

GloBox A/B Test Analysis Report

By

David Adeleye

November 2023

Table of Contents

Summary	3
Data and Approach	3
Context	4
Results	5
Summary	10
Recommendation	10
Appendix	10

Summary

I strongly recommend not launching the banner yet because of the A/B test results. The results demonstrated a notable boost in conversion rates for the treatment group compared to the control group which would suggest that the banner is effective in enhancing awareness of the food and drink category and driving more conversions. However, it's crucial to note that the test didn't reveal any significant variance in the average amount spent. Continuous monitoring of the average amount spent as well as longer observation time is recommended to gain a more comprehensive understanding of the banner's impact on user spending behaviour.

Data and Approach

The GloBox data used consists of 3 tables stored in the PostgreSQL database. This database was accessed via Beekeeper IDE and SQL queries were used to extract the information needed for the A/B test. The methodology involved the following steps:

1. Data Extraction with SQL Queries:

This was the first step. It involved extracting the necessary data for the A/B test from the GloBox database. This database consists of 3 tables:

- Users: This table contains the users' ID and users' demographic information such as country and gender. The country and gender features were stored as text data types while the users' ID was stored as int4(32,0).
- Groups: this table contains the test group to which each user was assigned, their device type and the date they joined the test. The date they joined the test was stored as a date data type while the group and device features were stored as text data types.
- Activity: this table contains the users' purchase activity. It captured the users' ID, purchase date, device type and the amount spent. The purchase date was stored as date data type, the amount spent was stored as int4(32,0) data type while the device type was stored as text data type.

I made use of SQL queries which I ran using Beekeeper IDE to gather relevant information for the A/B test analysis of the 2 groups – control and treatment groups.

2. Hypothesis Testing with Spreadsheet:

After gathering the data, I proceeded to conduct hypothesis testing to compare the conversion rates and average spending between Group A (control) and

Group B (treatment). For testing these parameters, I established the null hypothesis (H_0) suggesting no disparity in conversion rates and average spending between the two groups, while the alternative hypothesis (H_1) posited the presence of a difference.

I employed both the t-test and z-test with a significance level set at 5%, utilizing Google Spreadsheets to assess the statistical significance of the findings. I employed the t-test for the difference in the average amount spent because the mean of the 2 groups was being compared. I employed the z-test for the difference in the conversion rates because the difference in two proportions was being analysed.

3. Data Visualisation with Tableau:

In order to delve into user behaviour and spending trends more comprehensively, I utilized Tableau for data visualization. I generated visual representations to examine the distribution of average spending, conduct analyses based on gender, device usage, and explore country and region patterns. These visualizations proved instrumental in pinpointing trends and patterns, which, in turn, informed my recommendations.

Context

GloBox, an e-commerce company known for its distinctive and high-quality products, executed an A/B test with the aim of boosting awareness and revenue within its food and drink category. This test was specifically implemented on the mobile website, targeting users who landed on the main GloBox page. Upon joining the experiment, users were randomly assigned to either the Control or Treatment group based on their join date. For the treatment group, a banner showcasing food and drink products was presented, while the control group did not encounter this banner. The tracking of user conversions was conducted, defining conversion as a user making a purchase either on the day they joined or on a subsequent date.

Examining this dataset enabled an exploration of how the presence of the food and drink category banner influenced both conversion rates and the average spending per user. Furthermore, the dataset presented a chance to delve into user behaviour, considering variables like gender, device type, and geographical

location (country or region). These insights, in turn, provided valuable information for crafting targeted marketing strategies.

Results

Group	Sample Size	Average Amount Spent	Conversion Rate
A (Control)	24,343	\$3.37	3.92%
B (Treatment)	24,600	\$3.39	4.63%

Average Amount Spent:

The Null Hypothesis (H_0) posited that there is no disparity in the average amount spent per user between Group A and Group B, while the Alternative Hypothesis (H_1) asserted that a difference exists in the average amount spent between the two groups.

