Food and Drinks Landing Page Banner A/B Test Analysis

Goal: Increase Revenue

A report presented by: Jennifer Kengmo Nkemubuh

On this day: Nov 2023

TL:DR
RECOMMENDATION: DO NOT LAUNCH

After conducting an A/B test for an e-commerce website's new homepage with food and drinks banner, there was a 26.1% difference in conversion rate between control and treatment groups, no significant difference in average amount spent between both groups. The results impacted female IOS users more, Users in some European countries experienced a change in average spending while difference in conversion rate was quite commendable for all regions, the sample size used to conduct this experiment was relatively very small. My recommendation is not to launch this new food and drinks banner on the homepage based on the fact that revenue is not yielded.

GLOBOX

Table of Contents

1. Summary

The A/B test conducted on Globox homepage is aimed to access a new feature. Users are randomly assigned to both groups: group A(Control group) do not see the new food and drinks banner on the home page while group B(Treatment group) are exposed to the food and drinks banner on the home page. The test is conducted based on two key metrics: conversion rate and average amount spent per user.

- **Conversion rate:** The conversion rate In A(control) group is 3.92% and in B(treatment) group is 4.63% giving percentage of change of 18.1%. This is quite a noticeable change. Additionally, a two-sample t-test with pooled proportions was conducted and the probability that our results are due to chance is 0.0001, this result suggests that there is a significant difference in conversion rate between both groups.
- **Average amount spent per user**: On the other hand, the difference in average spending between both groups is 0.59% which is very minimal. A two-sample z-test with unequal variance was conducted in the case and the probability our results are due to chance is 0.94 this is way higher than the confidence threshold of 0.05. This suggests that we do not have enough evidence to conclude that there is a difference in average spending between both groups.

Further insights were drawn from a more detailed breakdown of the results by gender, device type and country.
Females were the most impacted among all other genders by a 5.8% of change conversion rate and -$2.9 change average spending.
Males on the other hand experienced a 3.79% conversion rate and $2.60 average spending.
IOS users converted at a 6.47% compared to Android users at a 3.52%. There was a minimal 2.96% change in average spending for IOS users.

A power analysis was also conducted which provided information on the sample size needed to run this experiment, it suggested a 186k sample size, however our sample size was limited to 49k. Although the conversion rate made it to the 10% threshold, the average spending did not.
A larger sample size is recommended if we are to run a similar experiment in the future.

A novelty effect was also conducted to understand and visualize trends on conversion rates and average spending on a daily basis. A 6 months forecast was introduced which demonstrated a dramatic fall in conversion rate for the control group in 6 months while conversion rate in the treatment group remained constant at 4.663%. Average spendings dropped slightly after 6 months, however the difference between both groups is not very noticeable.

2. Context

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, their food and drink offerings have grown tremendously in the last few months, and the engineering and growth team consider running an A/B test with the aim of analyzing and understanding the effectiveness of the landing page with food and drinks banner on the homepage in the currently expanding food and drinks sector.
Their primary objective is to yield revenue. Some other aims of this include: boost awareness and conversion rates while meeting up with customer's wants and satisfying them.

2.1 TEST GROUPS

Users were randomly assigned to two groups:

- Control group(A): Consist of a total of 24343 users and 955 converted users.
- Treatment group(B): A total of 24600 users and 1139 converted users. Users in this group are exposed to the new homepage with food and drinks banners.

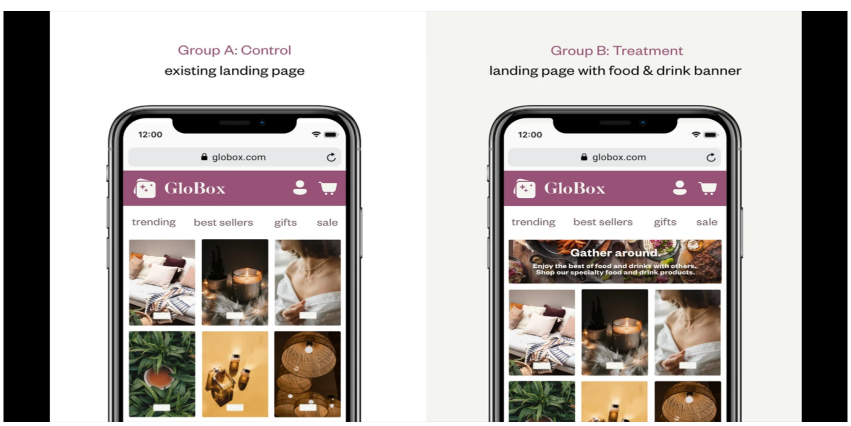

Fig 1.Showing website's homepage before and after addition of food and drinks banner.

