

Suggestion for Improving GloBox's current webpage based on the Experiment

Author: Rutuja Deshmukh

Contents

1. Context:.....	3
2. Executive Summary:.....	3
3. Dataset:.....	3
4. Insights on the Experiment:	4
5. Statistical Analysis:.....	5
5.1 Hypothesis Testing Result:	5
5.2 Confidence Interval:	6
6. Additional findings of the Experiment:	7
7. Power Analysis:	10
7.1 Conversion Rate:.....	10
7.2 Average Amount Spent:.....	10
8. Novelty Effects:	11
8.1 Average Amount Spent	11
8.2 Conversion Rate:	11
8.3 Average Amount Spent by Converted Users:	12
8.4 Cumulative Amount Spent for Converted Users:	12
8.5 Number of Converted Users per Date:	13
9. Recommendation:.....	13
10. Appendix:	14

1. Context:

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, its 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. To ensure effective promotion and increase awareness of this product category, the Growth team has devised an A/B test strategy.

The experiment was conducted for 13 days, showing part of customer the existing home webpage and part of customer the webpage with food and drinks category banner. The data was collected through the experiment giving the information of users who purchased the items and the amount spent on it.

The report outlines the methodologies used to gather the insights for our test metrics and, to increase the revenue whether the improvement in existing home webpage will require or not.

2. Executive Summary:

An A/B Test was conducted between 25th January 2023 - 6th February 2023 (13 days), to measure conversion rates and average spend with a total of 48,943 users, both conversion rate (+18%) and average spend (+0.6%) reported increases as compared to control group. The average amount spent for females seems to have decreased in treatment group while the conversion rate is increased from 5.1% to 5.4%. Male Conversion Rate and Average spend increased by 46% and 15.5% respectively than the control group.

The conversion rate for treatment group is increased by 25% in Android device, whereas it is increased only 10% in IOS device. Similarly, average spent in Treatment group is increased by 6% in Android device but the average spent decreased by 2.95% in IOS device. The statistical analysis of the data indicates a noteworthy rise in the rate of conversion; however, there is no significant increase observed in the average amount of money spent. The statistical evidence from the A/B test is not strong enough to support proceeding with the release of the current banner design. Also, the power analysis indicates that the available total sample size of 48,943 participants was insufficient to achieve the desired statistical power of 80%. So, I recommend to re-run the test with larger sample size in order to achieve the desired statistical power.

3. Dataset:

The dataset comes from the GloBox relational database.

The data is covering 48943 distinct users from 10 known countries with different genders and using different devices. The users are divided into 2 groups during the activity.

The fields mentioned in the dataset are:

- id: the user ID
- country: ISO 3166 alpha-3 country code
- gender: the user's gender
- group: the user's test group
 - A = Control group (Existing landing page)
 - B = Treatment Group (Landing page with Food and Drink Banner)
- conv_table: the user who make a purchase
 - 1 = the user who make one or more purchases
 - 0 = the user who did not make any purchase
- device: the device the user visited the page on

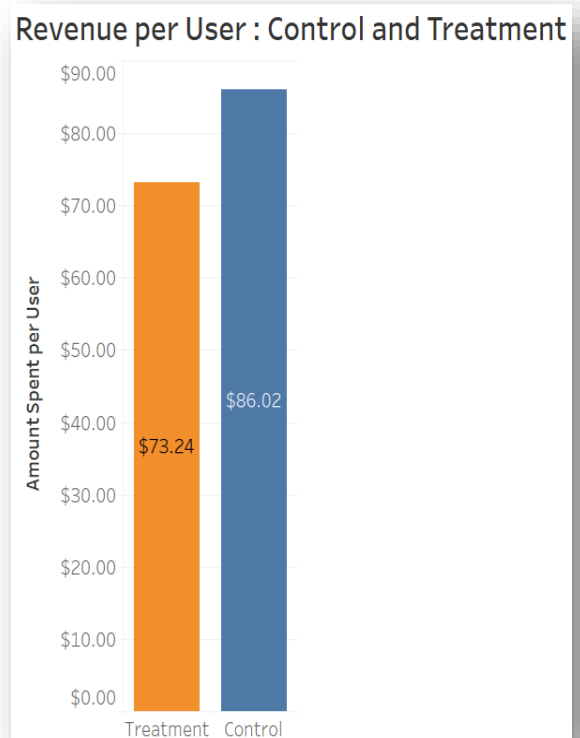
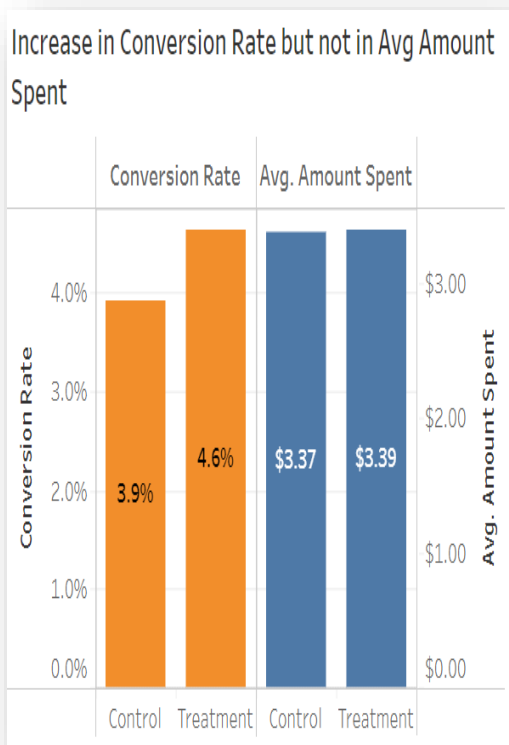
- total_spent: the purchase amount in USD
- Metric:
 - Average amount spent - Mean of total_spent
 - Conversion rate - Percent of the users who make one or more purchases

