Vanessa Kimana
2023-09-11

Globox

A/B Test and Data Analysis Report

Group A: Control

existing landing page

Group B: Treatment

landing page with food & drink banner

Vanessa Kimana
2023-09-11

1. Launching the experiment: Summary

This report provides a summary of an A/B test conducted to analyze the effects of adding a banner promoting the food and drink category on GloBox's mobile website. Moreover, the A/B testing was used to measure the performance of two different versions of the website. The report aims therefore to present findings and recommendations for all stakeholders involved in the project. The stakeholders involved in this decision include the Growth Product & Engineering Team, Leila Al-Farsi (Product Manager, Growth), Alejandro Gonzalez (User Experience Designer, Growth), and Mei Kim (Head of Marketing).
Following the comprehensive statistical analysis, the A/B test showed a statistically significant increase in conversion rates for users who saw the banner. whereas there was no significant difference in the average amount spent between the control(A) and treatment groups(B). **Based on these findings, it is recommended to launch the exiperiment with careful consideration of other implicated variables.**

2. Project Background

GloBox, an online boutique, especially known for its boutique fashion items and high-end decor products among its customers. Following the significant growth of food and drink offerings in recent months, the growth team decided to run an A/B test to raise awareness of its food and drink category. The mobile website was chosen as the platform for this experiment, with the primary goal of increasing revenue.

3. Context and A/B Test Setup

A/B tests are essential tools for businesses to measure the effectiveness of two different versions of the website. In the case of Globox, the A/B test is structured as follows:

- The experiment exclusively applies to the mobile website version.
- When a user accesses the GloBox main page, they are randomly allocated to either the control or test group. This assignment is recorded as the user's join date.
- If the user is assigned to the test group, the webpage displays the banner; however, if the user belongs to the control group, the banner does not appear.
- The user may or may not make purchases on the website. These purchases, whether occurring on the same day they join the experiment or in the days following, are considered “conversion.”

4. Experiment Context

The experiment aimed to analyze the performance of two groups, Group A and Group B, to determine if there are significant differences in conversion rates and average amounts spent per user between these two groups. The data collected includes information on User_id, user’s country, user’s gender, test group assignment (A or B), user’s device (A or I) and the amount spend per user.

Vanessa Kimana
2023-09-11

This dataset was subjected to various calculations to facilitate statistical analysis aimed at assessing whether there exist notable disparities in conversion rates and average amount per user between the two groups.

The following calculations served as the basis for conducting statistical analyses.

Row Labels	Sum of Total_spent	Average of Total_spent2	Sum of Converted	StdDev	Count of Test_group
A	82145.90306	3.374518468	955	25.93639	24343
B	83402.86237	3.390498084	1139	25.41433	24599
Grand Total	**165548.7654**	**3.382550068**	**2094**	**25.67506**	**48942**

Parameters of the Test:

Hypothesis Test for Conversion Rate:
In this test, we are comparing the conversion rates between two groups, Group A and Group B. The null hypothesis (H0) states that there is no difference in the conversion rates between these two groups, meaning that they are equal. The alternative hypothesis (H1) suggests that there is a difference in conversion rates between the groups, indicating that they are not equal. The level of significance (α) is set at 0.05, which means we will consider results statistically significant if the p-value of observing such results by random chance is less than 5%.

Null Hypothesis (H0): P^ = P^2
Alternative Hypothesis (H1): P^1 ≠ P^2
Level of Significance (α): 0.05

In this test, we are comparing the average amount of money spent per user between two groups, Group A and Group B. The null hypothesis (H0) states that there is no difference in the average amount spent between these two groups; their averages are equal. The alternative hypothesis (H1) suggests that there is a difference in the average amount spent between the groups, indicating that they are not equal. The level of significance (α) is again set at 0.05, which means we will consider results statistically significant if the p-value of observing such results by random chance is less than 5%.

Hypothesis Test for Average Amount Spent:

Null Hypothesis (H0): μ1 (Average Amount Spent for Group A) = μ2 (Average Amount Spent for Group B)
Alternative Hypothesis (H1): μ1 ≠ μ2 (There is a difference in Average Amount Spent)
Level of Significance (α): 0.05

Vanessa Kimana
2023-09-11

5. Result

Result Summary

Group B appears to have a higher conversion rate and slightly higher average spending compared to Group A. Difference in conversion rates is more noticeable than the difference in average amounts spent, which suggests that improving the conversion rate in Group A may generate more revenue than increasing the average spending in Group B.