The t-test yielded a p-value of 0.944, and the confidence interval was from -0.439 to 0.471.

Considering the p-value, I found no grounds to reject the null hypothesis. This suggests that there is no significant statistical distinction in the average amount spent per user between the two groups.

Conversion Rate:

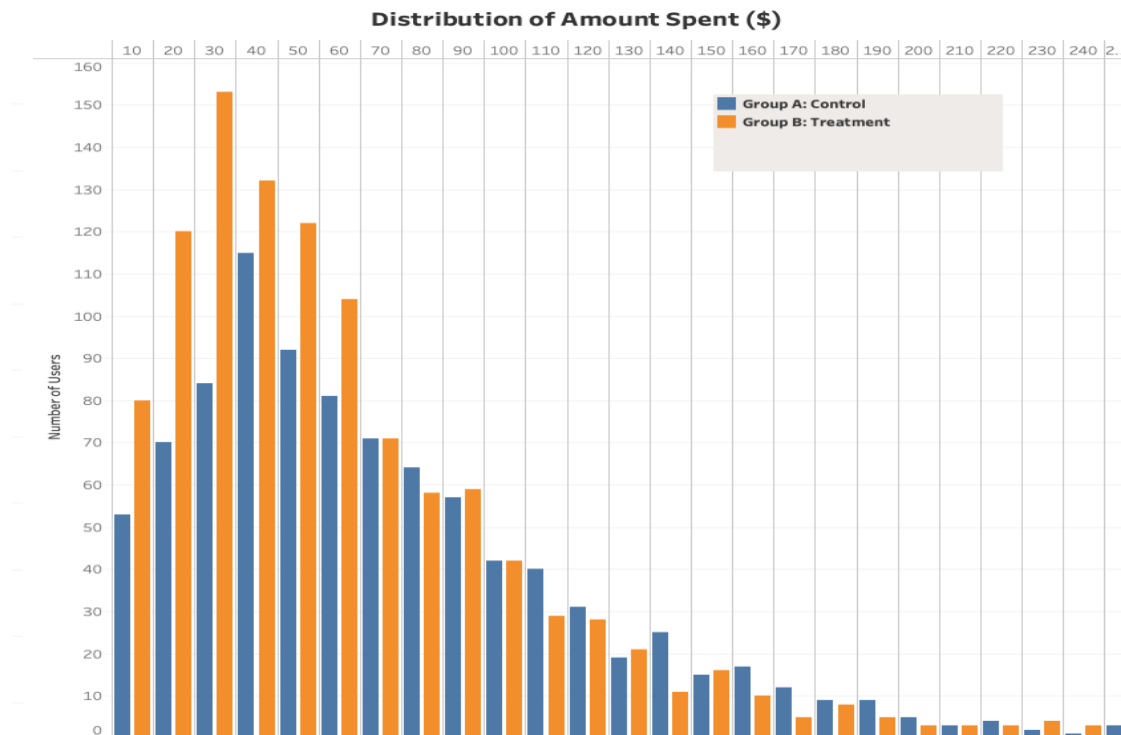
The Null Hypothesis (H_0) posited that there is no difference in the conversion rate between Group A and Group B, while the Alternative Hypothesis (H_1) asserted that a difference exists in the conversion rate between the two groups.

A significance level of 0.05 was chosen, and the computed p-value was 0.0001. The 95% confidence interval for the difference in conversion rates was found to be 0.0035 to 0.0107.

The statistical analysis led to the rejection of the null hypothesis, indicating a significant distinction in conversion rates between the two groups. The positive values within the confidence interval suggest that the treatment group likely demonstrates a higher conversion rate compared to the control group. However, it's crucial to correctly interpret the confidence interval. The 95% confidence interval implies that if we were to repeat this experiment 100 times, we would

anticipate the true difference in conversion rates to fall within 0.35% and 1.07% in approximately 95 of those experiments.