2.2 KEY METRICS

The key metrics were carefully selected for this experiment and they are:

- **Conversion rate:** It is expressed in % and It represents the percentage of the proportion of users who made a purchase to the total number of users who partook in the experiment.
- **Average amount spent per user($):** This provides insights on the total revenue generated. It is the average of the total amount spent per user by all users in the experiment and it includes all those who either made a purchase or not.

2.3 Testing Variables

The A/B test ran from January 25th to February 6th making a total of 12 days. Users were randomly assigned to both groups, however we tried to apply a 50/50 split ratio between both groups to prevent bias in the experiment process.
The variables used for analysis were:
Gender(Males, Females, Others)
Device type(Mobiles were used, specifically IOS AND ANDROID)
Regions(Asia, North America, South America, Europe, Australia). Globox users were in specific countries in the above regions and some of these countries are; CAN, USA, GER, FRA, GBR, BRA, MEX, ESP, TUR, AUS

3. Research Methodology.

This experiment employed an A/B test methodology to understand the trends and behaviors of users with and without the food and drinks banner on the website's homepage. Users were randomly assigned to both groups, users in the control group(A) could only see the existing homepage, however, users in the treatment group could see the existing homepage with the food and drinks banner.
This experiment ran for 12 days and our focus was on mobile platforms only for users in the 10 countries from the various regions mentioned above.

3.1 OBJECTIVE

The main motive behind this is to understand the impact of the food and drinks banner on the homepage on conversion rate and average spending per user. The experiment aims at understanding how the food and drinks banner affects users' behavior, overall revenue and user engagement especially in the food and drinks sector.

3.2 COLLECTION OF DATA

The entire dataset used was collected from mobile platforms of the 48943 users from the 10 different countries for a 12-day period
This dataset was cleaned and relevant data extracted using sql queries on beekeeper.
The data obtained was exported to tableau for visual understanding of trends and users behaviors.

3.3 Key Metrics

The two success metrics used to conduct this A/B Test are:

Conversion rate: This is the percentage of the proportion of converted users to the total population. In our case, there were only 2094 converted users, 955 users from the control group and 1139 from the treatment group.

Average amount spent per user: The average spending was taken based on the entire population, that is average spending was calculated for all users regardless of whether they converted or not.

3.4 COLLECTION OF STATISTICS

The data extracted from beekeeper was exported to Google sheets for statistical analysis. For the conversion rate, a hypothesis test was conducted to get the resulting p-value. A significance level of 0.005 was used as a guide for the p-value, normal distribution with pooled proportions, null and alternative hypothesis were also generated.

This was similar with the average spending, only that we assumed unequal variances.
PS: See spreadsheet for Null and Alternative hypothesis for both success metrics.

3.5 UNDERSTANDING NOVELTY EFFECT

The novelty effect was checked to understand the users' conversion rate and average spending on a daily basis. However, there was no clear trends observed for both success metrics.

3.6 DETERMINING THE SAMPLE SIZE NEEDED

A power analysis was performed to provide insights on the sizes needed to run this A/B test at a 10% minimum noticeable difference. Some variables used to conduct the power analysis include:

- **A/B Split ratio = 0.5 that is 50/50**
- **Significance level = 0.005**
- **Test sensitivity = 0.8**

The sample size recommended is quite large compared to our 49k sample size used to run this A/B test.

4. RESULTS AND FINDINGS

4.1 Overall

Generally, the food and drinks banner positively impacted the users' conversion rate, but the effect on average spending was almost unnoticed.

a. Conversion rate and Hypothesis test for both test groups

The conversion rate started off at 3.92% for the control group and 4.63% treatment group(after exposure to the banner). This signifies 18.1% change, which implies that this banner positively affected users' conversion rate.

The hypothesis test conducted for the conversion rate rejected the null hypothesis that; "The conversion rate in A is equal to the conversion rate in B" in favor of the alternative hypothesis that; "The conversion rate in A is different from the conversion rate in B". The probability that this result is due to chance is 0.0001. At a 0.05 confidence level this probability is almost insignificant.

