GLOBOX

A/B BANNER TEST - UPDATED REPORT

15 NOVEMBER 2023 / BY DOUGLAS KIBAZOHI / DATA ANALYST

TL;DR

I recommend the Growth Team to continue reiterating the banner/user-experience. Results show little-to-no revenue increase but test group conversions are very positive, hence room for improvement.
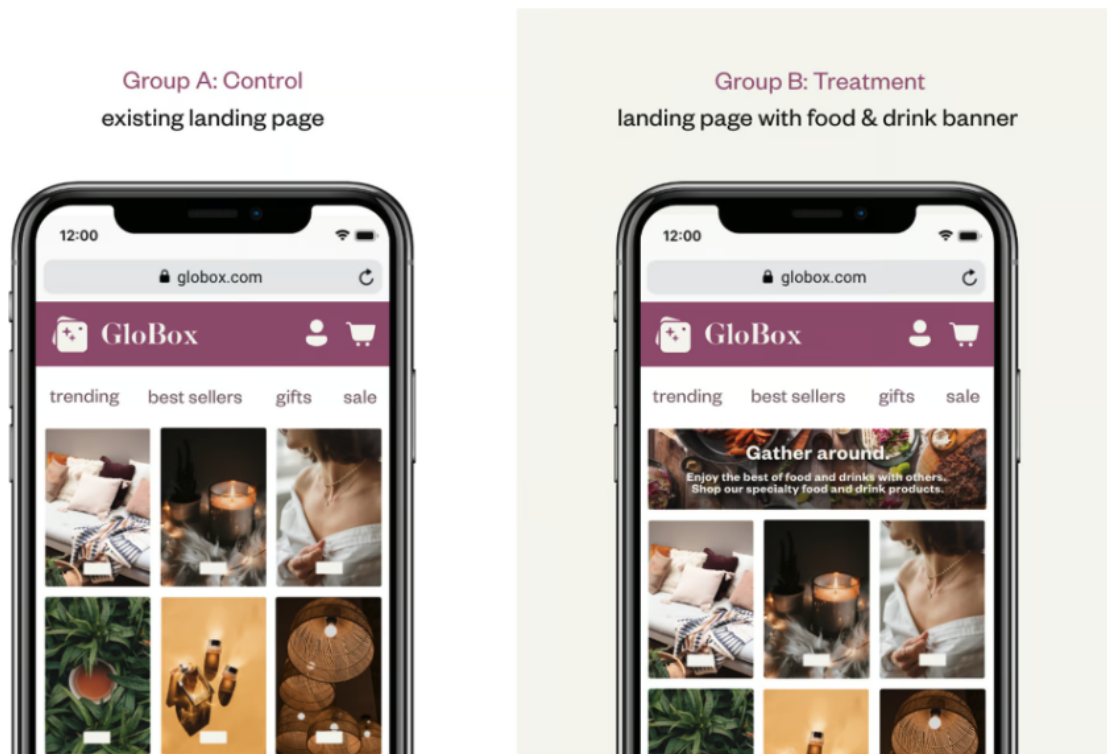
SUMMARY

Analysis of A/B Test showed 18% increase in conversions – higher than our expected minimum of 10%. Revenue on the other hand increased by 0.48% – a stark difference, which we could attribute to the fact that we are boosting sales in a low-revenue category. This gives us room to improve, reiterate, test reselling customers and create further updates to the user experience or banner itself in order to take better advantage of this opportunity.

Our sample size was not large enough to provide enough data to confirm whether or not we can release the banner now, however what we have collected so far is enough to inform our next modifications.

CONTEXT

Motivation:

While GloBox is commonly known for selling boutique fashion items and high end decor products, their **food and drink category has grown significantly.** The goal of this test is to see if the website banner will **increase awareness** for this category and **boost revenue.** The Growth Team decided to run an A/B test to

Group A: Control
existing landing page

Group B: Treatment
landing page with food & drink banner

examine results from introducing a new banner on their website to highlight key products in GloBox's food and drink category. The test group is shown the banner and the control group is not.

Test parameters:

- Sample Size of Control Group (A): **24343**
- Sample Size of Test Group (B): **24600**
- Total Sample Size: 48943
- Duration of Test: 13 Days (Jan 25 – Feb 06, 2023)

Data:

Our dataset consists of three tables:

- **Users:** user id, country, gender
- **Groups:** user id, group (A or B), join date and device (eg: Android, iOS)
- **Activity:** user id, date of purchase, device, amount spent

Unfortunately, I did not have access to 'product specific' data and that is one limitation in the analysis.

