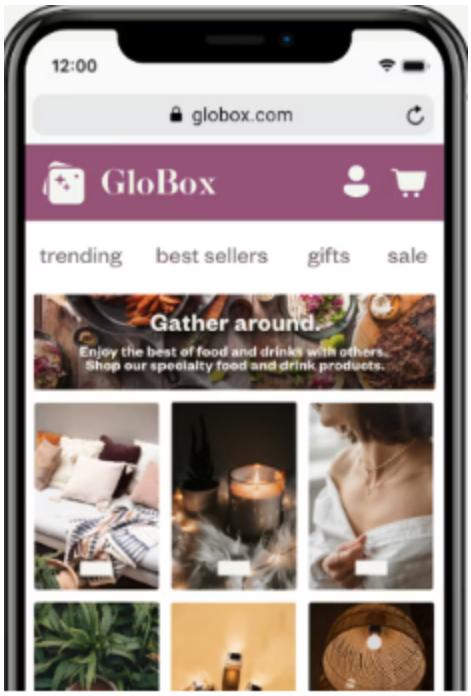
# GloBox A/B testing report

Idan Marcus

July 24th 2023.


**Summary**

We conducted an A/B test over a period of 13 days to evaluate the impact of featuring key products from the food and drink category as a top banner on the website.

The control group did not see the banner, while the test group was shown the banner as described earlier. We evaluated two key metrics: conversion rate and average spend per user.

Based on the results of the A/B testing, I recommend extending the test for a few more weeks. Although the test group showed a substantial 18% improvement in conversion rate, we did not observe a corresponding increase in average spend as anticipated. Given the 13-day duration of the test, it is advisable to extend it to obtain more reliable results and gather additional insights.

## Context

In an effort to optimize our website performance and drive higher user engagement, we initiated an A/B test to assess the impact of a top banner featuring key products from the food and drink category. By presenting this banner exclusively to the test group while keeping the control group free from any banner exposure, we aimed to gauge the influence of this visual element on key metrics such as conversion rate and average spend per user. This test serves as an opportunity to identify potential enhancements that can positively impact our users' purchasing behavior.

After the cleaning process was done, the dataset contained a csv file with total 48944 users splitted to 2 groups, control and treatment, that contained 24343 and 24600 users accordingly.

The assignments of participants were done randomly and distributed evenly between the users' variations.

The observations within each group are independent of each other.

To measure the conversion rate, we calculated the proportion of website visitors who made a purchase. The average user spend represents the average amount of money spent per user purchase.

To analyze the results, we employed several statistical techniques using SQL, spreadsheet and tableau.

