GloBox

A/B Test Experiment:

To increase revenue, by improving GloBox's online marketplace homepage

Report by: Mahlet S. Sisay
Data Analyst
Growth Product & Engineering Team
Dec 2023

Table of Contents

Executive summary

GloBox is an e-commerce company. Among its featured products, GloBox's 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. An A/B test experiment was conducted to understand the impact level.

An A/B test experiment focused on two key hypotheses—user conversion rate and average spending per user—was conducted. The experiment found a significantly higher conversion rate in the treatment group compared to the control group, indicating that the treatment group is more effective at encouraging users to take desired actions on the platform. However, the analysis revealed no significant difference in the average amount spent per user between the two groups. This suggests that while the treatment group achieves higher conversion rates, it may not necessarily generate significantly more revenue per user than the control group.

Several important observations were made concerning user demographics. Despite a lower number of iOS users participating, they exhibited substantially higher conversion rates and average spending per user. The user counts between male and female users were similar. However, females demonstrated both higher conversion rates and greater average spending per user. The United States and Canada displayed the highest conversion rates and average spending per user, while Australia had the lowest conversion rate and average spending.

In conclusion, the A/B test results suggest that the treatment group is more effective at driving user conversion, particularly among female and iOS users. However, the lack of a significant difference in average spending per user raises questions about the revenue potential of this group.

Several limitations were identified during the analysis, including a non-representative sample size and experiment duration. By acknowledging these limitations and planning for future assessments, GloBox can make data-driven decisions and maximize the revenue potential of its expanding food and drink product category.

We recommend proceeding with the launch of the food and drink category banner but with a clear understanding of the current limitations. To obtain more reliable insights, a follow-up experiment is strongly recommended. This follow-up should address the issues, ensuring a fair and balanced sample size and an adequate experiment duration. For more detailed insights and a comprehensive understanding of the experiment, please refer to the full report.

Introduction

Company Background:

GloBox is an e-commerce company. It's an online marketplace specializing in sourcing unique and high-quality products worldwide. GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products.

Methodology

Objective and hypothesis

Objective:

A/B testing: This is an experimentation technique used by businesses to compare two versions of a webpage, advertisement, or product feature to determine which one performs better. By randomly assigning customers or users to either the A or B version, the business can determine which version is more effective at achieving a particular goal.

Among its featured products, GloBox's 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. An A/B test experiment was conducted to understand the impact level.

Hypothesis:

General hypothesis: Adding a banner, that highlights key products in the food and drink category, at the top of the landing page of the website, can increase GloBox's revenue.

Refined hypothesis: Adding a banner, that highlights key products in the food and drink category, at the top of the landing page of the website, can increase the purchase rate of website visitors.

Hypothesis test I: The conversion rate hypothesis was used in conducting the A/B test experiment.

- *Null Hypothesis (H0):* Adding a banner, that highlights key products in the food & drink category, at the top of the landing page of the homepage, *will not* significantly impact the conversion rate compared to the existing landing page without a top banner.
- *Alternative Hypothesis (H1):* Adding a banner *will* significantly impact the conversion rate compared to the existing landing page without a top banner.

Hypothesis test II: For more refined decision-making, another hypothesis test was made to see whether there is a difference in the *average amount spent per user* between the two groups.

- *Null Hypothesis (H0):* Adding a banner *will not* significantly impact the average amount spent per user compared to the existing landing page without a top banner.
- *Alternative Hypothesis (H1):* Adding a banner *will* significantly impact the average amount spent per user compared to the existing landing page without a top banner.

Experimental design

A/B Test Setup:

The Growth team decided to run an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the test group sees it as shown below:
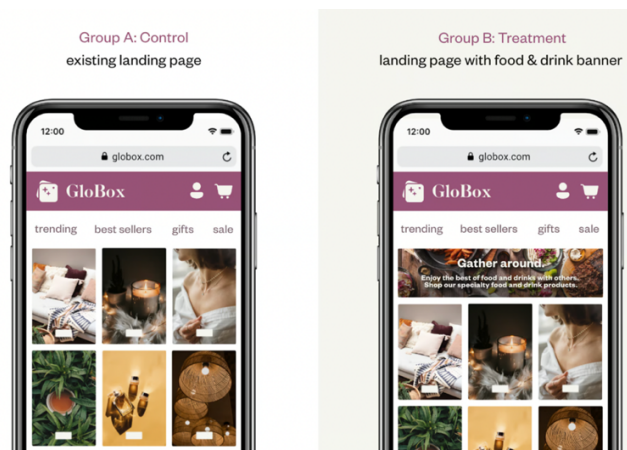

Figure 1 A/B Test setup

- **Scope of the experiment:** The A/B test is limited to the mobile version of the GloBox website.
- **User Assignment:** When a user visits the main page of the GloBox mobile website, they are randomly assigned to one of two groups: the control group or the test group.
- **User Purchases & conversion:** 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".