4. Insights on the Experiment:

This experiment will focus on improvement that might happen after the new food and drink banner launch on the existing webpage. To understand the changes in both metrics (mean and percent), I have performed statistical analysis and hypothesis testing.

Tool used for the analysis is Tableau and Google Spreadsheets, see appendix for tableau files.

The results show that CR increased from 3.9% for control group to 4.6% for treatment which is a 18% increase while the average amount spend has a relatively small difference of 0.6%.



As per the revenue per user visualization (2nd graph above), total amount spent per user is higher in Control group which is \$86.02 than the treatment group which is \$73.24. Since only graphical representation is not enough to make the inference, I performed below statistical analysis to come up with the significant conclusion.

5. Statistical Analysis:

The primary goal is to observe the statistically significant and meaningful increase in the user conversion rate and to determine whether new banner influences the average spent amount per user within the food and drink category compared to the control group.

Key finding of A/B test:

Metric	Group A: Control	Group B: Treatment	Total	p Value
No. of Users	24343	24600	48943	N/A
Conversion Rate	3.9%	4.6%	4.28%	0.00011
Average Amount Spent	\$3.37	\$3.39	\$3.38	0.944

The above data shows that the Treatment group has a higher conversion rate than the Control group which is almost 18% higher than it. However, average amount spent is quite similar for both groups, resulting in a negligible difference of 0.6%.

Further analysis and calculation were conducted in the form of Hypothesis Testing and Confidence Intervals to determine the statistical significance of the A/B Test Data.

5.1 Hypothesis Testing Result:

Hypothesis test for proportion – Conversion Rate:

Null hypothesis (H₀): There is no difference between conversion rates of Group A (control group) and Group B (treatment group).

Null hypothesis (H₀): $\mu_A = \mu_B$

Alternative hypothesis (H_a): There is a difference in the conversion rate between Group A and Group B.

Alternative hypothesis (H_a): $\mu_A \neq \mu_B$

Significance level = 0.05 | 5%

See the spreadsheet for hypothesis calculations.

Conclusion:

With $p = 0.00011$, we reject the null hypothesis that there is no difference between conversion rates of Group A (control group) and Group B (treatment group) in favor of the alternate hypothesis that there is a difference in the conversion rate between Group A and Group B.

Hypothesis test for mean – Average amount spent:

Null hypothesis (H₀): There is no difference between average amount spent per user of Group A (control group) and Group B (treatment group).

Null hypothesis (H₀): $\mu_A = \mu_B$

Alternative hypothesis (Ha): There is a difference in the average amount spent per user between Group A and Group B.

Alternative hypothesis (Ha): $\mu_A \neq \mu_B$

Significance level = 0.05 | 5%

See the spreadsheet for hypothesis calculations.

Conclusion:

With $p = 0.944$, we fail to reject the null hypothesis that there is no difference between average amount spent per user of Group A (control group) and Group B (treatment group) in favor of the alternate hypothesis that there is a difference in the average amount spent per user between Group A and Group B.

5.2 Confidence Interval:

Metric	Lower CI	Point	Upper CI
Conversion Rate	0.35%	0.71%	1.07%
Average Amount Spent	-0.4386	0.01635	0.4713