The Dataset can be accessed through Beekeeper using the link below via <u>Import from Url</u> option:

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

ANALYSIS AND RESULTS

Data was analyzed on the database, Spreadsheets and Tableau Public.

Main Export (Conversions and Amount Spent per Customer)

Since this test is focused on seeing whether there is a change in conversion rate and revenue between the control and test groups, the first challenge was to:

1. combine all relevant information in one table (user id, group, device, gender, total spend)
2. create a new column to indicate whether a website guest had converted into a buying customer (conversion_status)
3. clean duplicate values by summing the spend amount per unique users. *Multiple purchase activities from the same customer still count as one conversion.

See SQL code

Next, the data was exported as a CSV file for further analysis in Spreadsheets.

Statistical Analysis in Spreadsheets

Two pivot tables were created in spreadsheets from exported data. One dedicated to conversion rate results, the second to analyze the average spend per user. Below is a table summarizing the findings from both.

Group	Conversion Rate	Average Spend per User $
Control (A)	0.0392	3.37
Test (B)	0.0463	3.39

Figures for Average spend were calculated with SQL, See Code.

To check the validity of the results, I performed a two-tailed z-test to check the p-value of the results under the following null and alternative hypotheses (significance level being 0.05).

- **NUll Hypothesis:** There is **no change** in (conversion rate or avg spend) between the groups.
- **Alternative Hypothesis:** There is a **change** in (conversion rate or avg spend) between the groups.

Below is a summary of results from the statistical analysis:

Analysis For:	P-value	Conclusion
Conversion Rate	0.000111 (p-value < 0.05)	Reject null hypothesis. There exists a significant difference between the groups.
Avg. Spend	0.944194 (p-value > 0.05)	Fail to reject null hypothesis. There is no significant difference between the groups.

Before accepting the data, I made sure to check whether the sample size of the groups was large enough for the A/B Test, to truly be able to measure changes accurately with 95% confidence level, 10% MDE and 80% power. This was verified with the help of online statistical calculators.

Power: Probability to correctly reject the null hypothesis. 80% is the standard for power tests so I decided to use it.

MDE: Minimum detectable effect is the smallest possible change we hope to detect in our results. Assuming this was selected before our test, I would go with 10% of what we see in the control group.

Results:

- **Conversion Rate:** Sample size needed for each variant = **38500**

Calculated with https://www.statsig.com/calculator

Baseline Conversion Rate is represented by the conversion rate of our control group.

- **Avg. Spend:** Sample size needed for each variant = **92734**

Calculated with https://statulator.com/SampleSize/ss2M.html#

Difference between means is represented by 10% of the mean of the control group which is the minimum difference we expect to see. The MDE comes into play here. The Expected Standard Deviation is that of the control group also.

Results from our power analysis tests show that our total sample size (48943) was not large enough for our A/B Test since the required sample sizes are beyond what we had. This information was taken into account in my final recommendation.

Further analysis was done to calculate the Confidence Interval Difference between the groups. These results will be posted in the next section, 'Visualizations'.

Click for complete statistical analysis (Spreadsheets).

Note: Standard Deviations for Average Spending were calculated in Beekeeper using SQL, NOT in Spreadsheets. *See Code*.
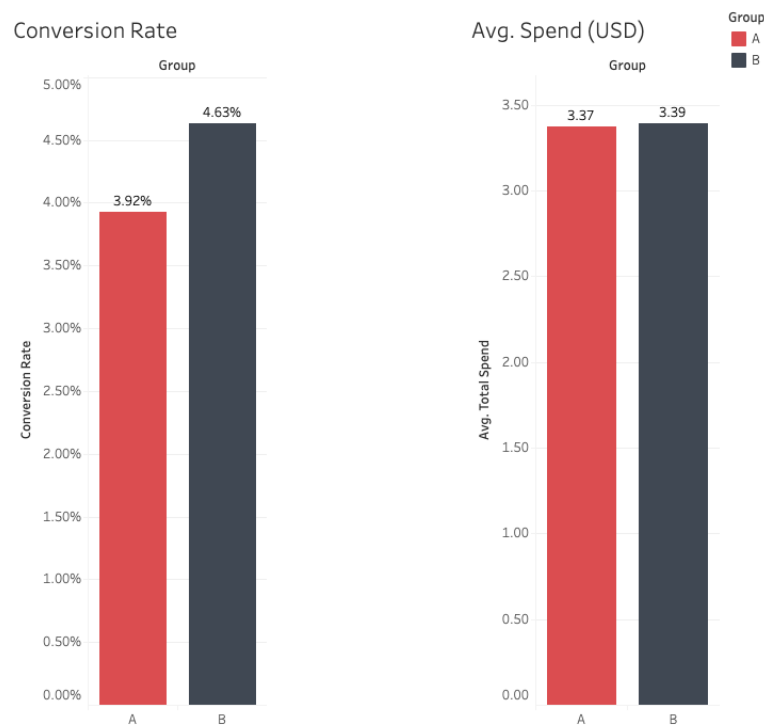
Visualizations and Further Analysis in Tableau