Sample size and duration:

	Number of users in the experiment
Group A: Control group	24,343
Group B: Treatment group	24,600
Total	48,943

(**Refer to: Appendix A, 1 & 2**)

The A/B test experiment duration was approximately around 2 weeks. It run from January 25th, 2023 - February 6th, 2023. Which was exactly a total of 13 days.

A/B test experiment duration	
Start date	**End date**
January 25th, 2023	February 6th, 2023
Total: 13 days	

(**Refer to: Appendix A, 3**)

A/B Test Analysis

To analyze the A/B test results an inferential statistics analysis method was conducted. The necessary calculations were done using spreadsheets.

Metrics:

Conversion rate:

Conversion rate is a metric commonly used in marketing and e-commerce to measure the effectiveness of a specific action or marketing campaign in terms of converting potential customers into actual customers.

$$Conversion\ Rate = \frac{Number\ of\ Conversions}{Total\ Number\ of\ users} \times 100\%$$

Group Category	Number of Conversion	Total Number of users	Conversion Rate
Group A (control group)	955	24,343	3.92%
Group B (treatment group)	1139	24,600	4.63%
Total (All users)	**2094**	**48,943**	**4.28%**

(Refer to Appendix A, 5)

The average amount spent per user:

Calculating the average amount spent per user helps assess the effectiveness of the A/B test in terms of revenue or profitability. The average amount spent per user of both the control group (A) and the treatment group (B) was calculated. And users who did not convert were included in the metric to measure the impact of revenue.

Group Category	Average amount spent per user (USD)
Group A (control group)	3.375 $
Group B (treatment group)	3.391 $

(Refer to Appendix A, 6)

Analysis results (using a spreadsheet.)

For hypothesis test I:

The first hypothesis tested was to see whether there was a difference in the conversion rate between the two groups.

Ho = There is no significant difference in the conversion rate between the two groups

H1 = There is a significant difference in the conversion rate between the two groups

Result: Hypothesis test I		
p-value	p val	0.00011
significance level	alpha	0.05
	p val < alpha	**REJECT Ho**

(Refer to Appendix C, 1)

Conclusion: The resulting p-value was 0.00011, which was significantly less than a significance level of 0.05. Therefore, we reject the null hypothesis (Ho) that there is no difference in the user conversion rate between the control and treatment groups.

95% confidence interval		
Calculation	**Notation**	**Value**
margin of error	MOE	0.0036
lower bound	LB	0.0035
upper bound	UB	0.0107
Confidence interval = (0.0035, 0.0107)		

(Refer to Appendix C, 1)

For hypothesis test II:

The first hypothesis tested was to see whether there was a difference in average spent per user between the two groups.

Ho = There is no significant difference in average spent per user between the two groups

H1 = There is a significant difference in average spent per user between the two groups

Result: Hypothesis test II		
p-value	p val	0.944
significance level	alpha	0.05
	p val > alpha	**FAIL TO REJECT Ho**

(Refer to Appendix C, 1)

Conclusion: The resulting p-value was 0.944, which was significantly higher than a significance level of 0.05. Therefore, we fail to reject the null hypothesis (Ho) that there is no difference in average spent per user between the control and treatment groups.

95% confidence interval		
Calculation	**Notation**	**Value**
margin of error	MOE	0.455
lower bound	LB	-0.439
upper bound	UB	0.471
Confidence interval = (-0.439, 0.471)		

(Refer to Appendix C, 1)

Result visualization (using Tableau)

Quick insight:

To clearly understand the hypothesis test results of both the conversion rate and average spent per user, visualizing was done using vertical bar charts.

*Conversion rate vs average amount spent per user:*

Test group comparision_conversion rate vs average amount spent per user

- Based on the above chart, it's shown that the conversion rate for the control group is 3.92% and the treatment group is 4.63%. which shows that the conversion rate is significantly higher for the treatment/test group.

- The average amount spent per user for the control group is $3.37 and for the treatment group, it is found to be $3.39, which shows that there is very little difference between the test groups.