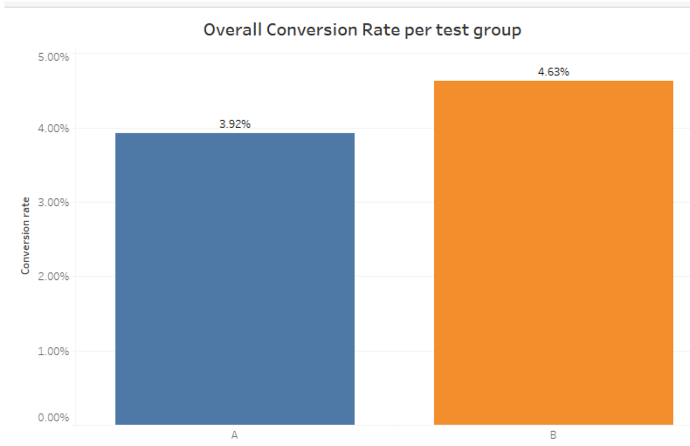

Fig 2. Showing the total conversion rate of users in both test groups

b. Average Spending and Hypothesis Test for test groups

Initially, the total average spending per person before exposure to the banner was $3.37, after exposure the average spending per user is $3.39. The percentage of change here is 0.59%. This change is very minimal.

Looking at the hypothesis test, it is quite difficult to come up with a conclusion, however, based on the results we have, we fail to reject the null hypothesis that; "There is no difference in average spending between both groups" in favor of the alternative hypothesis that; " There is a difference in average spending between both groups". This result is due to a probability of chance of 0.944 and with a confidence level of 0.05, the result obtained is quite commendable.
Also, users in the control group(A) spent more totally with a total spent of $1659.40 compared to users in the treatment group with the highest total spent at $1546.30

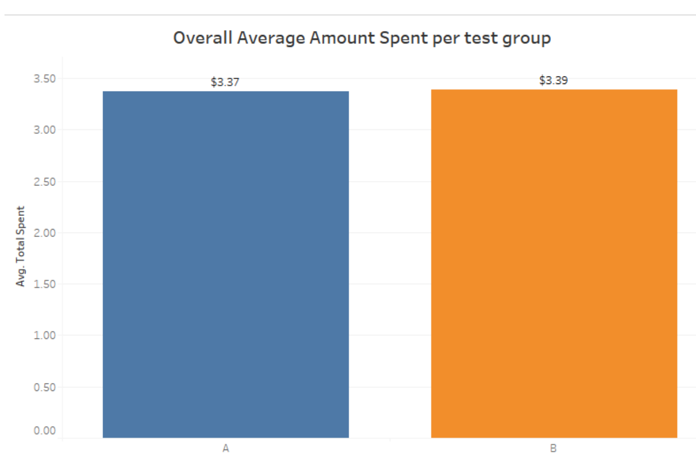

Fig 3. Showing the total average spending per user in both test groups

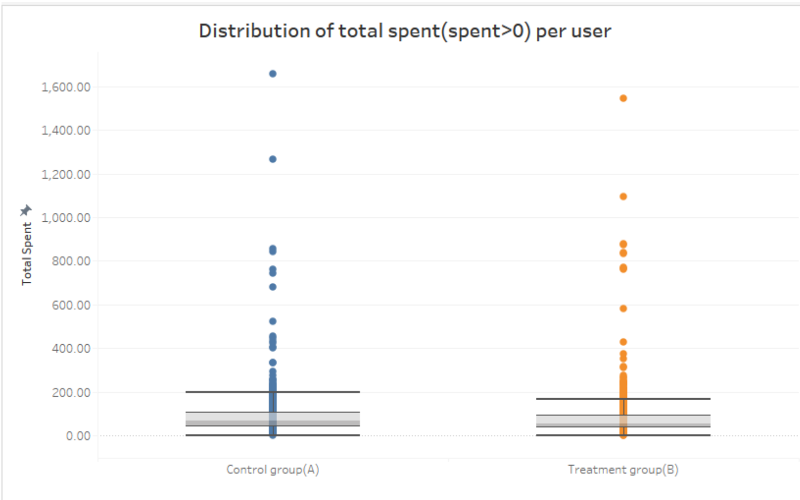

Fig 4. Showing the distribution total spent of all users in both test groups

c. Confidence Interval for Conversion Rate

In this A/B test for the conversion rates, we are quite sure that the true values fall within a range of 3.5% and 1.07%. The observed difference in the outcomes between the A and B groups is higher than zero, and I can be reasonably confident (at the specified confidence level of 95%) that this difference is real and not due to random chance.

