# GloBox

## A/B Testing Report

Prepared By Jaypal Jadeja

# Content

# TL;DR

I suggest the Growth Team persist in refining the banner and user experience. Although revenue growth has been minimal, the test group's conversions have shown positivity, indicating potential for enhancement.

# Summary

The A/B Test analysis revealed an 18% increase in conversions, surpassing our expected minimum of 10%. However, the revenue growth was only 0.48%, notably lower. This disparity could be linked to our focus on a low-revenue category. This situation indicates a scope for improvement, suggesting the need to iterate, test reselling strategies to existing customers, and further refine the user experience or banner design to better leverage this opportunity. Although our sample size was insufficient to conclusively determine the banner's readiness for release, the data collected thus far is adequate to guide our next modifications and decision-making process.

# INTRODUCTION

In this project, we embark on a journey with GloBox, an innovative e-commerce company known for its unique selection of boutique fashion items and high-end decor products.

GloBox is primarily known amongst its customer base for boutique fashion items and high-end deco 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.

As the Data Analyst for GloBox, we are tasked with dissecting the results of a crucial A/B test designed to elevate the visibility of their rapidly Increasing food and drink category. This test involves showcasing key products in the food and drink section through a banner on the mobile website.

The control group does not see the banner while the test group sees it as shown below

Group A: Control

existing landing page


Group B: Treatment

landing page with food & drink banner

Our role extends to collaborating with the Growth Product & Engineering Team, led by Leila and Alejandro, along with Mei, the Head of Marketing, to derive data-driven insights.

As we delve into the user-level aggregated dataset extracted using SQL, we will uncover patterns in user behavior, conversions, and purchases, drawing upon statistical methods and visualization tools.

With the impending decision on whether to implement the new experience across the board, our findings and recommendations will serve as the cornerstone of this pivotal choice.

# Test Parameters

In this experiment, two distinct test groups have been established: Group A, also referred to as the Control Group, and Group B, designated as the Treatment Group. To ensure the impartiality of the results, the assignment of individuals to these groups has been executed randomly. Group B, the Treatment Group, will experience the mobile website with a prominently displayed banner showcasing products, while Group A, the Control Group, will encounter the existing mobile website without this banner.

Data:

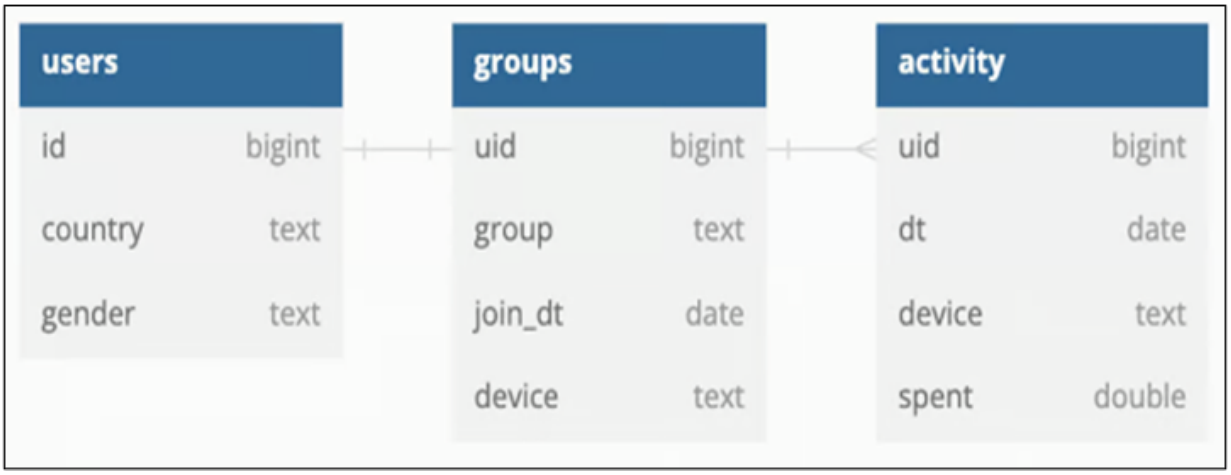
- Sample Size of Control Group (A): 24343
- Sample Size of Test Group (B): 24600
- Total Sample Size: 48943
- Duration of Test: 13 Days (Jan 25 - Feb 06, 2023)

Two key metrics are at the heart of this experiment's evaluation. The first metric is the conversion rate of customers, which serves as a critical indicator of whether customers are making purchases because of the experimental banner.

The second metric, the average purchase amount, is pivotal in assessing the overall revenue implications of the experiment. Together, these metrics provide a robust foundation for evaluating the experiment's impact on both customer behavior and the company's financial performance.

# Dataset Overview

The dataset comprises three tables: the users table, groups table, and activity table. The users table contains demographic information about the users, while the groups table focuses on assigning users to two different groups for the A/B test. The activity table records user purchase activity.




The table reveals the presence of four distinct data types, namely bigint, text, date, and double

- .bigint is employed for handling larger-scale integer data or whole numbers.
- text is utilized for storing textual information, such as words or sentences.
- date data type serves the purpose of storing date-related information.
- double data type is designated for storing floating-point numbers or decimal values.

**Regrettably, The dataset lacked to have 'product specific' data, presenting a limitation in the analysis.**

# Hypothesis Test for Conversion Rate (A & B groups)

To test whether there's a difference in conversion rates between the two groups, a two-sample z-test was carried out.

The Set Parameters:

Null Hypothesis (H0) = No difference in the conversion rate between control and treatment group.

