# GloBox A/B Test Analysis

Experiment: GloBox Website Optimization

Testing New Banner for Globox

Author: Neha Varshney | Date:July 2023

Abstract

**Recommendation: do not launch the experiment 🗷**

The banner lead to a significant lift in conversion rate (+4.28%) but virtually no change in the average amount spent per user. Any further recommendation would need further data analysis into the results and larger sample size.

SUMMARY

A/B Test was conducted to determine the effectiveness of a new feature on Globox's platform, using two metrics: conversion rate and average amount spent per user. The results showed that there was a statistically significant difference in the conversion rate between the two groups i.e. control and treatment groups, with the treatment group showing a higher conversion rate. However, there is no statistically significant difference in the average amount spent per user between the two groups.

A more detailed breakdown of the results by device, gender and country showed a higher increase in conversion rate for Android (+27%) than iOS (+10%) and a higher increase in conversion rate for male users (+44%) than female users (+6%). However, we examined that the average amount spent was pretty flat across all segments.

A power analysis showed that we did not reach the minimum sample size to detect a 10% change in both the metrics. Based on the Power Analysis (with 3.92% as baseline conversion rate at 0.05 significance level and 95% statistical power), we would need 710 users in total, and the test only had 49k users. Any further iteration of this banner should involve an A/B test with a larger sample size.

Based on these findings, I recommend that GloBox should not launch the new feature to all users even though there were some promising results that show we could possibly make changes to the banner experience and get better improvement next time. Maybe, some further data analysis is required to understand this better, or we need a larger sample size to make a confident recommendation.

CONTEXT

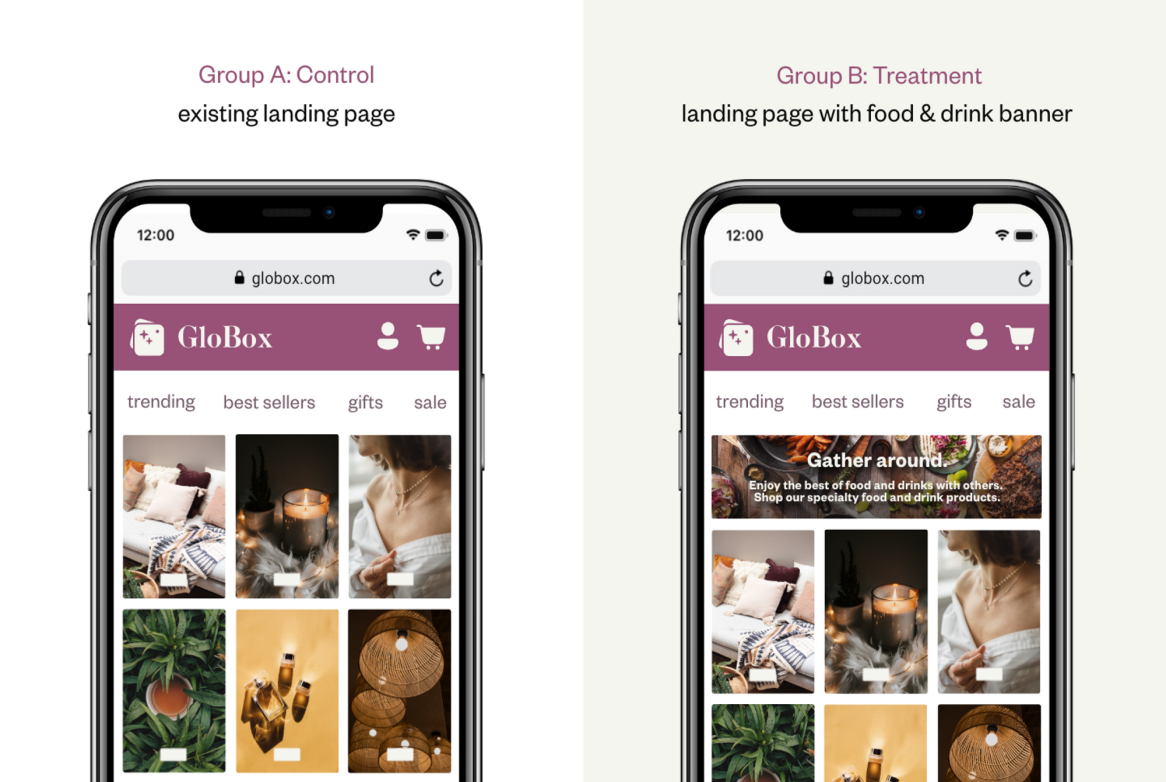
**Motivation**

The growth team ran this experiment to test the effectiveness of the banner at the top of the website that highlights key products in the food and drink category to bring awareness to the product category. The inventory in this category have grown tremendously in the last few months. The team hopes that showcasing this product category will increase the revenue.

