Globox A/B Test Experiment

By **Tega odjugo**

Performed on 2023-08-20

Summary

An experiment was conducted on the Globox mobile website to determine if a banner that highlights key products in the food and drink category as a banner at the top of the website should be launched on the website based on an increase in purchases. Two groups were created with one being showed the website change and the other was not. The result of the experiment showed sufficiently strong evidence that there would be a difference in the conversion rate of users between the two groups. However, there is no sufficiently strong evidence that there would be a difference in the average amount spent per user between the two groups.

I recommend we do not launch the banner because even though there is an increase in conversion rate of users, there is no increase in the average amount spent per user. The banner takes up high-value real estate on the main page, so we should only be comfortable launching it if it leads to more than a 10% increase in revenue per user.

Context

Globox is an online marketplace specializing in sourcing unique and high-quality products from around the world. They are running an A/B test experiment to improve their homepage.
The setup of the A/B test is as follows:

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

2. The experiment begins on Jan 25th, 2023 and ends Feb 6th, 2023.

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

4. 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.

5. 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".

Globox stores their data in a relational database, which is accessed using beekeeper studio and the link.

The Globox data has 3 tables which are; users, groups and activity. The tables have the following columns;

- **users**
  - **id**
  - **country**
  - **gender**
- **groups**
  - **uid**
  - **group**
  - **join_dt**
  - **device**
- **activity**
  - **uid**
  - **dt**
  - **device**
  - **spent**

Results

The data extracted and analysed using SQL gives the following answers;

Total of experiment	Control group (A)	Treatment group(B)
48,943 users	24,343 users	24,600 users
4.28% conversion rate	3.92% conversion rate	4.63% conversion rate
$3.382 average amount per user	$3.375 average amount per user	$3.391 average amount per user
20130 females, 20289 males, 6855 without a specified gender, 1669 other genders	10069 females, 10054 males, 3412 without a specified gender, 808 other genders	10061 females, 10235 males, 3443 without a specified gender, 861 other genders
30289 use android, 18360 use apple, 294 did not have a device specified	15054 use android, 9142 use apple, 147 did not have a device specified	15235 use android, 9218 use apple, 147 did not have a device specified

To determine if there is a difference between the conversion rate of the two groups, a hypothesis test is conducted. A two-sample normal distribution test is conducted on the test data. The null hypothesis is that there is no difference between the conversion rate of the two groups and the alternative hypothesis is that there is a difference between the conversion rate of the two groups. The significance level is assumed as 5%. After the calculation conducted the p value is determined to be 0.0001. This is statistically significant since p<0.05 therefore we reject the null hypothesis that there is no difference in the user conversion rate between the control and treatment groups. The 95% confidence interval for the difference in the conversion rate between the treatment and control group using normal distribution and unpooled proportions for the standard error is 0.0035 to 0.0107.

To determine if there is a difference between the average amount spent per user of the two groups, a hypothesis test is conducted. A two-sample student's t distribution test is conducted on the test data. The null hypothesis is that there is no difference between the average amount spent per user of the two groups and the alternative hypothesis is that there is a difference between the average amount spent per user of the two groups. The significance level is assumed as 5%. After the calculation conducted the p value is determined to be 0.944. This is statistically insignificant since p>0.05 therefore we fail to reject the null hypothesis that there is no difference in the average amount spent per user between the control and treatment groups. The 95% confidence interval for the difference in the average amount spent per user between the treatment and control group using t distribution and assume unequal variance is -0.439 to 0.471.

Tableau is used to visualize data set extracted from SQL and the following visuals are the results;
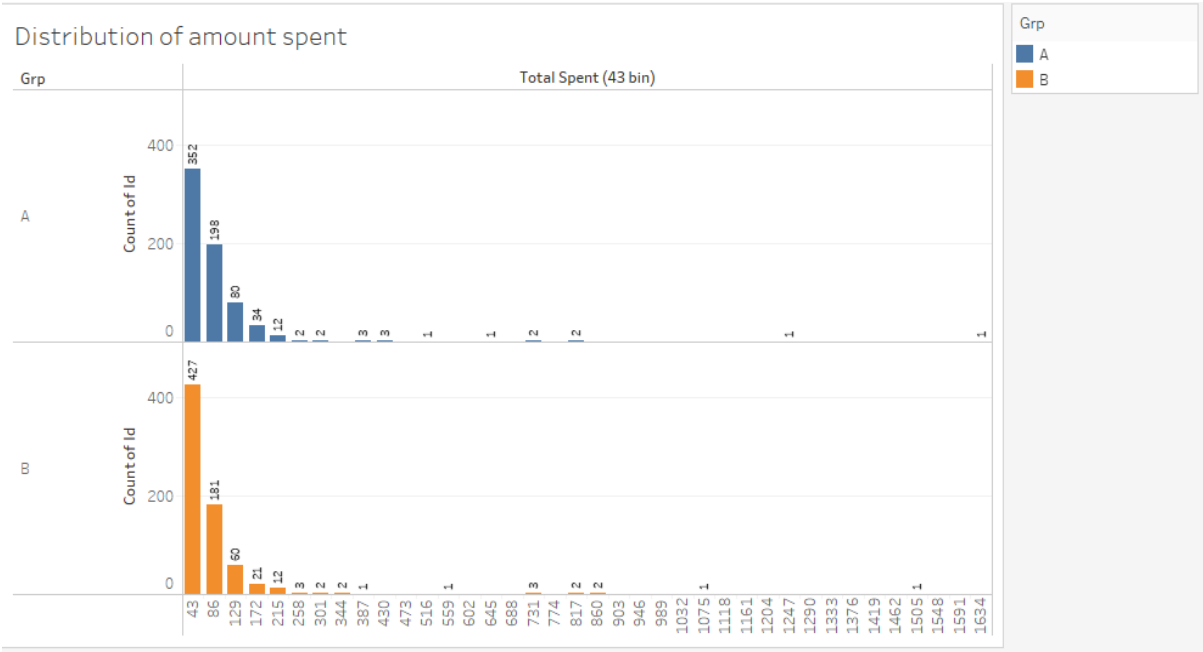