To further analyze patterns in our data, I used Tableau Public to help visualize the information to find patterns.

### 1. Comparison of Key Metrics Per Group

Analysis of Conversion Rate and Average Spend per user per group is available in the previous section.

2. Confidence Interval Difference

Different charts were made to analyze the confidence interval difference between the groups on conversion rate and average spend.

Confidence Interval Difference - Conversion Rate

1.07%

0.71%

0.35%
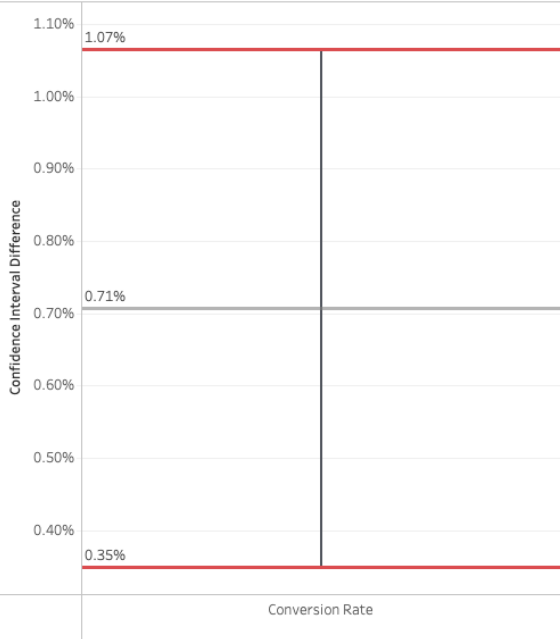
Conversion Rate

Confidence Interval Difference - Avg Spend USD

0.47

0.02

-0.44
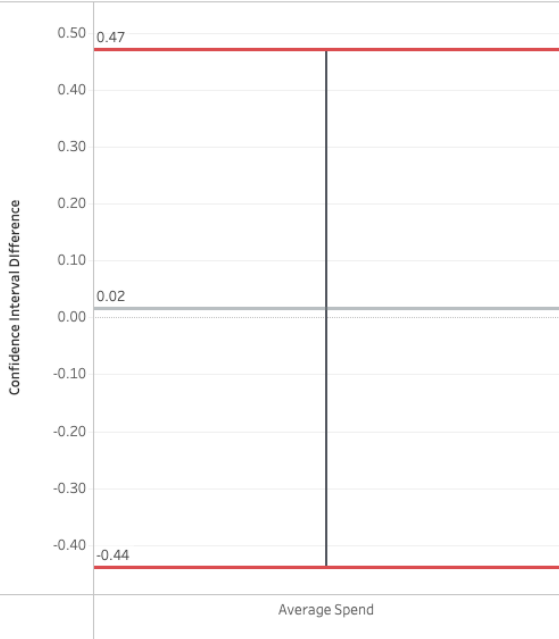
Average Spend

Confidence Interval Levels - Conversion Rate

Group
- A
- B

4.89%
4.63%
4.37%

4.17%
3.92%
3.68%

A   B
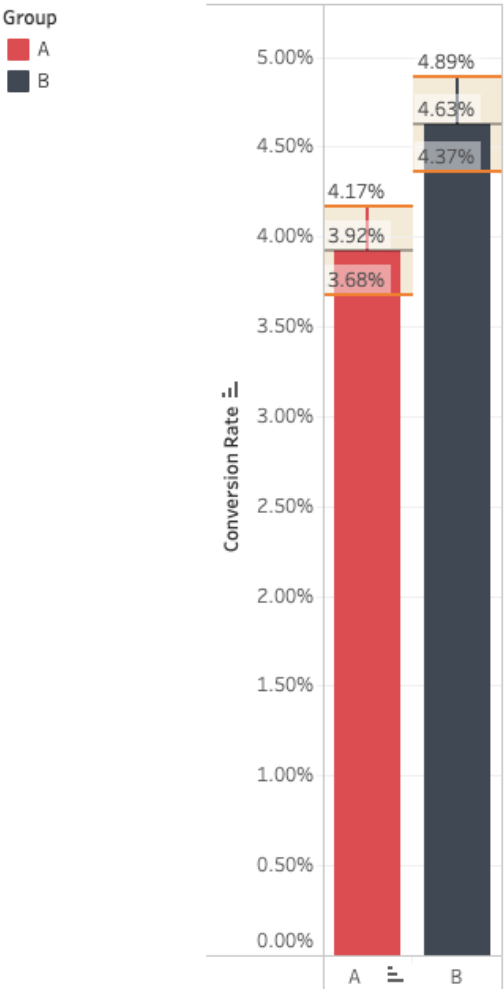
Confidence Interval Levels - Average Spend (USD)

3.70   3.71

3.37   3.39

3.05   3.07

A   B
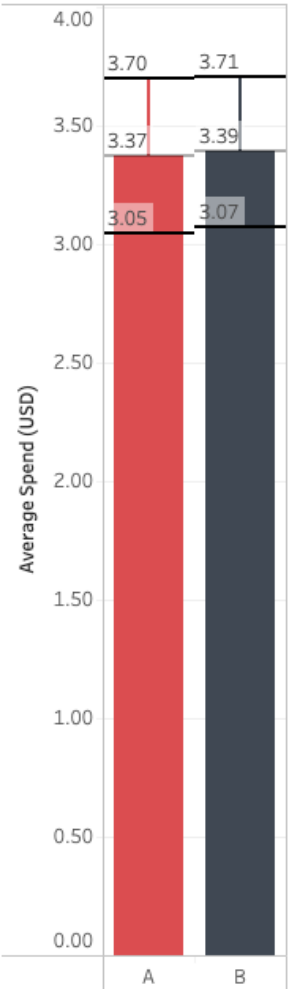
Confidence intervals are calculated with a 95% confidence level. Below are the conclusions:

➢ We can be confident that 95% of values for conversion rate will fall between **3.68-4.17%** for the control group and **4.37-4.89%** for the test group (those seeing the banner).

➢ We can also be confident that 95% of values for average spend per user will fall between **$3.05-$3.70** for our control group, and for those shown the banner, values will fall between **$3.07 - $3.71**. This continues to prove that we are not seeing any significant revenue growth as most values overlap.

3. Novelty Effect

Another visualization was made to check if our findings might be a result of a novelty effect. A SQL code was made to extract the difference in conversion rates and average spending of users by date and the results were visualized on the line chart below.

See SQL Code

Difference in Key Metrics Between Groups (B-A)

The line chart shows that positive results in conversion rate and average spend values are not a novelty effect, since the results do not decline with time. Instead, results continue in a positive trend, especially for the conversion rate where this can be detected very clearly. The average spending struggled a bit in the beginning but towards the end of the test period, it showed better results. However, it is important to reiterate that a much larger sample size would be more appropriate for our analysis.

