GloBox A/B Test Analysis

Result and Recommendation Report

LIEZYL JUGALBOT

OCTOBER 2023

CONTENTS

EXECUTIVE SUMMARY

The A/B test for GloBox's food and drink banner on the mobile website showed a statistically significant increase in the conversion rate but no significant change in the average amount spent per user. Therefore, I recommend continuing iterating on the banner experience and conducting further analysis before considering a full-scale launch, ensuring a better understanding of its impact on revenue and user experience.

CONTEXT

GloBox is an e-commerce company specializing in sourcing unique and high-quality products from around the world. While it's renowned for boutique fashion items and high-end decor products, the company has experienced significant growth in its food and drink offerings over the last few months.

MOTIVATION

The primary motivation behind conducting an A/B test is to increase awareness of the food and drink product category and, consequently, boost revenue. GloBox's mission is to bring the world's treasures to its customers' doorsteps, offering everything from exotic spices to handmade jewelry.

A/B TEST SETUP

The A/B test aimed to assess the impact of prominently featuring critical products in the food and drink category on the website. Users who visited the GloBox main page were randomly assigned to either the control or treatment group, marking their join date. For users in the treatment group, the landing page displayed the banner highlighting food and drink items, whereas the control group saw the standard website without this feature.

Group A: Control
existing landing page

Group B: Treatment
landing page with food & drink banner
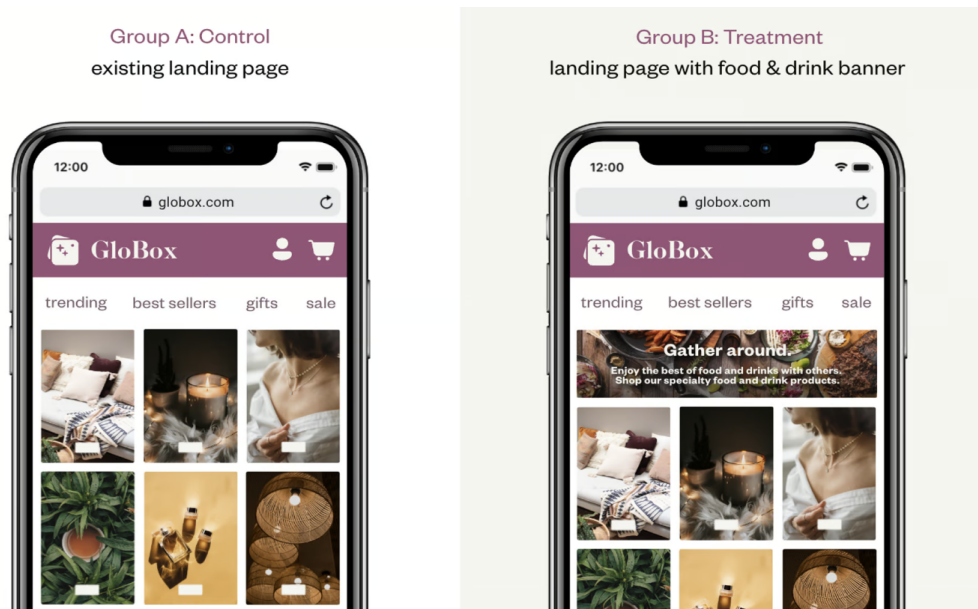
The experiment was executed for a 13-day period from January 25 to February 6, 2023. There were 24,343 users in the control group and 24,600 users in the treatment group, a total of 48,943 users.

The subsequent actions of users, such as making purchases on the same day they joined the experiment or in the days that followed, were tracked as conversions.

DATA OVERVIEW

The dataset used in this analysis contained users' demographic information, their group assignments, and whether or not they made purchases after viewing the website with or without the food and drink banner.

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

EXTRACTING AND CLEANING THE DATASET

I wrote a SQL query code that returns the user ID, the user's country, the user's gender, the user's device type, the user's test group, whether or not they converted (spent > $0), and how much they spent in total ($0+). Downloaded the data in CSV and Excel formats and used it for the following analysis phase.

RESULTS

HYPOTHESIS TEST

- The null hypothesis (H0) stated that there is no difference in the conversion rate/average amount spent between the two test groups.
- The alternative hypothesis (H1) stated that there is a difference in the conversion rate/average amount spent between the two test groups.