WHILE

Alternative Hypothesis (Ha) = There is difference in the conversion rate between control and treatment group.

The calculated z-score was -3.86429177, and the corresponding p-value was determined to be 0.000111412 which is less than 0.05 (P-value<0.05).

The obtained p Value <0.05 is statistically significant & hence there is sufficient evidence to reject the Null Hypothesis (H0) indicating that the control and treatment groups did not share the same conversion rate.

Confidence Interval for Conversion Rate: With a 95% confidence interval, we determined that the difference in conversion rates fell between [0.00349, 0.01065] or [0.35%, 1.07%].

# Hypothesis Test for Average Amount Spent (A & B groups)

To test whether there is a difference in the average amount spent per user between the Control (A) and Treatment (B) groups, a two-sample t-test for the difference in means was carried out.

Null Hypothesis (H0) = No difference in the average amount spent between the control and treatment groups.

WHILE

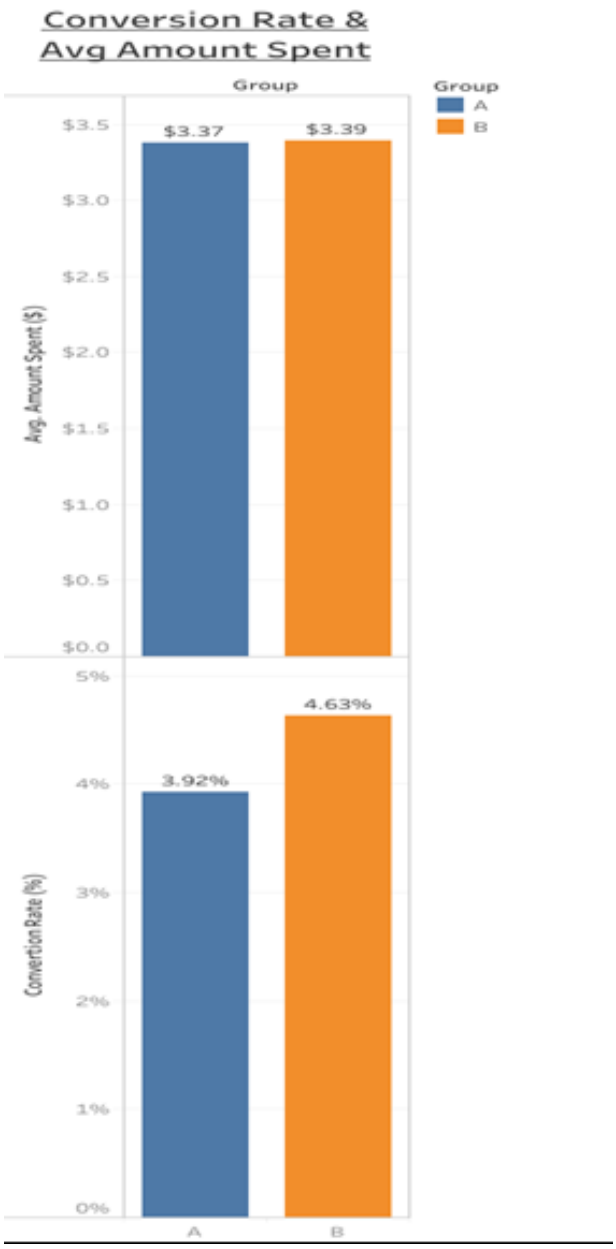
Alternative Hypothesis (Ha) = There is a difference in the average amount between the control and treatment groups.

To determine whether to reject or fail to reject the Null Hypothesis, a p-value was calculated.

The resulting P-value is 0.94386 > 0.05 & hence it's not statistically significant therefore we fail to reject the null hypothesis. In other words, there is insufficient evidence to reject the Null Hypothesis, suggesting that there is no significant difference in the average amount spent between the Control(A) and Treatment (B) groups.

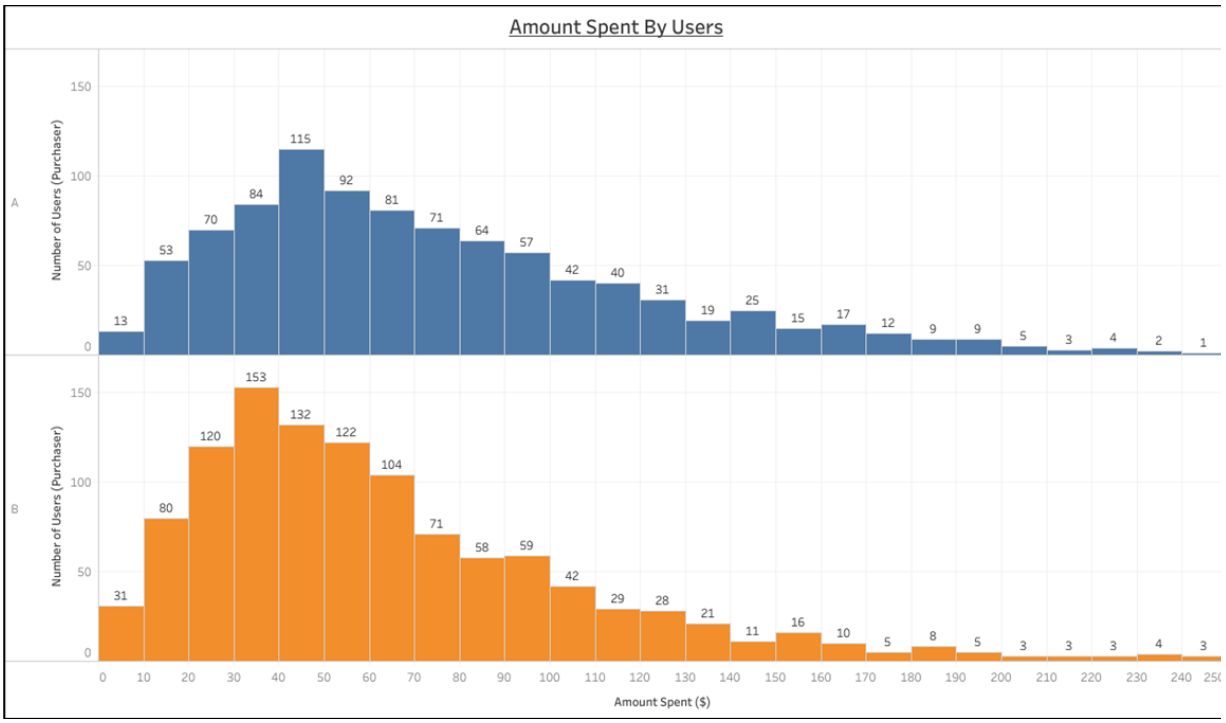
Confidence Interval for Average Spending: With a 95% confidence interval, we established that the difference in average spending per user lies in confidence interval [-0.4387, 0.4714].