4. Distribution of Amount Spent by Users per Group

Distribution of Amount Spent in USD by $30 bins
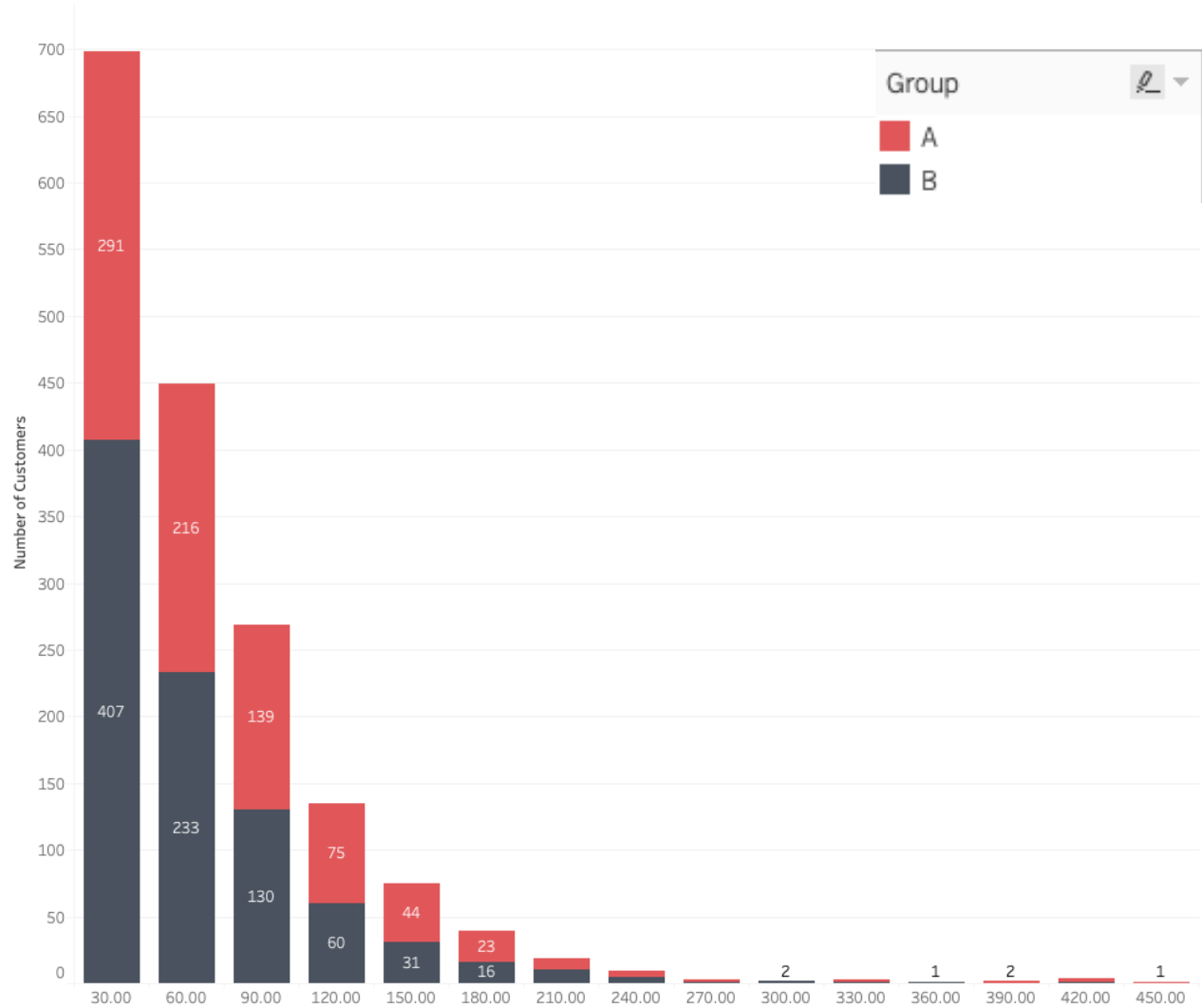
This distribution shows that the ratio of customers spending varies per group according to amount.

At $30 or under, Group B converts 1.4x more users than Group A. But that changes at higher prices. By the $30-$60 bin, Group A has caught up and surpasses Group B starting from the $60-$90 bin. This could imply that the food and drink category that has boosted conversion rates sell for much cheaper prices compared to other product categories.

Note: Outliers and customers who did not purchase anything were excluded from the distribution.