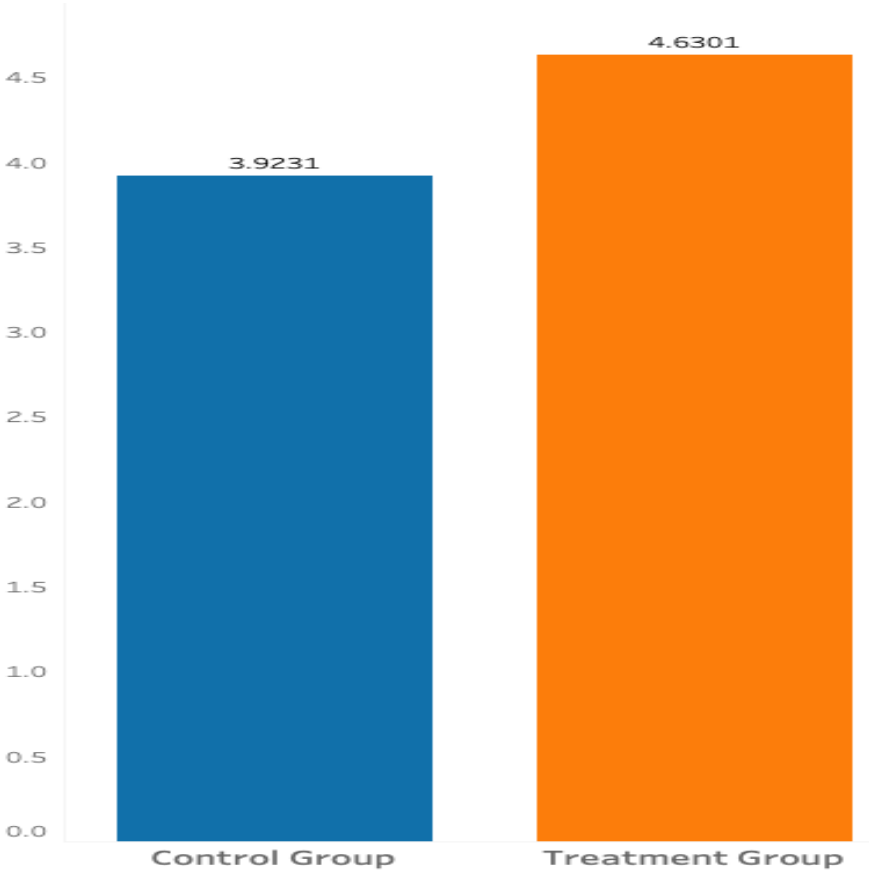
## Results

After cleaning the row data using SQL and pulling out the needed metrics as conversion rate and average spend per user.  We moved on to the next phase which is analyzing the data using spreadsheets.

| | AVG spent | Conversion rate |
|---|---|---|
| **Control Group** | $3.37 | 3.923 |
| **Treatment Group** | $3.39 | 4.630 |

- The image above indicates the key metrics we examined, AVG spend and conversion rate.

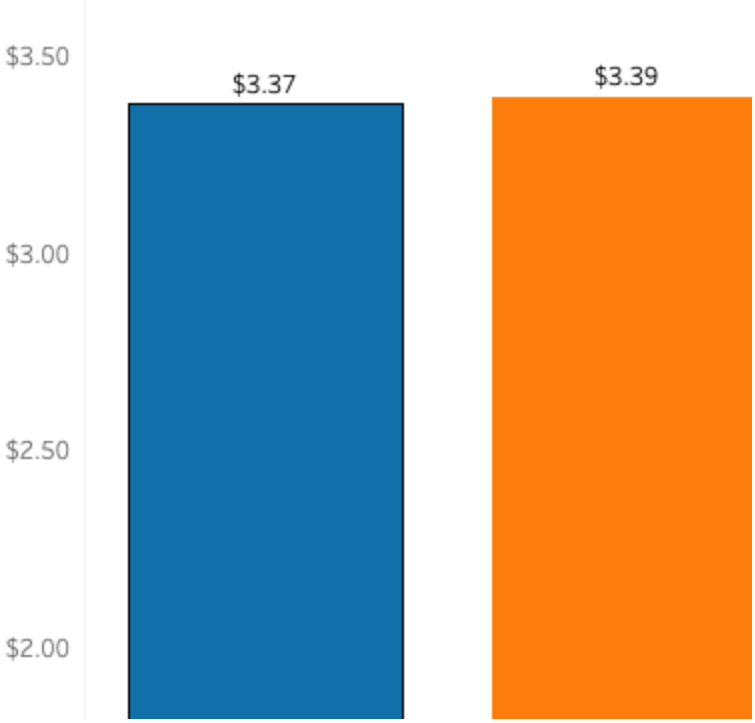
For the conversion rate:




- Null Hypothesis (H0): There is no difference in the conversion rate between the two groups.
- Alternative Hypothesis (HA): There is a significant difference in the conversion rate between the two groups.
- We used a 5% significance level and the normal distribution.

- The p-value for the conversion rate comparison was found to be 0.0001, leading us to reject the Null Hypothesis, indicating a significant difference in the user conversion rate between the groups.
- We tested the 95% confidence interval for the difference in conversion rate between the groups which is (0.0035, 0.0107), this considered a narrow result which gives us a more precise estimate.

