversion rates and spending patterns, providing valuable insights for targeted strategies and decision-making in optimizing user experiences.



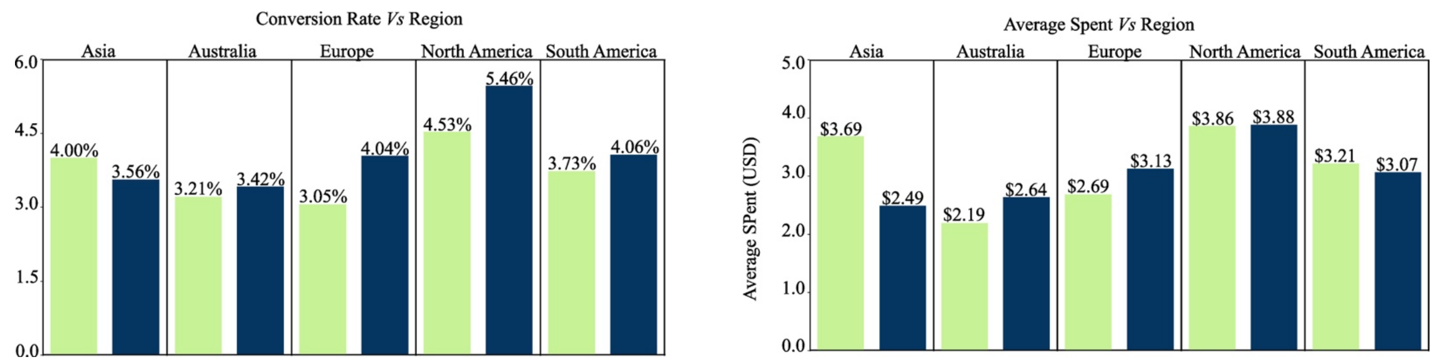
(a) (b)

Figure 9: Illustration of statistical parameters with gender. (a) Conversion Rate, (b) Average Spent.

The country-wise analysis in Figure 10 highlights North America's dominance, leading in both the control group with a 4.53% conversion rate and the treatment group with 5.46%. Similarly, North America shows consistency in average spending, recording 3.86% in the control group and 3.88% in the treatment group. Overall, the statistical parameters exhibit positive trends across regions, with the exception of Asia, suggesting areas for improvement in that region. Notably, South America experiences a decrease in average spending.

In terms of relative improvement, Europe stands out with a substantial positive increment of 32.45% in conversion rates, indicating noteworthy progress. Additionally, Australia leads in average spending, demonstrating a significant relative improvement of 20.54%.

These insights provide a nuanced understanding of regional variations, enabling targeted strategies to enhance conversion rates and spending behaviors.



(a) (b)

Figure 10: Illustration of statistical parameters with region. (a) Conversion Rate, (b) Average Spent.

Novelty Effect

The novelty effect examines user behavior in response to a new change, gauging the initial reactions of users. These early responses have the potential to introduce bias into the results. Consequently, it is prudent to assess and account for the novelty effect in the context of A/B testing to ensure the accuracy and reliability of the outcomes.

To thoroughly assess the novelty effect, a more in-depth analysis has been conducted. Figure 11 displays the distribution of conversion rates over time, with the weighted curve representing the number of users. While there are noticeable fluctuations in the curve, the overall trend remains consistent. This observation leads to the conclusion that no novelty effect is evident in the conversion rate. The conversion rate for the Control group is

3.92%, while the Treatment group shows a slightly higher conversion rate at 4.63%. Additionally, the trend line indicates a consistent increase in the conversion rate within the treatment group over time. This analysis suggests a stable and positive trajectory, further supporting the absence of a novelty effect in the conversion rate.

