28 Oct 2023

Glo-Box Analysis Results

GloBox A/B Test Analysis

Presented by Harleen Saini
MASTERSCHOOL PROJECT 1

Executive summary

GloBox, a renowned provider of boutique fashion and high-end decor products, has recently ventured into the realm of food and beverage offerings. To boost the visibility and profitability of this new product category, an A/B test was conducted to evaluate the impact of introducing a dedicated food and drink banner on the company's mobile website.

The A/B test examined the impact of a food & and drink product category landing page banner, revealing a noteworthy increase in the conversion rate from 3.92% in the control to 4.63% in the treatment group (+18%). However, there was an insignificant change in the average amount spent per user, standing at $3.37 in the control and $3.39 in the treatment group (+0.5%). This discrepancy suggests a potential cannibalization effect on other product categories, warranting further investigation for confirmation.

A detailed breakdown across devices, genders, and regions uncovered a more substantial increase in the conversion rate for Android users (+27%) compared to iOS users (+10%). Similarly, male users exhibited a higher conversion rate increase (+44%) than female users (+6%). Despite these variations, the average amount spent remained relatively consistent across all examined segments. No evidence of a concerning novelty effect was identified. However, a power analysis indicated that the test did not reach the minimum sample size required to detect a 10% change in both metrics. The test included 49k users, falling short of the necessary 186k users. While not a significant concern, any future iterations should involve an A/B test with a larger sample size.

Given the minimal impact on the average amount spent per user, I recommend proceeding with the launch of this experiment. For further refinement, it is advisable to delve into the results by product category and consider rerunning the test with an increased sample size if there is a commitment to further iteration.

Introduction:-

GloBox takes pride in its commitment to curating a selection of products that inspire a sense of adventure and transport the world to the doorsteps of its customers. While the company's existing customer base is well-acquainted with its boutique fashion items and premium decor products, the recent expansion into the food and beverage category presents a unique opportunity to diversify revenue streams. To ensure an effective promotion and heightened awareness of this new product category, the Growth team has developed an A/B test strategy.

The A/B test will involve two distinct groups: the control group and the treatment group. In the control group, users will experience the standard browsing experience on the GloBox website, without any banners highlighting the food and beverage category. Conversely, the treatment group will be exposed to a prominently displayed banner at the top of the website, featuring key food and beverage products (as illustrated in Figure 1).

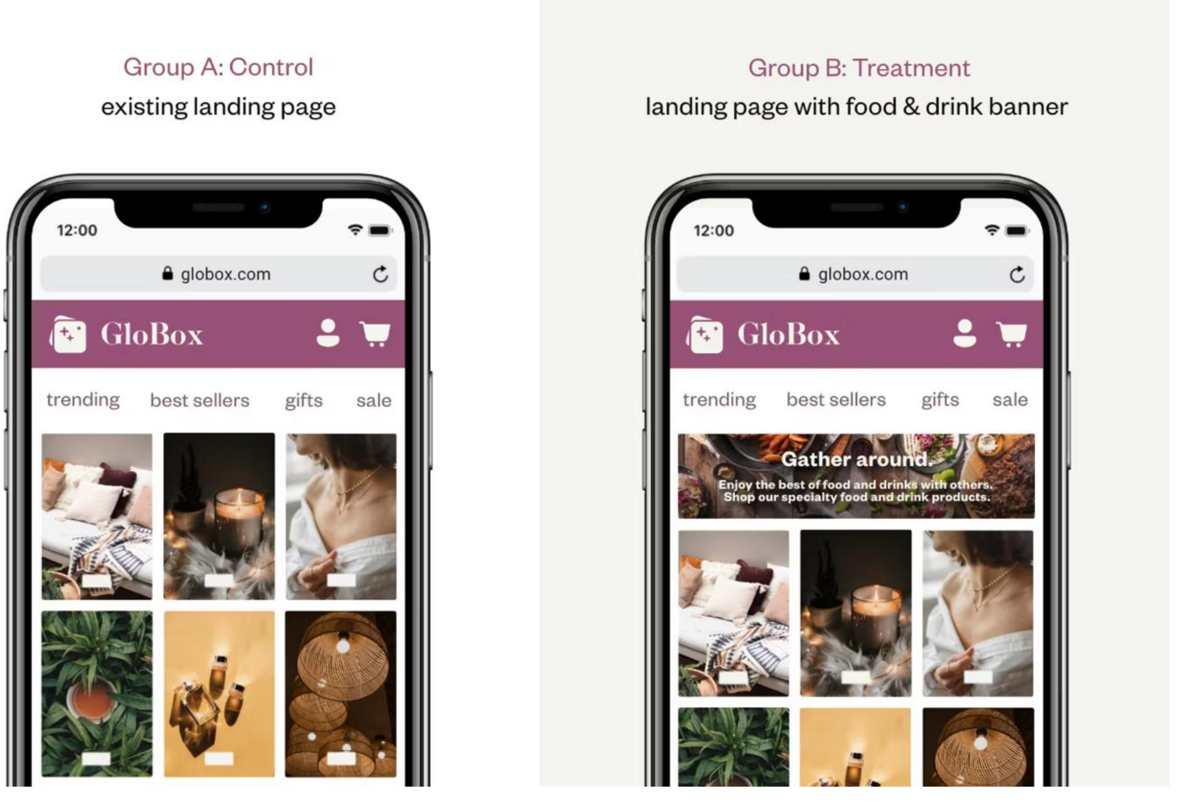


Figure -1 (Globox A/B Test Analysis)