For the average spend per user:

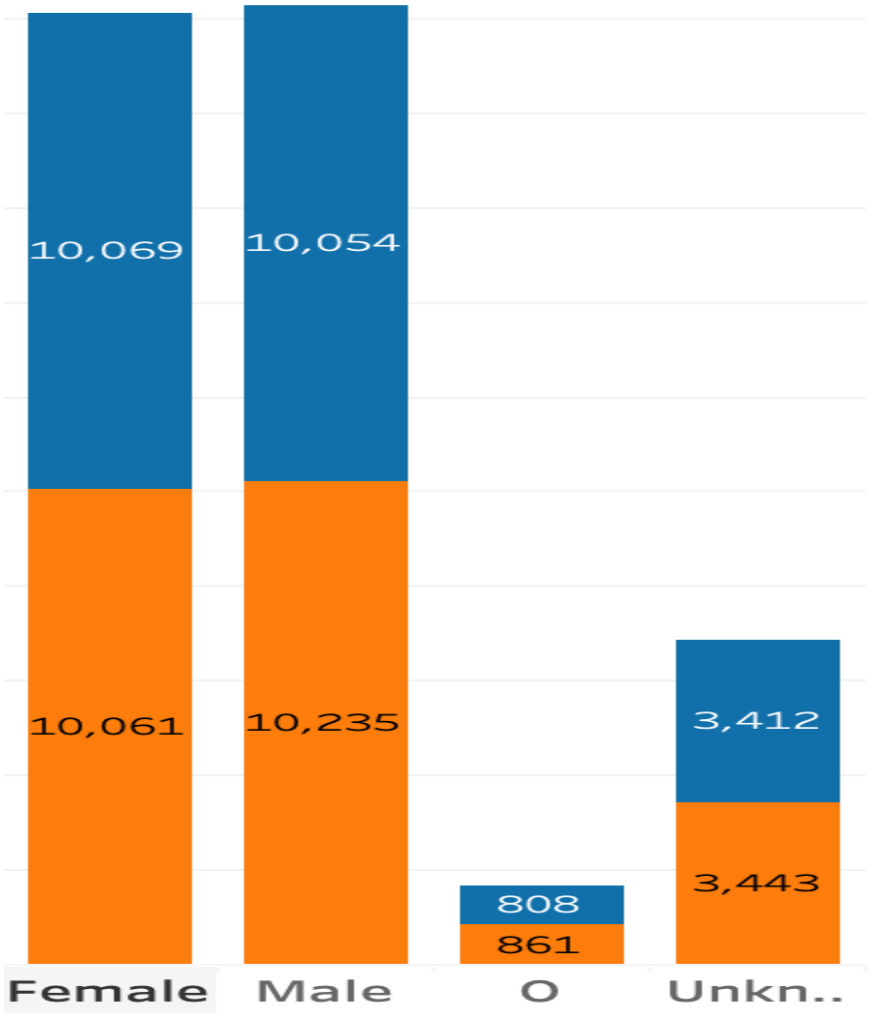



- Null Hypothesis (H0): There is no difference in the average spend per user between the two groups.
- Alternative Hypothesis (HA): There is a significant difference in the average spend per user between the two groups.
- We used the normal distribution and a 5% significance level.
- The p-value for the average spend comparison was calculated as 0.944, which does not provide enough evidence to reject the Null Hypothesis. Thus, there is no significant difference in the user average spend rate between the groups.

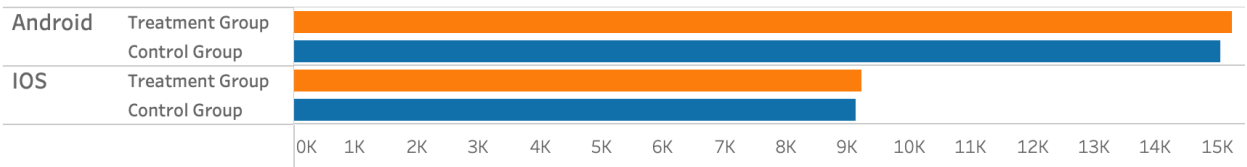
- The 95% confidence interval for the difference in average user spent between the groups is (-0.439, 0.471), this considered a wide result which gives us uncertainty regarding the average spend per user results.