**Test Groups**

The two test variants are as follows:

1. Control: existing landing page
2. Treatment: landing page with food and drink banner



**Test Parameters**

We ran this experiment with the following parameters:

|  |  |
| --- | --- |
| Date Range | Jan 25th- Feb6th, 2023 (12 days) |
| Traffic Split | 50/50 |
| Platforms | Mobile-only(iOS + Android) |
| Countries | AUS, BRA, CAN, DEU, ESP, FRA, GBR, MEX, TUR, USA |
| Total Users | 48,943 |

**Success Metrics**

We choose the following metrics to measure the success of this experiment:

* User conversion rate (%): the percentage of users in the experiment that made a purchase during the experiment.
* Average amount spent ($): the average amount USD spent for all users in the experiment (both converted or not) during the experiment. This measures the expected change in overall revenue.

RESULTS

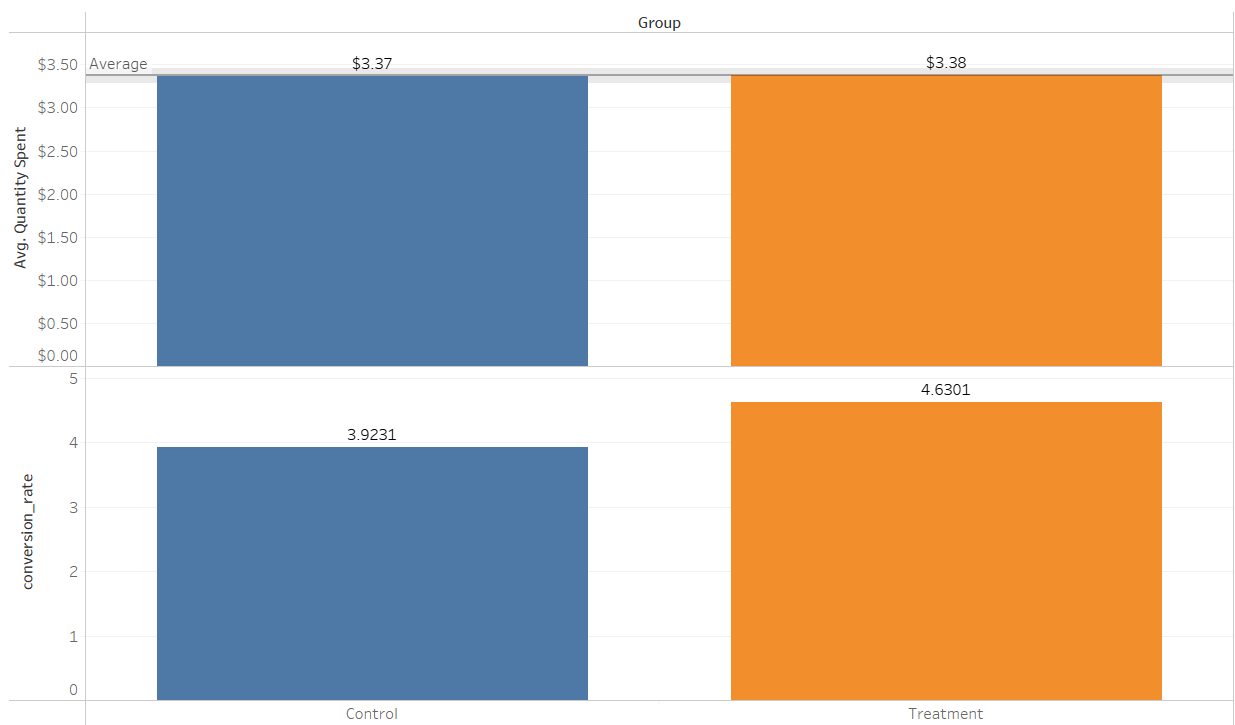
The A/B test results were primarily analyzed using Spreadsheets. Prior to the analysis, the datasets was extracted from a database using SQL in Beekeeper Studio. Visualization of the results was in Tableau.