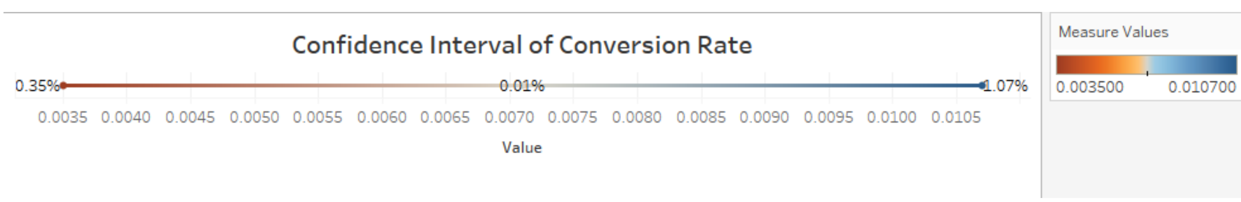

Fig 5. Confidence Interval for the conversion rate with point estimate

d. Confidence Interval for Average Spending

The confidence interval of the average amount spent at a 95% confidence interval, lies between -$0.439 and $0.471. This effect might be due to chance since it includes negative values and zero. However, the point estimate still provides valuable insights, indicating a small positive effect that could be explored further in subsequent analyses or experiments.

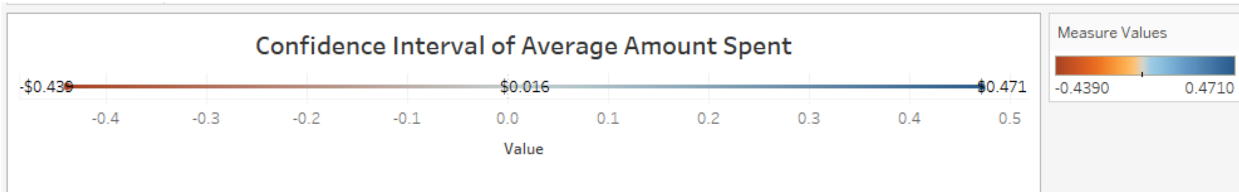

Fig 6. Confidence Interval for average spending with point estimate.

4.2 TEST VARIABLE FINDINGS(Results Breakdown)

After examining the various success metrics based on the test parameters used to conduct this A/B test, several observations and findings were made and are worth making note of please see below for the results breakdown.

a. Results by Gender

The A/B test conducted took three genders into consideration; males, females and others, however during my analysis, I excluded others and only considered males and females.

- **Conversion rate:** It was observed that males overall had the lowest conversion rate and average spending across all genders. They started off at 2.63% for the A group and 3.79% B group giving a percentage of change of 44%. This is quite noticeable. Females on the other hand converted more across all genders. However, the percentage of change between both groups was not very large, this means that females regardless of the group they were in, they converted more. This is seen at 5.14% control group and 5.44% treatment group giving a percentage of change of 5%.

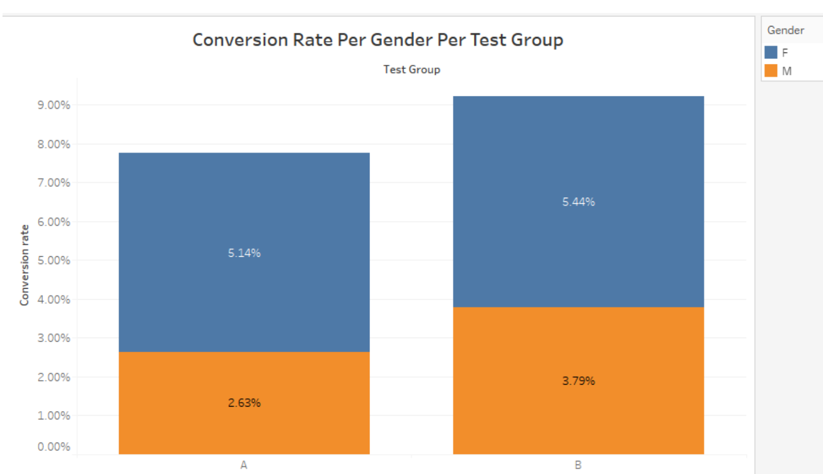

Fig 7. Conversion rate for males and females of both test groups

- **Average Spending:** The average spending for males was noticed at a 15% change, that is averagely they spent $2.25 for users in the control group and $2.60 for users in the treatment group.On the other hand, females in the control group spent more averagely than those in the treatment group leaving the percentage of change between both groups at -7%. Average spending left from $4.46 in the control group to $4.13 in the treatment group.