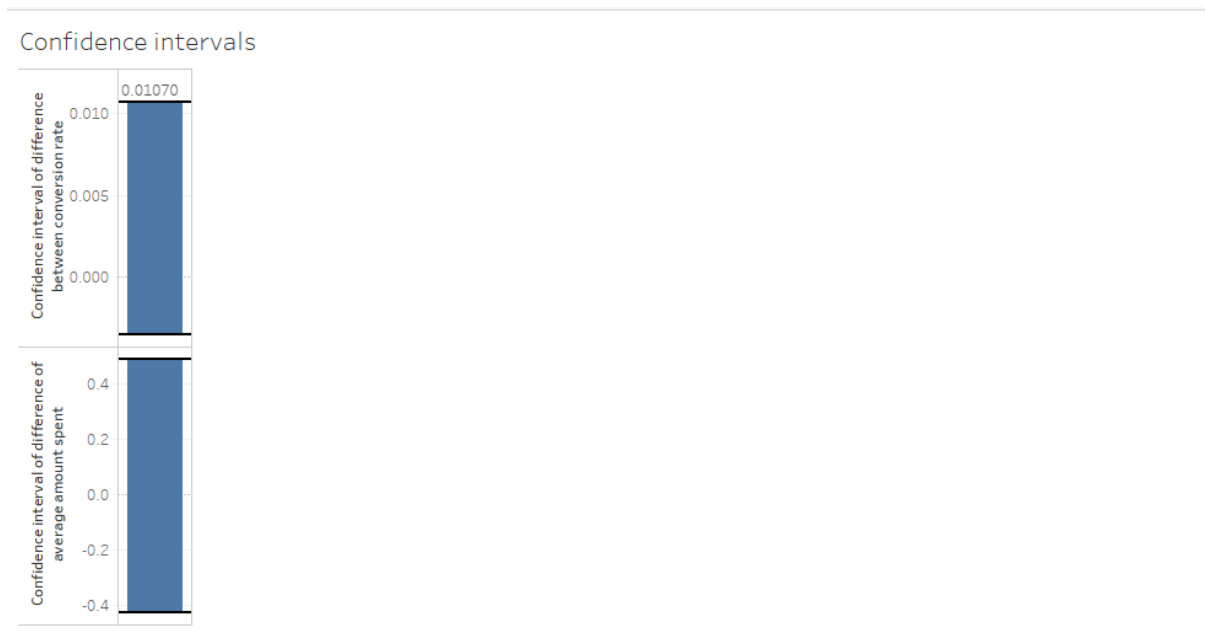Fig 1 Distribution of amount spent per user of each group