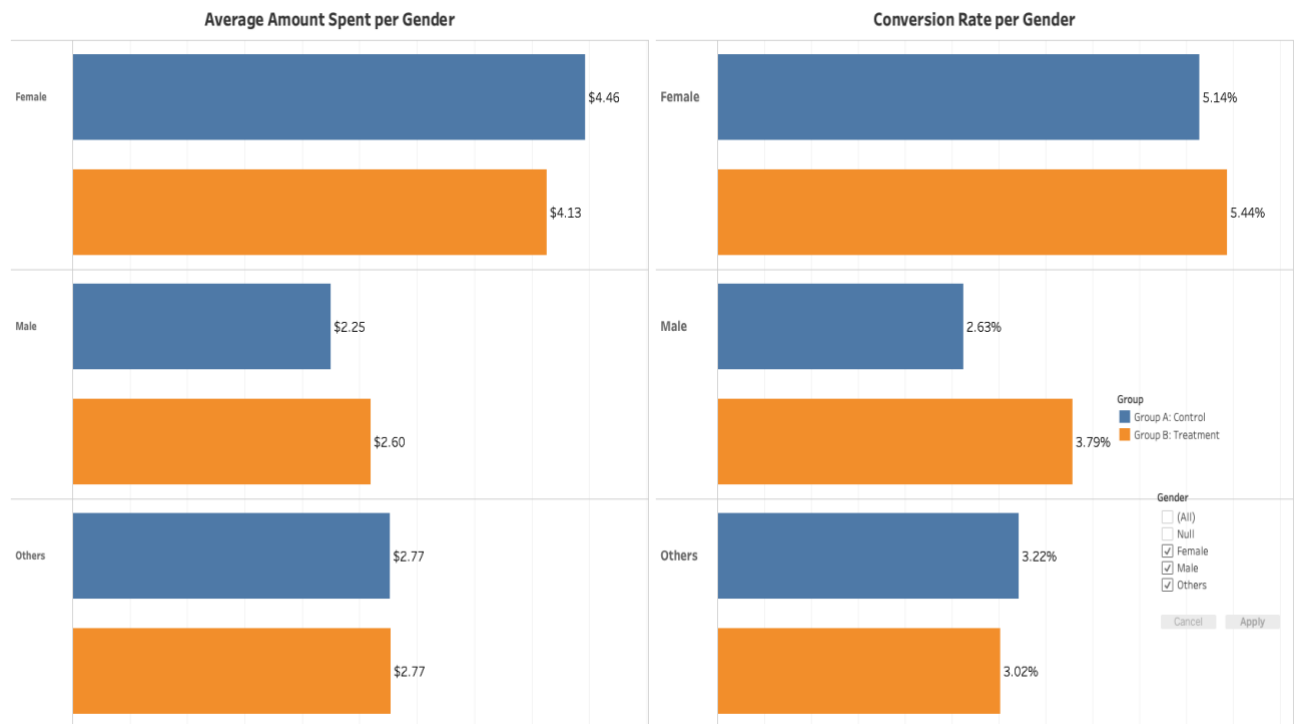
Distribution of Average Amount Spent:



Insights:

- The majority of both groups spent between \$30 and \$50.
- The distribution of the average amount spent is right skewed.