4

Hypothesis Test for the Difference in Conversion Rate

Two-sample **z-test** with pooled proportion

A - Control   B - Treatment

test_group	COUNT of user_id	SUM of conversion	AVERAGE of conversion
A	24343	955	0.0392
B	24600	1139	0.0463
Grand Total	**48943**	**2094**	**0.0428**

Calculation	Notation	Value
sample size (A-control)	n1	24343
sample size (B-treatment)	n2	24600
sample mean (A-control)	x1 bar	0.0392
sample mean (B-treatment)	x2 bar	0.0463
sample proportion	p-hat	0.0428
standard error	SE	0.0018
test statistic	T	-3.8643
p-value	pval	**0.0001**

The hypothesis test results provide statistically solid evidence of a significant difference in conversion rates between the control group and the treatment group.

With the p-value = 0.0001 < 0.05, we **reject the null hypothesis** that the conversion rate is the same between the two groups in favor of the alternative hypothesis that there is a difference in the conversion rate between the two groups.

Hypothesis Test for the Difference in Average Amount Spent

Two-sample **t-test** with unpooled variance

A - Control   B - Treatment

test_group	COUNT of user_id	AVERAGE of total_spent	STDEV of total_spent
A	24343	3.375	25.936
B	24600	3.391	25.414
Grand Total	**48943**	**$3.38**	**25.675**

Calculation	Notation	Value
sample size (A-control)	n1	24343
sample size (B-treatment)	n2	24600
sample mean (A-control)	x1 bar	3.375
sample mean (B-treatment)	x2 bar	3.391
sample std dev (A-control)	s1	25.936
sample std dev (B-treatment)	s2	25.414
standard error	SE	0.2321
test statistic	T	-0.07043
degrees of freedom	df	24342
p-value	pval	**0.94385**

(see Appendix for the spreadsheet link)

The hypothesis test results for the difference in the average amount spent between the two groups indicate **no statistically significant difference** in the average spending of the users.

With the p-value = 0.94 > 0.05, we **fail to reject the null hypothesis** that the average amount spent is the same between the two groups in favor of the alternative hypothesis that there is a difference in the average amount spent between the two groups.

The chart provides a visual comparison of the key metrics between the two groups:

The line chart compares the two groups' conversion rates. The treatment group showed a higher conversion rate of 4.63% compared to the control group.

The bar chart illustrates no significant difference in the average amount spent between the two groups.
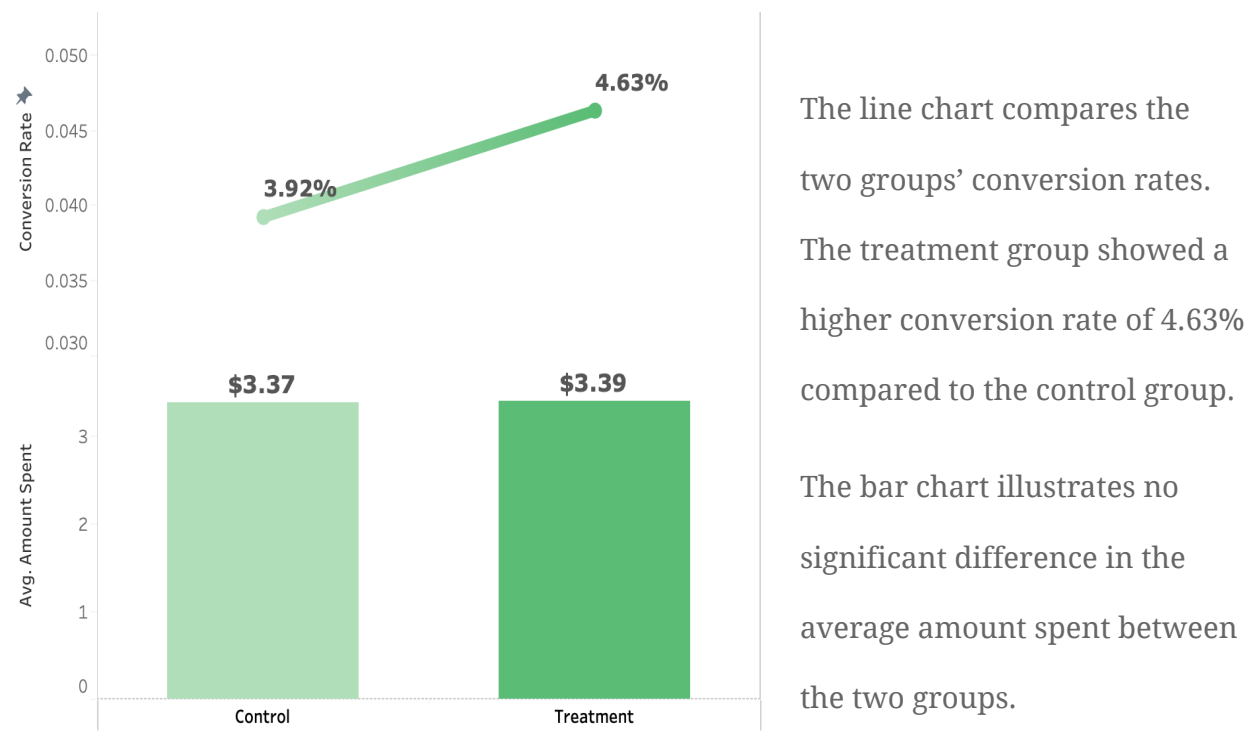
The hypothesis tests indicate that the presence of the food and drink banner did not have a statistically significant impact on the average amount spent by users. While the conversion rate showed a significant difference, suggesting that more users from the treatment group made purchases, the average spending per user remained consistent between the two groups. This result has implications for GloBox's revenue strategy, indicating that the banner primarily influences user engagement and conversion rather than the amount spent per transaction.