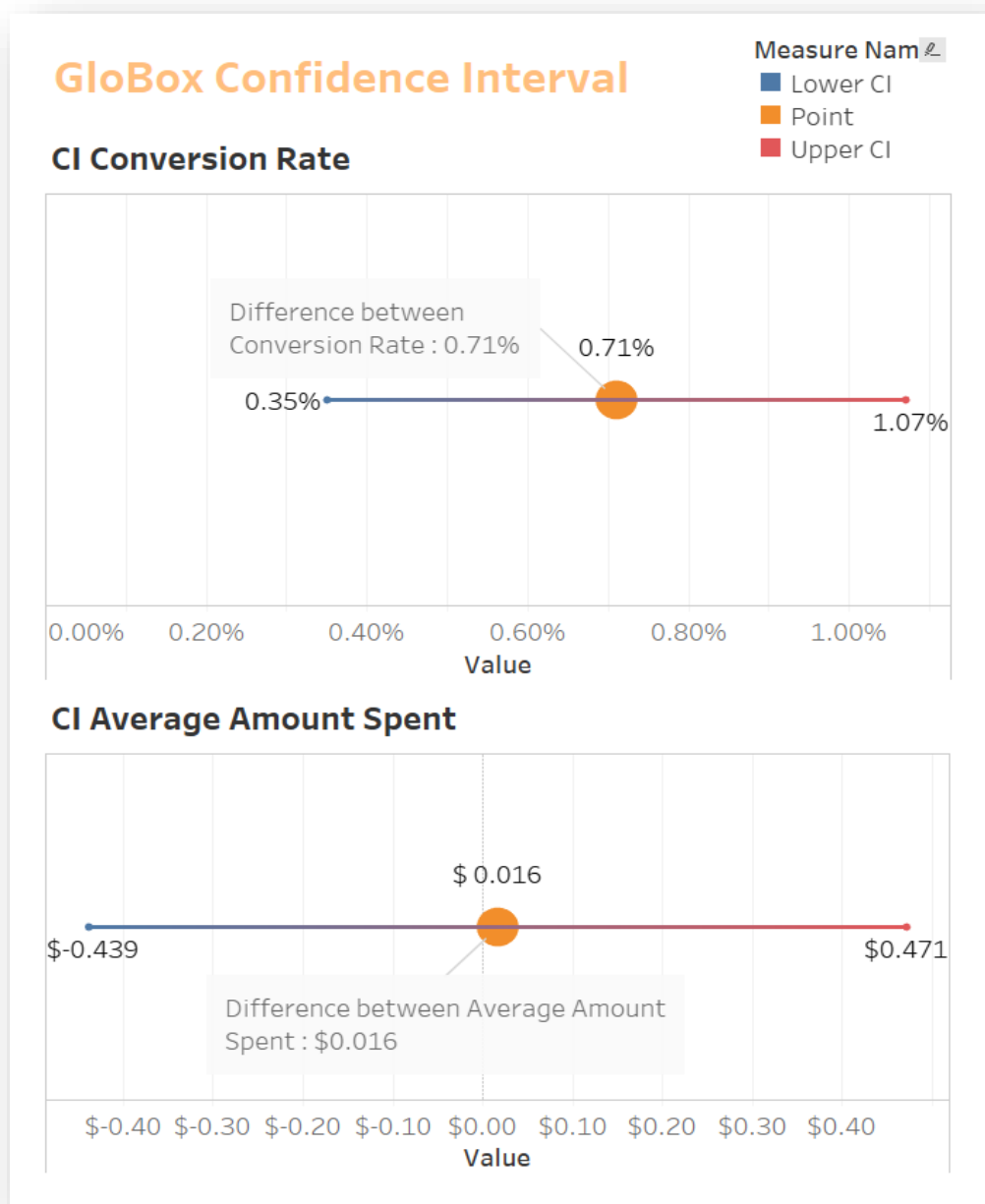
Average Amount Spent:

For the Control Group, the average amount spent is \$3.37 and the Treatment Group, average amount spent is \$3.39, with a confidence interval ranging from -\$0.0439 to \$0.0471. The 95% confidence interval means that if we were to repeat the experiment 100 times, we would expect the true difference to be within -\$0.0439 to \$0.0471. The data does not provide sufficient evidence to suggest that the treatment (Group B) had a significant impact on the average amount spent compared to the control group (Group A).

Conversion Rate:

For the Control Group, the conversion rate is 3.92% and for the Treatment Group, the conversion rate is 4.63%, with a confidence interval ranging from 0.35% to 1.07%. The 95% confidence interval means that if we were to repeat the experiment 100 times, we would expect the true difference to be within 0.35% to 0.71%. The data does provide sufficient evidence to suggest that the treatment (Group B) had a significant impact on the conversion rate compared to the control group (Group A).

See the spreadsheet for confidence interval calculations.



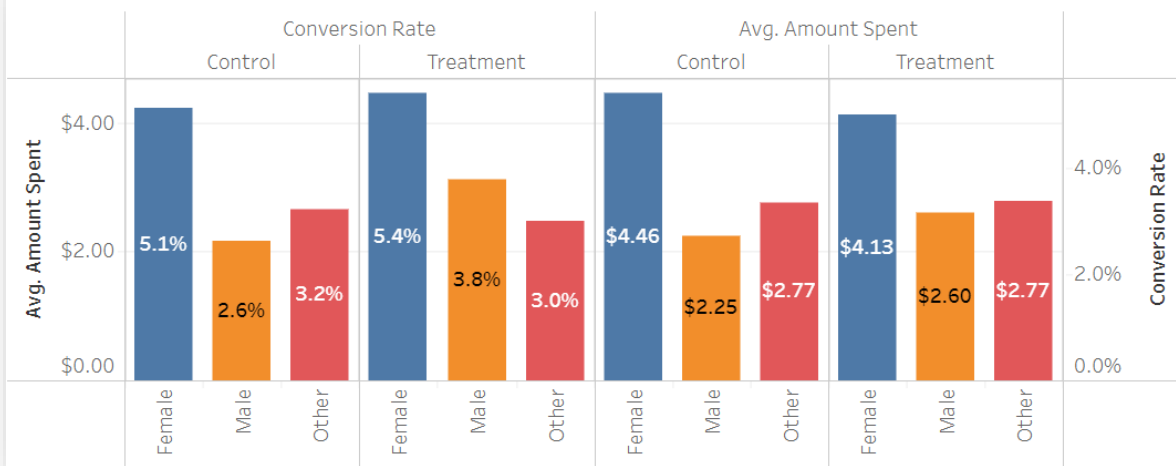
6. Additional findings of the Experiment:

Overall average amount spent and conversion rate is higher in female users in both the groups other than male and other gender. But the average amount spent for females seems to decreased in treatment group while the conversion rate is increased from 5.1% to 5.4%. On the other hand, male conversion rate saw a significant increase in both metrices of Treatment group.