**Overall Results**

We found out that the banner led to the significant increase in the conversion rate, but there was not much change in the average amount spent per user(revenue). The conversion rate was 3.92% for the control group and 4.63% for the treatment group which is statistically significant with p = 0.0001. The average amount spent per user was $3.37 in the control and $3.39 in the treatment group which is not statically significant with p = 0.94.

**Hypothesis Tests for the equality between the two groups**

* The conversion rate was 3.92% for the control group and 4.63% for the treatment group. We used Two-sample z-test with pooled proportion to find the difference between the conversion rates of the two groups. Our Hypothesis test results in p = 0.0001 which is less than the set significance level (α) of 0.05. Our results are statistically insignificant and so we conclude that there is no difference in the user conversion rate between the control and treatment group.
* The average amount spent per user was $3.37 for the control and $3.39 for the treatment group. We used Two-sample t-test with unpooled variance to find the difference in means amount spent per user between the two groups. Our Hypothesis test results in p = 0.944 which is greater than the set significance level (α) of 0.05. Our results are not statistically significant and so we conclude that there is no difference in the mean amount spent per user between the control and treatment.



We can visualize the confidence intervals for the difference between two groups for each of our metrics from the above tableau visualization.

* We are 95% confident that the difference in conversion rate between the treatment and control is between 0.0035 and 0.0107. We see that the result is statistically significant since the p value (0.0001) is less than the significance level (ɑ) of 0.05
* We are 95% confident that the difference in the average amount spent per user between the control and treatment is between -0.439 and 0.471. We see that the result is statistically insignificant since the p value (0.944) is greater than the significance level (ɑ) of 0.05.

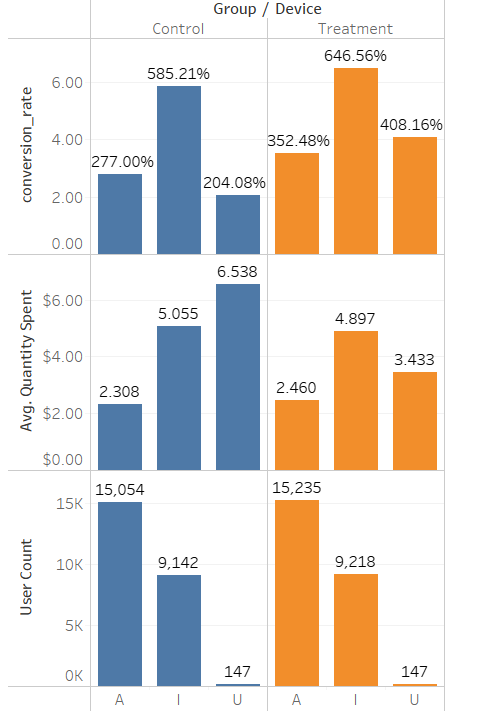
**Results Breakdown**

When we break down the results by device, gender, and countries, a few things stand out:

* Higher increase in conversion rate for Android (+27%) than iOS (+10%).
* Higher increase in conversion rate for male (+44%) than female (+6%)
* Slightly higher increase in conversion rate for America and Europe
* Average amount spent per user is pretty consistently flat across segments.

**Results By Device**

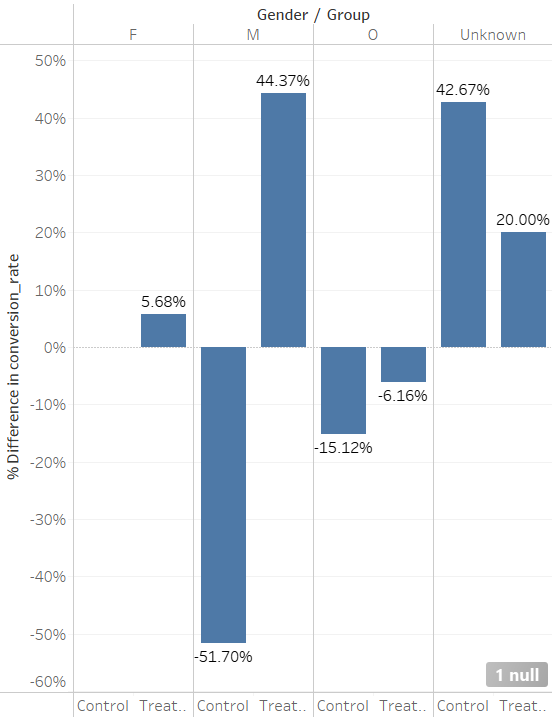
When we segment by device, we see that both platforms had a higher conversion rate with the treatment group.