5. Analysis by Gender

Conversion Rate by Gender

M - Male
F - Female
O - Other

Avg. Spend (USD) per Gender
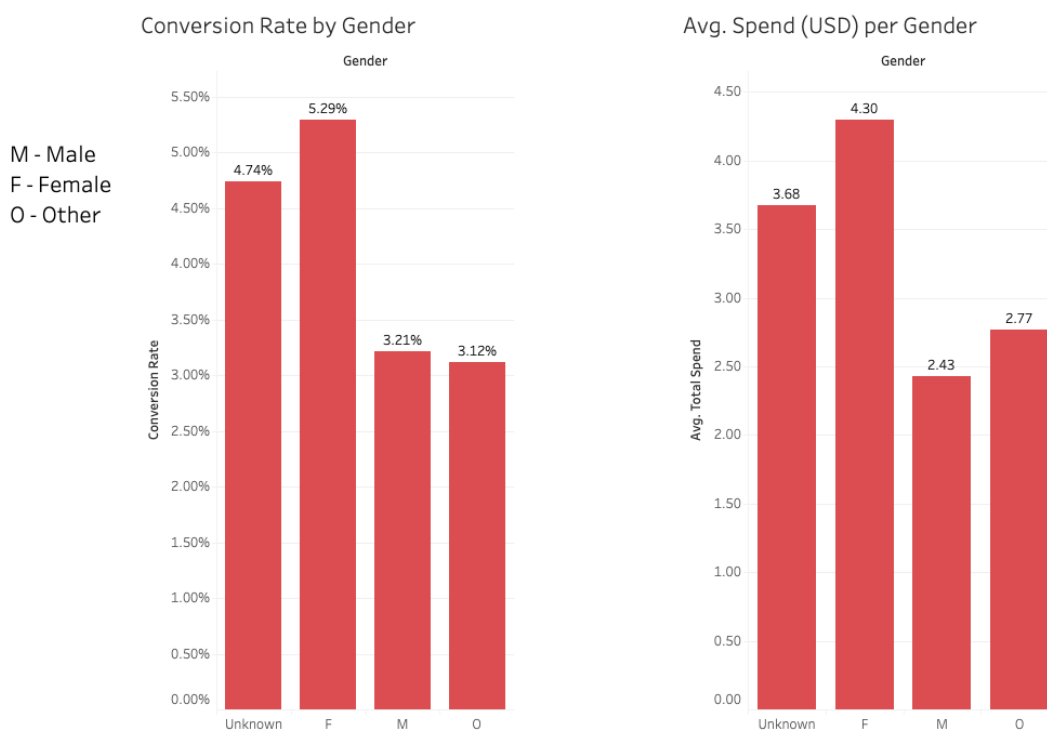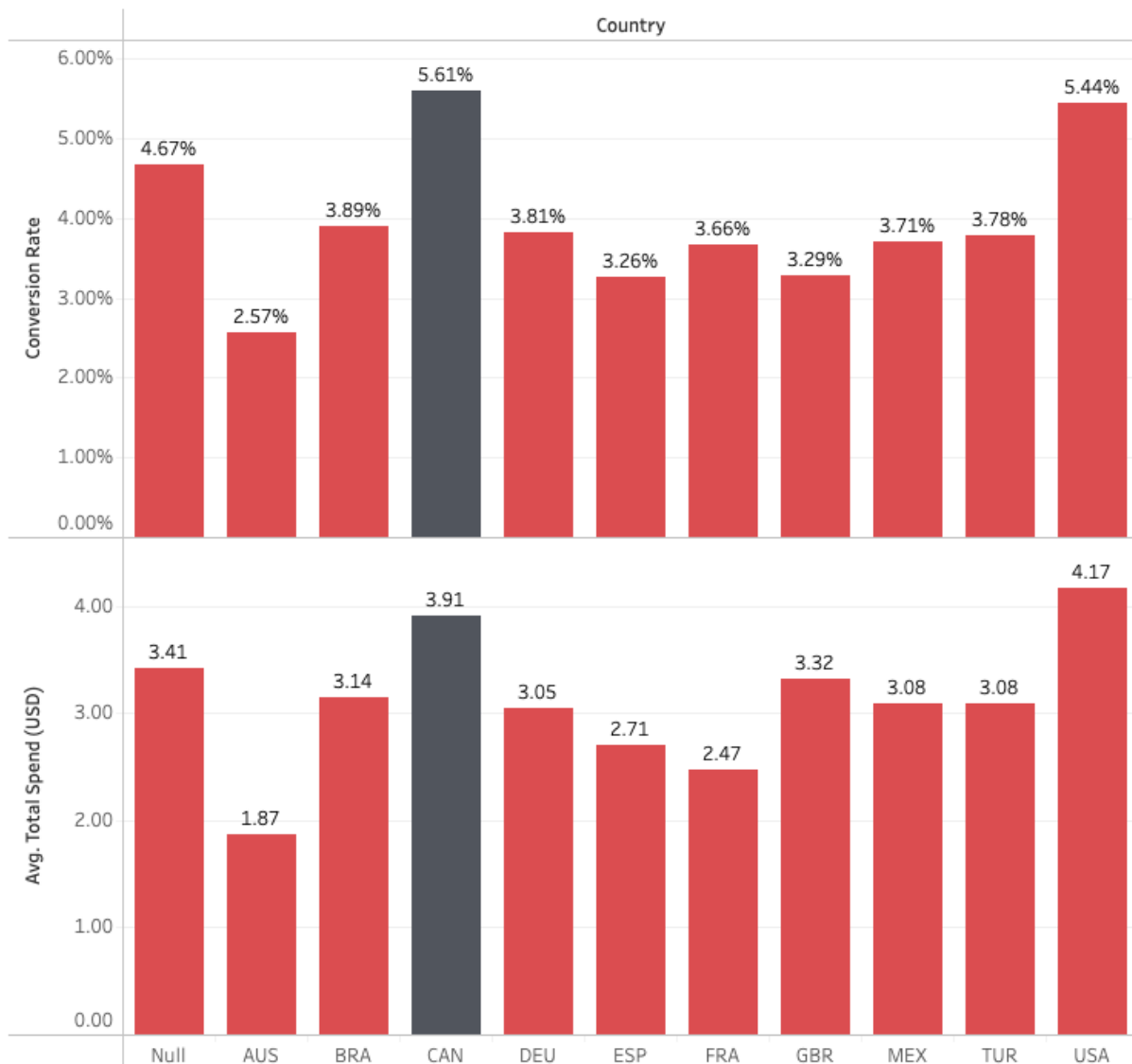
In this visualization Women lead in both metrics: 5.29% Conversion Rate and $4.30 average spent per customer. These values are well above our overall values, making Women our most valuable demographic by gender.