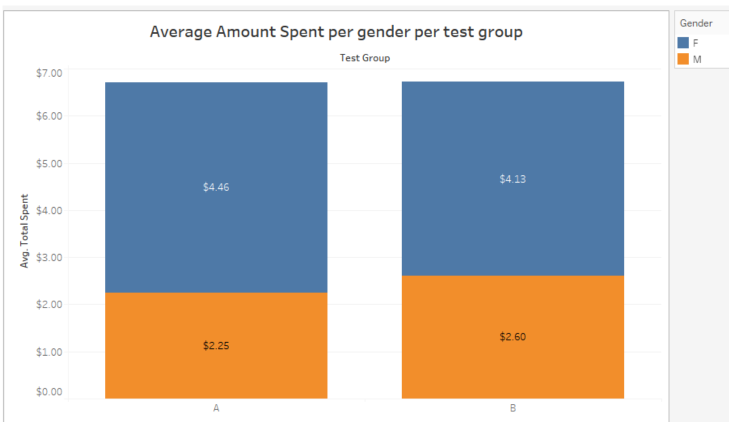

Fig 8. Average spending for males and females in both test groups

b. Results by Device Type

The A/B test conducted only considered data taken from mobile devices, specifically IOS and Android devices. After analyzing and visualizing the behavior of users from the data collected, I made several observations which I believe will help GloBox personalize the platform for each specific device.

- **Conversion Rate:** The food and drinks banner positively influenced the behavior of IOS users in the treatment group who experienced a rise of about 10.6%. That is 5.85% in the control group to 6.47% in the treatment group. Android users converted at a much lower rate than IOS users upon exposure to the banner. This is explained by a notable 27% difference between both groups, that is 2.77% control group and 3.52% treatment group.

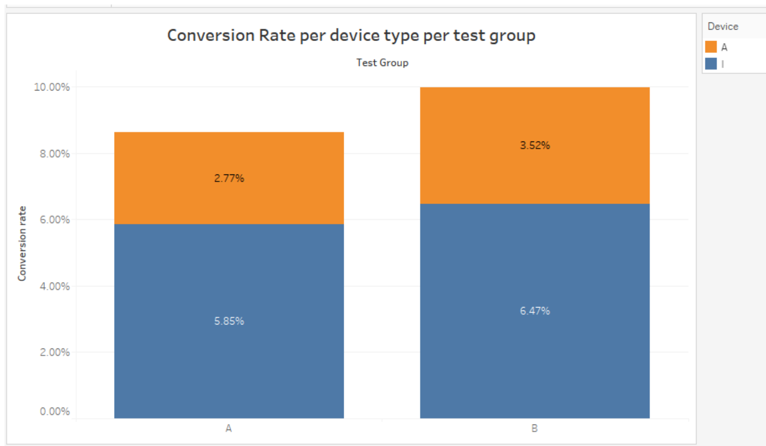

Fig 9. Conversion rate for Android and IOS users in both test groups

- **Average Amount Spent Per User.**

The food and drinks banner had a very slight impact on the average spending of IOS users and it is worth noting the 3% decrease. It started at $5.07 in the control group and dropped to $4.92 in the treatment group. Android users on the other hand experienced a 7% increase in average spending upon exposure to the banner. They started off at $2.31 in the control group to $2.47 in the treatment group.

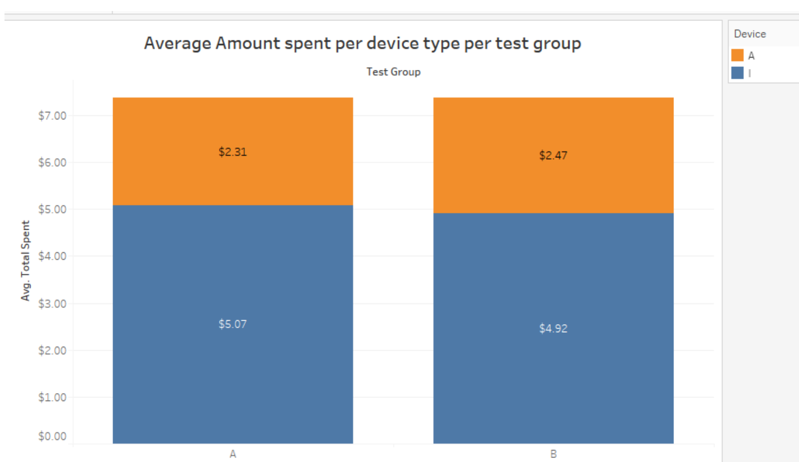

Fig 10. Average spending for IOS and Android users from both test groups

c. Results by Regions and Countries

Geographical data drawn from the dataset was used to understand the spending behavior and conversion rates of users from different geographical locations to enable the team make decisions and offerings to suit the diverse taste of its users.