# Conversion Rate & Average Amount Spent between the Test Groups



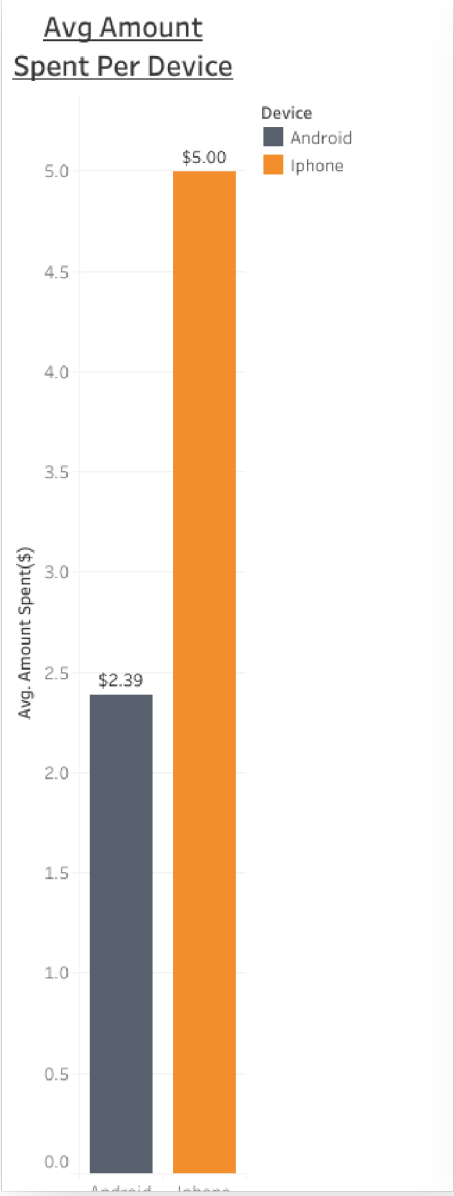

- The chart illustrates Avg Spending & Conversion Rate between the Test groups.
- ·For the Control (A) group average spending is $3.37, while for the Treatment (B) group is $3.39.
- ·The conversion rate for the Control group stands at 3.92%, whereas for the treatment group Its significantly higher at 4.63%.
- · Though there is a significant difference in the conversion rate, this difference did not led to a significant increase in the total revenue or average amount spent per user.

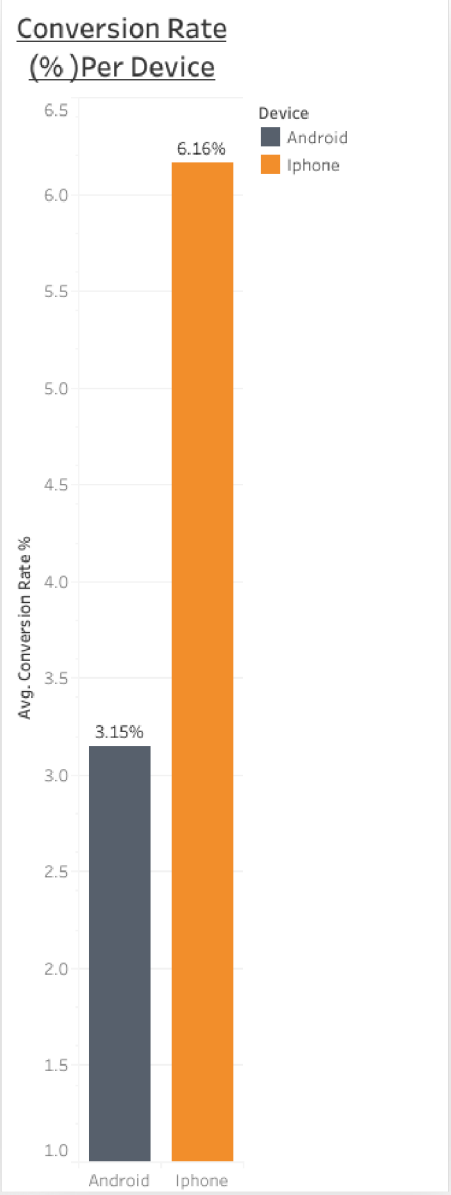
# Distribution of the amount spent per user for each group