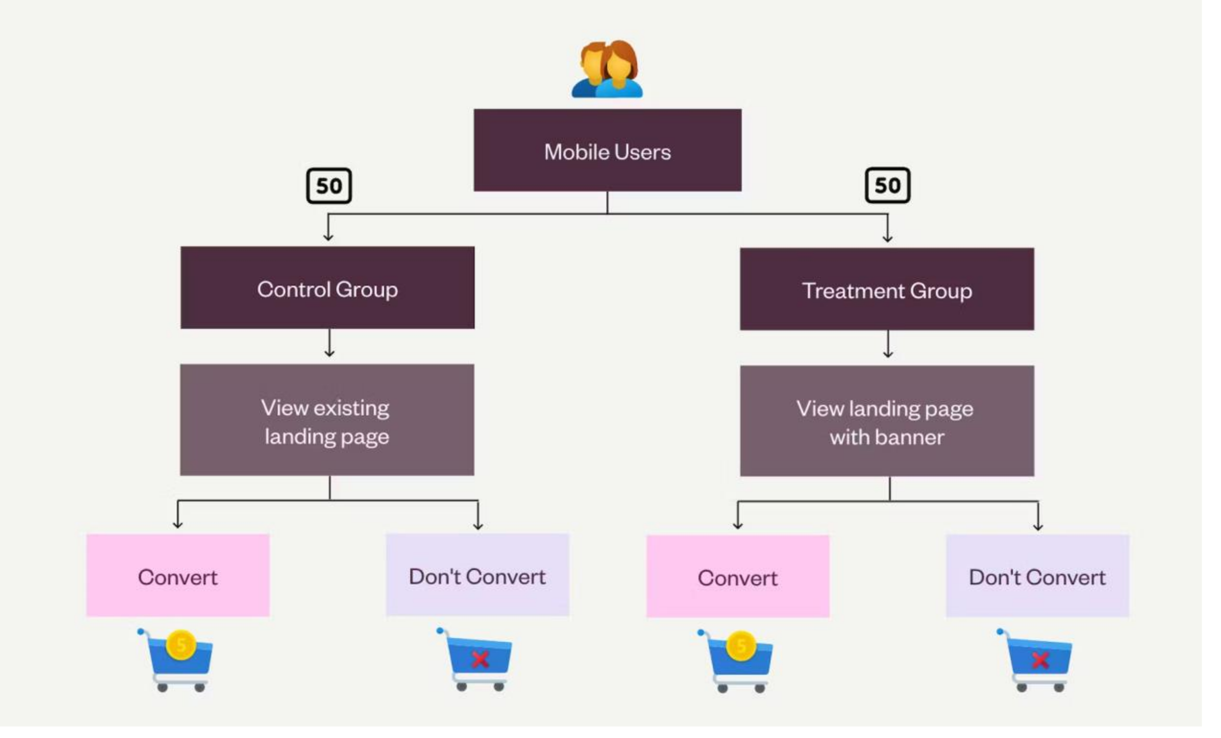


Figure -2 (Globox A/B Test Analysis)

Key Metrics and User Attributes

Key Metrics

1. Conversion Rate
2. Average Amount Spent per User

User Attributes

1. Gender (M = male, F = female, O = other, NA = unknown)
2. Device (I = iOS, A = Android, NA = unknown)
3. Country (ISO 3166 alpha-3 country code, NA = unknown

Project Process

The project process can be summarized as follows:

- **Data Extraction:**

Data was extracted from a relational database and stored in a CSV file for more comprehensive analysis queries were used to extract user-level aggregated data and consolidate relevant tables for further analysis.

- **Data Analysis:**

Extensive data analysis was conducted using SQL. Inferential statistical methods, including confidence intervals and hypothesis tests, were employed to identify areas of concern and opportunities for learning from successful outcomes.

- **Data Visualization:**

To enhance the initial analysis, data visualizations were created using Tableau. Visual representations were designed to effectively identify and highlight trends within the dataset.

- **Recommendations and Conclusion:**

Based on the analysis results, recommendations and conclusions were provided. It's important to note that the test parameters were not set by the data analysis team, and the recommendations were made in accordance with the findings. For more details, please refer to the Recommendations section of the report.

Overview of the dataset

The dataset comprises three tables: the users table, groups table, and activity table. The users table contains demographic information about the users, while the groups table focuses on assigning users to two different roups for the A/B test. The activity table records user purchase activity.

Data type

id	bigint
uid	bigint
country	text
gender	text
group	text
Join_dt	date
device	text
dt	date
spent	double

The table reveals the presence of four distinct data types, namely bigint, text, date, and double.

- bigint is employed for handling larger-scale integer data or whole numbers.
- text is utilized for storing textual information, such as words or sentences.
- date data type serves the purpose of storing date-related information.
- double data type is designated for storing floating-point numbers or decimal values.

The Business Goals

The relevance and importance of the changes seen during A/B testing are highly dependent on inferential statistics. The user conversion rate and the average amount spent per user are two crucial business objectives for GloBox that are the main subjects of this analysis. We want to determine whether the launch of the new food and drink banner is causing appreciable changes in these KPIs by doing rigorous statistical analysis.

GloBox's A/B test would be successful if it met the following criteria:

- **Increased User Conversion Rate:**
  The main objective is to see an improvement in the user conversion rate for the food and beverage category that is statistically significant. This would suggest that the banner's launch was successful in grabbing users' attention and encouraging them to make purchases from this category. The conversion rate for the test group should be noticeably greater than for the control group.

- **Improved Average Spend per User:**
  The new banner's impact on the typical amount spent per user in the food and drink category is another important goal that must be accomplished. A statistically significant and appreciable rise in the test group's average expenditure relative to the control group would indicate success. This would suggest that the ad not only attracted people but also encouraged them to spend more on food and drink goods.
  The proper statistical procedures, such as hypothesis testing or confidence interval estimates, will be used to evaluate these business goals. The findings will shed light on how well the food and beverage banner performs in terms of increasing user interaction, conversion rates, and total revenue. GloBox may increase revenue by producing notable increases in average expenditure and user conversion rate. The banner's effectiveness in raising consumer engagement and promoting revenue development within the food and drink sector allows GloBox to deem the A/B test a success.

Inferential Statistics

Metric	Group A: Control	Group B: Treatment	Grand Total	P Value
Total Number of Users	**24,343**	**24,600**	**48,943**	**N/A**
Conversion Rate	**3.92%**	**4.63%**	**4.28%**	**0.0001**
Average Amount Spend $	**$3.37**	**$3.39**	**$3.38**	**0.9438**

According to the available statistics, Group B, the treatment group, has a conversion rate of 4.63%, which is greater than that of Group A, the control group, which is 3.92%. However, the amount that consumers in both groups spend on average is relatively comparable, with Group A spending an average of $3.37 and Group B spending $3.39, for a little difference of $0.02. To ascertain the statistical significance of the A/B Test Data, additional analysis and computation in the form of hypothesis testing and confidence intervals were carried out.

Hypothesis Testing Outcomes

Hypothesis test 1 - Conversion Rate

Null and alternative hypotheses:

Null hypothesis (H0): The conversion rate is the same for Group A (control) and Group B (treatment). In other words, µA (the average conversion rate for Group A) is equal to µB (the average conversion rate for Group B).Alternative hypothesis (Ha): There is a difference in the conversion rate between Group A and Group B. In other words, µA is not equal to µB, indicating a difference in conversion rates.

Significance level: 0.05 (5%)

Conclusion:

The p-value is 0.0001, which is very small.

This small p-value indicates statistical significance.

Therefore, we reject the null hypothesis, meaning there is a statistically significant difference in user conversion rates between the control and treatment groups.

In simpler terms, the data shows that there is a real difference in how many users are converting between the control and treatment groups, and this difference is not due to chance.

➢ Hypothesis test 2 - Average Spend

Null and alternative hypotheses:

Null hypothesis (H0): There is no difference in the average amount spent per user between the Control and Treatment groups.