- **Conversion Rate:** European countries in the treatment group were positively impacted upon exposure to the food and drinks banner and experienced an uplift of 21.8%. Some European countries worth noting are; GBR with a 27% difference between both groups, ESP at 24%, DEU 38% and FRA 31%. It is only right to say that the banner positively influenced the conversion rate of users in Europe.

South American countries experienced a slightly higher difference than European countries in conversion rate between both groups. The percentage of change between both groups for South American countries is 30%

North American countries topped all other countries in both control and treatment groups, as a matter of fact, the most converted users overall are users from North American countries. The percentage of change between both groups here is 25%.

Australian users were greatly affected by the banner as seen in the 42% of change between both groups.

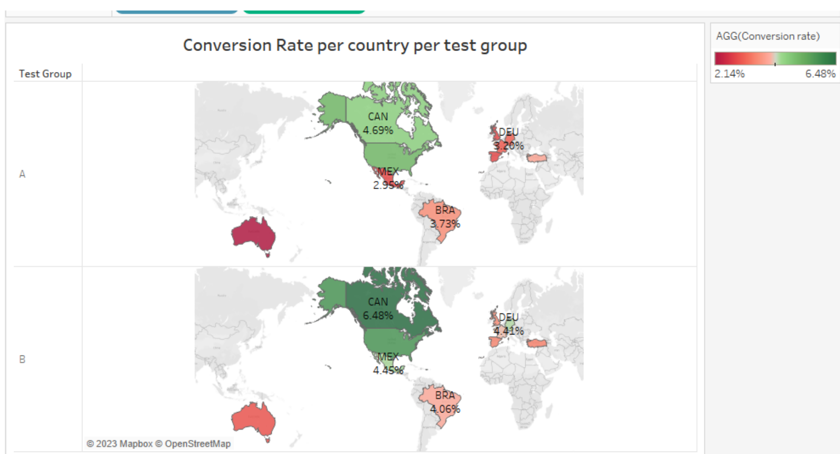

Fig 11. Showing the conversion rate of all countries in all regions for both test groups with the most converted countries coloured in green.

- **Average Spending**

When we look at the average amount spent, we realize that the banner didn't affect the average spending of most countries. As a matter of fact, the average spending of most countries in all regions was higher in the control group than in the treatment group. However, it is worth noting the massive change in average spending of users in GBR. In GBR, users spent averagely $2.11 in the control group and $4.50 in the treatment group giving a percentage of change of 113%. Users in ESP were also positively influenced with a 53% percent of change.

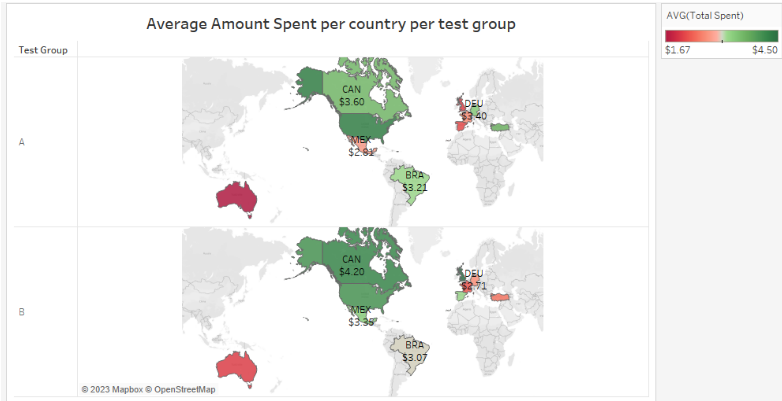

Fig 12. Showing the average spending of users in all countries from both test groups and countries in red indicate low average spendings.

The table below will give you a clearer understanding of the conversion rate and average amount spent of users in all countries in both test groups and their percentages of change.

Country	Conversion rate			Average Spending		
	Group A(%)	Group B(%)	%Change (%)	Group A($)	Group B($)	%Change (%)
CAN	4.69	6.48	38	3.60	4.20	17
USA	5.12	5.75	12	4.30	4.05	-6
MEX	2.95	4.45	51	2.81	3.35	19
BRA	3.73	4.06	9	3.21	3.07	-4
GBR	2.89	3.68	27	2.11	4.50	113
ESP	2.91	3.61	24	2.11	3.23	53
DEU	3.20	4.41	38	3.40	2.71	-20
FRA	3.20	4.18	31	2.68	2.27	-15
TUR	4.00	3.56	-11	3.69	2.49	-33
AUS	2.14	3.04	42	1.67	2.08	25

4.3 Further Analysis

Further analysis was carried out to understand the trends that exist between the various test parameters for both groups and some really interesting insights were made. These results will be presented through charts which have been designed to portray the information needed without saying much.

a. **Country VS Mobile Devices:** After analyzing the data, we can conveniently say that IOS users in both groups specifically from North and South American countries and European countries converted more than users in the treatment group from all countries using all other devices. Some IOS users in North America from the treatment group also had a high conversion rate. IOS users in North America had a quite commendable average spending for users in the control group, also, IOS users from Europe, some North and South American countries in the treatment group experienced high average spendings compared to users using other devices. Android users from all countries were the least affected.