- To conduct an analysis of expenditure between the groups, we omitted outliers where spending exceeded $250. This decision was made due to the less than 1% occurrence of users spending above $250 (considered outliers). Their spending ranged from $250 to $2000, potentially impacting the analysis.
- Within the Control Group (A), a significant number of users (115) predominantly spent within the $40-$50 range, while in the Treatment Group (B), 153 users spent within the $30-$40 range.
- Both groups exhibit a similar distribution pattern, with the majority of users favoring expenditures up to $100. This observation aids in comprehending user spending behaviors

# Comparison of Avg Amount Spent & Conversion Rate By User's Device

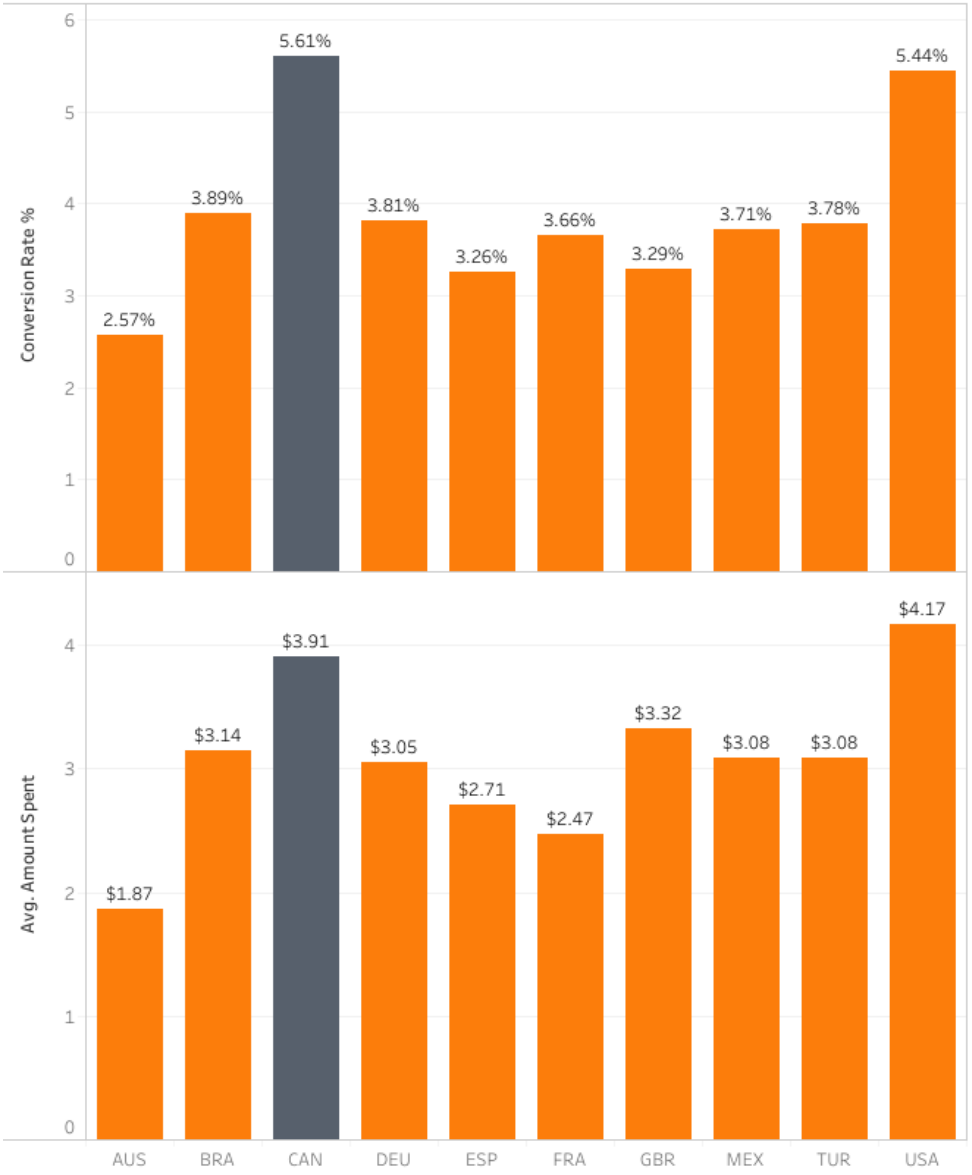



iOS users showed the highest conversion rates, approximately twice that of Android users. The average amount spent by iOS users is notably higher in comparison.

