

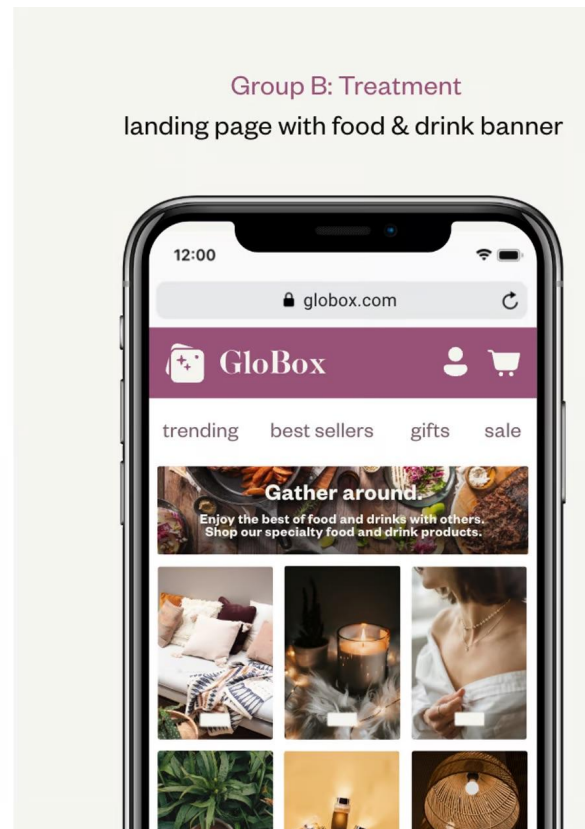
GLOBOX LANDING PAGE AB TEST

Food and drink niche awareness

Group A: Control
existing landing page



Group B: Treatment
landing page with food & drink banner



Introduction

GloBox is an online marketplace that specializes in sourcing unique and high-quality products from around the world. GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, their food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue.

For that purpose, Globox designed a banner that will be displayed in their mobile homepage to some of their users as part of their testing efforts.

Description

To bring awareness to their food and drinks products, Globox started testing a banner that would be printed in the landing pages of users assigned in the treatment group (B).

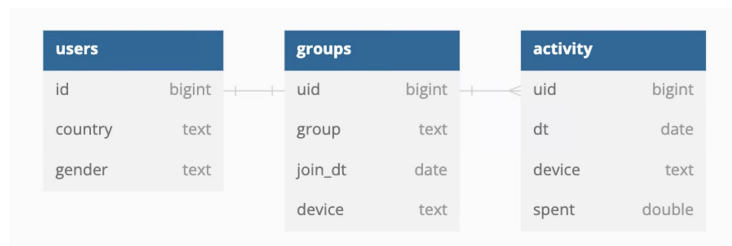
Users visiting Globox mobile website were randomly assigned to either the control or test group. The page loads the banner if the user is assigned to the test group, and does not load the banner if the user is assigned to the control group.

The test was run from 2023-01-25 to 2023-02-06 with a total duration of 13 days.

The key metrics to track will be:

1. Conversion Rate (number of conversions divided by the total number of unique users). A conversion is considered if a user makes one or more purchases.
2. Average Revenue per User or ARPU (sum of the total spent divided by the total users).

For the purpose of the analysis of results a database was created. The data was structured as shown in the following entity relationship diagram (ERD).



Two tables were extracted from the database:

Table 1

Table 1 joined together relevant fields to analyze the activity of users in control and treatment groups.

For the Table 1 some data preparation was needed:

- There were null values in activity.spent that were replaced by zeros.

- There were null values in user.gender that were replaced by “unknown”
- There were users that performed the action of buying more than one time so spent records were aggregated based on the user id into the field total_spent.
- To filter the users that converted a field “converted” was added with 1 representing a conversion and 0 the opposite. The final table had the following header:

id	country	gender	device	group	converted	total_spent
----	---------	--------	--------	-------	-----------	-------------

Table 2

Table 2 aggregated key metrics over time grouped by control (A) and treatment (B) groups. The main goal was to have a picture of the different metrics behaved over time.