CONFIDENCE INTERVALS

Confidence Interval for a Difference in Conversion Rate			
Two-sample **z-interval** with unpooled variance			
A - Control B - Treatment			
test_group	COUNT of user_id	SUM of conversion	AVERAGE of conversion
A	24343	955	0.0392
B	24600	1139	0.0463
Grand Total	**48943**	**2094**	**0.0428**

Calculation	Notation	Value
sample size (A-control)	n1	24343
sample size (B-treatment)	n2	24600
sample mean (A-control)	p1	0.0392
sample mean (B-treatment)	p2	0.0463
degrees of freedom	df	24342
critical value	c	1.96
standard error	SE	0.0018
margin of error	E	0.0036
lower bound	LB	**0.0035**
upper bound	UB	**0.0107**

The calculated confidence interval for the conversion rate is (0.0035, 0.0107). We can say that we are 95% confident that the difference in the conversion rate falls between 0.35% and 1.07%. Moreover, this confidence interval **does not include the value 0**, indicating that **there is a statistically significant difference** in the conversion rates between the two groups.

Confidence Interval for a Difference in Average Amount Spent

Two-sample **t-interval** with unpooled variance

A - Control B - Treatment

test_group	COUNT of user_id	AVERAGE of total_spent	STDEV of total_spent
A	24343	3.375	25.936
B	24600	3.391	25.414
Grand Total	**48943**	**$3.38**	**25.675**

| Calculation	Notation	Value
sample size (A-control)	n1	24343
sample size (B-treatment)	n2	24600
sample mean (A-control)	x1 bar	3.375
sample mean (B-treatment)	x2 bar	3.391
sample std dev (A-control)	s1	25.936
sample std dev (B-treatment)	s2	25.414
degrees of freedom	df	24342
critical value	c	1.96
standard error	SE	0.2321
margin of error	E	0.4550
lower bound	LB	**-0.4387**
upper bound	UB	**0.4714**

The confidence interval for the difference in average amount spent is (-0.4387, 0.4714).

With 95% confidence, the difference in the average amount spent falls between -0.4387 and 0.4714. More importantly, the confidence interval **includes the value 0,** indicating a chance that the actual difference in the average amount spent between the two groups is zero or, in other words, there is **no statistically significant difference** in the average amount spent.

VISUALIZING THE CONFIDENCE INTERVALS

**Confidence Intervals for the difference in conversion rate**
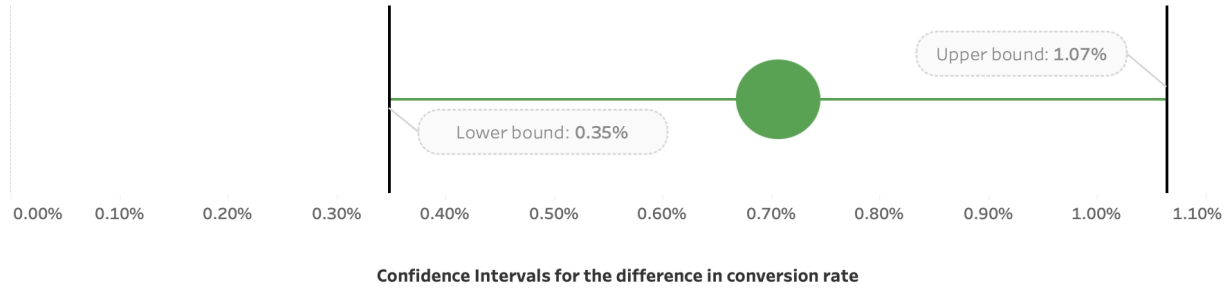
Considering practical significance being +5% MDE, the entire CI is less than the practical significance but there is statistical significance since 0 is not in the bound.