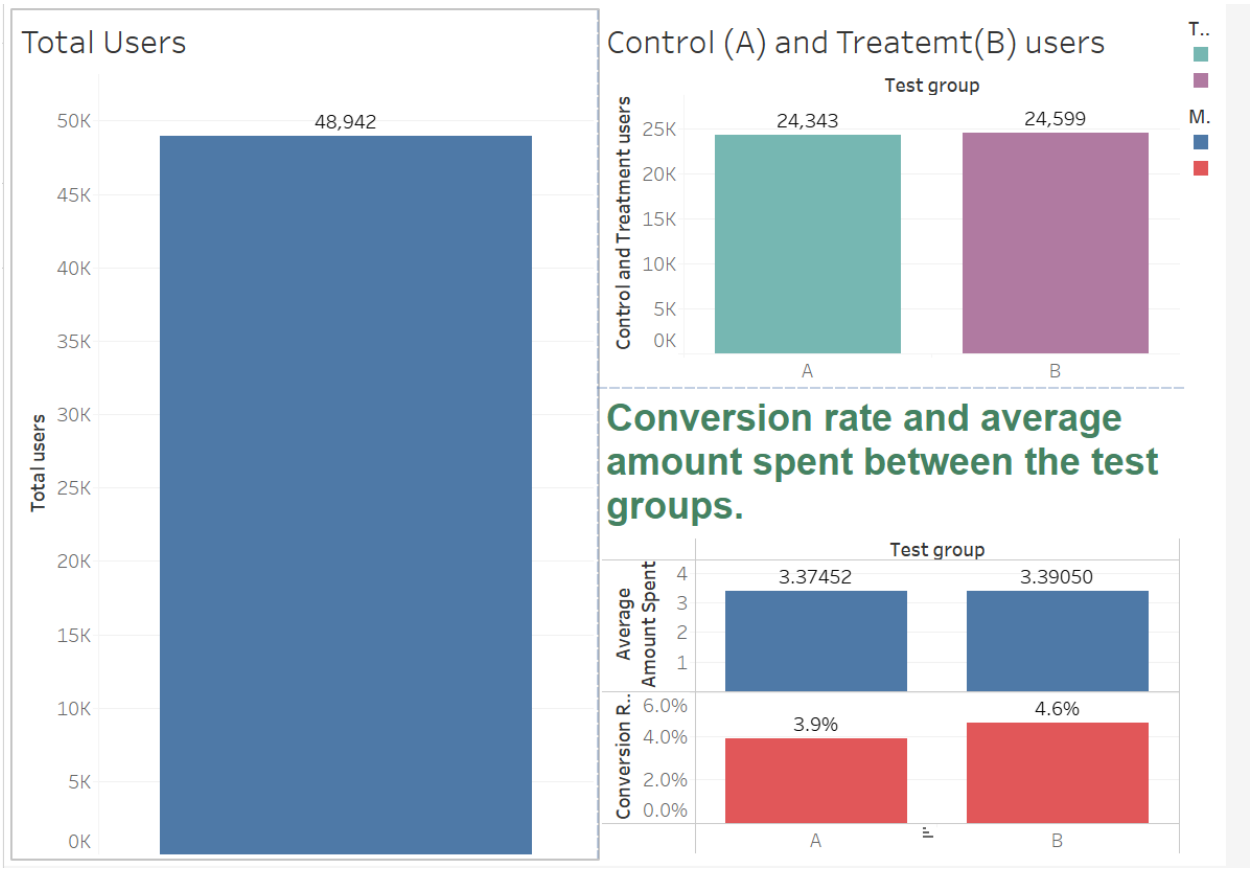

Fig 1: This shows the sample size for group A and B and the total size for all users and conversion and Average sepent between A and B.

Vanessa Kimana
2023-09-11

Average Amount Spent Results: The T-value for the difference in the average amount spent between Group A and Group B is -0.0688.
When compared to the critical value (z*) of 1.96 for a 95% confidence interval, the P-value is relatively high at 0.9451 (greater than the standard significance level of 0.05).
This means that there isn't enough evidence to conclude that the average amount spent per user is different between the two groups.
The 95% Confidence Interval suggests that the actual difference in means likely falls within a range of -$0.47 to $0.44.
We cannot therefore confidently say there is a significant difference in the average amount spent per user between Group A and Group B.