The below visualization shows the insights of conversion rate and average amount spent by the different users of different genders.

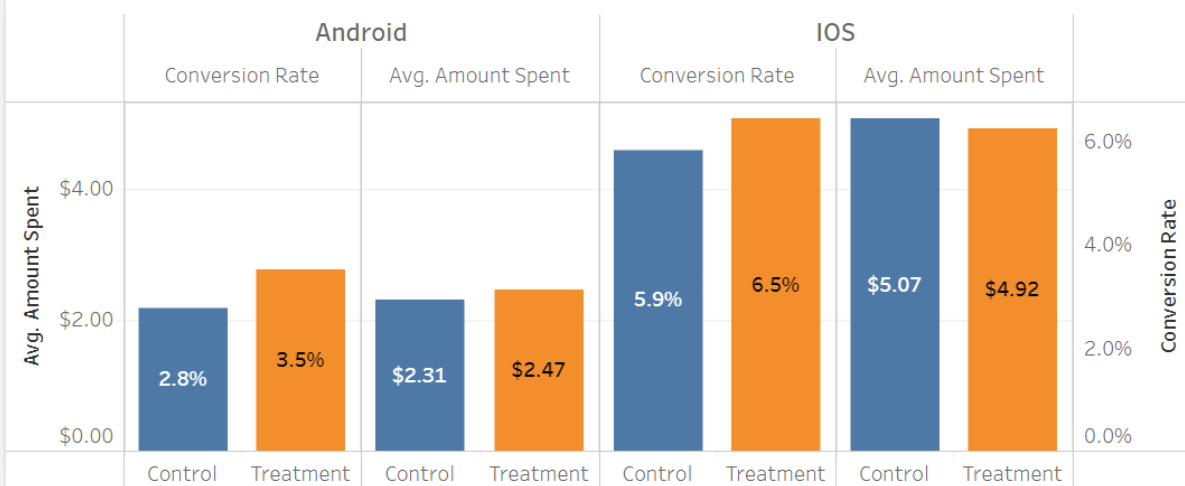
Note: Some users choose not to disclose the gender, device or country information. Average amount spent and conversion rate for such users is not included in any of the visualization.