In other words, μ1 (the average amount spent in the Control group) is equal to μ2 (the average amount spent in the Treatment group).

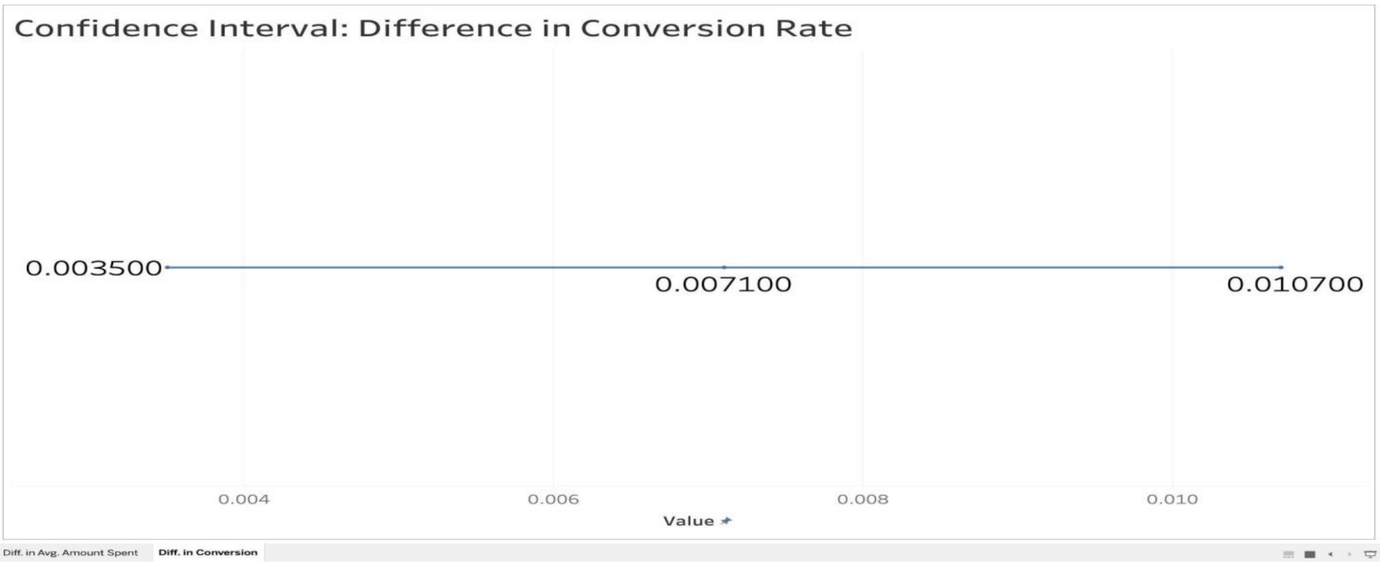
Alternative hypothesis (HA): There is a difference in the average amount spent per user between the Control and Treatment groups.

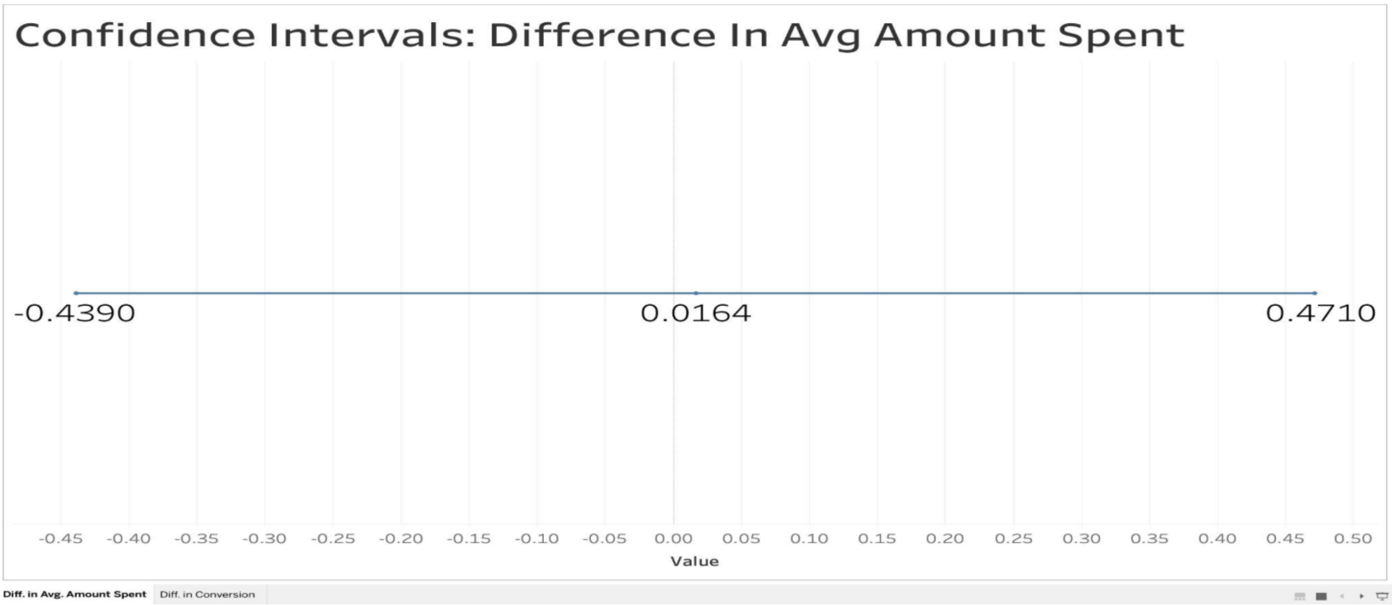
In other words, μ1 is not equal to μ2, indicating a difference in the average spending.

Significance level: 0.05 (5%)

The p-value is 0.944, which is relatively high. This high p-value indicates that the result is statistically insignificant. Therefore, we fail to reject the null hypothesis, suggesting there is no statistically significant difference in the mean amount spent per user between the control and treatment groups. In simpler terms, the data suggests that there is no real difference in the average spending per user between the control and treatment groups; any observed differences are likely due to chance.

Metric	Group Mean	Lower Bound	Upper Bound
Average Amount Spent	$0.016	$-0.0439	$0.0471
Conversion Rate	0.71%	0.35%	1.07%





Average Amount Spent: In the Control Group, people spent an average of $3.37, and in the Treatment Group, they spent an average of $3.39. The 95% confidence interval for the difference in spending ranges from -$0.0439 to $0.0471. This means that if we were to repeat the experiment many times, we would expect the actual difference in spending to fall within this range most of the time.

Based on the data, we can conclude that there is no significant difference in the average spending between the Control Group and the Treatment Group. In simpler terms, the numbers don't provide enough evidence to say that the treatment (Group B) had a big impact on how much people spent compared to the control group (Group A).