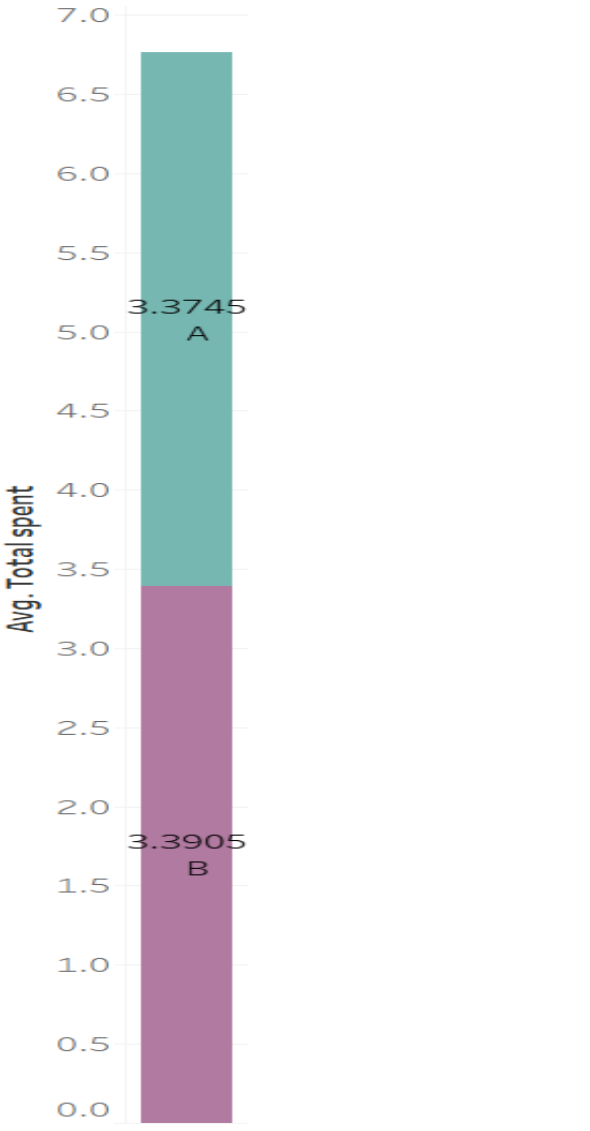

Figure 2: Difference in spending between the two groups was quite small.

Vanessa Kimana
2023-09-11

Conversion Rate Results: The difference in conversion rates between Group A and Group B is quite significant, with a Z Score of 3.8652.
When compared to a critical value of 1.96 for a 95% confidence level, the P-value is very low at 0.0001 (much less than the standard significance level of 0.05).
This means there is strong evidence to conclude that the conversion rates are different between the two groups.
Furthermore, a 95% Confidence Interval suggests that the true difference in proportions between the groups lies somewhere between 0.0035 and 0.0107.
We can therefore confidently say that there is indeed a significant difference in conversion rates between Group A and Group B.

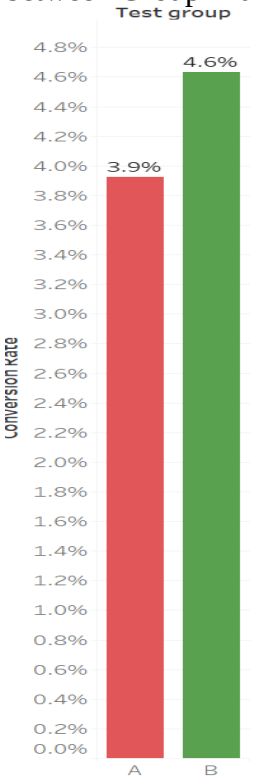

Figure 3: Group B is performing better in terms of conversions, with a true difference in conversion rates.