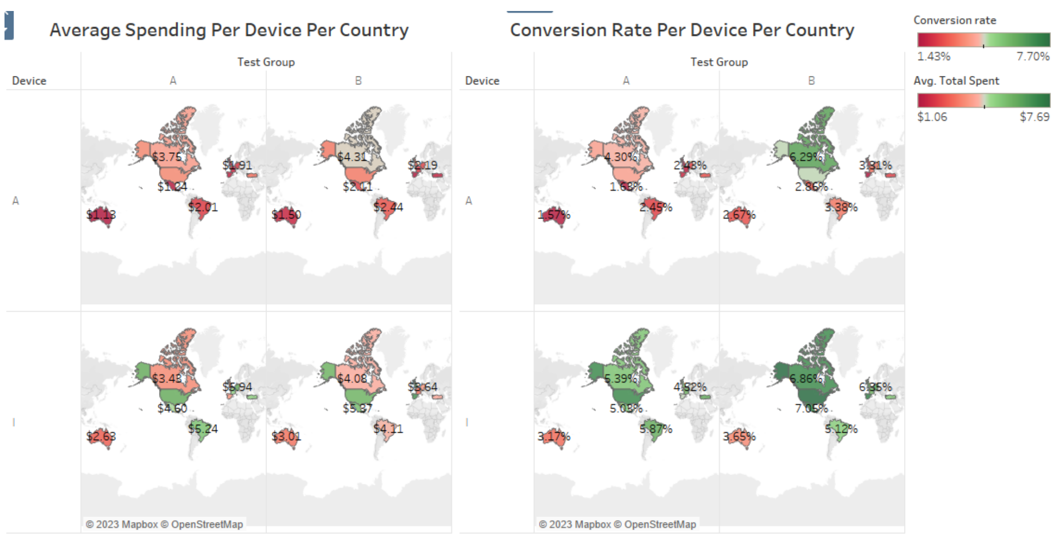

Fig 13. Showing the conversion rate and Average spending of users in both groups and from all countries using the various mobile devices. Red indicates low amounts.

b. **Country VS Gender:** I observed that females from North/South America, Europe and Asian countries had the highest conversion rate in both groups. Males in the treatment group from the USA and CAN also converted at a high rate.Looking at the average spending, females again in both groups and from all regions spent most averagely.