We observed that there is 27% relative increase in conversion rate for Android and 10% for iOS. It may be interesting to investigate why the increase was higher on Android. There was little to no change in the average amount spent per user. There were some users with unknown devices, but it is a small enough number that we can ignore them for the purpose of these results.

**Results By Gender**

When we segment by gender, we see that the conversion rate for the Male users have increased the most as compared to the female users.

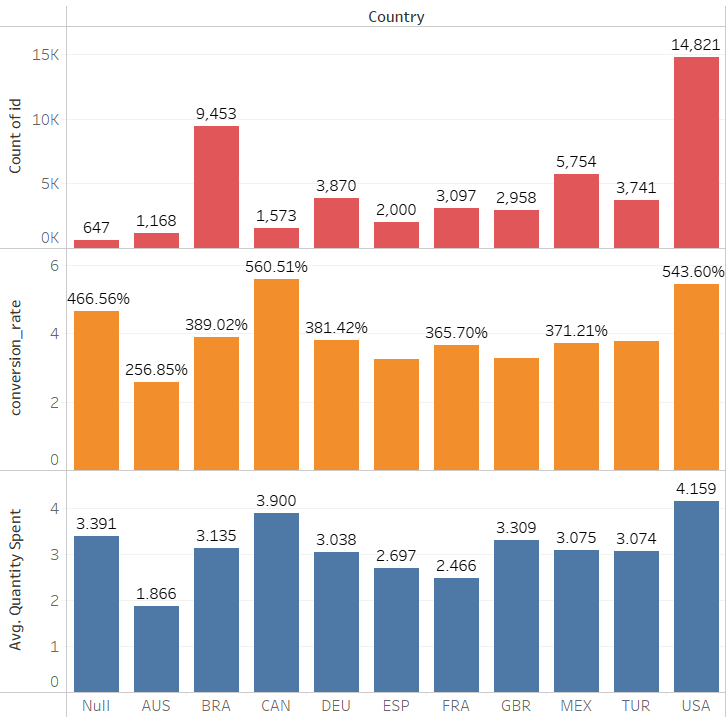


There was a +6% relative change for ‘female’, +44% increase for ‘male’, -6% for ‘other and +20% for ‘Unknown’.

It’s not clear why males converted more than females.

**Results by Countries**

There is an increase in conversion rate across the countries, and the average amount spent per user was pretty consistently flat.



**Power Analysis**

Based on the power analysis below, we did not reach the sample sizes necessary to detect a 10% change in both metrics.

We wanted to check if we have an adequate sample size to detect the desired change in conversion rate and the average amount spent.

We decided to go with a 10% relative change in each metric would be desired in order to launch the banner.

Sample sizes for 10% relative MDE

|  |  |  |
| --- | --- | --- |
| Metric | Minimum Sample size (total both groups) | Inputs |
| Conversion Rate | 77k | Baseline conversion rate: 3.92%  MDE: 10%  Two-sided, ɑ = 0.05, power=0.8 |
| Average Amount Spent | 710 | Difference between two means:0.337(10% of control)  Expected standard deviation:1.6  Two-sided, ɑ = 0.05, power=0.8 |

Calculators used: Statsig, Statulator

RECOMMENDATION

Based on all the results above, I recommend that we should not launch this experiment. There is definitely a lift in conversion rate but there is no significant change in average amount spent per user (revenue).

If we want to invest in this feature and want to continue iterate on it, then we should consider to carry out the experiment on the larger sample.

APPENDIX

SQL Queries

Query to get user-level analysis dataset

select users.id, users.country, users.gender, groups.device,coalesce(sum(activity.spent),0) as total\_amount\_spent,

case

when (sum(activity.spent) > 0) then '1'

else '0'

end as converted

from users

left join groups

on users.id = groups.uid

left join activity

on groups.uid = activity.uid

group by users.id, users.country,users.gender,groups.device;

Imported into excel for statistical calculations

Imported into Tableau for visualizations.

Analysis Files

Excel for statistical calculations

Tableau for visualizations

<https://public.tableau.com/views/Globox_ABTestVisualization_070723/Sheet1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

Spreadsheet A/B Testing

<https://docs.google.com/spreadsheets/d/1rWw5keYqvkZV-9TyCTjPvbrxAB85KysgFpDD08pCHGw/edit#gid=2020343099>