Vanessa Kimana
2023-09-11

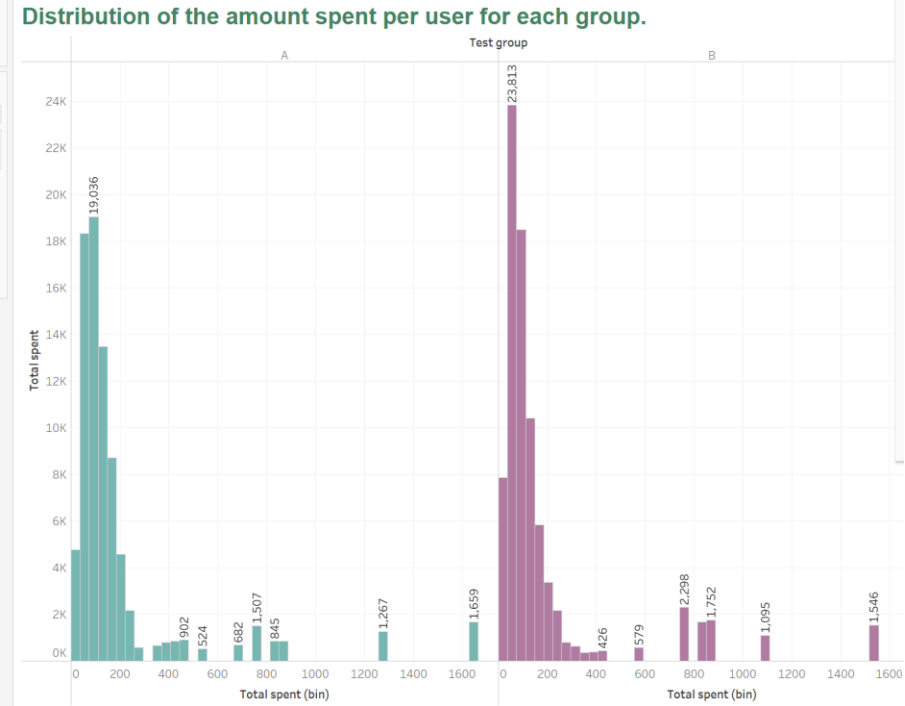

Fig 4: The graph above the distribution of the amount spent per user of each group. It becomes apparent that Group B has a higher number of users who did not make any purchases when compared to Group A.

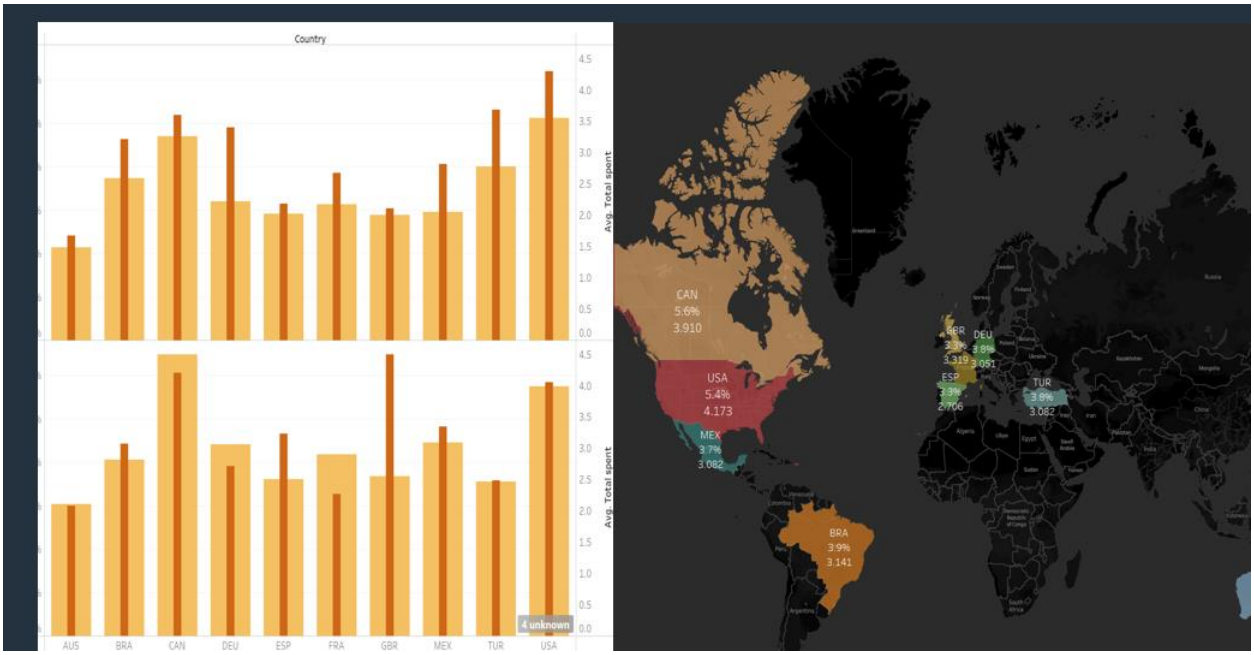

Fig 5: Both Group A and Group B show exceptionally higher conversion rates in both Canada and the US.

Vanessa Kimana
2023-09-11

6. Recommendation

The analysis above depicts a notable increase in conversion rates, with the treatment (B) group presenting a higher conversion rate compared to the control group (A). The introduction of the banner effectively captured user attention and encouraged conversions. However, it's worth mentioning that while conversion rates improved, the average amount spent by users in both groups remained relatively similar. This suggests that the banner is successful in prompting users to make purchases but does not significantly influence their spending habits. Based on the statistical analysis**, it is recommended to consider launching the experiment, we feel confident in releasing the food and drink category banner to all users.** However, it's essential to weigh various factors and engage with all project stakeholders to optimize and maximize the benefits of this experiment's launch.