Distribution of the amount spent per user:

Distribution of the amount spent per user for each group
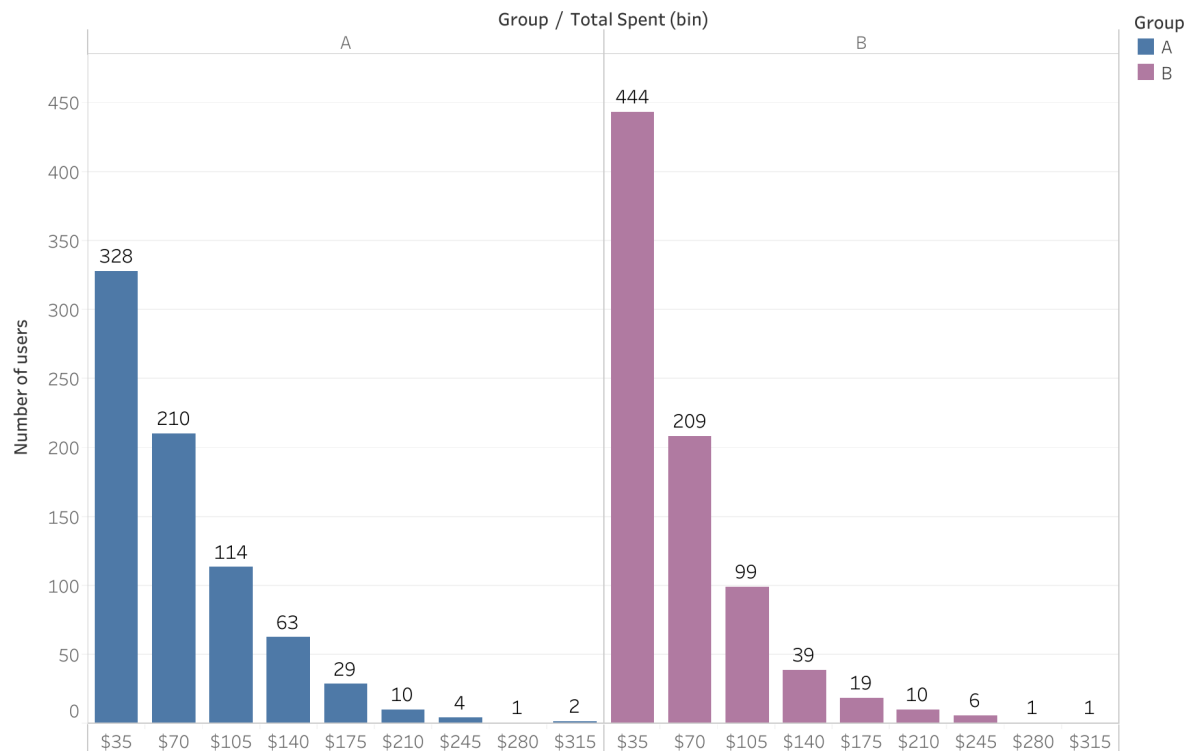
- The above histogram shows that the treatment group had a larger number of users in the 0-30$ bracket than the control group. In general, the overall distribution of the amount spent per user for both groups is similar with a slightly higher value from the treatment group.

Conclusion:

- The treatment group is more effective at getting users to take the desired action, such as visiting the site and making a purchase, compared to the control group. This seems to be a positive outcome.
- Although the conversion rate is higher, the lack of significant difference in the average amount spent per user suggests that the treatment group may not generate significantly more revenue per user than the control group.
- Therefore, it is important to analyze the data further by segmenting users to see if certain subgroups within the treatment group show a significant increase in average spending. This can help identify which user segments are driving the conversion rate improvement.

The A/B test experiment results were analyzed using three distinct segmentation metrics. The initial two metrics involved demographic segmentation, specifically examining the connection between the test metrics and users' device used, as well as the correlation between the test metrics and users' gender. Additionally, geographic segmentation was employed as another segmentation metric in the analysis.

The relationship between the test metrics and the user's device:

The relationship between the test metrics and the user's device
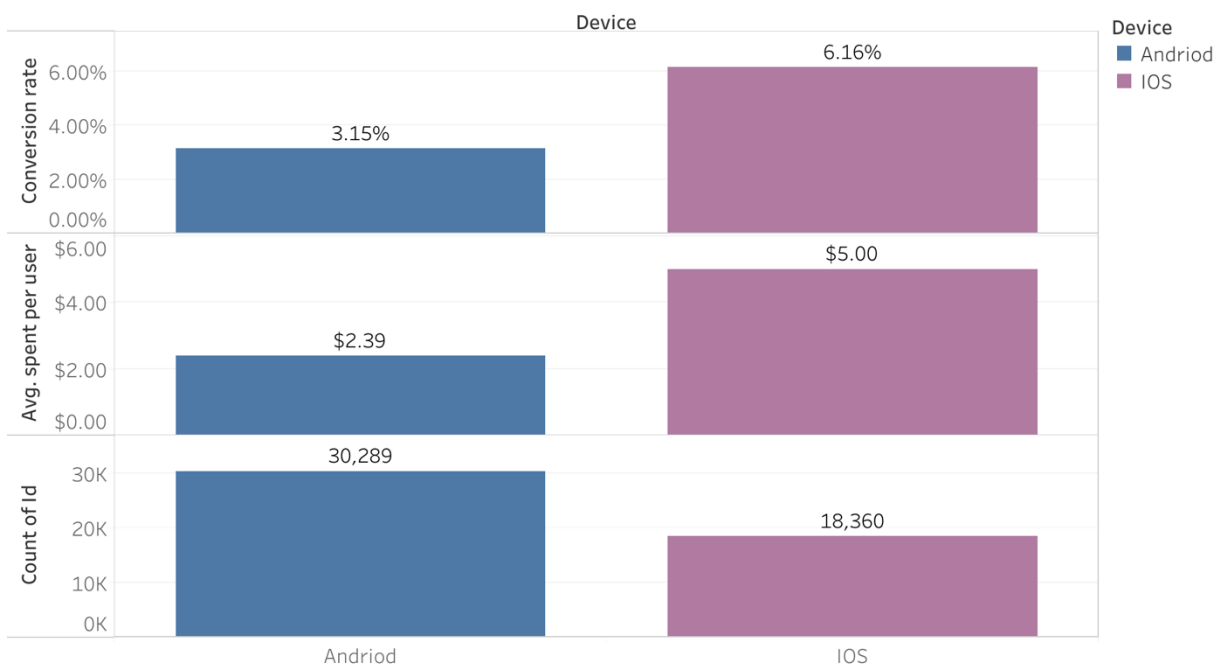
- Null values were omitted from the chart mentioned above due to their relatively minor impact. There were only 294 null values among the total 48,943 users, representing just 0.6% of the entire user dataset. As a result, they were excluded from the visualization chart.
- Despite the significantly lower number of IOS users participating in the experiment, both the conversion rate and the average spending per user among IOS users are notably higher.

*Relationship between test metrics and user's gender:*

Relationship between test metrics and user's gender
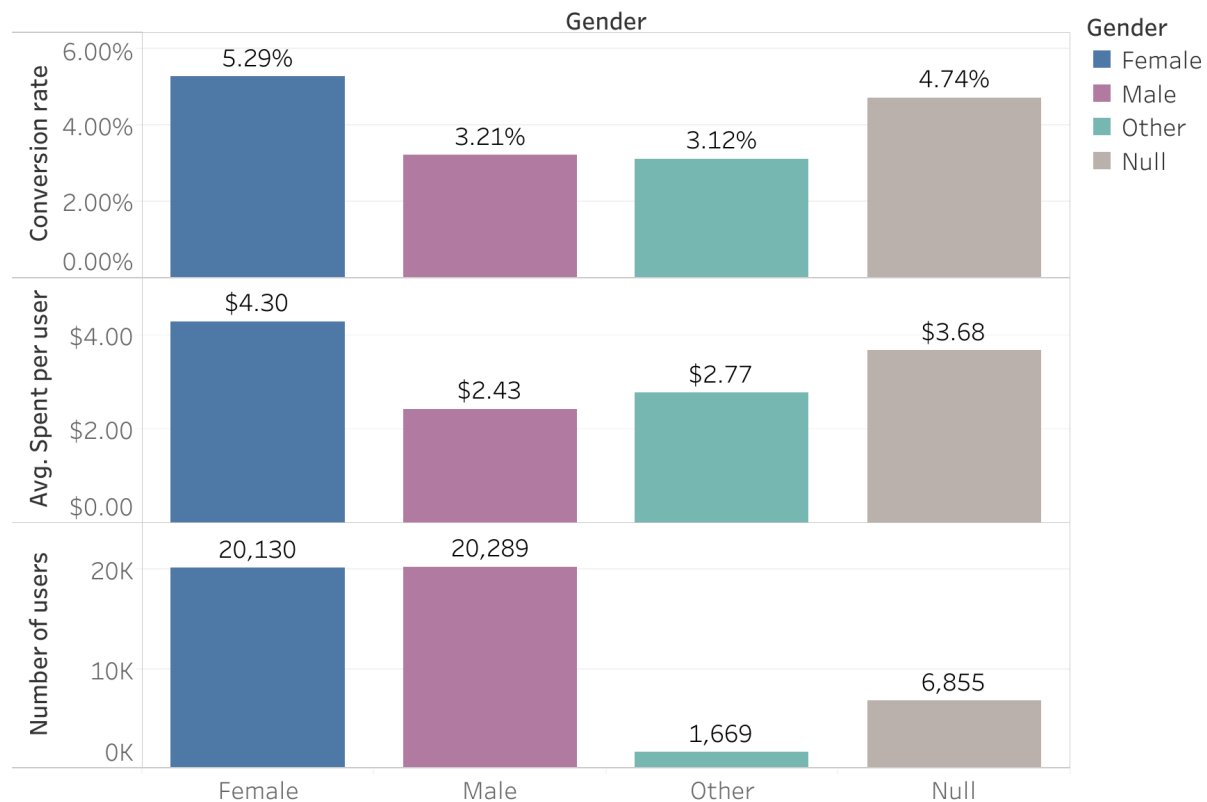
- The male and female user counts appear quite similar, suggesting a fair and balanced sample.
- Females exhibit both a higher conversion rate and a higher average spending per user.
- Users in the "other" and "unknown" categories collectively make up approximately 17% of the total user count. Therefore, it's important not to overlook these segments, as they exhibit higher values on both metrics being measured.