**(Note: unknown devices have been excluded from this study)**

# Analysis of Conversion Rate & Avg Amount Spent By User's Country

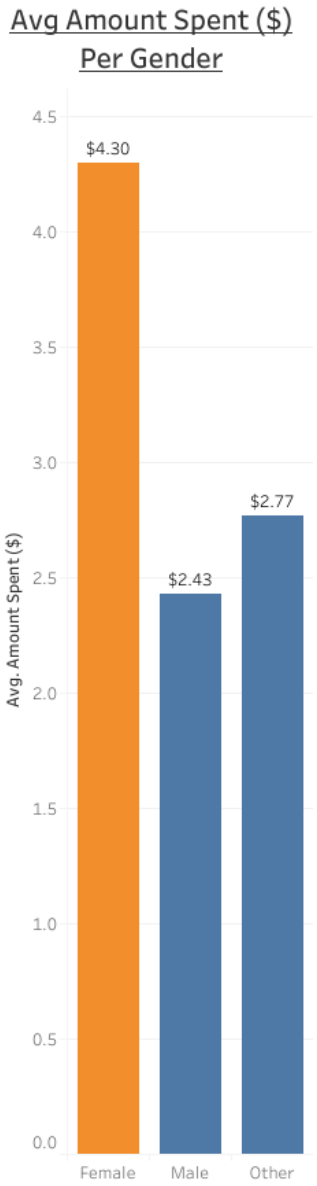


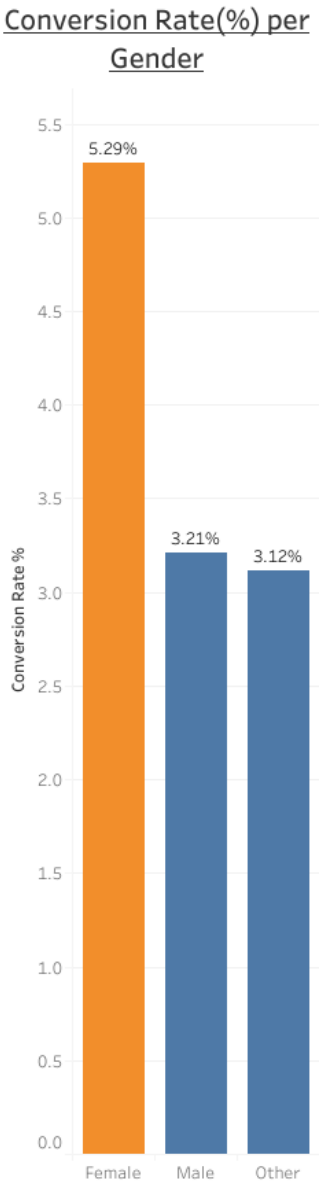

- United States has reported the highest Avg. amount spent per user $ 4.17 while Australia has the lowest at $1.87.
- The highest conversion rate is found in Canada at 5.61%, while the lowest is in Australia at 2.57%.
- Australia has the lowest conversion rate and average amount spent: 2.5% and 1.86$ respectively.

Meanwhile, Mexico, Brazil, Turkey, and Germany show nearly identical conversion rates and average spending figures.

The analysis by country highlights Canada with the highest conversion rate of 5.61% and ranking second, after the USA, in average spending per customer at $3.91. Despite Canada having one of the lowest populations among our targeted countries, leading to relatively fewer customers from this region, the noteworthy conversion rate suggests the potential value in focusing more on the Canadian market

# Comparison of Avg Amount Spent & Conversion Rate By User's Device



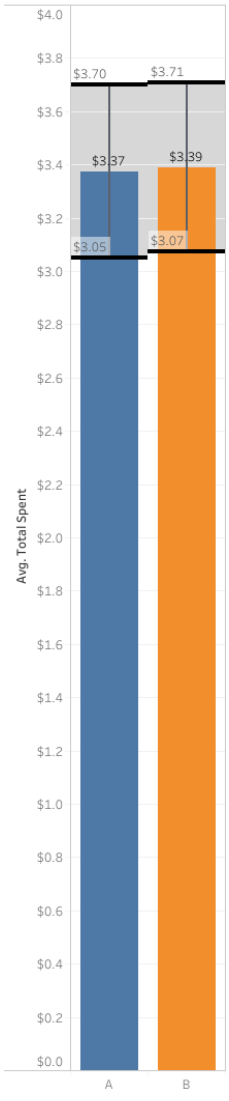




- Women are leading in both metrics: a 5.29% Conversion Rate and an average spend of $4.30 per customer. These figures notably surpass our overall averages, establishing women as our most valuable demographic in terms of gender.

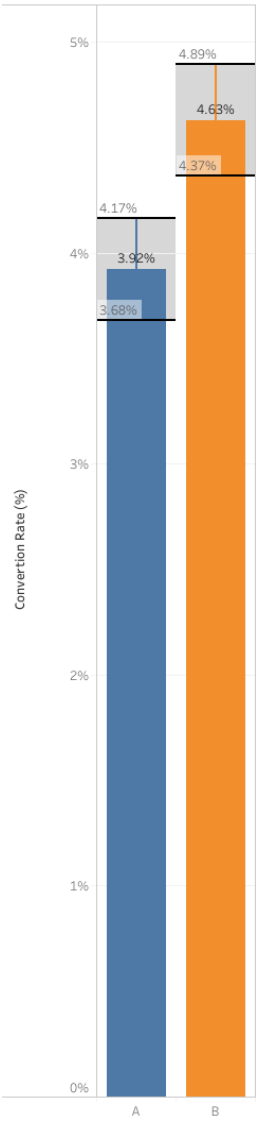
**(Note: unknown gender category has been excluded from this study)**

# Confidence Interval

Confidence Interval for Avg Amount Spent (USD)



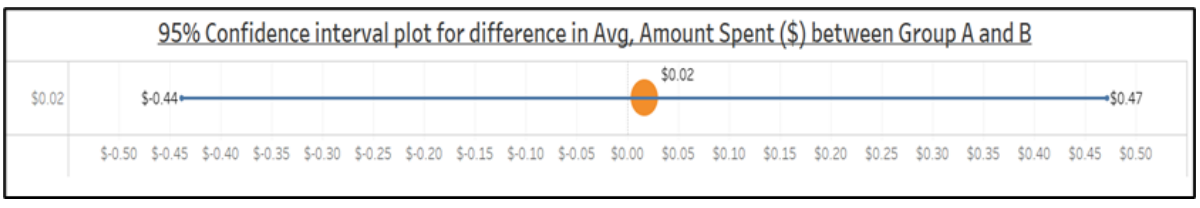

Confidence Interval for Conversion Rate (%)

