

A/B Test Report and Recommendations for GloBox

Date: August 21, 2023

To: [Product Manager], [Marketing Manager], [Head of Engineering]

From: Benneth Chukwuemeka

Subject: A/B Test Report and Recommendations

The Growth, Product, and Engineering Team,

I am pleased to present the findings and recommendations from the recent A/B test conducted to **evaluate the impact of placing a banner of the food category on our GloBox website and how this has affected product performance**. This test aimed to provide insights into user engagement, conversion rates, spending, and perhaps overall user experience. The cross-functional collaboration among the product, marketing, and engineering teams was essential to the success of this endeavor.

EXECUTIVE SUMMARY

The setup of the A/B test is as follows:

The experiment is only being run on the mobile website, with data from just a sample of the entire population.

A user visits the GloBox main page and is randomly assigned to either the control or treatment group. This is the join date for the user.

The page loads the banner if the user is assigned to the treatment group and does not load the banner if the user is assigned to the control group.

The user clicking on the page is termed "conversion" and may or may not purchase products from the website. It could be on the same day they join the experiment, or days later. A key thing to note is that purchase is for all categories and not just for the food product category and the assumption is that all sampled are rational customers.

The test ran from 2023-01-25 to 2023-02-06 (**12days**) and enlisted a total of **48943** who were randomly assigned to the two Group A (Control) **24343** and Group B (Treatment) **24600**. Data generated was

stored as CSV files and was analyzed using SQL, spreadsheet, and Tableau. (Refer to appendix). While the experiment showed that the new banner led to changes in conversion rate and that we can rely on the result that the true parameter is similar to the observed samples, the same cannot be said about the spending pattern in both groups as the banner did not affect the total spend in both groups, and uncertainty with the true parameters and observed samples.

The lack of clarity in defining indices like the category of the product purchased, optimizing sample size based on our chosen minimum detectable effect (10%), and baseline conversion rate, could potentially lead to ambiguity in interpreting the outcomes and making informed decisions based on the results. This situation could hinder our ability to draw meaningful insights from the experiment and subsequently impact the effectiveness of the strategies we develop. Having outlined this, I believe there is potential in the experiment's outcome if properly done, and there will be a need to **iterate and test again**. The presentations below will further elucidate this.

1. Test Overview:

Objective: The primary goal of the A/B test was to determine whether the new food and drink banner on the website would lead to improved user engagement by affecting the conversion rates and total spend

DATA EXTRACTION

By leveraging SQL's comprehensive functionalities, I efficiently extracted, transformed, and manipulated user-level aggregated dataset to provide valuable insights for informed decision-making. I answered some fundamental questions giving reasons. "In this query: We are used LEFT JOIN to ensure all users from the "users" table are included, regardless of whether they have matching records in the "groups" or "activity" table. The CASE statement is used to create the "converted" column, which will be 1 if the user spent more than \$0 and 0 otherwise. The SUM(a.spent) function calculates the total amount spent by each user.

Note: If there are multiple rows for a single user in the "activity" table, the GROUP BY clause ensures that the total spent is correctly aggregated for each user.

This query will provide the user ID, country, gender, device type, test group, whether the user converted (spent > \$0), and the total amount spent for each user."

```
SELECT
u.id,
u.country,
u.gender,
```

```

g.device,
g."group",
CASE WHEN a.spent > 0 THEN 1 ELSE 0 END AS converted,
COALESCE(SUM(a.spent), 0) AS total_spent
FROM
users u
LEFT JOIN
groups g ON u.id = g.uid
LEFT JOIN
activity a ON u.id = a.uid
GROUP BY
u.id,
u.country,
u.gender,
g.device,
g.group, converted;

```

Attached is a document detailing all my SQL queries.



GLOBOX A_B TEST -
SQL Queries (1).pdf

INFERENCEAL STATISTICAL ANALYSIS

Test Groups:

Group A (Control Group): Users experienced the existing land page of website.

Group B (Treatment Group): Users were exposed to the land page with food and drink banner.

Hypothesis: We hypothesized that for our analysis,

Null Hypothesis (H0): The new food and drink banner has no effect on the user conversion rate.

Alternative Hypothesis (H1): The new food and drink banner lead to changes in the user conversion rate.

To see the difference in the Hypothesis test for the conversion rate between Group A and Group B based on the assumptions that the data is independent and sample size large enough. We calculated the test statistics.

t= $(p_2 - p_1) / \sqrt{p(1-p)(1/n_1 + 1/n_2)}$ and then used it to calculate the p-value for pooled proportion as our standard deviation(probability of observing a difference of extreme as the one in the sample assuming (H0) is true.).

Pooled proportion was calculated as $(p1^{*}n1 + p2^{*}n2) / n1+n2$ and
p-value $(2*(1-NORM.S DIST(ABS(t))))$ using spread sheet

Null Hypothesis (H0): That there is no difference in the average amount spent between groups.

Alternative Hypothesis (H1): That there is a difference in the average amount spent between groups.