Relationship between test metrics and user's country:

Relationship between test metrics and user's country
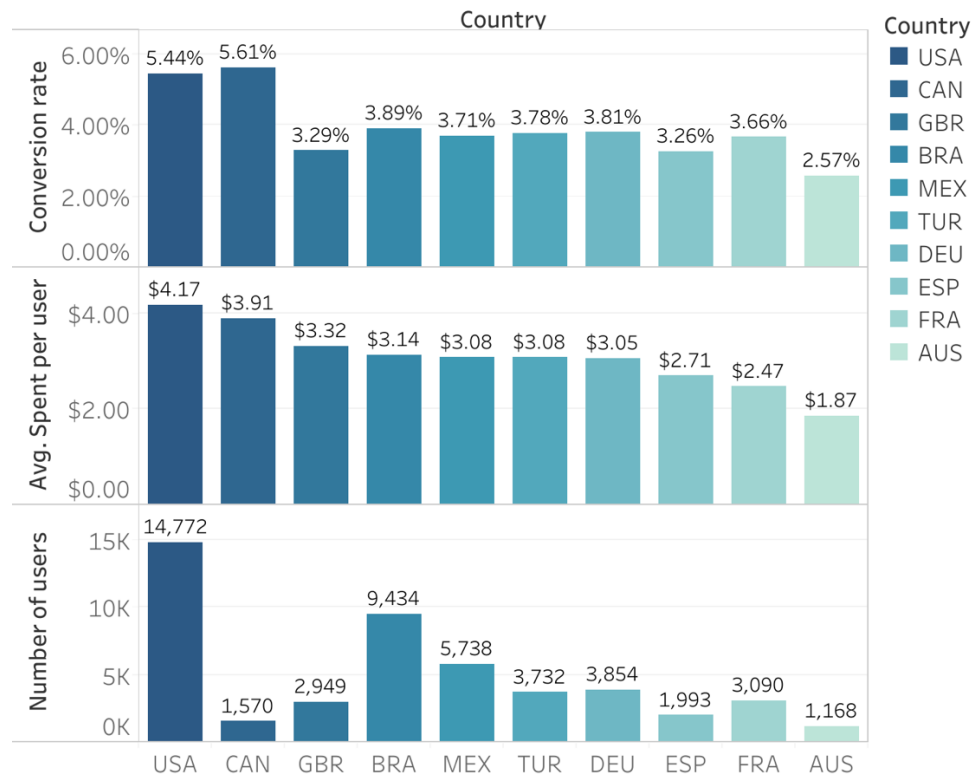
- The United States and Canada stand out with the higher conversion rates and the average amount spent per user. Specifically, users in the United States exhibit a conversion rate of 5.44%, with an average spending of $4.17, while users in Canada show a conversion rate of 5.61% and an average spending of $3.91.
- In contrast, Australia records the lowest conversion rate and the average amount spent. Users in Australia have a conversion rate of 2.57% and an average amount of spending of $1.87.

Summary of the A/B test results:

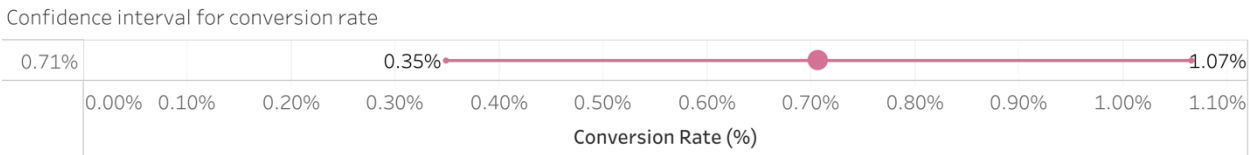key insights that provided valuable context for the A/B test results:

- While the conversion rate is higher in the treatment group, there isn't a significant difference in the average amount spent per user when compared to the control group. This suggests that the treatment group may not yield significantly more revenue per user.
- Despite a lower number of IOS users participating, they exhibit substantially higher conversion rates and average spending per user.
- Male and female user counts are similar, ensuring a fair and balanced sample. Females demonstrate both higher conversion rates and greater average spending per user.
- Notably, the United States and Canada stand out with the highest conversion rates and average spending per user, while Australia has the lowest conversion rate and average spending.