This table required a complex query with several steps:

- First, as users could have converted more than one time, it was needed to extract the earliest date of conversion (for conversion rate) with the total spent for both groups (for ARPU).
- In the second step, the date users joined the experiment and the total number of users were extracted. The results were grouped by join date.
- In the third step, the previous two tables were joined together.
- In the fourth step, key metrics were calculated with the previous table accumulating the needed variables by dates. This resulted in two tables with the same fields, each table representing each group.
- Finally, a union was used to join both groups' tables.

The final table had the following header:

group_name	date	conversion_rate	average_spent
------------	------	-----------------	---------------

The link to access the database, the description of the fields and the SQL queries used can be found in the [appendix](#).

Exploratory Data Analysis:

The experiment aggregated data of 48,943 users.

24,343 or 49.74% were in the Control group (A) and 24,600 or 50.26% were in the treatment group (B).

There were 2,094 total conversions and the average revenue per user was \$3.38.

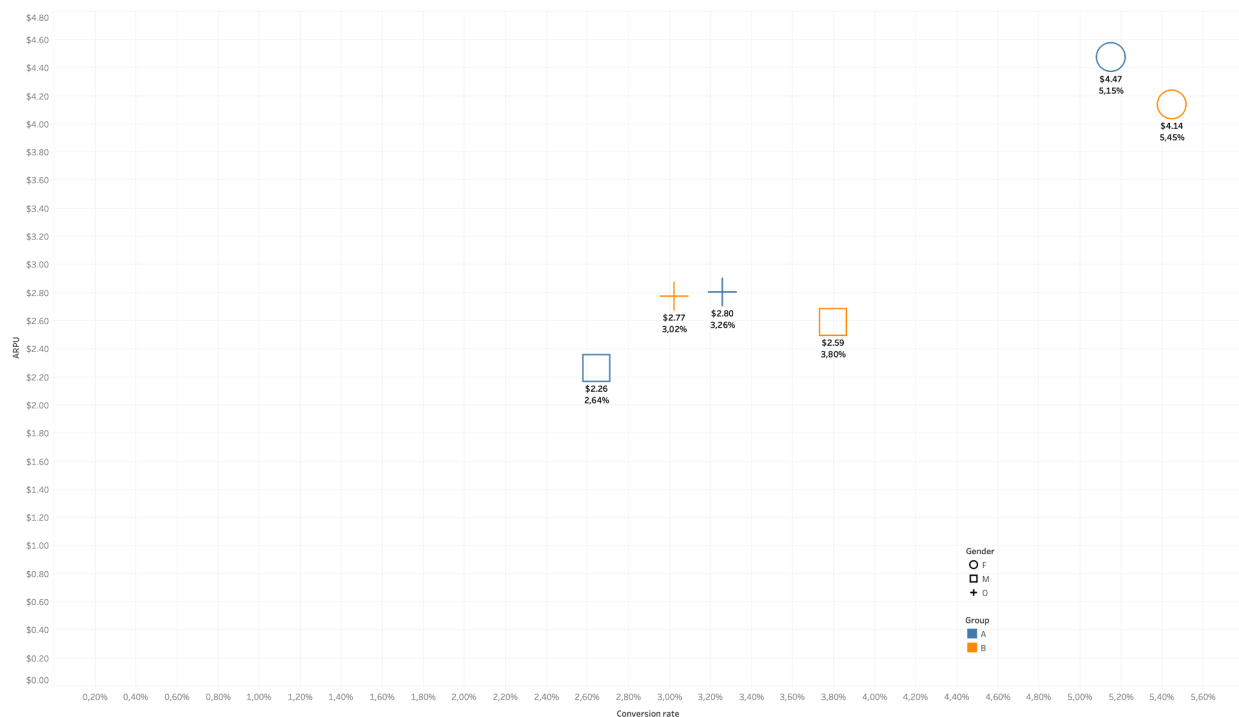
User demographics:

The dataset provided three demographic fields to better understand the users in both groups.