Confidence Intervals for the difference in average amount spent
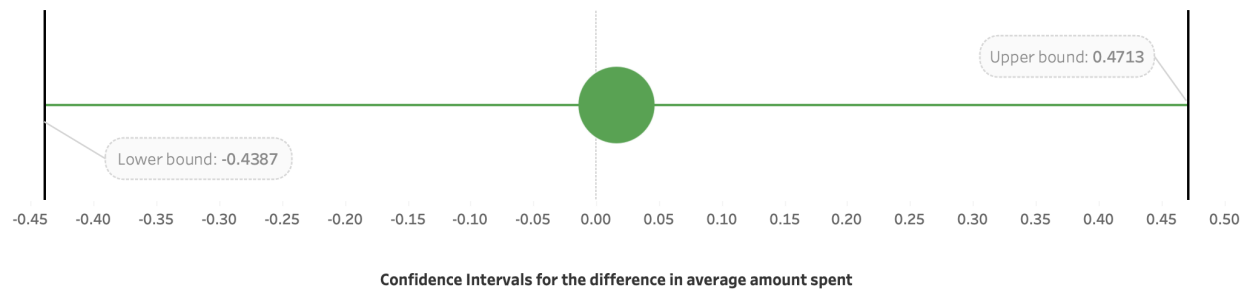
The entire CI for the difference in the average amount spent is less than the practical significance, and there's no statistical significance since 0 is in the bound.

NOVELTY EFFECT
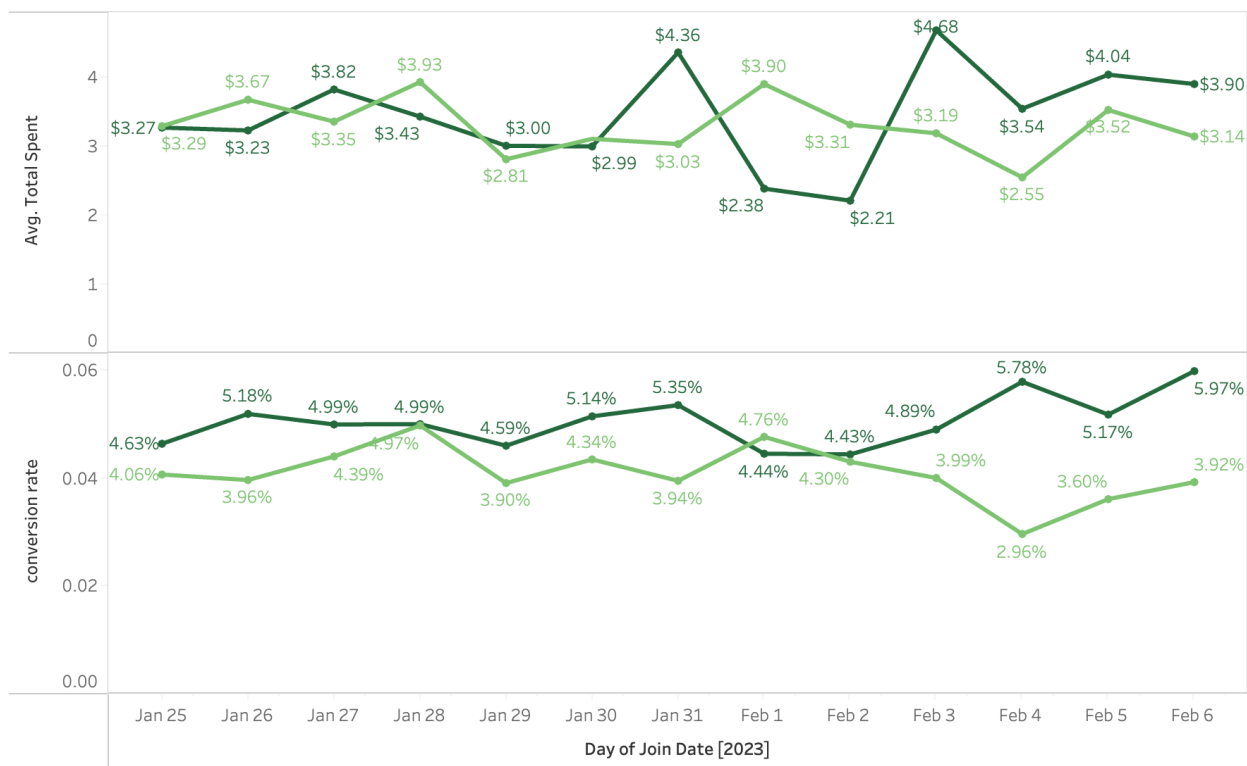
The introduction of the food and drink banner is related to an acceleration in the conversion rate, particularly in the later stages of the experiment. The statistics show **no detectable novelty effect** with the introduction of the banner, indicating that immediate and significant modifications in user behavior generally associated with novelty are not present. Nonetheless, the banner may have an impact on the unpredictable behavior exhibited in the average amount spent without demonstrating an obvious novelty effect. Overall, the lack of unexpected effects and the elusive nature of the novelty effect suggest that the banner integration has been smoothly integrated into the user experience, reducing disturbances and aligning with user preferences.