Confidence interval

A confidence interval is a range of values that provides a statistical estimate of the likely range within which a population parameter, such as a conversion rate or an average, falls. It quantifies the uncertainty associated with sample data and helps make inferences about the entire population based on that sample. Confidence intervals are essential for understanding the precision of A/B test results.
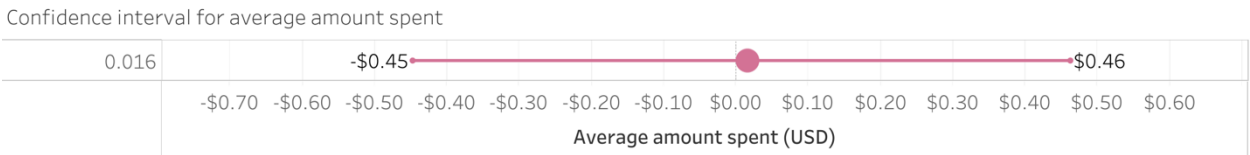
To visualize the confidence interval a new data source was created in the tableau workbook that contains the lower bound, point estimate, and upper bound for both intervals.

For conversion rate:

Confidence interval for conversion rate

| 0.71% | | | | 0.35% | | | | | | | | 1.07% |

| 0.00% | 0.10% | 0.20% | 0.30% | 0.40% | 0.50% | 0.60% | 0.70% | 0.80% | 0.90% | 1.00% | 1.10% |

Conversion Rate (%)

It suggests that there is a certain level of uncertainty associated with our estimate, but we can reasonably conclude that the true value of the parameter is likely to be within the range of 0.0035 to 0.0107 with a high degree of confidence.

For the average amount spent per user:

Confidence interval for average amount spent

| 0.016 | | | -$0.45 | | | | | | | | | $0.46 |

| -$0.70 | -$0.60 | -$0.50 | -$0.40 | -$0.30 | -$0.20 | -$0.10 | $0.00 | $0.10 | $0.20 | $0.30 | $0.40 | $0.50 | $0.60 |

Average amount spent (USD)

The confidence interval suggests that we have some level of uncertainty associated with our estimate. The interval spans both positive and negative values, indicating that there is no strong evidence to suggest a specific direction for the parameter. Instead, the true value is likely to be within the range of -0.439 to 0.471 with a high degree of confidence.

In A/B testing, the "novelty effect" refers to a temporary change in user behavior or response that occurs when users are exposed to a new or changed element in the user experience.

Novelty Effect
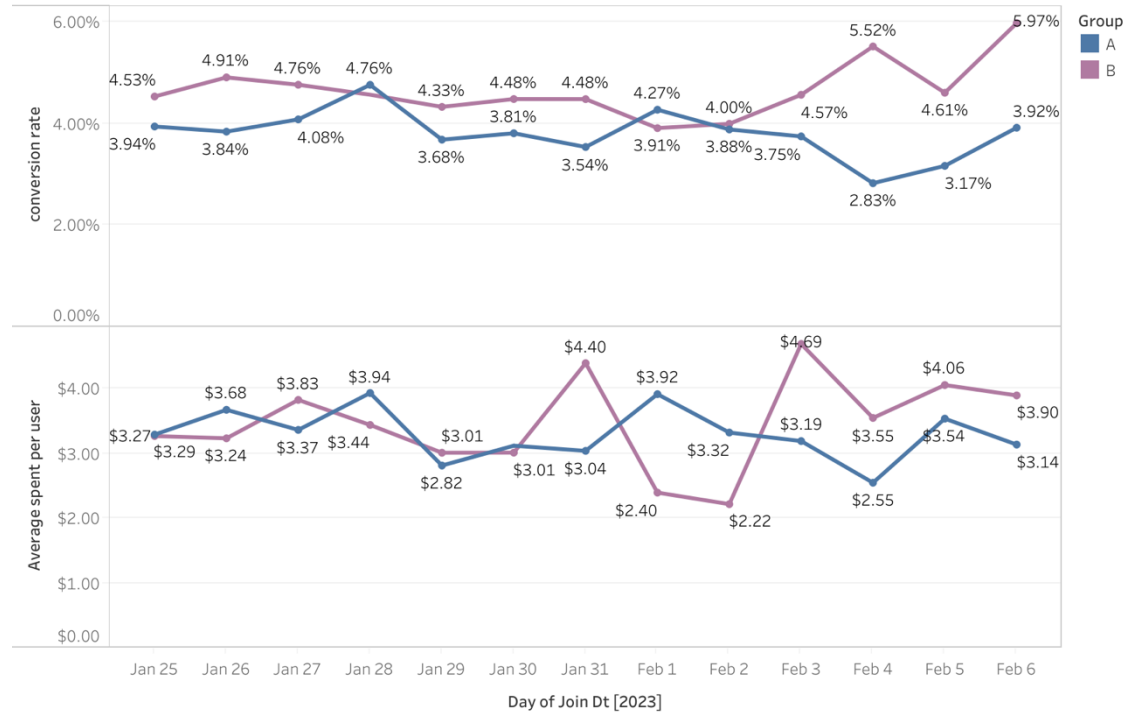
- Throughout the two-week experiment, the conversion rate of the treatment group consistently outperformed the control group. However, an intriguing development emerged during the second week, where the conversion rate exhibited a notable and steady increase. This is an encouraging sign, indicating that the changes implemented may be positively impacting user behavior. It's worth noting, though, that the relatively short duration of the experiment might not provide a comprehensive view of the complete effect.