Gender:

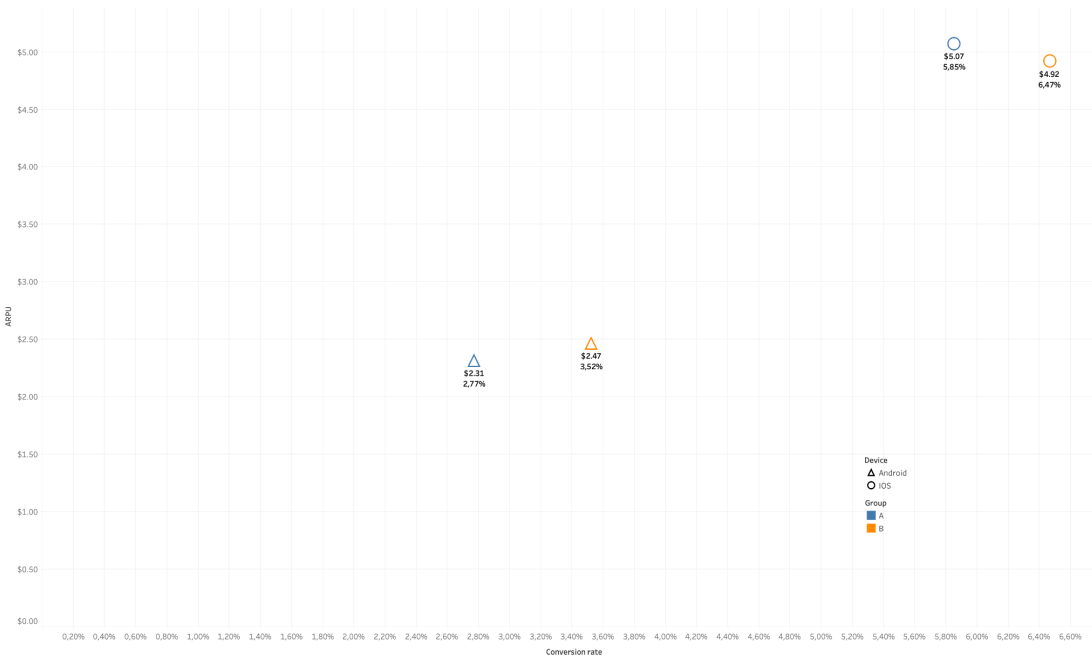
Around 41% of the total experiments were male and the other 41% were female. Around 3% identified with Other and there was a 14% of unknown gender.

Female gender had better results in both key metrics. Also, the biggest difference in conversion rate between test groups was accounted in Male users. As it can be observed in the following plot.



Device:

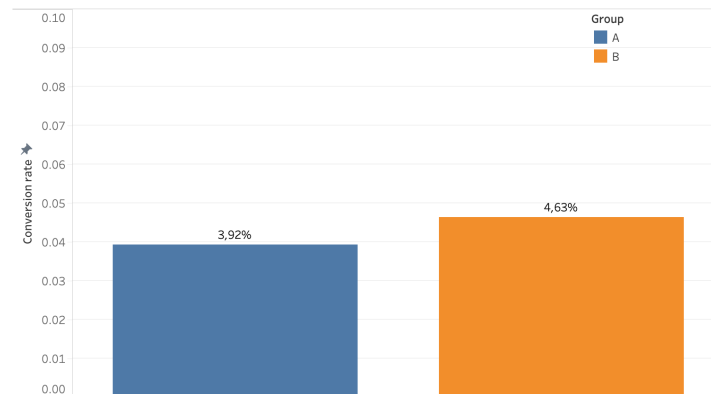
Around 62% of the devices of users in the experiment were Android users. The other 38% were IOS users. However, the average revenue per customer and the conversion rate is higher in IOS than in Android in both control (A) and treatment group (B) as shown in the following plot.



Results:

Changes in key metrics:

In terms of conversion rate, there was a difference of 0.71% in group B against group A.



A hypothesis test was conducted to check if the difference was statistically significant with a 5% significance level.

