Globox's New Banner A/B Testing Report

Author: Tiantian Gao   Date: 2023/7/6

Summary

This analysis revealed a significant difference in the conversion rates between the control and treatment groups (p-value=0.0001), indicating that the new banner has a significant impact on the conversion rate. Although we observe a higher conversion rate, we did not observe a higher average amount spent per user (p-value=0.944). Nevertheless, this marks a positive start in acquiring more paying customers. The improvements in our success metrics have provided us with the confidence to launch the banner for all users.

Project Background

Motivation of the experiment: Determine whether launching a food and drink banner on Globox's site will potentially boost revenue and conversions.
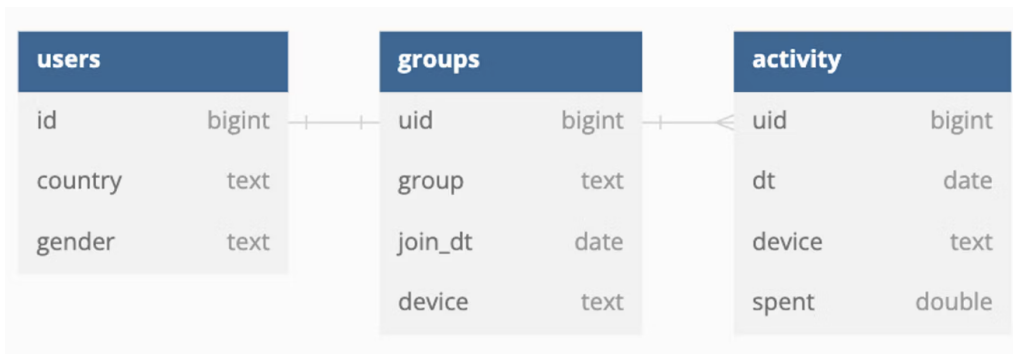
The parameters of the test: conversion rate, average amount spent per user

A/B Test Setup:

1. The experiment is only being run on the mobile website.

2. A user visits the GloBox main page and is randomly assigned to either the control group A or treatment group B.

3. The page loads the banner if the user is assigned to the test group B, and does not load the banner if the user is assigned to the control group A.

4. The user subsequently may or may not purchase products from the website. It could be on the same day they join the experiment, or days later. If they do make one or more purchases, this is considered a "conversion".
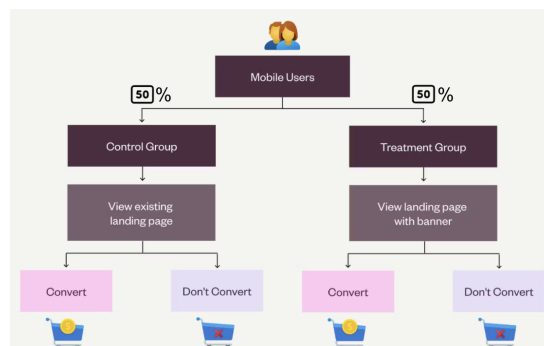
The dataset: Three tables of users' information collected during the experiment duration (Jan 25 to Feb 6)

users		groups		activity	
id	bigint	uid	bigint	uid	bigint
country	text	group	text	dt	date
gender	text	join_dt	date	device	text
		device	text	spent	double

Context and Results

1. Understand Globox's database and extract a user-level aggregated dataset in SQL

We first familiarized ourselves with the database tables. We achieved this by writing a SQL query that returns the following information: user ID, user's country, user's gender, user's device type, user's test group, conversion status (whether or not they spent > $0), and total amount spent ($). Subsequently, we downloaded the data as a CSV file.

2. Analyze the A/B test results using statistical methods in spreadsheets.

We imported the CSV file into a spreadsheet and conducted two hypothesis tests to determine the A/B test result.