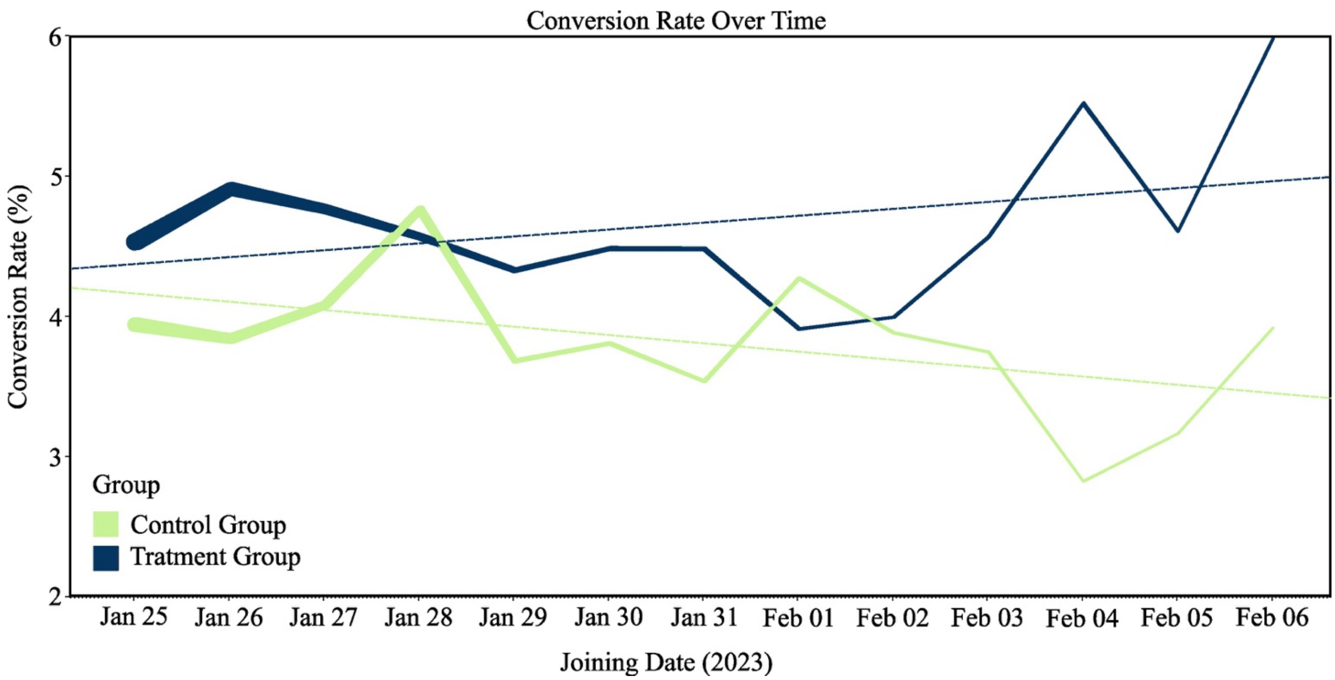


Figure 11: Conversion rate over time for control & treatment group.

Likewise, additional analysis was conducted with average spending over time (Figure 12) and average spending specifically for converted users over time (Figure 13). The average spending for the Control Group is $3.37, and for the Treatment Group, it is $3.39. The findings from both graphs indicate a lack of discernible novelty effect.