Male Users had a Better Increase in Conversion Rate



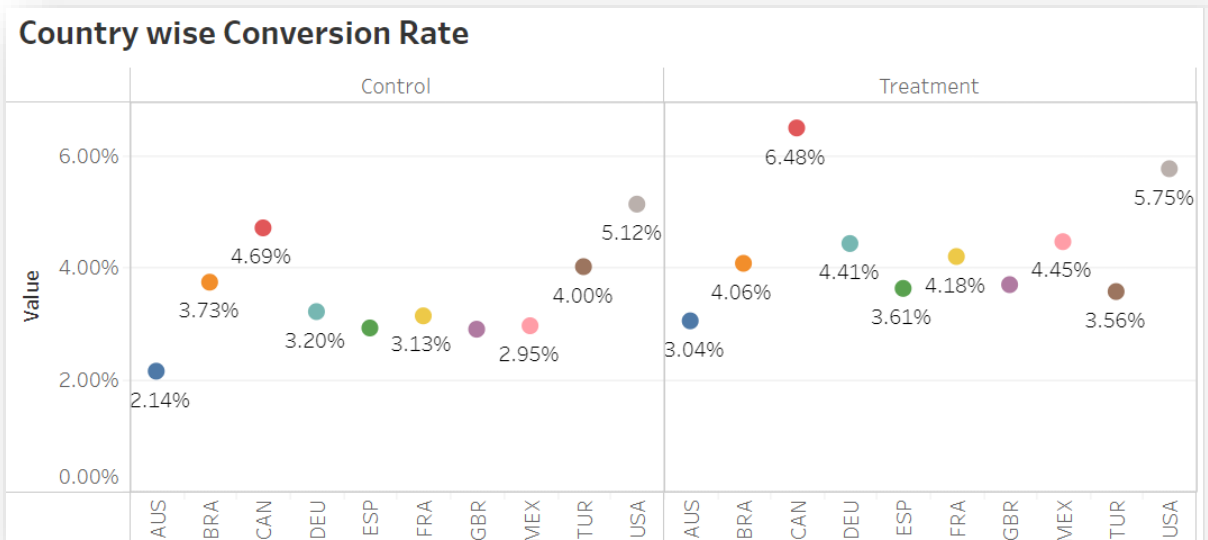
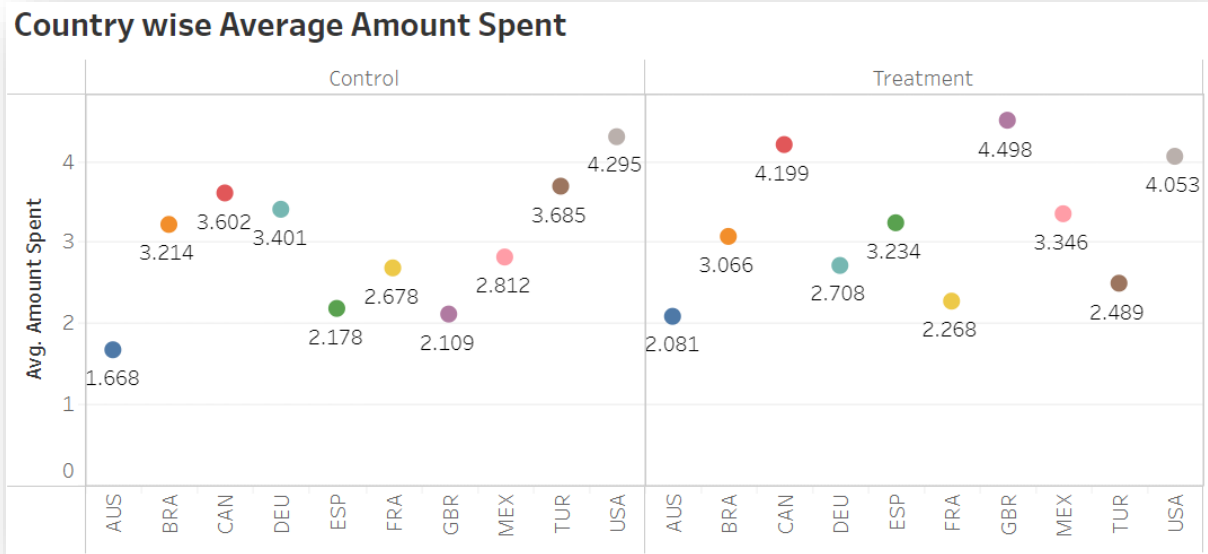
Taking more granular approach into the data, and checking the test metrics (Mean and Percent) with iOS and Android devices used by users. Even though Android has a higher number of users (30289 in total, including both groups) compared to IOS (18360 in total, including both groups), the conversion rate for IOS devices is higher at 6.16% compared to Android's conversion rate of 3.15%. In Android devices, the conversion rate for the treatment group has seen a 25% increase, whereas in IOS devices, it has only increased by 10%. In terms of average spending, the treatment group in Android devices has experienced a 6% increase, while in IOS devices, the average spending has decreased by 2.95%.

iOS Experienced a more Significant Improvement in Conversion Rate



Looking at the below graph, average amount spent stands higher for countries like Canada, USA, UK(GBR). Out of the 12 countries analysed, 6 countries experienced an increase in average spend in the Treatment group compared to the Control group, resulting in a net increase of 0.6% across all Countries. UK shows highest growth of 113%, indicating the positive impact on average spend and also 27% increase in the conversion rate.

Canada shows highest increase of 38%, indicating the positive impact on conversion rate and also 16.6% increase in the average spend.



The observation of conversion rate suggest that the treatment group had a positive impact of the new banner in multiple countries resulting in increase of conversion rate except for Turkey, showing 11% decrease in conversion rate as compared to control group.

Result:

According to our analysis, we found that the current food and drink banner feature does not show substantial enhancements in the success metrics we aim for, specifically in terms of how much users spend on average. Although there was a notable increase in the conversion rate, it was insufficient to justify proceeding with launching the existing banner design. Nevertheless, despite the lack of

significant improvements, we did notice some encouraging signs that modifying the banner experience could potentially lead to better outcomes in the future.

7. Power Analysis:

7.1 Conversion Rate:

A power analysis helps us understand the necessary sample size in order to achieve our desired minimum detectable effect and statistical power. Baseline conversion rate is determined from control group, which is 3.9% and MDE of 10% was chosen.

Available total Sample Size: 48,943

Significance (α): 0.05

A/B split ratio: 0.5

Statistical power($1-\beta$): 0.8

Calculator: Statsig

On a one-sided test, the power analysis indicates that a sample size of 60,900 participants would be necessary to achieve a statistical power of 80%. Whereas, on a two-sided test, the power analysis indicates that a sample size of 77,400 participants would be necessary to achieve a statistical power of 80%. However, considering the available total sample size of 48,943, it is evident that the test fails short of the desired statistical power in both the scenarios.

7.2 Average Amount Spent:

For this analysis, difference of mean is 0.02 and expected standard deviation of 25 was chosen.