Null hypothesis (H0) and Alternative hypothesis (Ha) were defined as follows:

H0: There is no difference in the conversion rate between the two groups

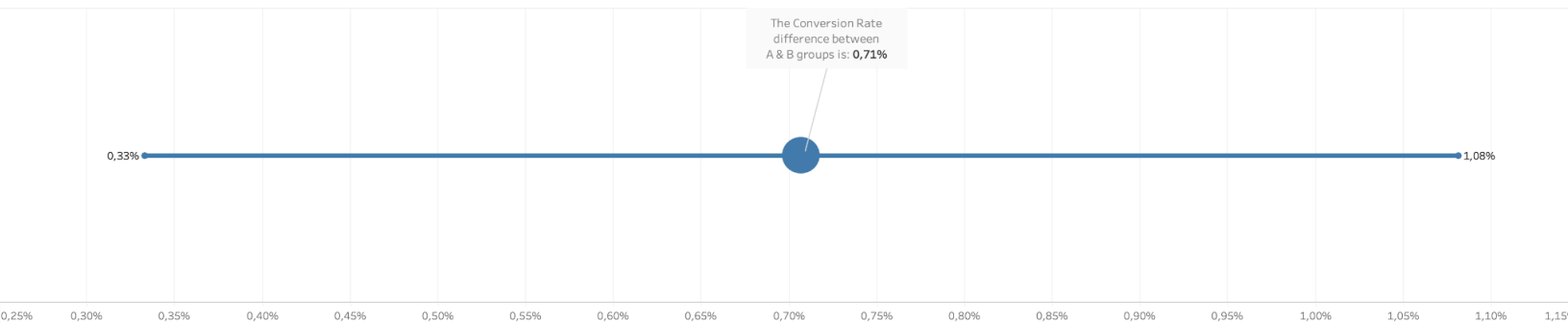
H0: A conversion - B conversion = 0

Ha: There is a difference in the conversion rate between the two groups

Ha: A conversion - B conversion \neq 0

With a p-value of $0.044 < 0.05$ the null hypothesis "There is no difference in the conversion rate between the two groups" was **rejected** in favor of the alternative hypothesis "There is a difference in the conversion rate between the two groups".

The 95% confidence interval for the difference in conversion rate between both groups is $\pm 0.37\%$ with an upper bound of 1.08% and a lower bound of 0.33%



This strongly suggests that the changes in the treatment group, in terms of conversion rate will most likely be positive.

In terms of the Average Revenue per User (ARPU), there was a difference of \$0.02 in group B against group A.



A hypothesis test was conducted to check if the difference was statistically significant with a 5% significance level.

Null hypothesis (H_0) and Alternative hypothesis (H_a) were defined as follows:

H_0 : There is no difference in the Average Revenue per User (ARPU) between the two groups

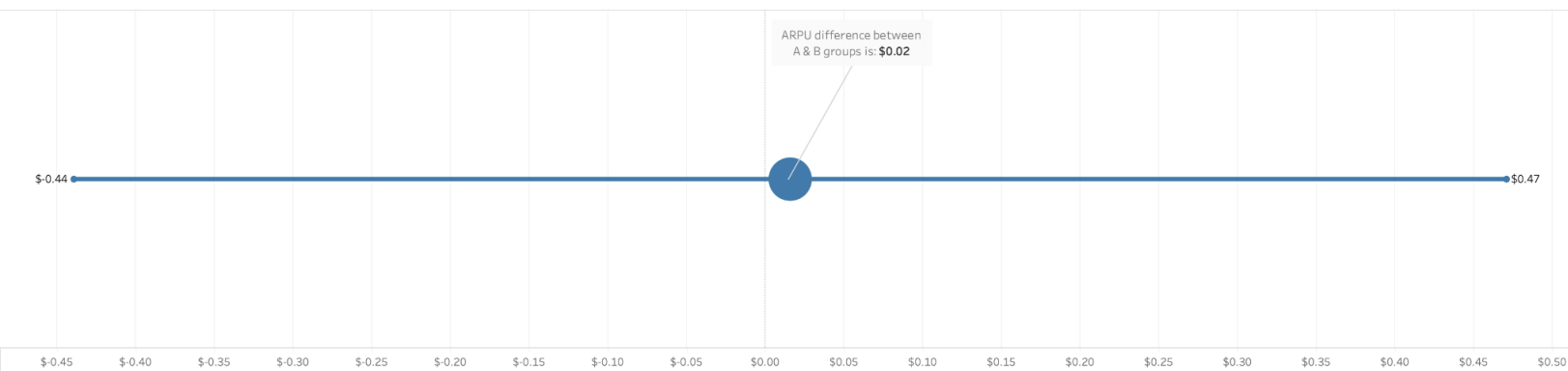
H_0 : $B \text{ ARPU} - A \text{ ARPU} = 0$