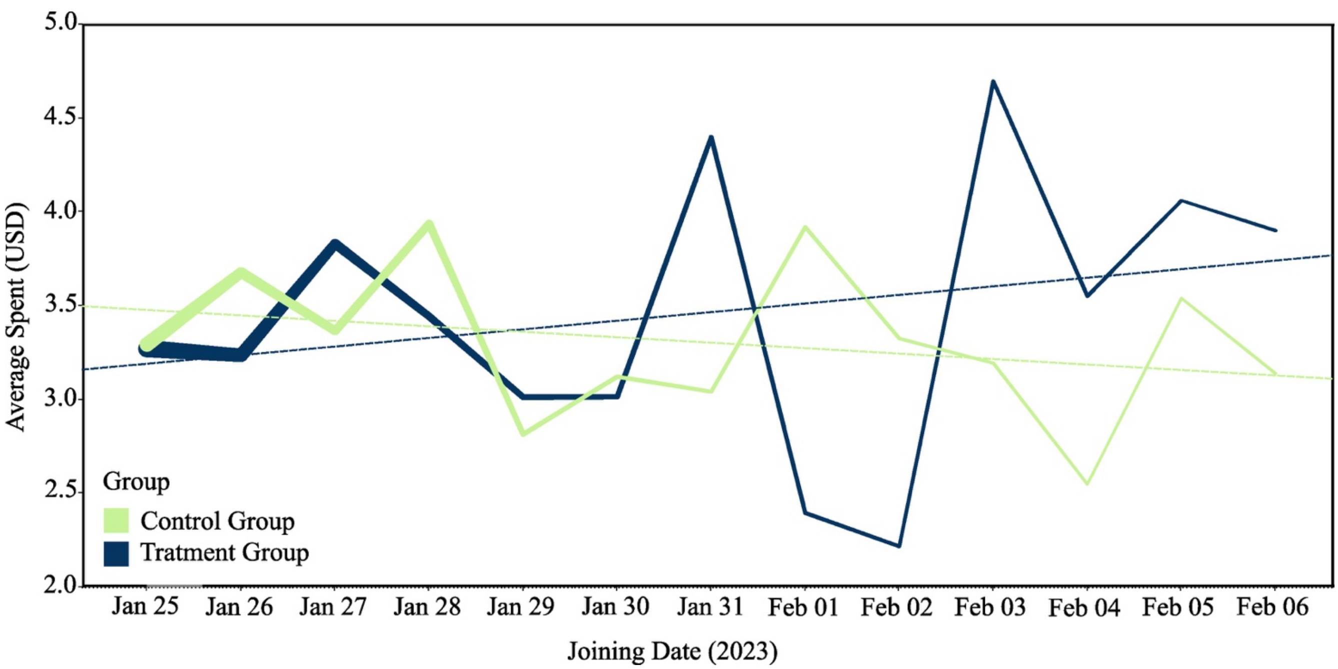


Figure 12: Average amount spent over time for control & treatment group.