Available total Sample Size: 48,943

Significance (α): 0.05

A/B split ratio: 0.5

Statistical power($1-\beta$): 0.8

Calculator: Statulator

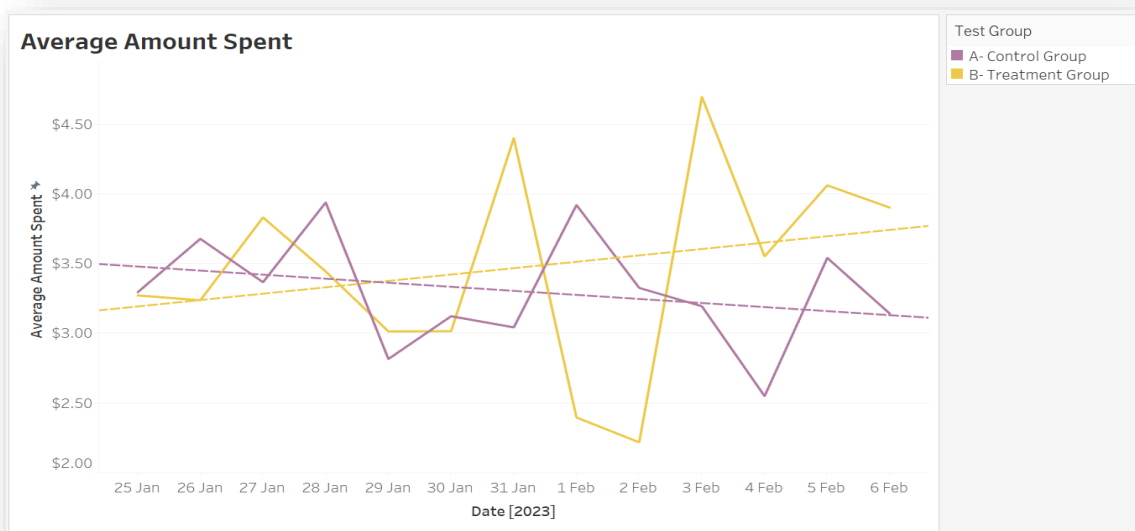
On a one-sided test, the power analysis indicates that a sample size of 19,320,492 participants for each group would be necessary to achieve a statistical power of 80% for detecting a true difference in means between the test and the reference group of 0.02 units. Whereas, on a two-sided test, the power analysis indicates that a sample size of 24,527,750 participants for each group would be necessary to achieve a statistical power of 80% for detecting a true difference in means between the test and the reference group of 0.02 units. However, considering the available total sample size of 48,943, it is evident that the test fails short of the desired statistical power in both the scenarios.

8. Novelty Effects:

In order to examine the existence of a Novelty Effect, I have examined 5 metrics: Average Amount Spent, Conversion Rate, Average Amount Spent by Converted Users, Cumulative Spend by Converted Users, and Time to Purchase to determine if there is a novelty effect.

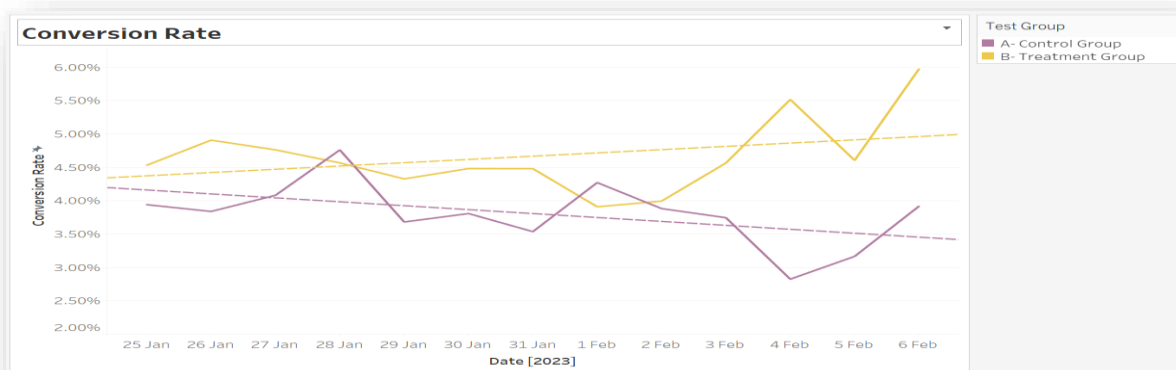
8.1 Average Amount Spent

Started with analysing the average spending of all users in both groups over time. Despite some fluctuations, it is essential to look at broader trends rather than fixating on individual days. By visualizing the data in a graph, we gained a better understanding of patterns. The graph demonstrated that there was no consistent and clear indication of a significant difference in average spending and evidence of a novelty effect on this metric. While some fluctuations occurred, the differences between the Control and Treatment groups were not consistently significant enough to imply a novelty effect.