The first hypothesis test conducted is a two-sample Z-test aimed at determining whether there is a statistically significant difference in the conversion rate between the two groups. Here are the steps:

Step 1: Define Population

Population	Users who visited the site of all time
Sample	Users who visited the site between Jan 25 and Feb 6 in 2023
Parameter	The conversion rate of users who visited the site of all time
Statistic	The conversion rate of users visited the site between Jan 25 and Feb 6 in 2023

Step 2: Define the null and alternative hypotheses

Null Hypothesis	No difference in conversion rate between the control group A and the treatment group B
	H0: P(A) = P(B), d=0
Alternative Hypothesis	There is a difference in conversion rate between the control group A and treatment group B
	HA: P(A) ≠ P(B), d≠0

Step 3: Compute the necessary summary statistics

User converted (X)	Xa=955, Xb=1139
Sample Size (n)	Na=24343, Nb=24600
Conversion Rate P	Pa=3.92%, Pb=4.63%
Estimate the difference d	d=Pb-Pa=0.0071
Pooled probability of a convert P	P=4.28%

Step 4: Calculate the test statistic

Standard Error (se)	0.0018
Test Statistic (z)	3.864

Step 5: Calculate the p-value and the confidence intervals

P-value (p)	0.0001
Critical Z-score(α:0.05)	1.96
Unpooled standard error	0.0018
center of C.I	0.0071
width of C.I (margin of error)	0.0036
Confidence interval of d	0.0035~0.0107

Step 6: Conclusion

Decision	Reject the Null hypothesis
Conclusion	With a p-value of 0.0001, which is less than the significance level (α), we reject the null hypothesis that there is no difference in the conversion rate between the control group A and the treatment group B. Instead, we accept the alternative hypothesis, indicating that there is a statistically significant difference in the conversion rate between the two groups

The second hypothesis test involved a two-sample t-test to determine whether there is a statistically significant difference in the average amount spent per user between the two groups.Here are steps:

Step 1: Define Population

Population	Users who visited the site of all time
Sample	Users who visited the site between Jan 25 and Feb 6 in 2023
Parameter(μ)	The average amount spent per user for users who visited the site of all time
Statistic($\bar{x}$)	Average amount spent per user of users who visited the site between Jan 25 and Feb 6 in 2023

Step 2: Define the null and alternative hypotheses

Null Hypothesis	No difference in average amount spent per user between the control group A and treatment group B
	H0: μ(A) = μ(B), d=0
Alternative Hypothesis	There is a difference in average amount spent per user between the control group A and treatment group B
	HA: μ(A) $\neq$ μ(B), d $\neq$ 0

Step 3: Compute the necessary summary statistics

Mean ($\bar{x}$)	$\bar{x}$a=3.375, $\bar{x}$b=3.391
Standard Deviation (s)	Sa=25.94, Sb=25.41
Sample Size (n)	Na=24343, Nb=24600
Mean Difference d	0.016

Step 4: Calculate the test statistic

Standard Error (se)	0.232
Test Statistic (T)	0.07
Degree of freedom (df)	24342

Step 5: Calculate the p-value and the confidence intervals

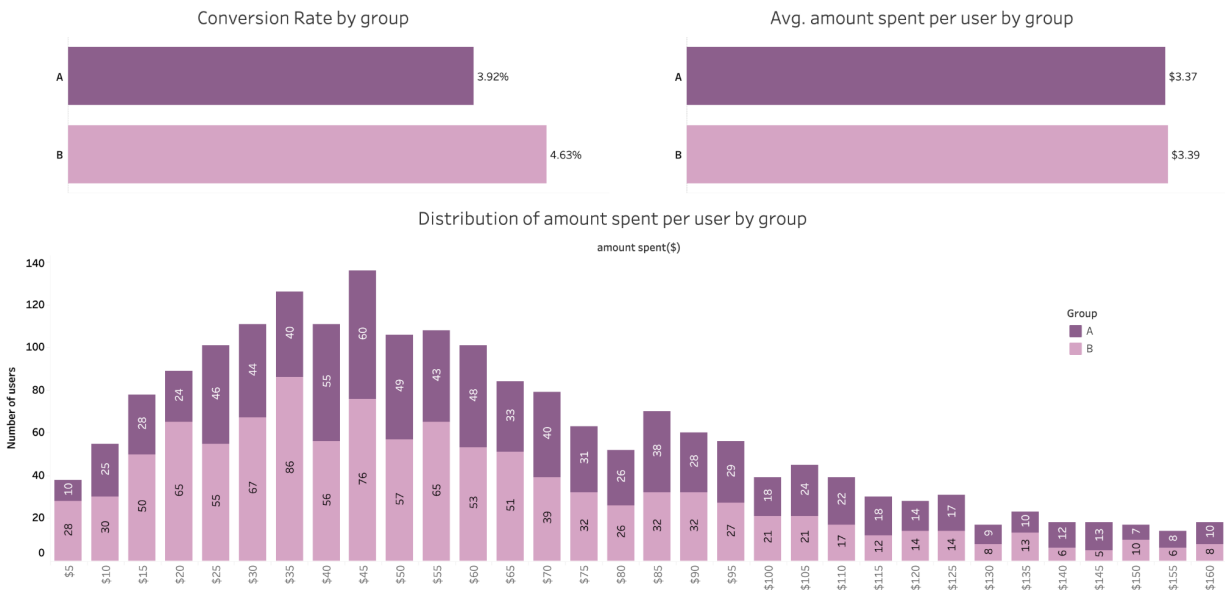| alpha α	0.05
P-value (p)	0.994
critical t value Tc	1.96
center of C.I	0.016
width of C.I (margin of error)	0.455
confidence interval of d	-0.439~0.471