In carrying out this test, we calculate the P-value, which is the probability that measures the likelihood of obtaining results as extreme as those observed, assuming H0 is true. Small p-value < 0.05 suggest that the observed results are unlikely to have occurred by random chance alone. This leads us to reject H0 Large p-value > 0.05 indicates that the observed results could plausibly occur by random chance, this led to failure to reject H0. Test statistics T was calculated using.

$T = (x_B - x_A) / \sqrt{(s_A^2/n1) + (s_B^2/n2)}$

p-value was calculated using T.DIST. 2T(t, df).

This alone doesn't tell you the size of the effect or the practical significance of the results, it only informs you of the likelihood of observing the results

CONFIDENCE INTERVAL

Range 95% confidence interval for the difference in the conversion rate between the treatment and control (0.0035, 0.0107)

Range 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control) (-0.44, 0.48)

2. Key Findings:

Find attached details of some of our statistical calculations and analysis.



GLOBOX A_B TEST -
Calculations (1).pdf

Engagement Metrics:

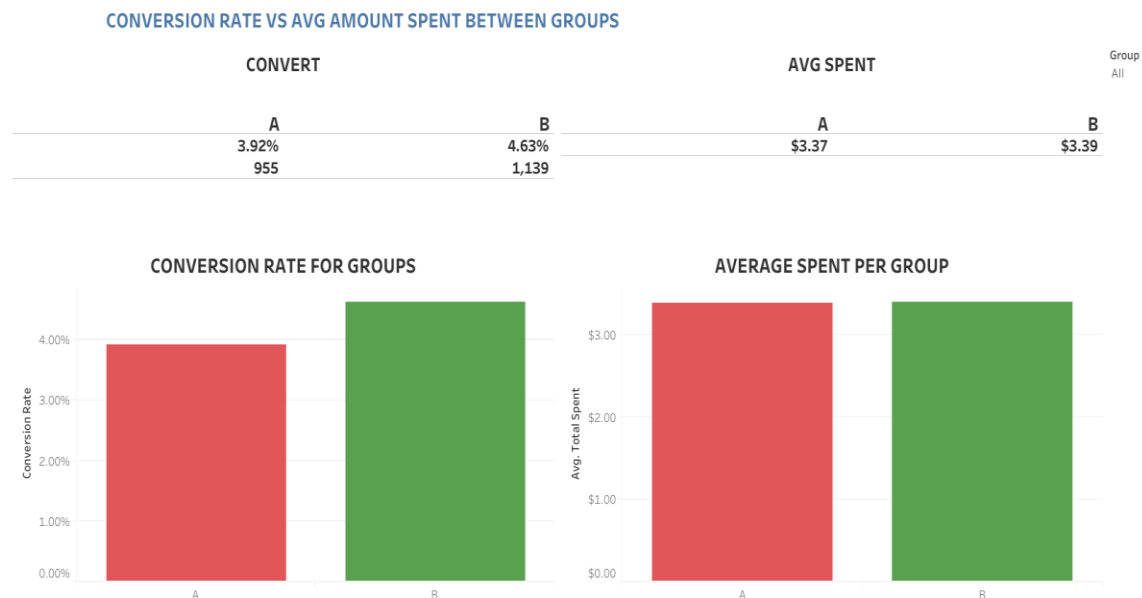
Metric 1: Conversion rate, determined by the click through rate of those who clicked on the food and drink category. Measured as [Control Group A: 3.92] vs. [Treatment Group B: 4.63]

With a p-value = 0.0001 < 0.05 is statistically significant, we reject the null hypothesis that the new food and drink banner has no effect on the user conversion rate. Hence we accept that the new food and drink banner leads to changes in the conversion rate. The confidence interval also gives certainty as the sample proportion (0.0071) is within the range of the confidence interval (0.0035, 0.0107).

Metric 2: Total spend, determined by total spend or purchase irrespective of the category. Averagely measured as [Control Group B: 3.37] vs. [Treatment Group 3.39]

p Value (0.9538) >= alpha (0.05) hence fail to reject H0 (There is no difference in the average amount spent per user between the two groups.) and there is uncertainty about the true parameter to that of the sample observed as the sample proportion (0.0131) was outside the range of the confidence interval

Find in fig1 below a summary of the metrics described above.



Demographic Pattern:

Result showed that we had more Female engagement which showed this is where our marketing should focus. The large Null group could be attributed to ease of use as many people didn't indicate gender both as converted and as the spenders. Most of the conversions and spending are from North America. But interesting also is the spending power of Great Britain which showed relatively greater spend as compared to the conversion rate.