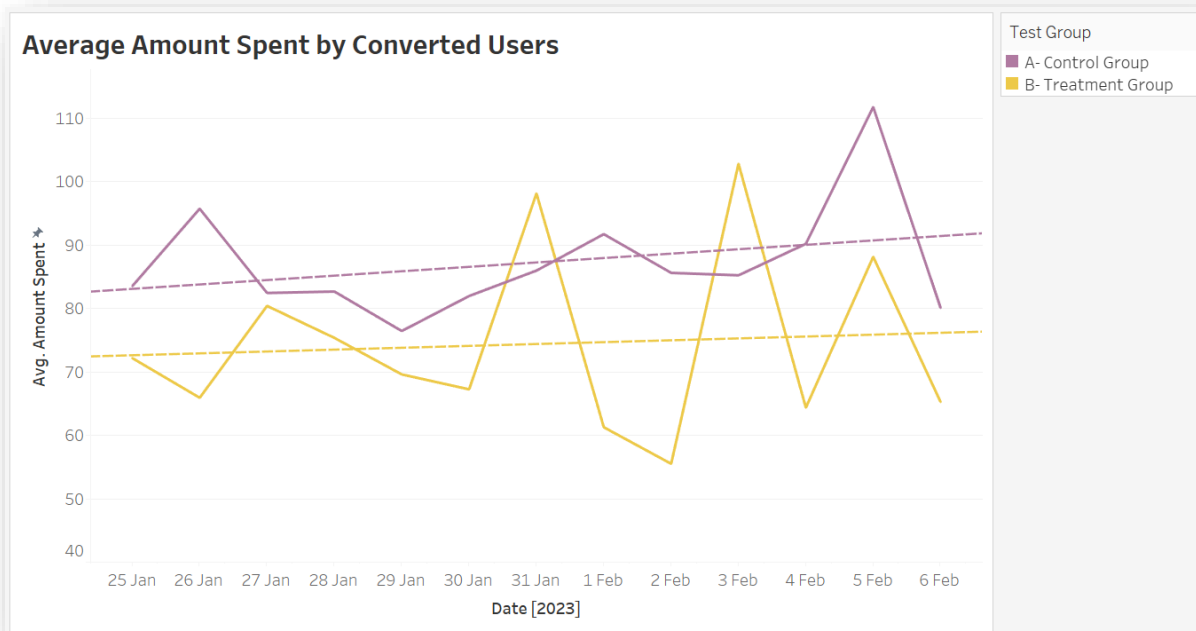
8.2 Conversion Rate:

The next metric analysed was the conversion rate for all users in both groups over time. Like the average amount spent, the conversion rates didn't show consistent and significant differences between the Control and Treatment groups. Despite minor fluctuations, there was no clear sign of a novelty effect significantly impacting the conversion rates.



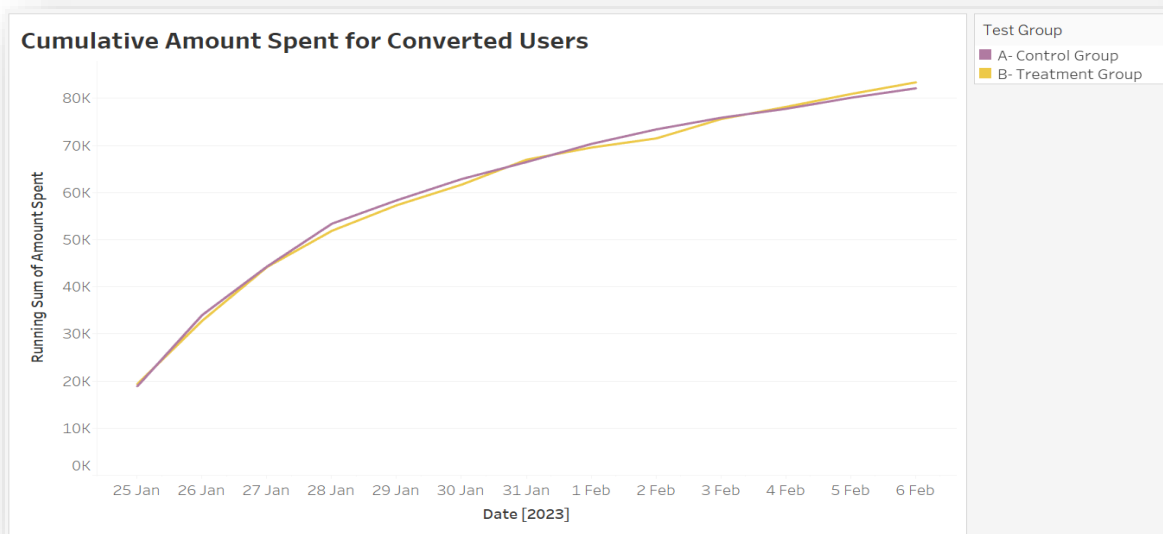
8.3 Average Amount Spent by Converted Users:

Our attention then shifted to analysing the average amount spent by users who were converted in both groups. The results consistently demonstrated that the Control group reported a higher average spend compared to the Treatment group. Although the accompanying graph displayed some variations, no discernible pattern emerged to indicate any novelty effect.