6. Analysis by Country

Metrics by Country

Analysis by Country revealed that Canada has the highest conversion rate at 5.61% and comes second to the USA in average spending per customer at $3.91. While Canada does have one of the lowest populations among our target countries which explains why we have comparatively few customers from the region, it may still be worthwhile to give special attention to the Canadian Market because of its impressive conversion rate.
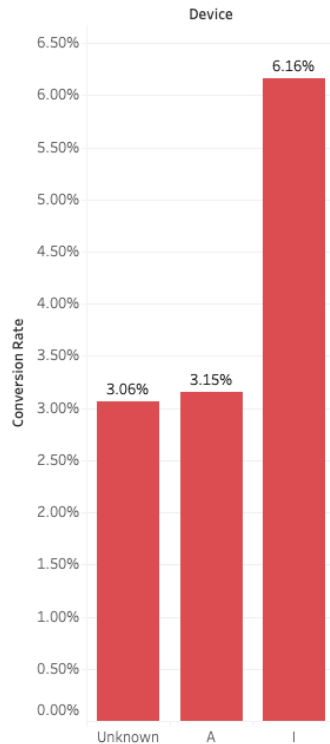
7. Analysis by User Device

Comparison of Key Metrics Per User's Device

A - Android
I - iOS

Conversion Rate by Device

Avg. Spend (USD) by Device
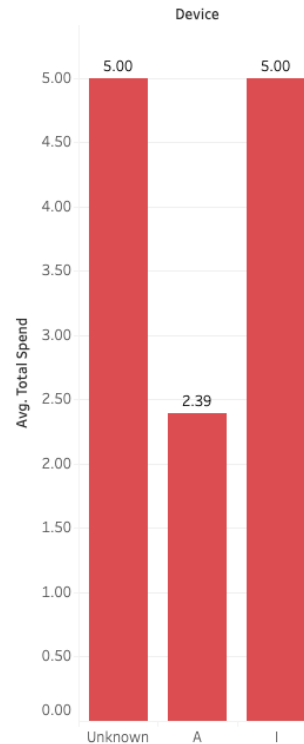
iOS users exhibited the highest conversion rates, about double that of Android users. Average Amount spent for iOS users is also high in comparison, but surprisingly devices categorized as unknown had the same average amount spent per user, despite not having a high conversion rate.

Links to all visualizations are available in the Appendix. Simply click here to jump.

RECOMMENDATION

Although there was a positive response in conversions for the test group, our sample size was insufficient for this test and our revenue increase is very low. This may also be due to the fact GloBox primarily sells high end decor and boutique fashion items which are usually more expensive compared to food and drink items.