- During the initial week of the experiment, spanning from January 25th to January 31st, the average amount spent per user in the treatment group did not exhibit significant differences compared to the control group. However, in the subsequent week, there was a noticeable upward trend in this metric. This suggests that while we may be observing the early stages of the novelty effect, the data collected thus far may not capture its full extent. Further observation and analysis are necessary to ascertain the lasting impact of the changes implemented.

Recommendation

Launch the experiment! and continue iterating.

Launching the experiment, involving the introduction of a food and drinks banner on the GloBox website's home page is recommended based on positive findings from a hypothesis test. The test results indicated a significant increase in the conversion rate, despite potential limitations in sample size according to power analysis. This banner implementation is relatively low-cost in terms of engineering and operations.

Even if only the conversion rate improves and not the average revenue per user, it's considered a positive step in attracting more paying customers, with the potential to increase purchases in the future.

However, it's acknowledged that the current sample size limitations may affect data representativeness and confidence in decision-making. The plan is to monitor and reevaluate results when the sample size becomes sufficient for the desired statistical power.

Additionally, the power analysis and awareness of the novelty effect highlight the non-representative sample size and experiment duration, emphasizing the need for a follow-up experiment to address these issues and obtain more reliable insights.

In summary, the recommendation is to proceed with the banner launch, understanding the current limitations and planning for future assessments with more robust data. A follow-up experiment is strongly recommended to ensure accurate conclusions.

Appendix

Appendix A

1. Dataset:

Globox stores its data in a relational database and the data can be accessed through Beekeeper Studio. Database link: - postgres://Test:bQNxVzJL4g6u@ep-noisy-flower-846766-pooler.us-east-2.aws.neon.tech/Globox
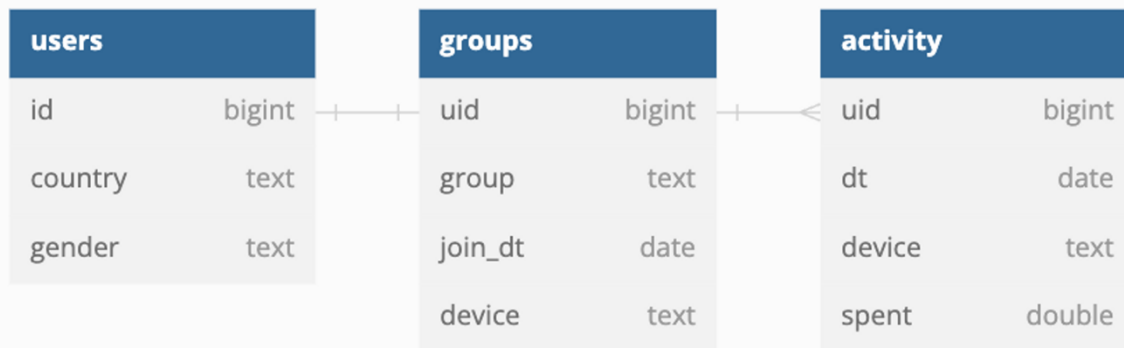
2. Database schema

Figure 2 Database schema

Appendix B: SQL query results

1. Total number of sample size

```
SELECT COUNT(uid) AS Total_number_of_users
FROM groups;
```

2. Sample size of each group

```
SELECT CASE WHEN t2.group = 'A' THEN 'Control_group'
        ELSE 'Treatment_group' END AS Group_category,
        COUNT(*) as number_of_users
FROM groups AS t2
GROUP BY t2.group
```

3. Experiment duration

```
SELECT MIN(dt) AS start_date,
        MAX(dt) AS end_date
FROM activity;
```

4. Query for A/B test analysis

```
--Query from online_users_database
--For A/B test analysis
SELECT t1.id,
       t1.country,
       t1.gender,
       t2.group,
       t2.device,
       COALESCE (SUM(t3.spent),0) AS total_spent,
       CASE
          WHEN SUM(t3.spent) > 0 THEN 1
          ELSE 0 END AS Converted
FROM users AS t1
LEFT JOIN groups AS t2
ON t1.id = t2.uid
LEFT JOIN activity AS t3
ON t1.id = t3.uid
group by t1.id, t2.group,t2.device
```