H_a : There is a difference in Average Revenue per User (ARPU) between the two groups

H_a : $B \text{ ARPU} - A \text{ ARPU} \neq 0$

With a p-value of $0.054 > 0.05$ the null hypothesis "There is no difference in the conversion rate between the two groups" was **not rejected**. It can be concluded that there is not a statistically significant change in ARPU in group B against group A.

The 95% confidence interval for the difference in Average Revenue per User (ARPU) between both groups is $\pm \$0.45$ with an upper bound of \$0.47 and a lower bound of $-\$0.44$.

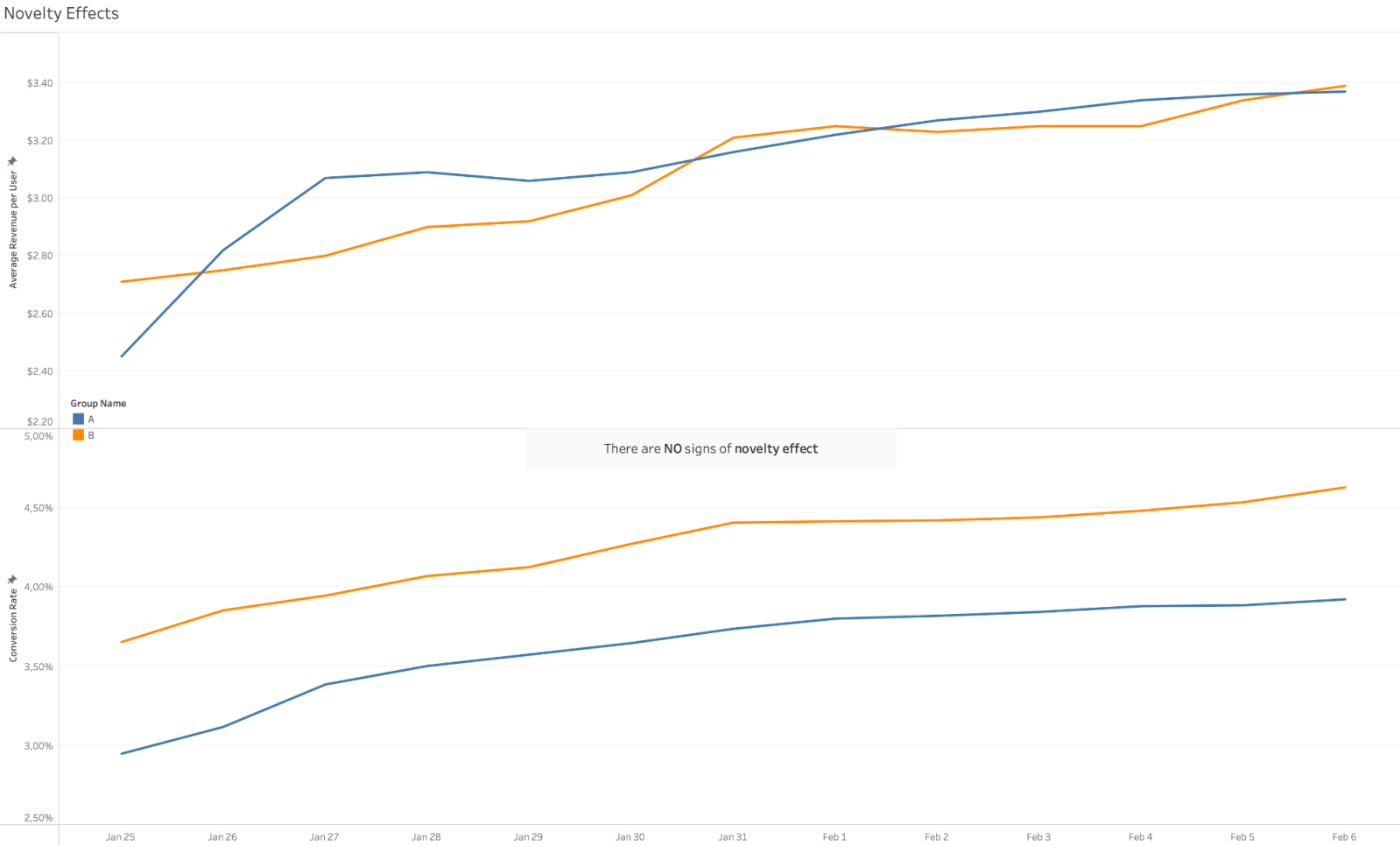


This shows the uncertainty around the Average Revenue per User (ARPU) outcome. It is as likely to have a positive or a negative outcome.

Details on p-value and confidence intervals calculations can be found in spreadsheets in the Appendix

Novelty effects:

Novelty effects could be occurring as users may find the new banner intriguing. Therefore, observing how both key metrics were performing during the test was needed to discard these novelty effects that could be influencing our results.



Plotting Table 2 in Tableau made it possible to visualize the performance of key metrics in both groups. For checking novelty effects, the orange line representing treatment group (B), had to be observed.

In both metrics the treatment group behaves in a way that is not showing peaks in the first days of the experiment having a downturn days after so it can be concluded that there are no signs of novelty effects.

Power analysis:

To ensure the accuracy of our hypothesis test results a power analysis was conducted.

Knowing that:

- The sample size was 48,943,
- The baseline conversion rate, or conversion rate of control group (A) was 3.92%
- The difference between conversion rates between control and treatment group was 0.71%
- Aiming a sensitivity or statistical power of 0.2 (standard)

We would need a much larger sample size to be sure the hypothesis tests conducted were accurate. This is a sample size of around 6.1 million users.

This results signals that the hypothesis test conducted earlier could not be accurate.