7. Appendix

```
--Can a user show up more than once in the activity table? Yes or no, and why?
-- SELECT uid, dt, spent FROM activity;
-- When selecting all data from activity, it shows 2233 rows

SELECT uid, COUNT(*)more_than_once FROM activity
GROUP BY uid
HAVING COUNT(*) > 1;
```

```
--What SQL function can we use to fill in NULL values?
SELECT u.id, COALESCE(a.spent, 0) FROM users u
LEFT JOIN activity a ON
 u.id = a.uid;
--The COALESCE function is employed to replace NULL values with designated values in fields involved in an operation;
-- an integer can interact with NULL values EX:COALESCE(a.spent, 0)
```

```
--What are the start and end dates of the experiment?
SELECT MIN(join_dt), MAX(join_dt)
FROM groups
```

```
----How many total users were in the experiment?
SELECT COUNT(*) FROM users;
--From the users table, it shows that the total users in the experiment is 48943
```

```
--How many users were in the control(A) and treatment(B)groups?
```

Vanessa Kimana
2023-09-11

```
SELECT g.group, COUNT(g.uid) AS tot_users_per_group FROM groups g
GROUP BY  g.group;
-----------------------------------------------------------------------------------------------------------------

SELECT
   SUM(CASE WHEN a.uid IS NOT NULL THEN 1 ELSE 0 END) AS successful_conversion,
   COUNT(u.id) AS total_users,
   ROUND(SUM(CASE WHEN a.uid IS NOT NULL THEN 1 ELSE 0 END) * 1.0 /
COUNT(u.id), 4) AS conversion_rate
FROM (SELECT DISTINCT id FROM users )u
LEFT JOIN(SELECT DISTINCT uid FROM activity )a ON a.uid = u.id;

-------------------------------------------------------------------------------
--What is the user conversion rate for the control and treatment groups?

SELECT g.group,
   SUM(CASE WHEN a.uid IS NOT NULL THEN 1 ELSE 0 END) AS successful_conversion,
   COUNT(u.id) AS total_users,
   ROUND(SUM(CASE WHEN a.uid IS NOT NULL THEN 1 ELSE 0 END) * 1.0 /
COUNT(u.id), 4) AS conversion_rate
FROM (SELECT DISTINCT id FROM users )u
LEFT JOIN(SELECT DISTINCT uid FROM activity ) a ON a.uid = u.id
LEFT JOIN groups g ON g.uid = u.id
GROUP BY g.group;

SELECT
   g.group,
   AVG(COALESCE(a.spent, 0)) AS average_amount_spent
FROM groups g
LEFT JOIN activity a ON g.uid = a.uid
LEFT JOIN users u ON a.uid = u.id
GROUP BY g.group;
-- SELECT
--   distinct u.id AS user_id,
--   u.country,
--   u.gender,
--   g.group AS test_group,
--   g.device,
--  CASE WHEN coalesce(a.spent, 0) > 0 THEN 1 ELSE 0 END as converted ,
--  SUM (coalesce(a.spent, 0)) AS total
--  FROM users u
--  LEFT JOIN groups g ON u.id = g.uid
--  LEFT JOIN activity a ON u.id = a.uid
-- --WHERE a.spent > 0
--  GROUP BY u.id, u.country, u.gender, g.group,g.device,a.spent
```

Vanessa Kimana
2023-09-11


```
SELECT
  u.id AS user_id,
  u.country,
  u.gender,
  g.group AS test_group,
  g.device,
  CASE WHEN cnt_purchases > 0 THEN 1 ELSE 0 END AS converted,
  SUM(COALESCE(a.spent, 0)) AS total_spent
FROM users u
LEFT JOIN groups g ON u.id = g.uid
LEFT JOIN (
  SELECT uid, COUNT(*) AS cnt_purchases
  FROM activity
  WHERE spent > 0
  GROUP BY uid
) a_count ON u.id = a_count.uid
LEFT JOIN activity a ON u.id = a.uid
GROUP BY u.id, u.country, u.gender, g.group, g.device, cnt_purchases;
```

---

https://public.tableau.com/app/profile/vanessa.kimana/viz/GloboxABtestingAnalysis/Story1

Excel File Attached in ZipFile.