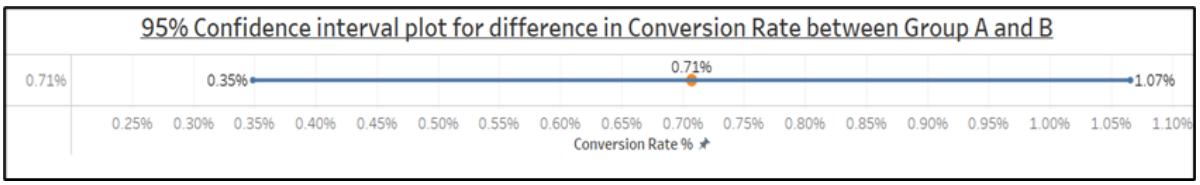



The calculated confidence intervals, at a 95% confidence level, yield the following conclusions:

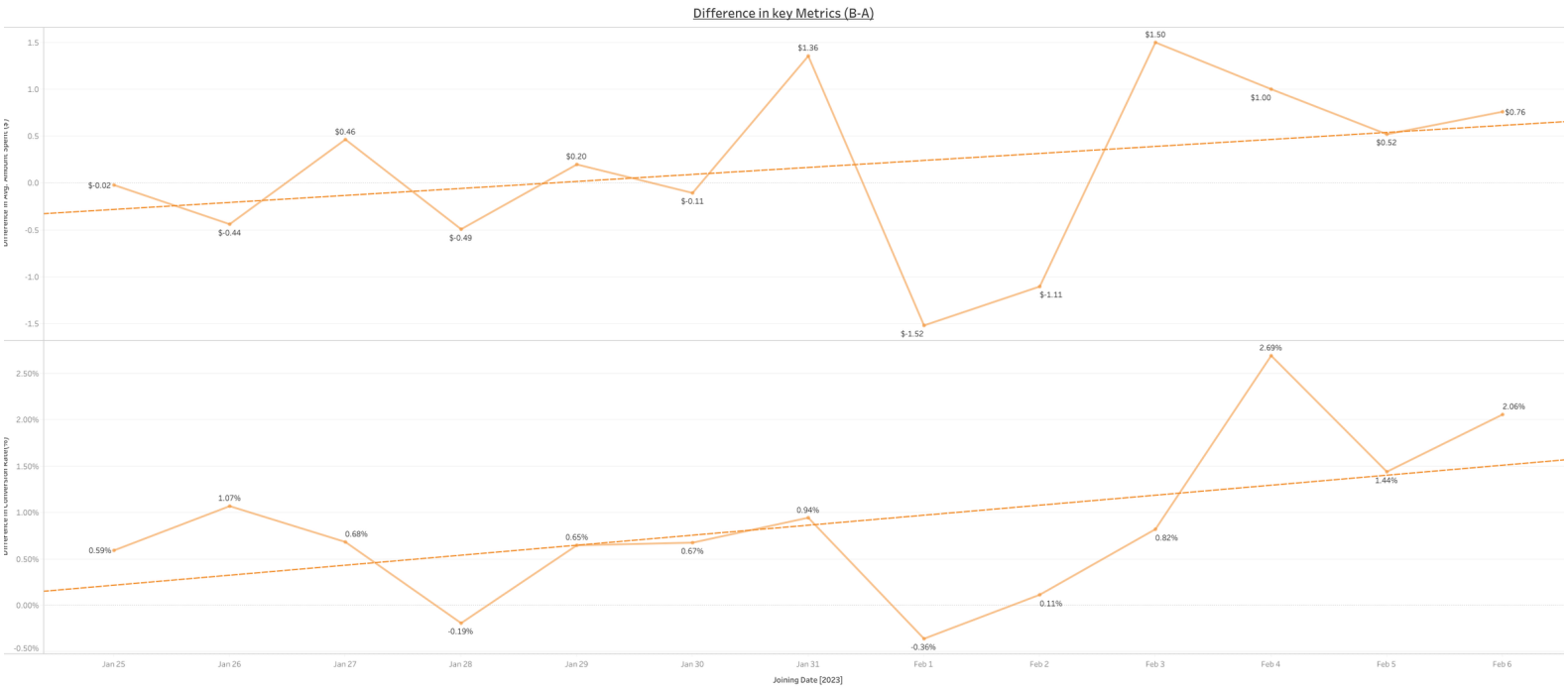
- For the conversion rate, we can be confident that 95% of values will range between 3.68% and 4.17% for the control group and between 4.37% and 4.89% for the test group.
- Regarding the average Amount spent per user, we can also be confident that 95% of values will fall between $3.05 and $3.70 for our control group. Similarly, for those exposed to the banner, values will fall between $3.07 and $3.71.

- These results reinforce the observation that we are not witnessing significant revenue growth, as most values overlap between the control and test groups.



# Novelty Effect




The line chart indicates that the positive outcomes in conversion rate and average spend values aren't due to a temporary novelty effect, as these results do not decrease over time.

Instead, they show a sustained positive trend, notably in the conversion rate, which exhibits a clear upward trajectory. While the average spending encountered some initial challenges, it displayed improved results towards the conclusion of the test period.

However, it's crucial to emphasize that a significantly larger sample size would provide more accurate and reliable insights for our analysis.

# Power Analysis : Conversion Rate

A power analysis helps us understand the necessary sample size in order to achieve our desired minimum detectable effect and statistical power. If we find that we did not have enough sample size for our test to be sufficiently sensitive, we could recommend that we run the test again at a larger scale.