The screenshot displays a 'Sample Size Calculator' interface. On the left, under 'Advanced Settings', the 'Baseline Conversion Rate (%)' is set to 3.92, and the 'Minimum Detectable Effect (%)' is set to 1. The 'Hypothesis' section has 'One-sided Test (Recommended)' selected. At the bottom, 'A/B Split Ratio' is 0.5026, 'Significance (α)' is 0.05, and 'Statistical Power (1 - β)' is 0.8. On the right, the 'Results' section shows 'TEST SIZE' and 'CONTROL SIZE' both as 3M, and a large blue box indicating a 'TOTAL SAMPLE SIZE' of 6.1M. A 'Share Link' button is also present.

Link to the sample size calculator can be found in the appendix.

Recommendations:

After the power analysis it seems clear that with the experiment conducted there is not sufficient evidence that the changes in metrics would be repeatable, this is that there is no proof that the banner would change conversion rate or ARPU.

With these uncertain results growth and marketing teams have to make a judgment based on other factors and take the risk of:

- Launching without enough proof that there will be a change in key metrics.
- Repeating the experiment with a much larger sample size. This sample size should be of at least 3 million unique sessions for each group based on the previous statistical power analysis.

In case of deciding to repeat the experiment. I would suggest increasing the granularity of data, including categories to conversion data to deepen the understanding of the activity of users. The goal of the AB Test was to know more about the awareness of the food & drink niche, so knowing conversions of this niche turn out to be essential to get the responses we are aiming for.

Appendix:

Database access:

```
postgres://Test:bQNxVzJL4g6u@ep-noisy-flower-846766-pooler.us-east-2.aws.neon
.tech/Globox
```

Fields description:

- users: user demographic information
 - id: the user ID
 - country: ISO 3166 alpha-3 country code
 - gender: the user's gender (M = male, F = female, O = other)
- groups: user A/B test group assignment
 - uid: the user ID
 - group: the user's test group
 - join_dt: the date the user joined the test (visited the page)
 - device: the device the user visited the page on (I = iOS, A = android)
- activity: user purchase activity, containing 1 row per day that a user made a purchase
 - uid: the user ID
 - dt: date of purchase activity
 - device: the device type the user purchased on (I = iOS, A = android)
 - spent: the purchase amount in USD

[SQL Queries](#)

[Statistical Calculations](#)

[Sample Size Calculator](#)