5. Conversion rate calculation

```
SELECT t2.group AS group_category,
    COUNT(DISTINCT (t3.uid)) AS number_of_conversion,
    COUNT(DISTINCT(t2.uid)) AS total_users,
    CONCAT(ROUND(COUNT(DISTINCT (t3.uid))*100.0/
          COUNT(DISTINCT(t2.uid)),2), '%') AS conversion_rate
FROM groups AS t2
LEFT JOIN activity AS t3
USING (uid)
GROUP BY t2.group
-- summary raw that sums the results of both groups
UNION ALL
SELECT
    'Both_groups' AS group,
    SUM(number_of_conversion) AS total_number_of_conversion,
    SUM(total_users) AS total_total_users,
    concat(round(SUM(number_of_conversion * 100.0) /
          SUM(total_users),2), '%') AS total_conversion_rate
FROM
```

```
(SELECT t2.group,
        COUNT(DISTINCT t3.uid) AS number_of_conversion,
        COUNT(DISTINCT t2.uid) AS total_users,
        COUNT(DISTINCT t3.uid) * 100.0 /
        COUNT(DISTINCT t2.uid) AS conversion_rate
    FROM groups AS t2
    LEFT JOIN activity AS t3
    USING (uid)
    GROUP BY t2.group ) AS subquery;
```

6. Average spent per user calculation

```
--Common Table Expression (CTE) to calculate total spent per
user
WITH userspending AS (
    SELECT
        t1.id,
        t2.group,
        COALESCE(SUM(t3.spent), 0) AS total_spent
    FROM users t1
    LEFT JOIN activity t3 ON t1.id = t3.uid
    LEFT JOIN groups t2 ON t1.id = t2.uid
    GROUP BY t1.id, t2.group )
-- Main query to calculate the average spent per user
SELECT t2.group,
CONCAT(ROUND(AVG(u.total_spent),3),'$')AS average_spent_per_user
FROM userspending AS u
LEFT JOIN
    groups t2 ON u.id = t2.uid
    GROUP BY t2.group;
```

7.Query for checking the novelty effect

```
--query for checking Novelty Effects
SELECT t1.id,t1.country,t1.gender,t2.device,t2.group,t2.join_dt,
        COALESCE(SUM(t3.spent), 0) AS total_spent,
        CASE
         WHEN SUM(t3.spent) > 0 THEN 1 ELSE 0
         END AS conversion
FROM users AS t1
LEFT JOIN groups AS t2
ON t1.id = t2.uid
LEFT JOIN activity AS t3 ON t1.id = t3.uid
GROUP BY t1.id, t2.device,t2.group,t2.join_dt;
```

Appendix C: spreadsheet results

1. A/B test Analysis

https://docs.google.com/spreadsheets/d/1lWzjKYOtGUiFxcZlSza1-RLOzkeoRYYmovjxLeZHrKE/edit#gid=1443676962

2. Query result for novelty effect check

https://docs.google.com/spreadsheets/d/13a68_EL49R-WTjTkstQXwoar80Xun-aaY4H7NmO94-s/edit?usp=sharing

Appendix D: Tableau visualization results

1. A/B test result

- Conversion rate vs average amount spent per user:

  https://public.tableau.com/authoring/GLOBOXPROJECTDECEMBER2023/Sheet1/Conversion%20rate%20vs%20average%20amount%20spent%20per%20user#1
- Distribution of the amount spent per user:

  https://public.tableau.com/authoring/GLOBOXPROJECTDECEMBER2023/Sheet1/Distribution%20of%20the%20amount%20spent%20per%20user#1
- The relationship between the test metrics and the user's device:

  https://public.tableau.com/authoring/GLOBOXPROJECTDECEMBER2023/Sheet1/The%20relationship%20between%20the%20test%20metrics%20and%20the%20user%E2%80%99s%20device#1
- The relationship between the test metrics and the user's gender:

  https://public.tableau.com/authoring/GLOBOXPROJECTDECEMBER2023/Sheet1/Relationship%20between%20test%20metrics%20and%20user%E2%80%99s%20gender#1
- The relationship between the test metrics and the user's country:

  https://public.tableau.com/authoring/GLOBOXPROJECTDECEMBER2023/Sheet1/Relationship%20between%20test%20metrics%20and%20user%E2%80%99s%20country#1

2. Novelty effect result

https://public.tableau.com/authoring/Noveltyeffect_16956545690180/NoveltyEffect#1