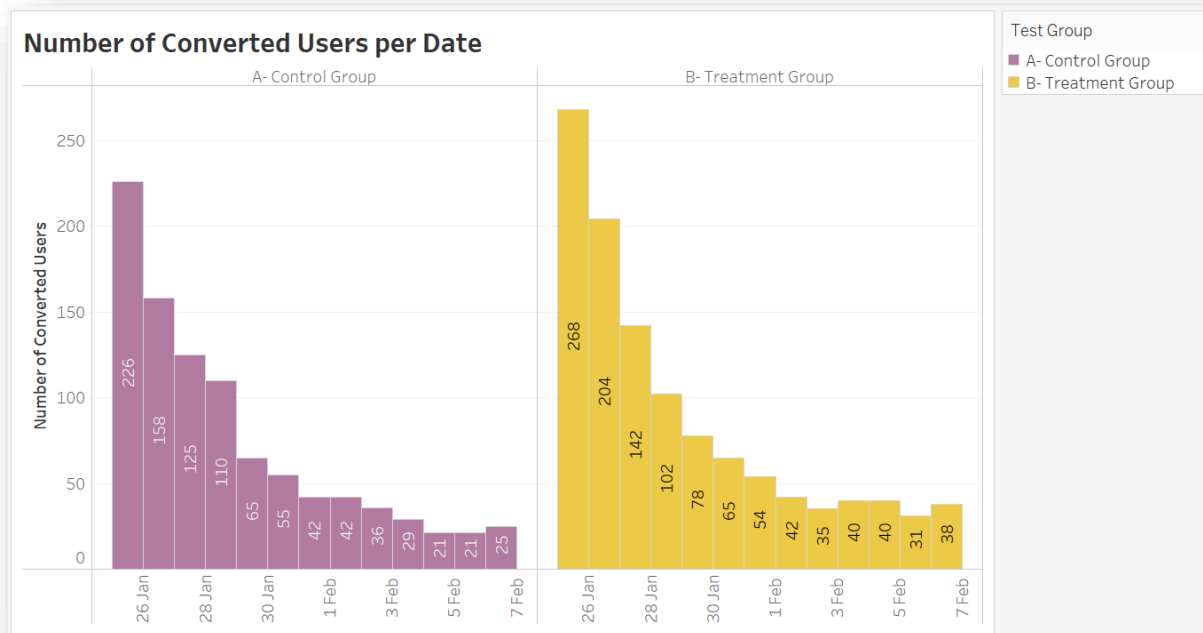
8.4 Cumulative Amount Spent for Converted Users:

I also analysed the cumulative amount spent by the users who converted in both the Control and Treatment groups. However, this metric did not reveal any significant differences, as it consistently tracked during the A/B test. Although the Treatment group reported a small difference of \$1,269, it was not substantial enough to suggest a novelty effect.



8.5 Number of Converted Users per Date:

Lastly, I examined the duration it took for users in both groups to complete a purchase after participating in the experiment. The findings failed to uncover any significant disparities in the time it took to make a purchase between the Control and Treatment groups. In fact, the Treatment group exhibited a heightened speed of purchase. While minor variations were observed, such as the number of users who made a purchase on the initial days or within a few days, these discrepancies did not suggest a novelty influence.



9. Recommendation:

Based on the result of A/B test, we can conclude that the new banner had a significant impact on the conversion rate, showing 18% increase compare to control group. However, when considering the average amount spent per user, it only shows \$0.02 increase which is not a significant difference to consider it.

The analysis of the five key metrics and the results of the novelty effect provide no clear evidence of a novelty effect in the Treatment group compared to the Control group during the A/B test. Therefore, it can be concluded that the novelty being tested had little to no significant impact on user behaviour throughout the testing duration. It is important to acknowledge that the lack of a novelty effect can be attributed to several factors. These include the specific test context, the characteristics of the user base, and the duration of the A/B test, which was conducted over a period of 13 days. In order to obtain a more thorough understanding of user behaviour and the potential impact of novelty on the product or experience, further research and testing may be necessary.

Based on the results of Power Analysis, it is evident that there is a noticeable difference in the conversion rate, while the change in revenue is minimal. However, the total sample size of 48,943 participants is inadequate to reach the desired statistical power of 80%. I recommend that we should conduct the same experiment again with a larger sample size in order to arrive on a conclusion whether the permanent changes on the home page can be done or not.

10. Appendix:

- SQL Code for extracting data set

```
with all_user as (  
  select distinct u.id as id,  
    u.country,  
    u.gender,  
    g.group as test_group,  
    g.device,  
    coalesce (a.spent, 0) as spent  
  from groups as g  
  
  left join activity as a  
    on a.uid = g.uid  
  left join users as u  
    on u.id = g.uid  
)  
  select id,  
    country,  
    gender,  
    test_group,  
    device,  
    case when spent > 0 then 1  
    else 0 end as conv_table,  
    sum(spent) as total_spent  
  from all_user  
  group by 1,2,3,4,5,6
```

- SQL Code for extracting data set for Novelty Effect

```
with all_user as (
  select distinct u.id as id,
    g.join_dt as activity_date,
    u.country,
    u.gender,
    g.group as test_group,
    g.device,
    coalesce (a.spent, 0) as spent
  from groups as g
  left join activity as a
    on a.uid = g.uid
  left join users as u
    on u.id = g.uid
)
select id,
    country,
    gender,
    test_group,
    device,
    activity_date,
    case when spent > 0 then 1
    else 0 end as conv_table,
    sum(spent) as total_spent
  from all_user
  group by 1,2,3,4,5,6,7
```

- Tableau Visualization:
 - [GloBox A/B Test Analysis Metrics Visualizations](#)
 - [GloBox Confidence Interval](#)
 - [GloBox Novelty Effect](#)
- Google Spreadsheet:
 - [GloBox A/B Test Analysis Spreadsheet](#)
 - [GloBox Confidence Interval Spreadsheet](#)
- Power Analysis Calculator:
 - [Statsig](#)
 - [Statulator](#)