Conversion Rate: In the Control Group, the conversion rate is 3.92%, and in the Treatment Group, it's 4.63%. The 95% confidence interval for the difference in conversion rates ranges from 0.35% to 1.07%. This means that if we were to repeat the experiment many times, we would expect the actual difference in conversion rates to fall within this range most of the time.

Based on the data, we can conclude that there is a significant difference in the conversion rates between the Control Group and the Treatment Group. In simpler terms, the numbers strongly suggest that the treatment (Group B) had a notable impact on the conversion rate compared to the control group (Group A).

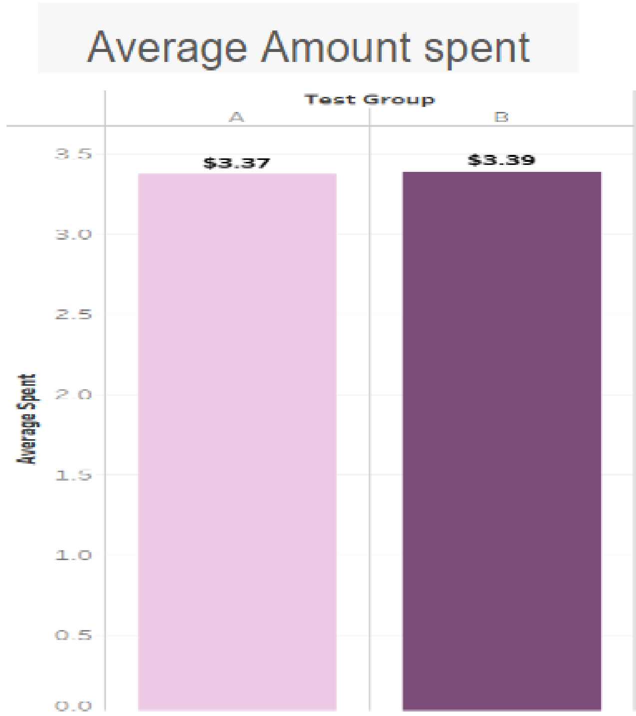
The results of the A/B test indicate that the new feature had a substantial impact on the conversion rate, with the treatment group achieving a conversion rate of 4.63%, which outperformed the control group's rate of 3.92%. However, when evaluating the average amount spent per user, there was no notable difference between the two groups, with both groups averaging $0.02.

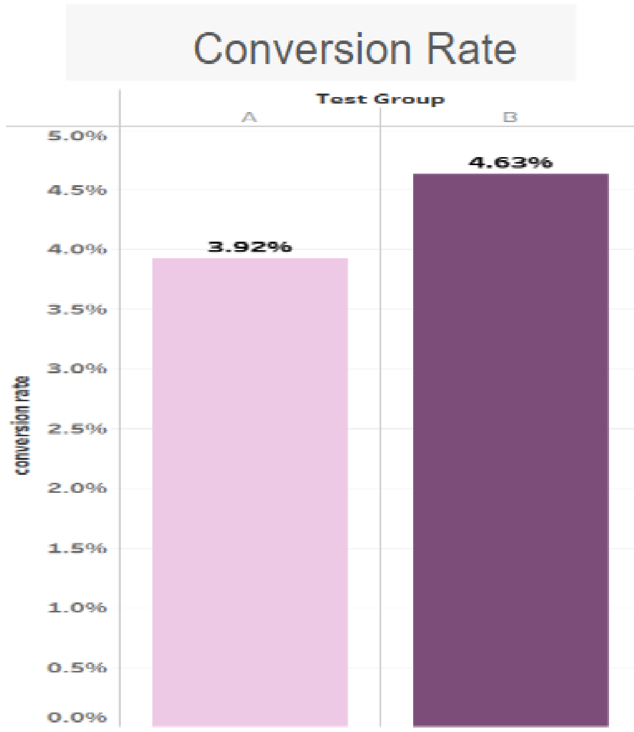
In summary, although the current status of the food and drink banner did not meet our expectations in terms of statistically significant improvements, the observed signs of potential success offer valuable insights for future enhancements. Through an ongoing process of refining and optimizing the banner experience based on these discoveries, we can work towards achieving the desired enhancements in user conversion rates and average spending, ultimately leading to revenue growth for GloBox.

Visualize the Results in Tableau

- **Visualization to compare the conversion rate and average amount spent between the test groups**

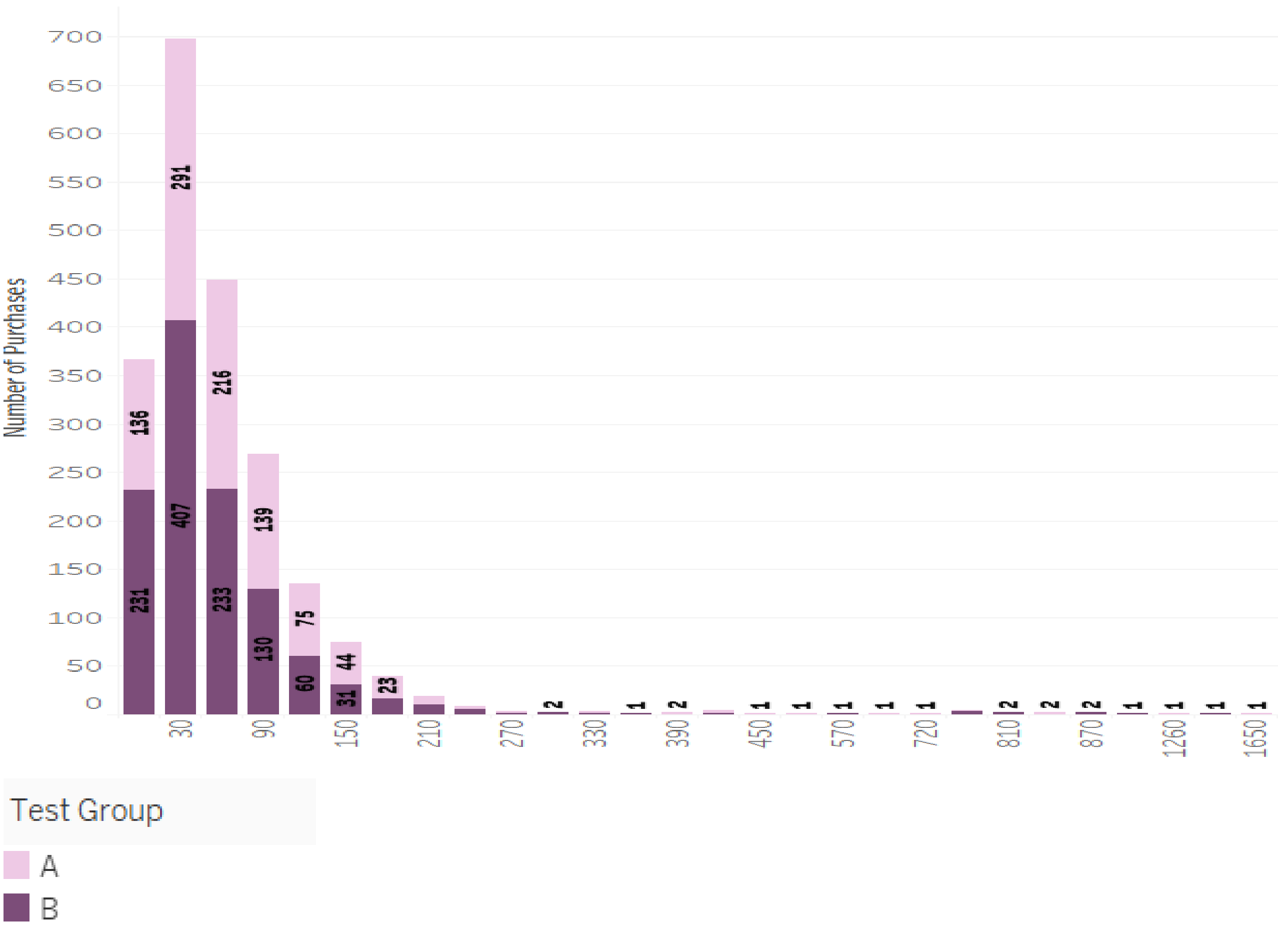
The chart illustrates that the control group's average spending is $3.37, while the treatment group's is $3.39. Additionally, the conversion rate for the control group is 3.92%, whereas the treatment group records 4.63%. These results highlight that the treatment group demonstrates higher average spending and a greater conversion rate compared to the control group. This emphasizes the effectiveness of displaying the banner, as depicted in the chart.