Additional analysis that were made to ensure the test accuracy:

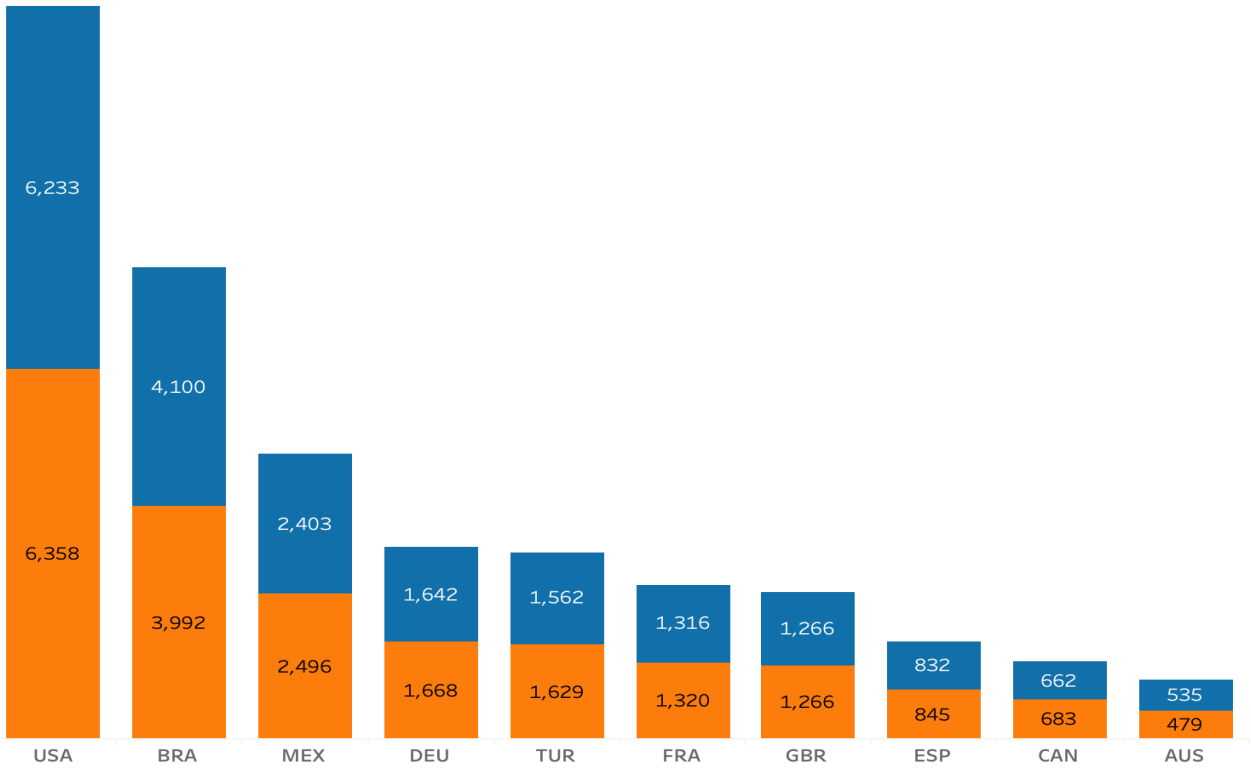
- Gender distribution - We ensured that the test results are not biased or affected by unequal representation. Our findings assure that the distribution is equal.




- Users device distribution - Checking device OS distribution is important to avoid biased results from differences in user behavior across operating systems. Our findings assure that the distribution is equal.



- Checking user country distribution in an A/B test is essential to account for potential regional variations in user behavior. The chart below clearly shows that users from every country were evenly split between the test groups.




## Recommendation

Based on these results, we recommend extending the duration of the test for a few more weeks to gather additional data and ensure the reliability of the findings.
Since we saw a statistically significant increase in only one success metrics, we don't feel confident enough that the banner will lead to more revenue.
The banner is not causing any loss of users or revenue so continuing with the test won't damage the business.

A longer test duration would help capture potential variations in user behavior over time and provide a more comprehensive analysis. It would allow us to assess whether the observed increase in conversion rate is sustained and whether there are any further impacts on average spend per user. This extended timeframe will provide us with a more robust understanding of the relationship between the banner and user behavior. By doing so, we can make more informed decisions and refine our strategies to maximize the potential benefits of the implemented changes.

## Appendix

SQL queries:
https://docs.google.com/document/d/1duKn0cd5OyroxPqpCWA2njXEWUphc2AX28EYRR6SYt8/edit?usp=sharing

Tableau story:

https://public.tableau.com/shared/B7WYSYRF7?:display_count=n&:origin=viz_share_link

Spreadsheet:

https://docs.google.com/spreadsheets/d/1GsiZ9oPs1EyDl-XbWI4Ed2LogTLAdWx2FgEOxO-VOiY/edit?usp=sharing