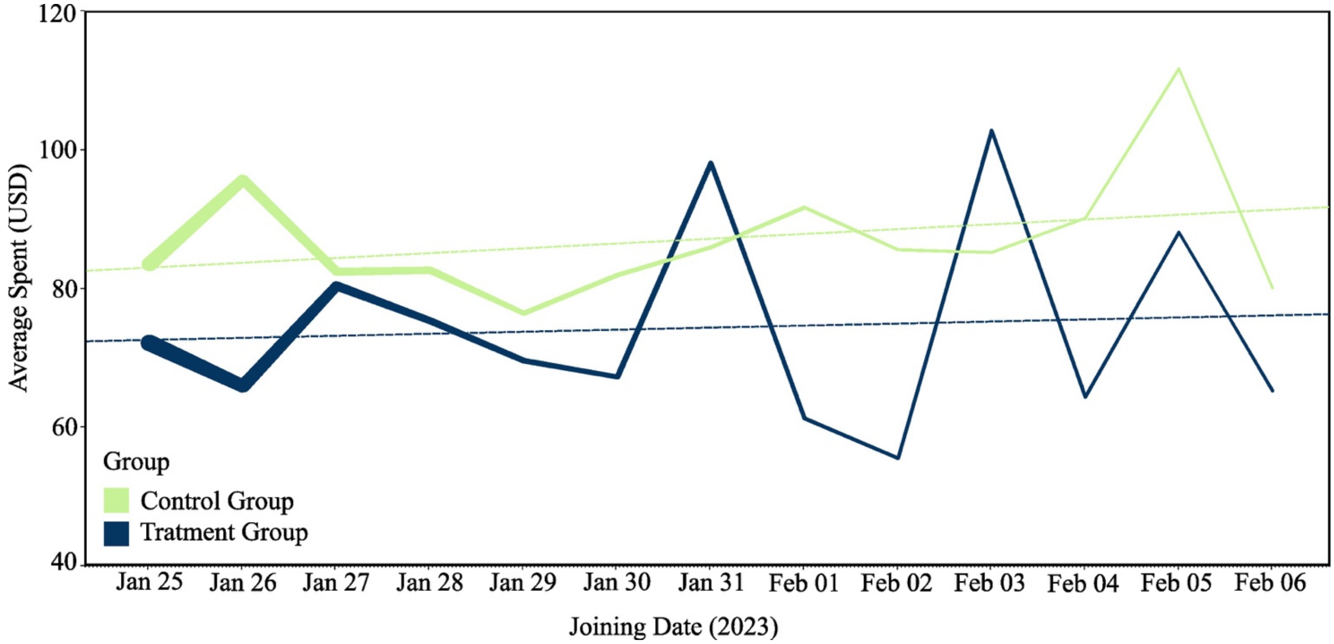


Figure 13: Average amount spent over time for control & treatment group with converted users.

Additionally, the analysis of the time taken for users to be converted as shown in Figure 14, further supports the absence of a novelty effect in the data. While there is a noticeable increase in same-day conversions compared to other days, this pattern is consistent across both groups, indicating no significant shift in user behavior. Therefore, a confident conclusion can be drawn that there is no evidence of a novelty effect in this study, evident in both the conversion rate and average spending within both groups.

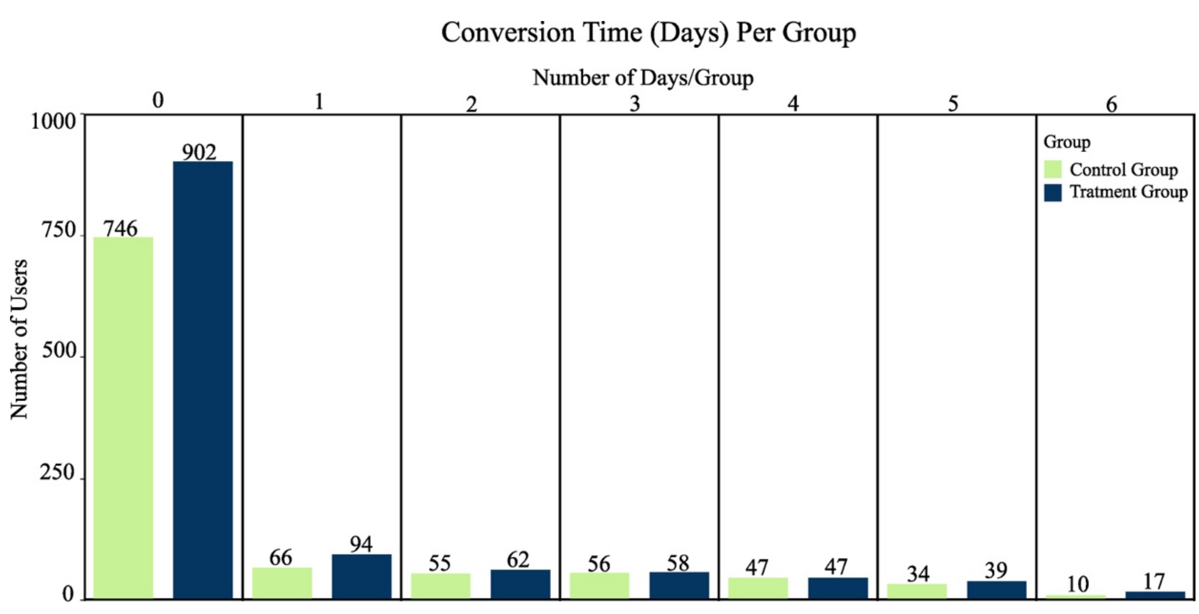


Figure 14: Number of days taken by users to be converted.

Power Analysis

Power analysis in A/B test estimates the probability of detecting the true effect if exist. The online calculator has been used to estimate the sample size both for conversion rate and average amount spent.

We started with a baseline conversion rate of 3.92% for the control group and aimed to detect a minimum effect of 10%, requiring a sample size of 77.4K for 80% power. After observing an 18% difference in conversion rates, recalculating based on this figure gave us a sample size of 24K. The match between this recalculated size and our actual sample size of 24,343 boosts our confidence in stating that our conversion rate results are accurate.