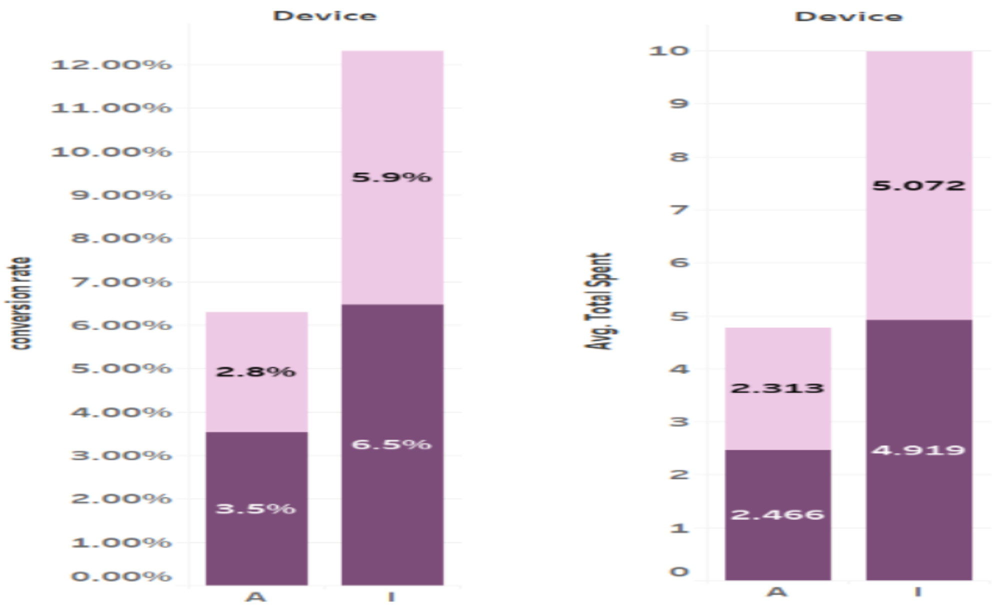


- **The distribution of the amount spent per user for each group**

For all categories, the bulk of purchases fell into the 30- to 60-dollar range, and most were under $150. A few purchases of up to 1650 USD do occur, however they are less frequent. Even if they happen less frequently, we still need to comprehend these high-value transactions.



Metrics by Device

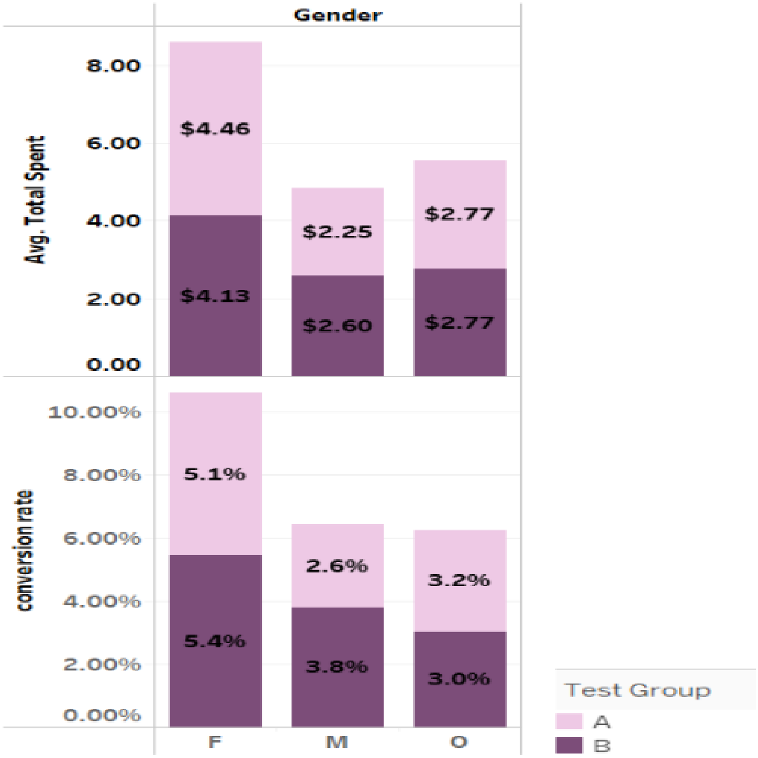


The bar chart below presents user metrics categorized by device. While the overall user count is highest among Android users, there is a significant difference in conversion rate and average spending per user between iOS and Android users.

Android users beat iOS users. With a 6% overall conversion rate, iOS users had the highest rate. Although this was lower in the treatment group, they also had the greatest overall average spend of $5 USD.

Android users who saw the banner spent 7% more and converted 25% more than those who did not, similar to how men behave.

Android users have had a greater conversion rate and average spending, which is interesting given that their sample size is larger than that of the iOS category. As this study only has access to the iOS banner, it would be wise to compare the two iterations and look for any variations in the banners' user interface and accessibility.