POWER ANALYSIS

48,943 mobile users participated in the A/B test, an approximately 50/50 split of 24,343 users in the baseline and 24,600 users in the test group. Given the predefined significance level (0.05) and statistical power (0.80), the test is **sufficiently sensitive to detect effect size in conversion rates but insufficient for the average amount spent**.

The sample size of around 12.7K users per group ensures that the test has reasonable power to identify meaningful changes in conversion rates. On the other hand, assuming a pooled standard deviation of 25, the test would require a sample size of around 24M for each group to achieve a power of 80% and a level of significance of 5% (two-sided) for detecting a true difference in means between the two groups.
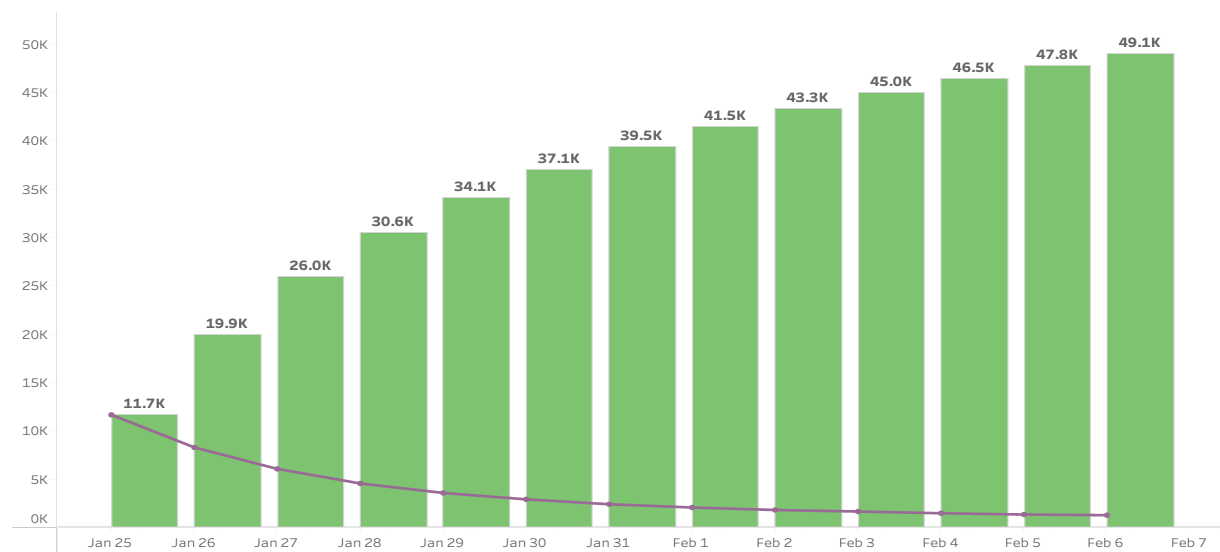
The line on the chart indicates the number of users joined daily. **Sample size reaches sufficiency in 3 days to detect changes in conversion rate**. We need to ramp up more for the average amount spent.

RECOMMENDATION

My recommendation is to **continue iterating** on the food and drink banner experiment rather than immediately launching it to all users. We needed to see more improvement in our success metrics to be confident in releasing the feature in its current state. However, some promising results show we could make changes to the banner experience and get better improvement next time. We could do further data analysis to understand this better or need a larger sample size to make a confident recommendation.

The banner positively affects user engagement, but launching based on one success metric would take time and effort. It's essential to evaluate its impact on profit in addition to revenue. We should also consider costs before the banner launch. It is necessary to evaluate adverse effects on user experience or significant resource requirements if clear financial benefits are absent.

REFERENCES

1. 🔷 SQL Query Codes for Globox Project_liez
2. 🟩 Statistics in Spreadsheets_liez
3. [Calculator - Estimated sample size for proportions](#)
4. [Calculator - Estimated sample size for means](#)
5. [Tableau Visualizations](#)