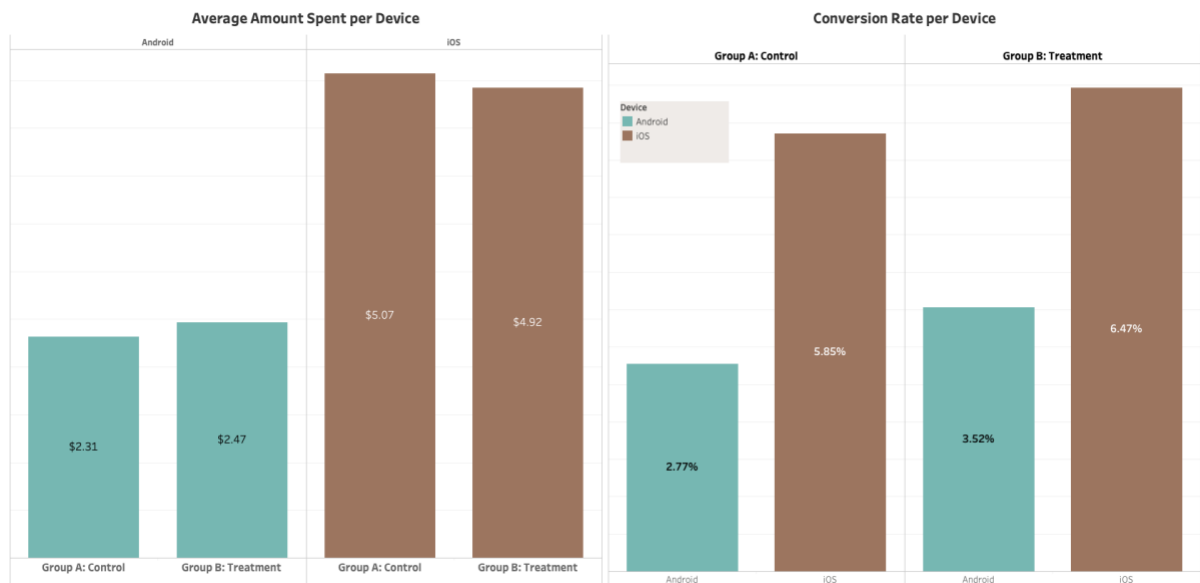
Gender-wise Analysis:



Insights:

- For the average amount spent, none of the genders show a significant difference between the control and treatment groups. However, it is observed that the average amount spent by the female gender far exceeds the average amount spent by other genders.
- For the conversion rate, each of the genders shows a significant difference between the control and treatment groups. The conversion rates are significantly higher for the treatment group across the male and female. Also, for each group, the female conversion rate is higher than for other genders.

Device-wise Analysis

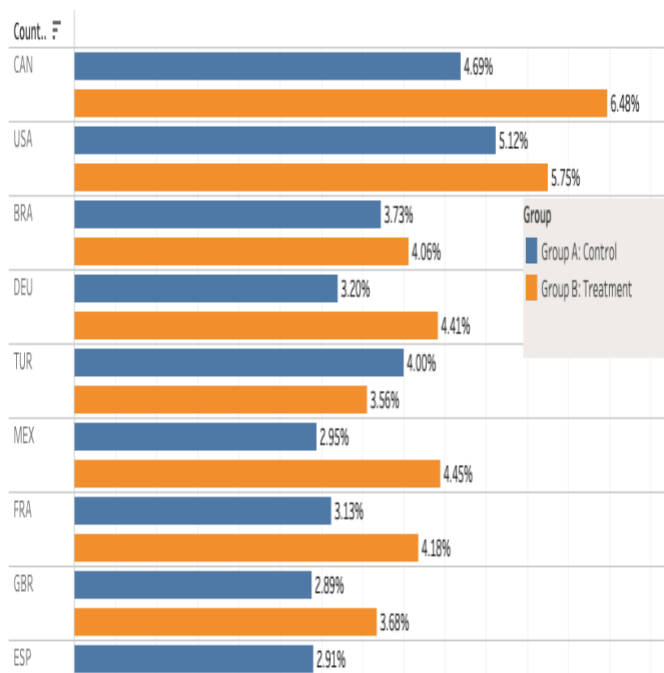


Insights:

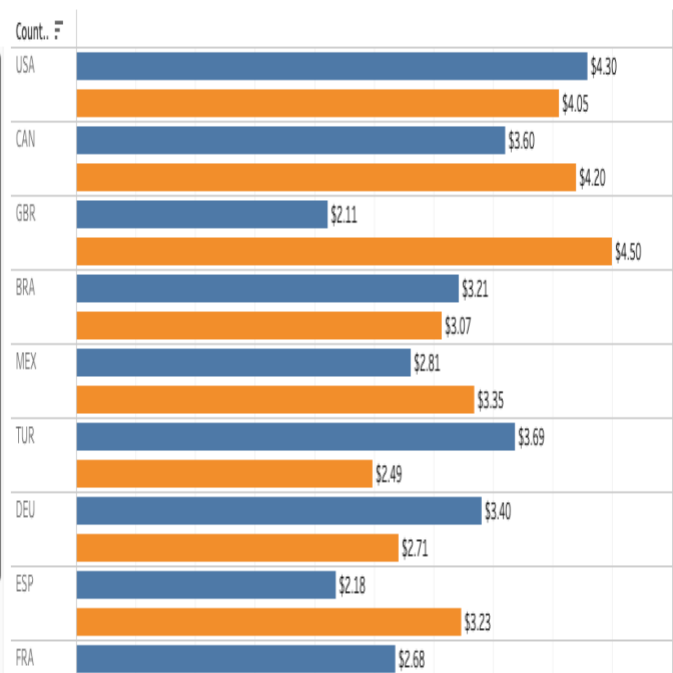
- For the average amount spent, there was no significant difference between the difference in spending by Android users to difference in spending by iOS users. Hence, we can say that the device type did not play a significant role in the A/B test for the average amount spent.
- For the conversion rate, while there was a significant increase in conversion rate between the treatment and control groups, the increase was similar for both Android and iOS users. Hence, we can also say that the device type played no significant role in the A/B test for conversion rate.

Country-Wise Analysis:

Conversion Rate per Country



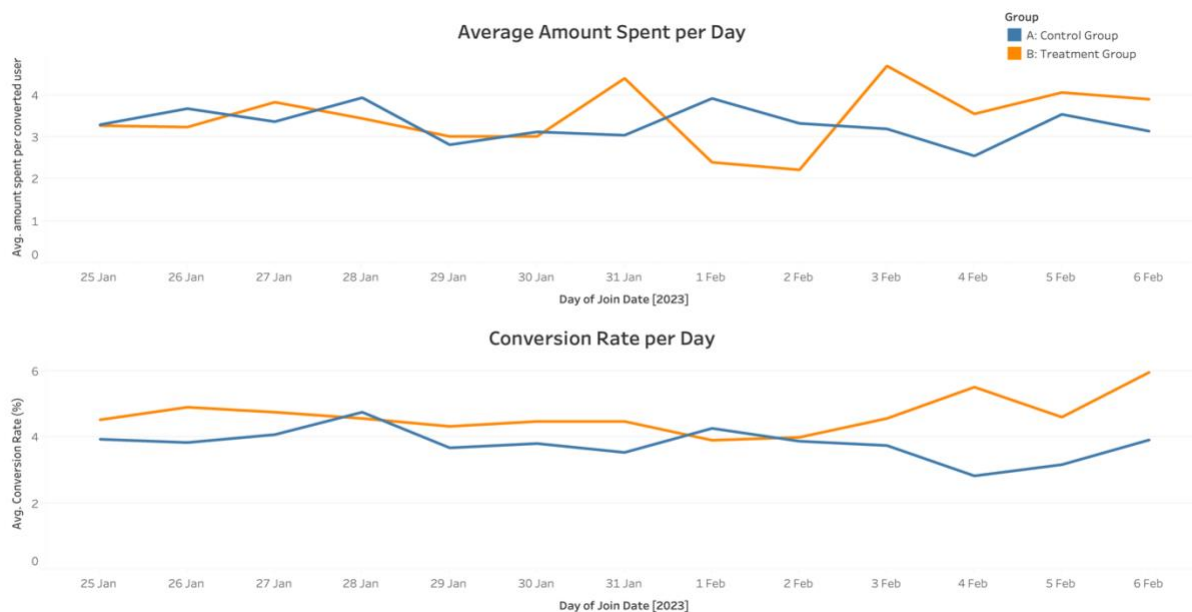
Average Amount spent per Country



Insights:

- For conversion rates, it was observed that the conversion rates were higher for Canada and the US
- For the amount spent, it was observed that once Canada, US and UK were the leaders.