Step 6: Conclusion

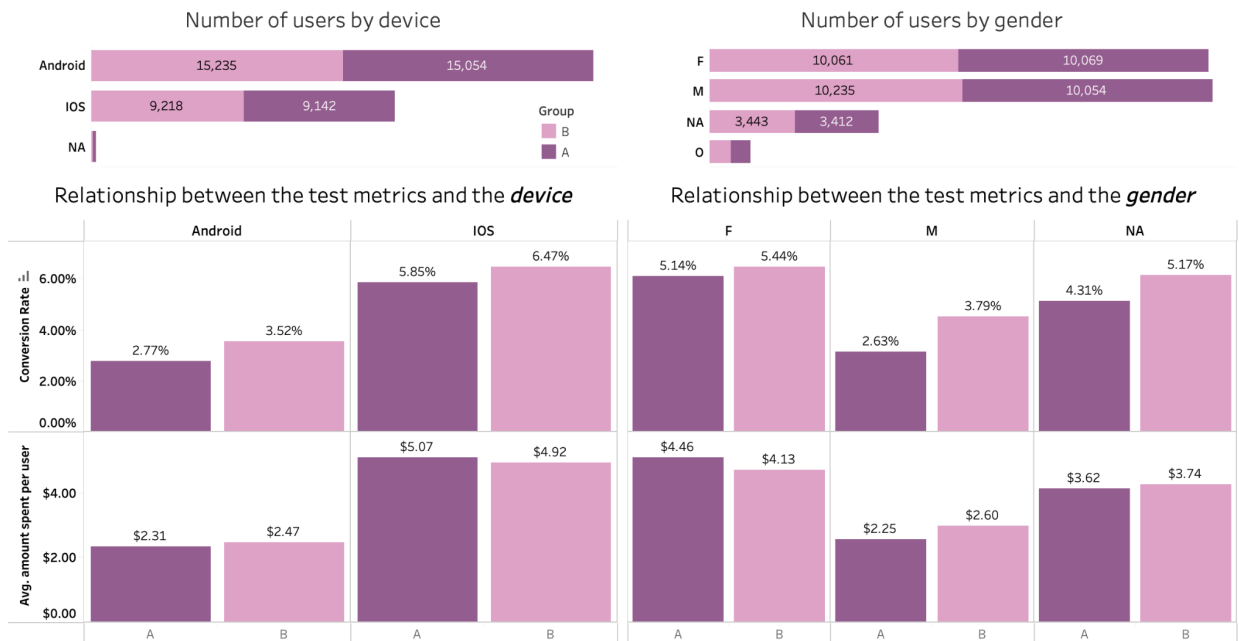| Decision	Fail to reject the null hypothesis
Conclusion	With a p-value of 0.994, which is greater than the significance level (α) of 0.05, the result is statistically insignificant. Therefore, we fail to reject the null hypothesis, indicating that there is no significant difference in the average amount spent per user between the control group A and the treatment group B. Instead, we accept the alternative hypothesis that there is a difference in the average amount spent per user between the two groups.