Metrics by Gender

The male and female segments exhibit the highest user counts. Although female users consistently maintain a higher conversion rate in both groups, male users demonstrate the most substantial increase in conversion rates within the treatment group compared to the control. Additionally, female users exhibit significantly higher spending levels in both groups.

The most significant increase in the conversion rate was seen in males (+44%), while the only decrease was in the 'other' group (-6%).

Compared to other genders, male and female users were significantly more prevalent in the test. In both groups, female users had the greatest total conversion rate (5%). The therapy group saw a decrease in the average spend of 4 USD, which was the largest among female consumers.

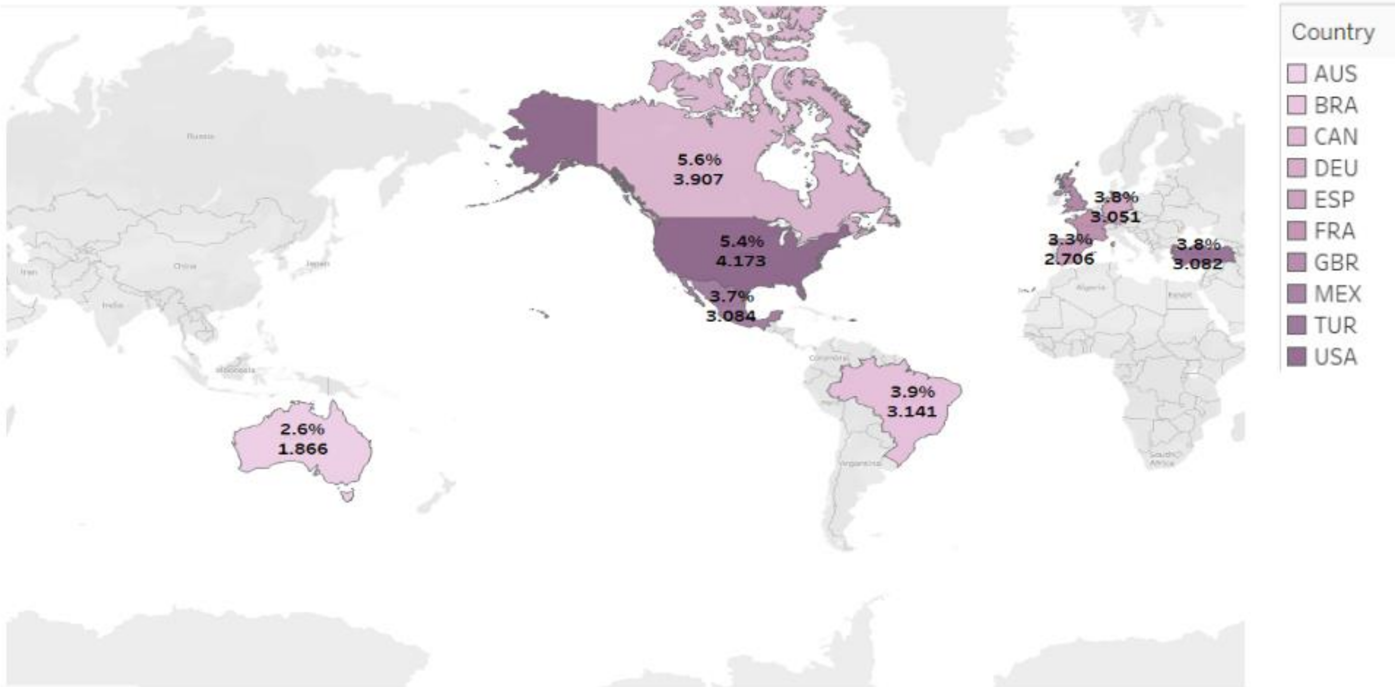
On the other hand, male users in the treatment group spent 15% more and converted 44% more than those in the control group.