Parameters Used:

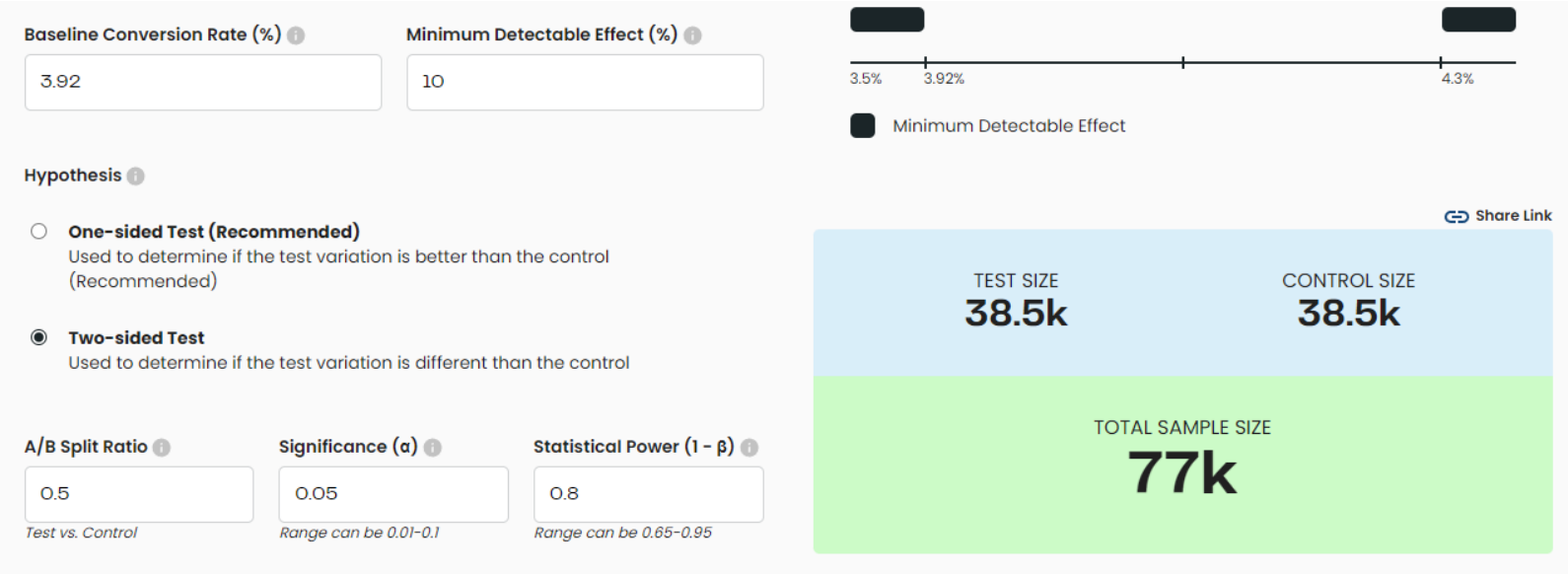
- Baseline Conversion Rate of Group A: 3.92%
- Minimum Detectable Effect (MDE): 10%
- Statistical Power (1-β): 0.80
- Significance Level (α): 0.05

Results of Power Analysis:

The power analysis determined that a sample size of 38.5 K participants is needed for each group (Group A and Group B) for a total of 77 k participants to achieve a statistical power of 0.80 at a significance level of 0.05.

This sample size calculation ensures that the test has an 80% probability of detecting a minimum detectable effect of 10% if it truly exists