3. Analyze the A/B test results using visualizations in Tableau

After importing the CSV file into Tableau, we were able to uncover conversion rates and spending patterns by group. Additionally, we conducted further analysis to explore user behavior across different variables such as device, gender, and region. Here are some insights we discovered:
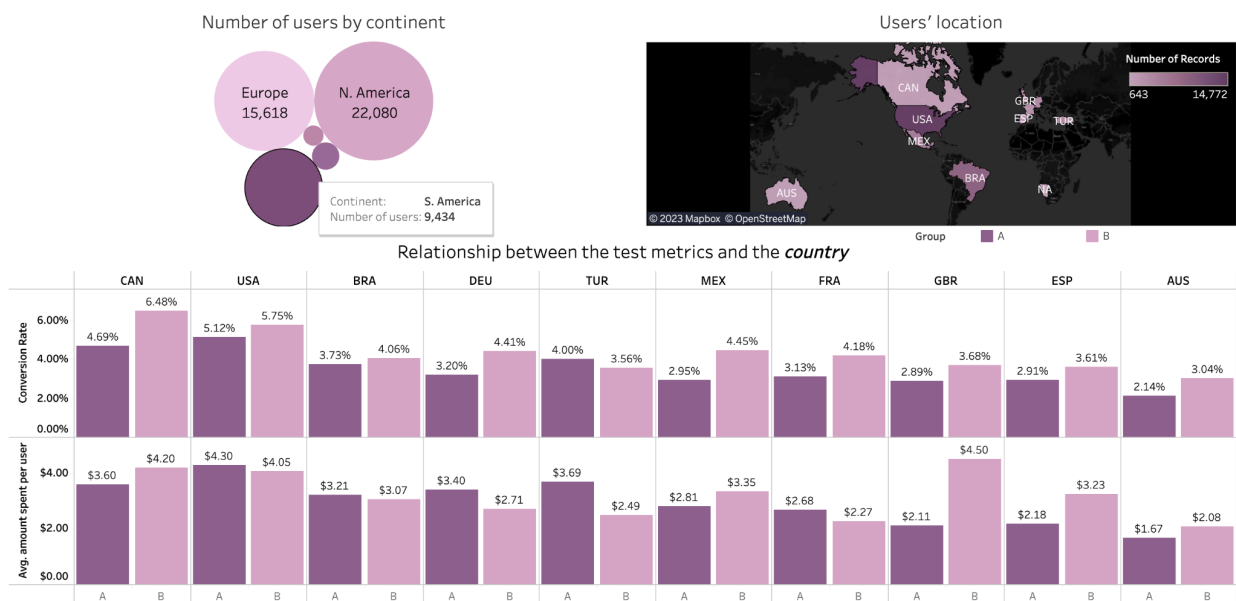
Insights 1: Test group B clearly exhibits a higher conversion rate but not a higher average amount spent per user, whereas control group A consists of a larger number of individuals who spend on higher-priced items. A possible cause for this discrepancy could be attributed to the new banner's focus on promoting food and drink products, which deviate from Globox's typical high-end and luxury offerings.

Conversion Rate by group

A	3.92%
B	4.63%

Avg. amount spent per user by group

A	$3.37
B	$3.39

Distribution of amount spent per user by group

amount spent($)

Group
A
B

Insights 2: We observe a consistent improvement in conversion rates across users' devices and genders in test group B. However, there is no significant increase in the average amount spent per user. Once again, this observation suggests a possible correlation between the banner's content and the observed results.

Number of users by device

Android	15,235	15,054
IOS	9,218	9,142
NA		

Group
B
A

Number of users by gender

F	10,061	10,069
M	10,235	10,054
NA	3,443	3,412
O		

Relationship between the test metrics and the *device*

Relationship between the test metrics and the *gender*

Insights 3: We observe that in test group B, there is a consistent improvement in the conversion rate across all countries except Turkey. We have especially noticed a significant drop in the average amount spent per user for Turkey. We wonder if this could be attributed to the content of the new banner, which promotes drinks, and may not align with religious beliefs in Turkey. To gain a better understanding, we can collect more data and conduct further analysis. Despite this, it is promising to see an overall improvement in the conversion rate across the majority of countries.