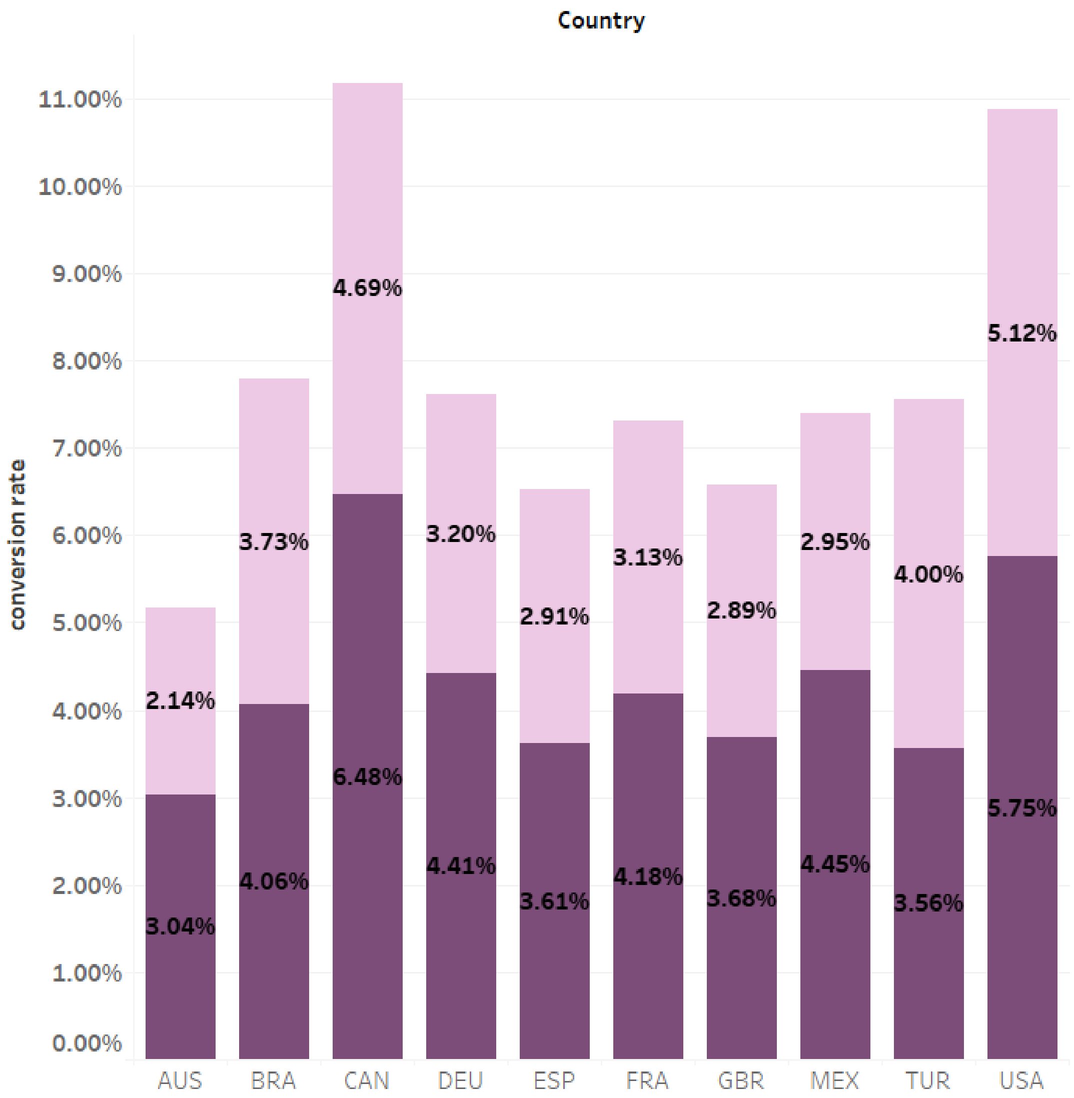
Metrics by Location

Ten countries took part in the test across America, Australia, and Europe.

The data reveals notable variations in both conversion rates and average spending across different countries. Canada stands out with a high conversion rate of 5.605%, while Australia lags at 2.568%. In terms of average spending, users in the USA lead with $4.173, while Canadian users spend an average of $3.907. Similarly, the highest average spending is attributed to USA users at $4.173, with the lowest reported in Australia at $1.866.

Turkey is the only country with both a decreased conversion rate and average spending in the treatment group. Germany, France, and Brazil also saw a decrease in average spending.





Advanced Task

Novelty Effect

What is the Novelty effect?

The novelty effect refers to the temporary increase in interest, engagement, or performance that individuals may exhibit when exposed to something new or novel. This heightened response is often a result of the novelty's initial appeal and the curiosity it generates. However, over time, the novelty effect tends to diminish as individuals become accustomed to the new stimuli. (**Loewenstein, G. (1994)**).

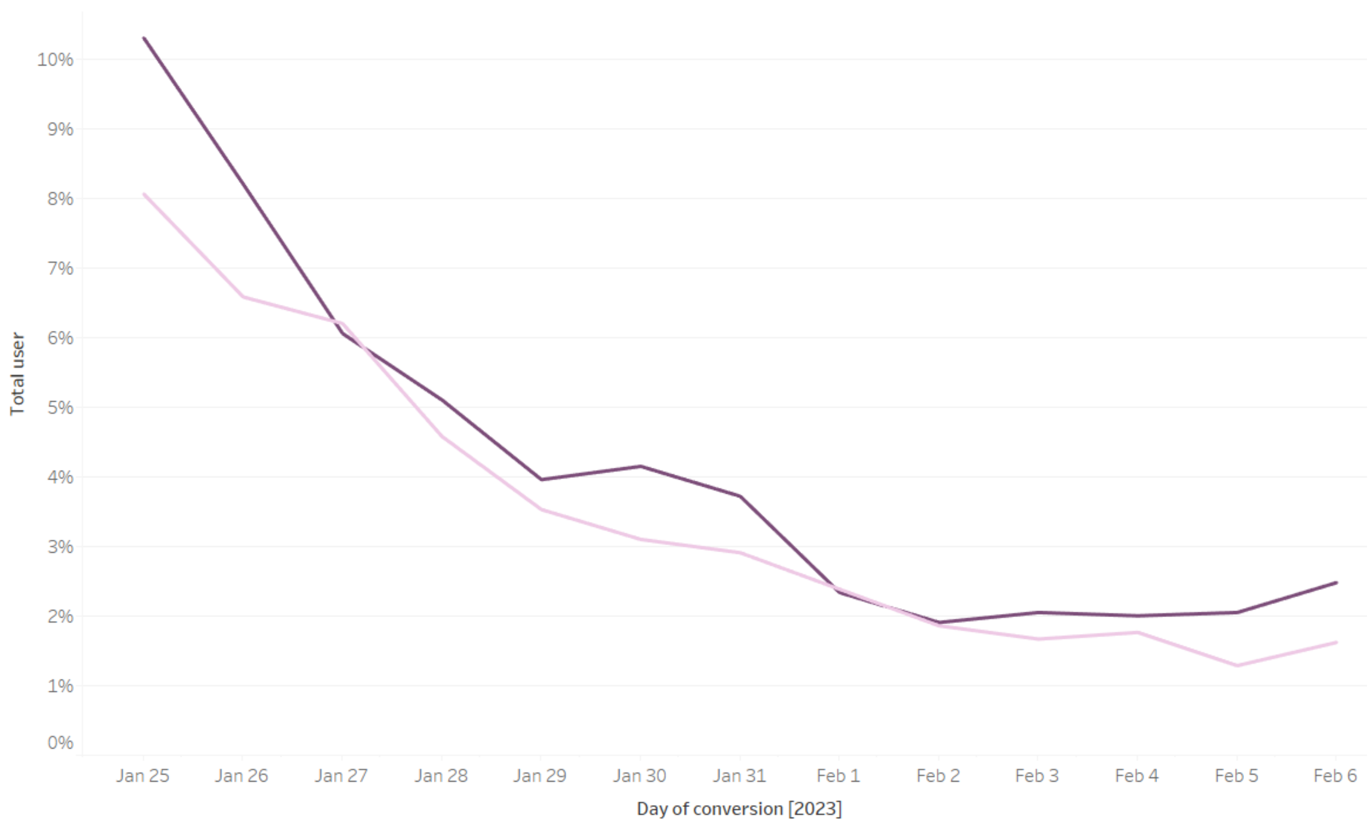
Imagine a company introducing a new fitness app to its users. When the app is initially launched, users may be curious and excited to explore its features, try out new workouts, and engage with the platform.

This initial surge in activity and interest is a result of the novelty effect. Users are drawn to the novelty of the new app.

However, as time passes, users become more familiar with the app's functionalities, and the initial excitement starts to wane. The frequency of app usage, enthusiasm for trying new features, and overall engagement may decline. This is a typical pattern associated with the novelty effect.

It is crucial to distinguish between the novelty effect and the banner's genuine long-term effectiveness. Additional studies or data analysis is typically needed to determine whether the banner indeed positively influences user engagement, conversion rates, or other relevant metrics.