Novelty Effect:



Insights:

- For the average amount spent, there is no clear novelty effect.
- For the conversion rate, it was observed that after the 2 Feb, the gap between the two groups began to widen. This suggests that there could be a novelty effect on the A/B test for conversion rates between the two groups. However, this is not conclusive. Further days would be required to be sure of this.

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.

For conversion rate, we assume we want a commonly used power value of 0.8. That means we want to have 80% probability that we will correctly reject the null hypothesis. The threshold for significance will be $\alpha = 0.05$. We want to have a 10 percent difference in conversion rate between the treatment and control groups. With these parameters, the total sample size should be 77K with 38.5k users in each group. This was calculated using this [calculator](#).

For the average amount spent, suppose we aim for a 10% difference in the mean of both groups. With a pooled standard deviation of 25.94, the study would need

a sample size of 93,008 for each group. In simpler terms, by randomly selecting 93,008 users for each group and observing a 0.337 (10%) difference in mean between the two, there's an 80% chance of confirming that the means of the two groups are different. This was calculated using this [calculator](#).

Summary

1. When comparing the key metrics of the control and treatment groups:
 - There was a significant increase in the conversion rates of users who saw the banner.
 - There was no significant difference in revenue.
2. Only Turkey showed a reduction in the conversion rate.
3. On average, the female gender spent the most as well as had the highest conversion rate.
4. IOS users spent the most.
5. No novelty effects were detected.

Recommendation

The banner is easy to launch and maintain. However, it takes up high-value real estate on the main page. It is essential to have a high degree of confidence regarding the impact of the banner before launching it.

As only one of the objectives was met, I recommend that we iterate the test with the following considerations:

1. A sample size of at least 77K users split equally for sufficient power.
2. A longer duration of six weeks to ensure that conversion and revenue increases are not due to the timing of payday.
3. Include the category and time of purchase in data collection.
4. Equal number of Android and iOS users.
5. Equalised number of users joining every day to check for novelty effects.

Appendix

A. SQL Queries

-- *The following query has been used for the test statistics

```
SELECT
    u.id AS user_id, u.country, u.gender, g.device,g.group,
    SUM(COALESCE(a.spent, 0)) AS total_spent_usd
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;
```

-- * Novelty Effect Analysis:

-- Converted Users Average Amount Spent Over Join Date:

```
SELECT
    g.join_dt AS join_date,
    g.group,
    COUNT(DISTINCT g.uid) AS total_users,
    COUNT(DISTINCT a.uid) AS paid_users,
    SUM(a.spent) AS total_spent
FROM
    groups AS g
LEFT JOIN activity AS a ON g.uid = a.uid
GROUP BY
    g.group,
    g.join_dt
ORDER BY 1;
```

-- All Users' Metrics Over Join Date:

```
SELECT
    n.join_date,
    n.group,
    ROUND(CAST(SUM(n.paid_users) / MAX(n.total_users) * 100 AS
    DECIMAL(10,2)), 2) AS conversion_rate,
```

```

ROUND(CAST(SUM(n.total_spent)/MAX(n.total_users) AS DECIMAL(10,2)),2)
AS
average_spent
FROM(SELECT
    g.join_dt AS join_date,
    g.group,
    COUNT(DISTINCT g.uid) AS total_users,
    COUNT(DISTINCT a.uid) AS paid_users,
    SUM(a.spent) AS total_spent
FROM
    groups AS g
LEFT JOIN activity AS a ON g.uid = a.uid
GROUP BY
    g.group,
    g.join_dt
ORDER BY 1) AS n
GROUP BY 1, 2;

```

B. Spreadsheet

https://docs.google.com/spreadsheets/d/1clt3NZ_yiSSDnTKEBlmDw25FcwmaGSECxCGQD-45yts/edit?usp=sharing

C. Tableau Visualisation

https://public.tableau.com/app/profile/david.adeleye/viz/Book3_17034452739680/GloBoxABTest?publish=yes