Fig 2 Confidence intervals for the difference in conversion rate and the difference in the average amount spent between the two groups
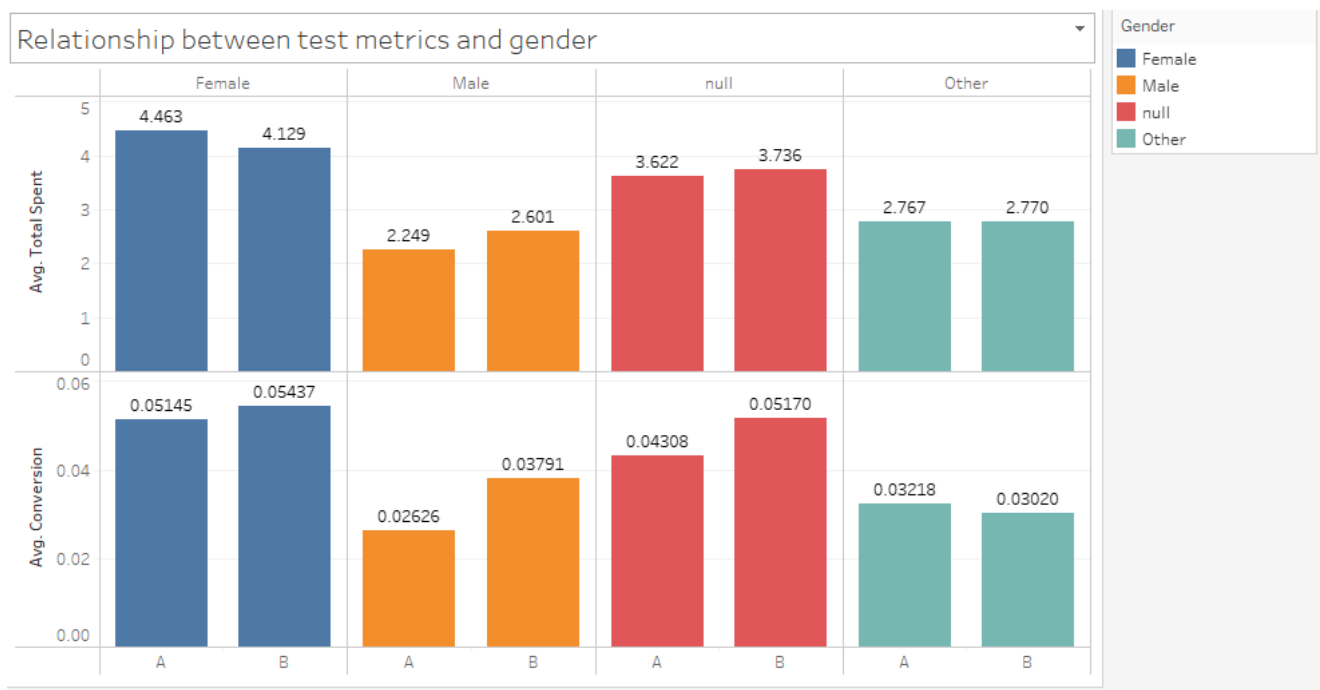

Fig 3 The relationship between the test metrics (conversion rate and average amount spent) and the user's gender.

Recommendation

I recommend we do not launch the banner because even though there is an increase in conversion rate of users, there is no increase in the average amount spent per user. The banner takes up high-value real estate on the main page, so we should only be comfortable launching it if it leads to more than a 10% increase in revenue per user. We can iterate on this experience and test with a larger sample size to determine if there would be a significant increase in a different season of the year.

Appendix

```
WITH d
    AS (SELECT id,
            country,
            gender,
            P3.device,
            P3.group  AS grp,
            Sum(COALESCE(spent, 0)) AS total_spent,
            CASE
              WHEN Sum(spent) > 0 THEN 1
              ELSE 0
            END AS Conversion
        FROM   users AS P1
            LEFT JOIN activity AS P2
                ON P1.id = P2.uid
            LEFT JOIN groups AS P3
                ON P1.id = P3.uid
        GROUP  BY p3.group,
            country,
            gender,
            p3.device,
            id)
SELECT *
FROM   d;
```

```sql
WITH sumtotal
    AS (SELECT id,
            country,
            gender,
            P3.device,
            P3.group,
            Sum(COALESCE(spent, 0)) AS total_spent
        FROM   users AS P1
            LEFT JOIN activity AS P2
                ON P1.id = P2.uid
            LEFT JOIN groups AS P3
                ON P1.id = P3.uid
        WHERE  P3.group LIKE 'A'
        GROUP  BY id,
                country,
                gender,
                p3.device,
                p3.group)
SELECT Avg(total_spent) AS Group_A_averageamountspent
FROM   sumtotal;


WITH sumtotal
    AS (SELECT id,
            country,
            gender,
            P3.device,
            P3.group,
            Sum(COALESCE(spent, 0)) AS total_spent
        FROM   users AS P1
            LEFT JOIN activity AS P2
                ON P1.id = P2.uid
            LEFT JOIN groups AS P3
                ON P1.id = P3.uid
        WHERE  P3.group LIKE 'B'
        GROUP  BY id,
                country,
                gender,
                p3.device,
                p3.group)
SELECT Avg(total_spent) AS Group_B_averageamountspent
FROM   sumtotal;
```

```sql
WITH d
   AS (SELECT id,
          country,
          gender,
          P3.device,
          P3.group,
          Sum(COALESCE(spent, 0)) AS total_spent,
          CASE
            WHEN Sum(spent) > 0 THEN 1
            ELSE 0
          END              AS Conversion
      FROM   users AS P1
         LEFT JOIN activity AS P2
             ON P1.id = P2.uid
         LEFT JOIN groups AS P3
             ON P1.id = P3.uid
      GROUP  BY id,
           country,
           gender,
           p3.device,
           p3.group)
SELECT Avg(conversion) AS Conversion_rate_for_all_users
FROM   d;
```

Hypothesis and CI Spreadsheet