I noticed that an increased ad budget may have likely been used at the beginning of the test, marked by a sharp increase in website visitors on Jan 25th, and this has been taken into consideration.

My recommendation is that we continue to reiterate. We could improve:

- the banner experience
- customer journey (including reselling)
- perhaps use this real-estate differently.

Our conversion rate (being well over our 10% MDE) indicates we can definitely use this space for something, but we need to experiment more to be sure of what and how.

With the low revenue returns we have seen from our A/B test, I am not in favor of extending our test duration just to collect more data *unless* we change something about the banner or user experience. Customers who bought cheaper items during the test period can be converted into future buyers of our high-end products. Or, food and drink items can be advertised as an ad-on to cart items while the banner space is used for something else.

There is potential here, and considering all investments made, it is best to find a solution that yields better returns.

NOTE: Revenue mentioned in this report is total revenue and not net-revenue.

APPENDIX

SQL Script for Main Export:

```
with updated_activity as (

  select *
  from activity
  right join groups using (uid)
  )

select distinct (gr.uid), country, gender, gr.device, gr.group,
          case when spent is not null then 'Converted'
    else null
    end as conversion_status,
    sum(coalesce(spent, 0)) as total_spend

from groups gr
left join users ON gr.uid = users.id
left join updated_activity ac using (uid)

group by 1,2,3,4,5,6
```

SQL Script for Standard Deviation of Average Amount Spent

*Calculated per unique user, for groups A and B

```
with converted as (

SELECT distinct gr.uid as uid_converts, gr.group, sum(coalesce(spent, 0)) as
total_spent
FROM groups gr

left join activity using (uid)
group by 1,2 )

select cv.group, stddev(total_spent)
from converted cv
group by 1
order by 1
```

SQL Script to Analyze Novelty Effect

```sql
with updated_activity as (
   select *
   from activity
   right join groups using (uid)
   ),

group_data as (
select gr.join_dt, gr.group, gr.uid,
            case when spent is not null then 'Converted'
   else null
   end as conversion_status,
   sum(spent) as total_spend

from groups gr
left join users ON gr.uid = users.id
left join updated_activity ac using (uid)
group by 1,2,3,4 ),

group_a as (

select distinct join_dt,
count(conversion_status)/count(distinct uid)::numeric as rate_a,
sum(total_spend)/count(distinct uid) as avgspend_a
from group_data gd
where gd.group = 'A'
group by 1 ),

group_b as (

select distinct gd.join_dt,
count(conversion_status)/count(distinct uid)::numeric as rate_b,
sum(total_spend)/count(distinct uid) as avgspend_b
from group_data gd
where gd.group = 'B'
group by 1 )

select distinct (ga.join_dt), (rate_b - rate_a) as rate_difference,
        (avgspend_b - avgspend_a) as avgspend_difference
from group_a as ga
left join group_b as gb using (join_dt)
```

SQL Script for Average Amount Spent per User:

```sql
with updated_activity as (

   select gr.uid, gr.group, coalesce(spent, 0) as new_spent
   from groups gr
   left join activity using (uid)
   )

select ua.group, sum(new_spent)/count(distinct(uid)) as average_spent
from updated_activity ua
group by 1
```

Spreadsheets Analysis:

https://docs.google.com/spreadsheets/d/13AGtvCLjot2dVnorJvSVGTaodmPUFLdiQWJYnfF8hj4/edit?usp=sharing

Tableau Public Visualizations:

Comparison of Key Metrics Per Group:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/MetricsbyGroup?publish=yes

Comparison of Key Metrics by User's Device:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/MetricsbyDevice?publish=yes

Comparison of Key Metrics by Gender:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/MetricsbyGender?publish=yes

Key Metrics by Country:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/KeyMetricsbyCountry?publish=yes

Confidence Interval Difference by Metrics:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/C_IDifferenceDashboard/C_IDifferenceDB?publish=yes

Comparison of Confidence Interval Levels:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/ConfidenceIntervalVisualization?publish=yes

Novelty Effect Analysis:
https://public.tableau.com/app/profile/douglas.kibazohi/viz/GloboxNoveltyEffectAnalysis/DifferenceinKeyMetricsBetweenGroupsB-A?publish=yes

Distribution of Amount Spent per Group
https://public.tableau.com/app/profile/douglas.kibazohi/viz/Globox-Anubis-Mean/DistributionofAmountSpentinUSDby30bins?publish=yes