USER PER DEVICE

	OTHERS	ANDROID	IOS
	3.06%	3.15%	6.16%
	\$5.00	\$2.39	\$5.00

USER PER GENDER

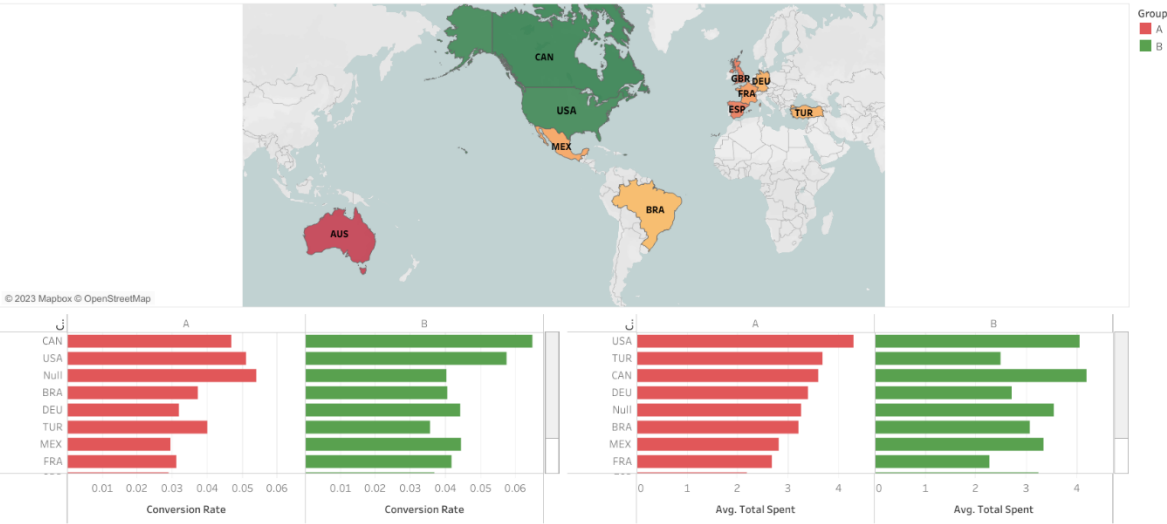
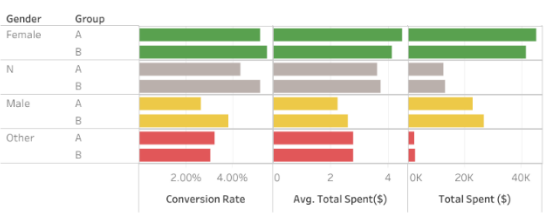
	N	Female	Male	Other
	4.74%	5.29%	3.21%	3.12%
	\$3.68	\$4.30	\$2.43	\$2.77

Device
All
Device
OTHERS
ANDROID
IOS
Gender
All

DEVICE CHART



GENDER CHART



3. Recommendations:

Based on the A/B test results and the insights gathered, I suggest we reiterate, and retest and I put forth the following recommendations:

Feature Optimization: Considering the positive feedback received from users regarding conversion rate especially for the female gender,

- we recommend refining and optimizing this feature based on usage patterns, so it is very female appealing and easy to use.

2. The amount spent should be specific to the food and drink offering and this will enable us to measure the impact of the rollout.
3. Easy to navigate to mitigate large null group in country conversion which was very significant. The percentage of gender that appeared as Null was also quite significant so maybe a tweak on the interface to appeal to gender Clear communication will help address potential user resistance.
4. To give adequate time span for the experiment or increase sample size as we require a sample size of 60600 to detect changes effectively and consistently in our sample.

On the flip side

Upon reviewing the experiment, as I stated in the summary, the lack of clarity in defining the amount spent could potentially lead to ambiguity in interpreting the outcomes and making informed decisions based on the results. This situation could be misleading and hence the desired outcome may not be reached.

Considering these observations, I recommend that we undertake the following steps to improve the experiment's design and the clarity of its measurement indices:

1. Review and Refine metrics and make it specific for the food and drink category.
2. Define Success Criteria for the experiment based on the revised measurement indices. These criteria should reflect not only the specific outcomes expected but also the magnitude of change that would be considered meaningful.
3. Increase sample size by 50%
4. Pilot testing the redesigned experiment on a smaller scale before full implementation. This will help identify any remaining ambiguities or issues and allow for adjustments to be made proactively.

Benefits of Redesign:

Redesigning the experiment to address the clarity of measurement indices will yield several benefits:

- 1.Improved Precision: Clear measurement indices will lead to more accurate and precise data collection, reducing the potential for misinterpretation.
- 2.Informed Decision-Making: Well-defined indices will provide actionable insights that can guide effective decision-making and strategy development.
- 3.Enhanced Experiment Value: A well-designed experiment with clear measurement indices increases its value by delivering more reliable and meaningful results.
- 4.I believe that by taking the time to refine the experiment design and ensure clarity in the measurement indices, we will significantly enhance the experiment's overall effectiveness and impact.

I welcome the opportunity to discuss this recommendation further and collaborate on implementing the suggested steps. Your input and insights are invaluable in ensuring that our experiments yield actionable and reliable results.

APPENDIX

Google spread sheet link.

<https://docs.google.com/spreadsheets/d/1BVnpeZ0D-YquJPVRA16qFSkG8n8bucaRrFuRNe2csFM/edit?usp=sharing>

Tableau link:

https://public.tableau.com/views/GloboxprojECTDSN/Countrymap?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

ATTACHMENTS

Appendix 1: SQL Queries

Appendix 2: Statistical calculations

Appendix 3: Power point slides