Conversion Rate: Sample size needed for each Group = 38500



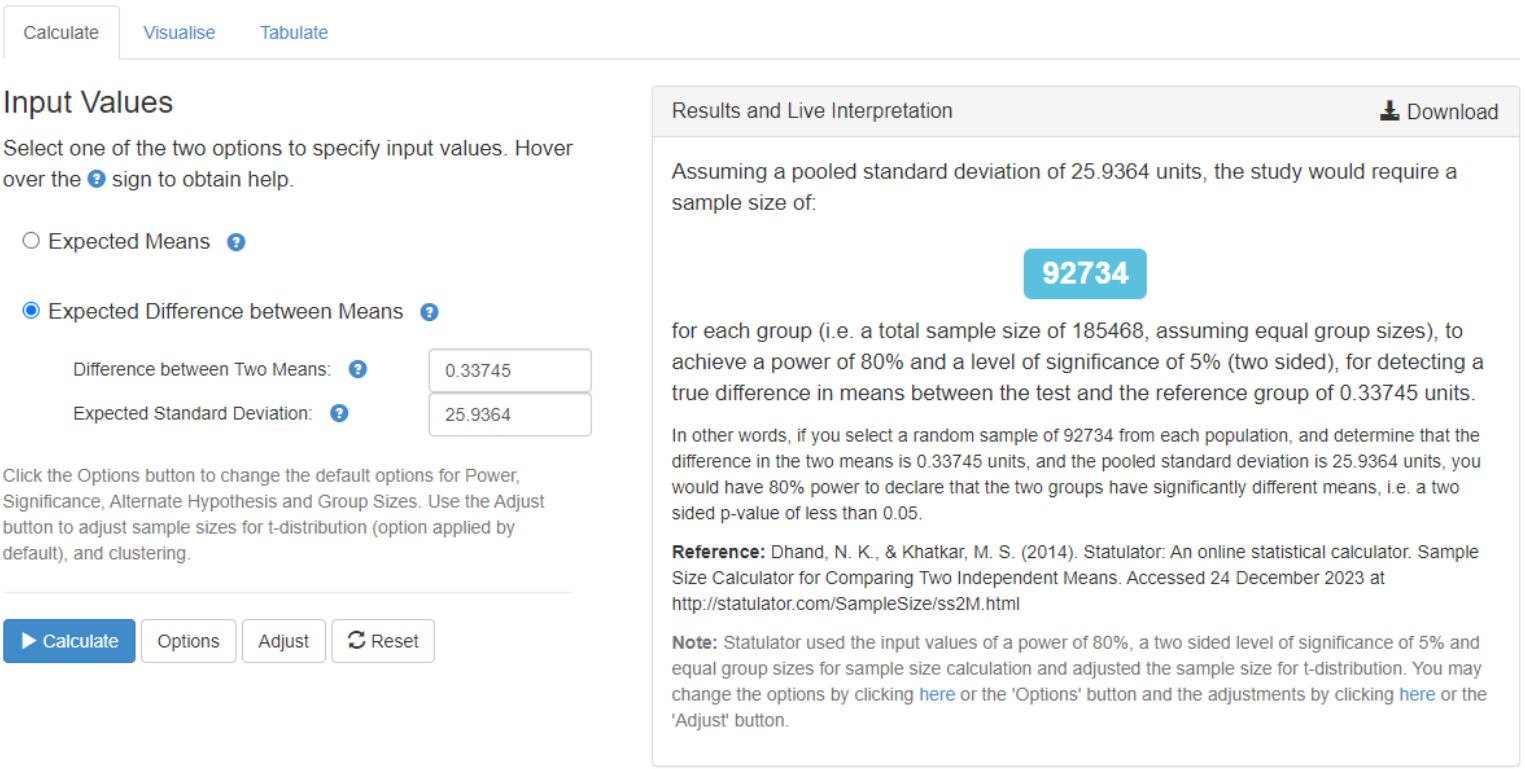
# Power Analysis : Avg. Amount Spend

The difference between the means is indicated by 10% of the mean of the control group (A), which represents the minimum difference we anticipate observing. The Minimum Detectable Effect (MDE) is relevant in this context. It's important to note that the Expected Standard Deviation aligns with that of the control group as well.

Parameters Used:

- Difference between Two Means: 0.33745( 10% of the mean of the control group (A))
- Expected Standard Deviation: 25.9364

Avg Amount Spend: Sample size needed for each Group = 92734



# Recommendation

After conducting a comprehensive analysis of the A/B test, it is recommended that we should refrain from launching the new banner on the mobile website. Even though there was a favorable response in conversions for the test group, our sample size was inadequate for this test, and the rise in revenue was minimal.

This could be attributed to GloBox's primary focus on selling high-end decor and boutique fashion items, which tend to be pricier compared to food and drink items

Additionally, It is recommended to continue retesting to evaluate and enhance the banner/user experience or explore alternative uses for that space aiming to achieve better results.

For further testing, it is advisable to use a larger sample size. (Avg Amount Spend: Sample size needed for each Group = 92734 & Conversion Rate: Sample size needed for each Group = 38500).

Our significantly higher conversion rate, exceeding our 10% Minimum Detectable Effect (MDE), indicates that there's potential in utilizing this space. However, further experimentation is required to precisely determine how and what to implement.

Given the lower-than-desired revenue outcomes from our A/B test, I'm hesitant about extending the test duration solely for additional data collection unless modifications are made to the banner or user experience. Customers who purchased lower-priced items during the test phase could potentially become future buyers of our high-end products. Additionally, promoting food and drink items as add-ons to cart purchases while utilizing the banner space differently is an option worth considering.

This situation holds promise, and considering the investments made, it's imperative to find a solution that generates better returns.

# APPENDIX

- Conversion Rate & Average Amount Spent between the Test Groups: https://public.tableau.com/views/FinalMS/ConversionRateAvgAmountSpent?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Distribution of the amount spent per user for each group : https://public.tableau.com/views/FinalMS/MoneyDistribution?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Comparison of Avg Amount Spent By User's Device: https://public.tableau.com/views/FinalMS/AvgAmountSpentbyDevice?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Comparison of Conversion Rate By User's Device: https://public.tableau.com/views/FinalMS/ConRatebyDevice?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Analysis of Conversion Rate & Avg Amount Spent By User's Country: https://public.tableau.com/views/FinalMS/ConRatebyCountry2?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Comparison of Avg Amount Spent & Conversion Rate By User's Device :
  - https://public.tableau.com/views/FinalMS/AvgAmountSpentbygender?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
  - https://public.tableau.com/views/FinalMS/ConRatebyGender?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Confidence Interval
  - https://public.tableau.com/views/FinalMS/CIAvgAmountSpent2?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
  - https://public.tableau.com/views/FinalMS/CIConversionRate2?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
  - https://public.tableau.com/views/FinalMS/95Confidenceintervalplotfordifferenceinaveragepurchasebetween GroupAandB?:language=en-
  - US&publish=yes&:display_count=n&:origin=viz_share_link
    https://public.tableau.com/views/FinalMS/95Confidenceintervalplotfordifferenceinconversionratebetween GroupAandB?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link
- Novelty Effect : https://public.tableau.com/views/FinalMS/DifferenceInKeyMetrics?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

# APPENDIX

## SQL Code

```
SELECT
u.id,
u.country, u.gender, g.device, g.group,

  CASE WHEN SUM(COALESCE(a.spent, 0)) > 0
THEN 1 ELSE 0 END as spent,
  SUM(COALESCE(a.spent, 0)) as total_spent

FROM users u

INNER JOIN
groups g ON (g.uid = u.id)

LEFT JOIN
activity a ON (a.uid = u.id AND a.device = g.device)

GROUP BY
1,2,3,4,5
```