Furthermore, a power analysis was carried out for the average amount spent, revealing a necessity for a sample size of 91K to detect a 10% difference in the mean. The calculation involved considering the difference mean as 10% of the control group mean, with a pooled standard deviation value of 25.67. This analysis was executed for a two-sided t-test at a significance level of 5%. However, given our observed mean difference of 0.016, the calculated sample size is considerably larger, leading to negligible practical significance.

In conclusion, the sample size used to assess the conversion rate effect was sufficient to validate the results. On the other hand, for capturing the difference in mean, the required sample size is impractically large, and the calculated mean difference would be negligible in practical terms.

Recommendation: Launch the banner

1. Based on the analysis of the average spending distribution graph (Figure 7), both the control and treatment groups exhibited spending in the range of $20-$30 and $30-$40, respectively and the conversion rate between the group (Figure 5) where control group shows 3.92% and treatment group has 4.63%. This insight suggests that implementing more effective marketing strategies could attract a higher number of users ultimately leading to increased revenue. Continuous monitoring of the spending patterns and conversion rate can provide ongoing insights for the development of more effective marketing strategies. Furthermore, this data can inform the selection of products to be featured in the new banner, optimizing its content for enhanced user engagement and potential conversion.

2. Android users are showing a big boost in conversion rates, up by 27%, while iPhone users are improving by 10.6% as shown in Figure 8. On the other hand, there is no significant change in relative spending. In Figure 8 the relative spending for android and iPhone users are 0.7% & 0.3% respectively. This suggests a chance for the team to pay extra attention to engaging iPhone users. Understanding the patterns of iPhone users is key. By putting these insights into the new banner, like making a dynamic banner that adjusts to the user's device and shows items based on their past purchases, we can step-up user engagement. This smart strategy is likely to grab the interest of iPhone users, leading to more interaction and, in the end, boosting revenue.

3. In the gender-based analysis depicted in Figure 9, there's a notable difference in spending patterns between males and females. In terms of average spending, males outperformed females, showcasing a 15.55% increase, while females exhibited a decrease of -7.4%. Similar, improvement was shown in conversion rates of 5.83% for females and 44.10% for males. To optimize engagement, the team should concentrate on understanding and catering to the spending preferences of male users. This could involve tailoring the banner to showcase products that align with male interests and purchasing behaviors. Conversely, for female users, incorporating more health-oriented items could be a strategic approach. By recognizing and adapting to these gender-specific preferences, the team can tailor the content of the banner to better resonate with both male and female users, ultimately enhancing user engagement and potentially increasing conversions.

4. The region-wise analysis presented in Figure 10 highlights distinctive trends. Europe emerges as a promising market with a substantial positive increment of 32.45% in conversion rates and 16.35% in relative average spending, signaling the need for thoughtful consideration of the products featured on the banner to enhance conversions in this region. Conversely, Asia exhibits a decline in conversion rates with recorded value of -12.35% as well as in average user spending with -32.12%, presenting an opportunity for the team to delve deeper into the Asian markets. Targeting more localized products, specifically tailored to each region's preferences, could be a strategic move. For instance, initiating connections with local food chains might serve as an effective starting point to resonate with the local audience and potentially boost conversions in Asia and South America.

CONCLUSION

In conclusion, the supported data regarding the conversion rate strongly favors the profitability of launching the banner. The statistical verification of the analysis and the additional insights suggest a considerable potential for increasing the conversion rate through the banner. While the analysis of average amount spent did not reveal a significant difference between groups, continuous monitoring and adjustments to the banner could contribute to revenue growth. The novelty analysis aligns with user behavior, indicating no discernible novelty effect. Additionally, the power analysis provides further confidence in the conversion rate analysis, affirming that the banner is likely to boost conversion rates with a high degree of certainty.

APPENDIX

1. **SQL:** All the SQL queries related to this project are uploaded to GitHub.
2. **Spreadsheet Link:** Statistical analysis are done using Excel Sheets and then copied to Google Spreadsheet.
3. **Tableau:** All the visualizations are saved on Tableau Public here.