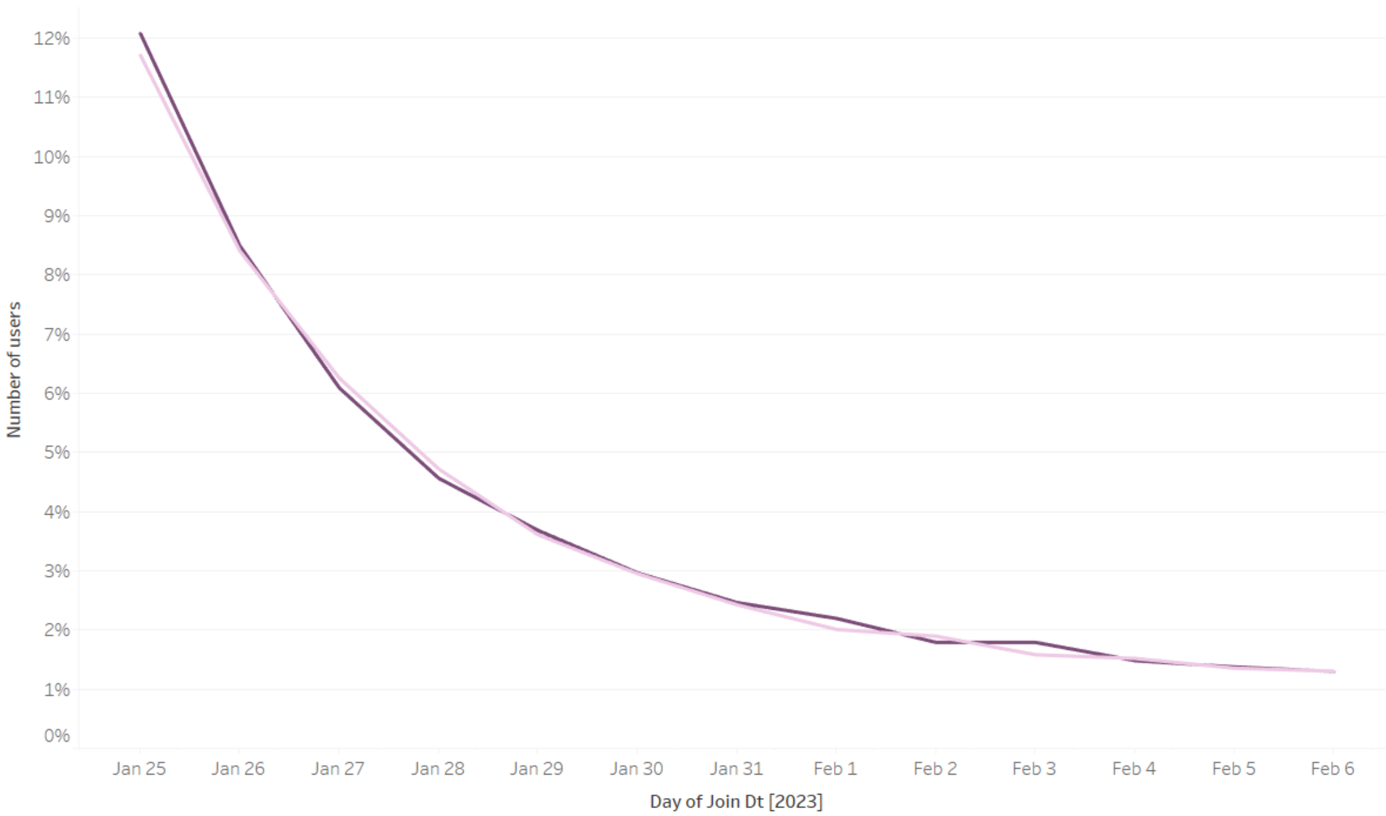
Conversion over time



In this study, regardless of whether a user saw the banner or not, nearly 50% of conversions happened in the first three days. While there are slight variations, control and treatment seem to follow the same trend with exceptions on the 30th and 31st of January. Conversion rates also seem to rise towards the end of the test duration

However, the novelty effect may be moderated by the number of users joining the test per day.

Users joining over time



A novelty effect is unlikely because the number of conversions seems to be proportional to the number of users joining the test per day.

Power analysis

What is power analysis?

Power analysis is a crucial process that ensures the reliability of our results. It is to test whether we have a sufficiently large sample size to accurately determine whether the differences we observe between two groups are meaningful or merely due to chance.

The test would need a sample size of 77k users, evenly divided between the two groups, for the conversion rate results to have adequate power. This calculator was used to compute the outcome.

To ensure that the average spending findings are statistically significant, the test would need to include 51,720,108 users, equally divided between the two groups. This calculator was used to compute the outcome.

Based on these calculations, we see that there were not enough users for the results to have a meaningful impact.

Analysis Summary and Recommendations

Summary:
The main metrics of the control and treatment groups were compared, and the results showed that people who saw the banner had much higher conversion rates.
The revenue did not differ significantly.
2. The most converted customers were Android users, men, and people from Mexico.

3. There were no noticeable novelty effects.

Recommendation:

It is simple to launch and maintain the banner. On the main page, though, it occupies valuable space. Before launching the banner, a high level of confidence regarding its impact is necessary. Since just one of the goals was achieved, I suggest that we repeat the test while taking the following into account:

1. For adequate power, a sample size of at least 77K users divided equally.

2. An extended period of six weeks to make sure that increases in revenue and conversion aren't the result of pay check timing.

3. When gathering data, make sure to include the type and time of purchase.

4. The same proportion of iOS and Android users.

5. An equal number of people sign up each day to monitor the impact of novelty.

Reference

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. Psychological Bulletin, 116(1), 75–98. doi: 10.1037/0033-2909.116.1.75

Appendix

SQL Query

```
SELECT
u.id,
u.country,
u.gender,
g.group,
g.device,
COALESCE(SUM(a.spent),0) AS total_spent,
CASE
WHEN SUM(a.spent) > 0 THEN 1
ELSE 0
END AS Converted
FROM groups AS g
JOIN users AS u
ON u.id = g.uid
LEFT JOIN activity AS a
ON g.uid = a.uid
GROUP BY u.id,
u.country,
u.gender,
g.group,
g.device;
```

Novelty Effect

```sql
WITH cte_conversion AS
 (SELECT g.uid, min(dt) AS dt, SUM(COALESCE(spent, 0)) AS spend,
 CASE WHEN SUM(COALESCE(spent, 0)) > 0 THEN 'converted'
  ELSE 'not_converted' END AS conversion
 FROM groups g
 LEFT JOIN activity a
 ON g.uid = a.uid
 GROUP BY 1)
SELECT c.uid, join_dt, dt, COALESCE (u.country, 'Unknown') AS country,
COALESCE (u.gender, 'Unknown') AS gender,
COALESCE (g.device, 'Unknown') AS device, g.group, c.spend, c.conversion
FROM cte_conversion c
LEFT JOIN groups g
ON c.uid = g.uid
LEFT JOIN users u
ON c.uid = u.id
```

Globox AB test Google sheet analysis Link

Globox Visualisation on tableau