Number of users by continent

Users' location

Relationship between the test metrics and the *country*

Insights 4: Contrary to the novelty effect, upon examining the difference in key metrics between the groups over time, we have observed that the effectiveness of the banner is not short-lived. This finding underscores the significance of launching the banner, as it continues to demonstrate positive results beyond the initial novelty period. (To achieve this, we rewrote the query to include order date data.)

Check for Novelty Effects
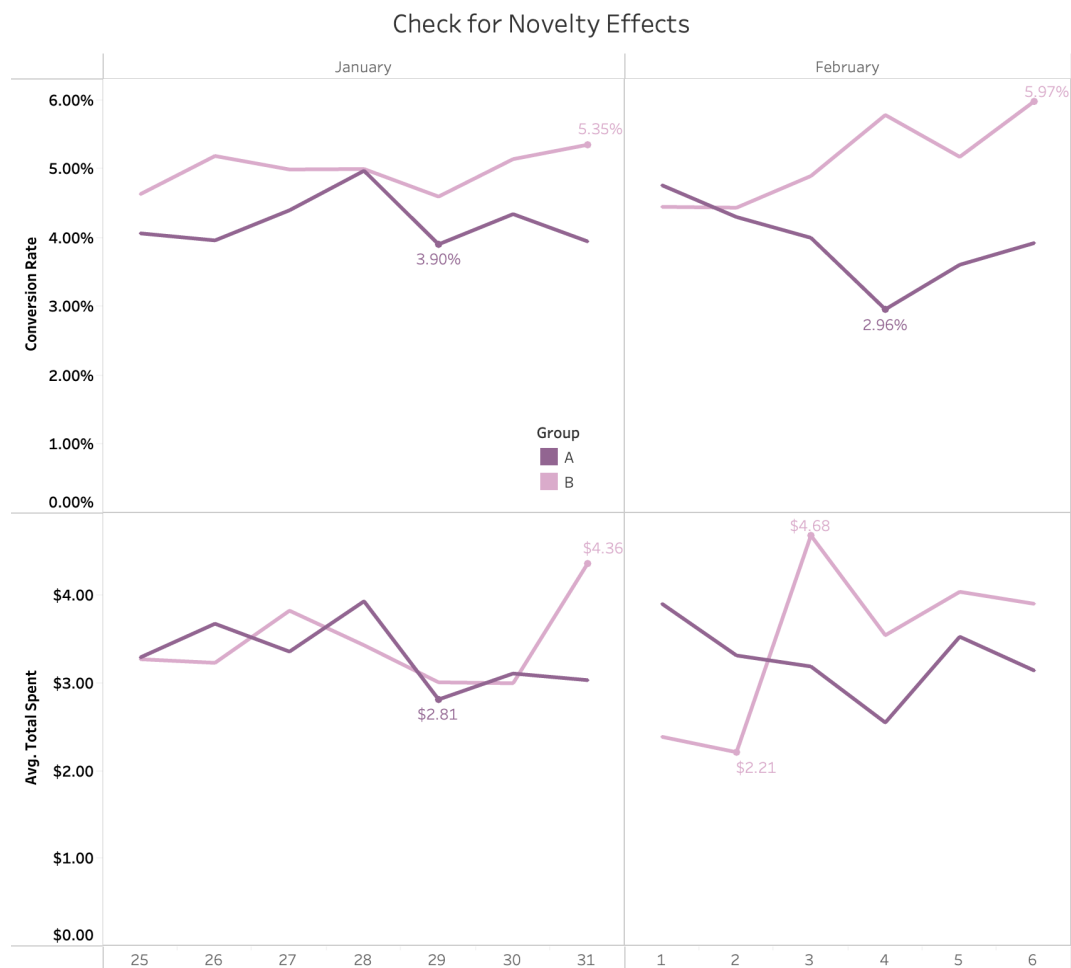
Insights 5: From the visualization of the 95% confidence intervals, we can be 95% certain that the difference in conversion rates falls within the range of 0.0035 to 0.0107, with an estimated point of 0.007. This confirms that the A/B test results suggest the new banner has a significant impact on the conversion rate, as demonstrated by group B, which shows a higher conversion rate. Conversely, we can be 95% certain that the difference in the average amount spent per user falls within the range of -0.439 to 0.471, with an estimated point of 0.016. Therefore, no significant difference exists between the two groups in terms of the average amount spent per user. This implies that while the new banner effectively increases conversions, it may not significantly influence the amount customers spend per order.

95% Confidence Intervals of difference in
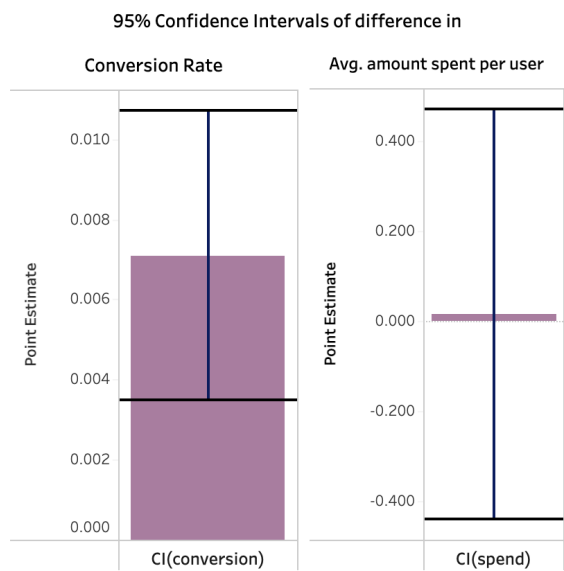
Insights 6: In order to achieve our desired minimum detectable effect and statistical power, we utilized a power analysis calculator to determine the necessary sample size. The analysis indicated that we had sufficient sample size to detect the conversion rate. However, for our average amount spent per user metric test, we found that our sample size was not large enough to ensure sufficient sensitivity. Therefore, we could conduct the test again on a larger scale.

Power Analysis (Conversion Rate)			Power Analysis (Average amount spent per user)	
Baseline Conversion Rate (%)	3.92		Power	0.8
Minimum Detectable Effect (%)	18%		Type II error	0.2
Significance (α)	0.05		difference between mean	0.016
Statistical Power (1 – β)	0.8		Expected Standard Deviation	25.94
Results			**Results**	
Total Sample Size	18,700		Total Sample Size	82,521,602

Recommendation

We strongly recommend launching the new banner. In spite of not observing a higher average amount spent per user, the spend didn't decrease and significantly we observed a higher conversion rate. This discrepancy could be attributed to the focus of our new banner, which emphasizes food and drink products rather than Globox's typical boutique fashion and high-end offerings. As a result, the food and drink section generates less revenue per order compared to our typical products. Something that could be verified with further testing.

Nevertheless, the higher conversion rate is a positive sign in attracting more paying customers. It presents an opportunity to engage these customers and potentially encourage additional purchases in the future. By conducting further analysis of users' conversion rates and spending patterns across different devices, genders, and regions, we have observed significant improvements in our success metrics. This gives us the confidence to release the banner to all users.

Overall, while the average amount spent per user may not be higher initially, the positive impact on conversion rate and the potential for future purchases make the launch of the banner a promising step towards acquiring more paying customers. We believe that with continued engagement and targeted strategies, we can successfully draw users back to make additional purchases. Therefore, we confidently recommend launching the banner to all users.

Appendix

Video Presentation

Tableau dashboards

Spreadsheets

SQL query:

```sql
SELECT
    u.id AS id,
    Coalesce(country,'NA') AS country,
    Coalesce(gender, 'NA') AS gender,
    (CASE WHEN g.device='I' THEN 'IOS'
        WHEN g.device='A' THEN 'Android'
    ELSE 'NA'
    END) AS device,
    g.group AS group,
    (CASE WHEN spent>0 then 1 ELSE 0 END) AS converted,
    CAST(SUM(Coalesce(a.spent,0)) AS DECIMAL(10, 2)) AS total_spent
FROM users as u
LEFT JOIN groups as g
ON u.id=g.uid
LEFT JOIN activity as a
ON u.id=a.uid
GROUP BY u.id, u.country, u.gender, g.device, g.group, converted
```