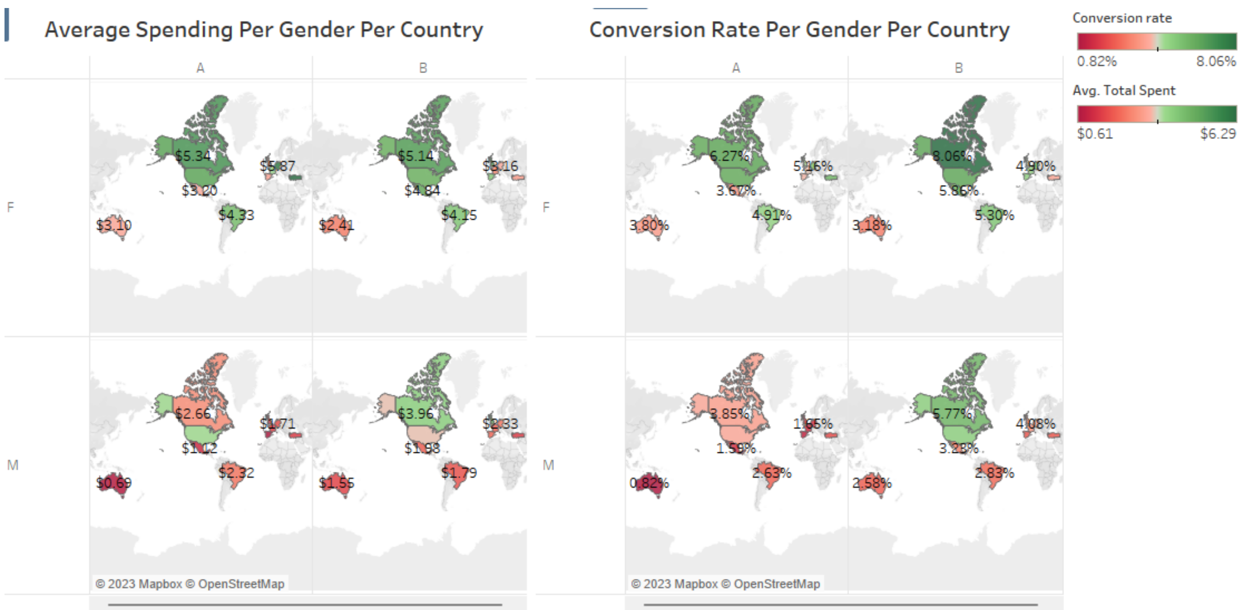

Fig 14. Showing the conversion rate and average spending of users of all gender in all countries/region

C: Gender VS Device: Females IOS users converted more and spent more averagely across all other genders and devices.

Conversion Rate per gender per device

Device	Test Group	Gender F	Gender M
A	A	4.04%	1.34%
	B	4.49%	2.63%
I	A	6.97%	4.80%
	B	7.05%	5.67%

Average Spending per gender per device

Device	Test Group	Gender F	Gender M
A	A	$3.39	$1.20
	B	$3.20	$1.74
I	A	$6.25	$4.02
	B	$5.71	$3.96

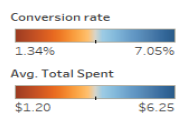

Fig 15. Showing the conversion rate and average spending of all genders using the various mobile devices.

5. UNDERSTANDING THE NOVELTY EFFECT

- **Conversion rate:** After conducting a novelty effect on the conversion rate, we realized that there is no clear cut trend in conversion rate for users in both A and B groups. Thus, there is no visible effect of the novelty effect on the conversion rate. However, further analysis was carried out to understand the behavior of users' conversion rate in the next 6 months. This is done to verify if launching this food and drinks banner on the homepage will continue to prompt users even in the long run or if it will only motivate the users for a short time. The results of this forecast shows that the conversion rate of users in the control group will fall drastically in the long run, however, conversion rate of users in the treatment group stays constant at 4.663% throughout the 6 months, thus I came to the conclusion that the food and drinks banner has a positive impact on users' conversion.

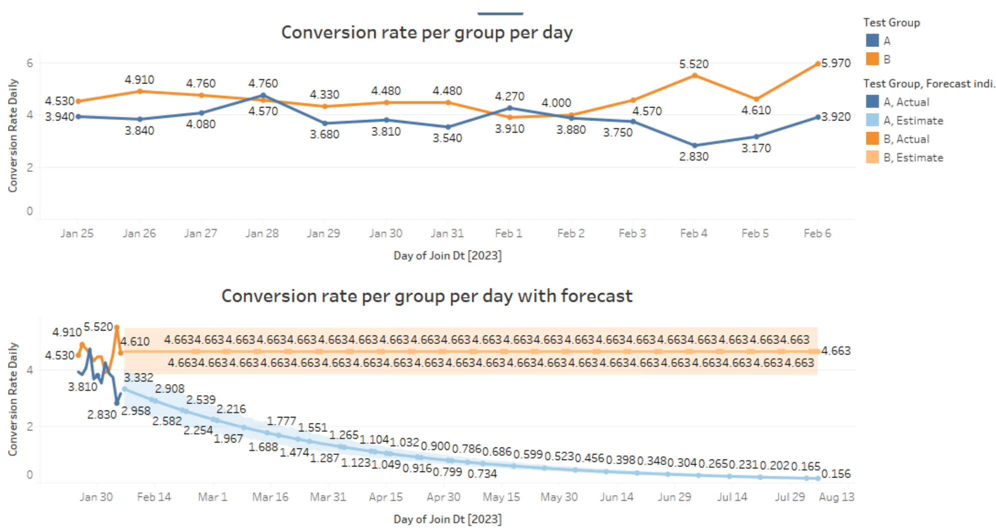

Fig 16. Showing the conversion rate of users on a daily level and in 6 months time.

Average Spent: The impact of the novelty effect on the average spending of users is worth paying attention to. This is because as the days go by, on every day of the experiment, there is no visible upward or downward trend in users' average spendings. That is, the banner doesn't affect the spendings of users either positively or negatively, the trend is constant. The forecast is then introduced to visualize the spendings of users in 6 months time, but this time around we observe that there is a slight fall in average spending of users in both test groups. At the end of the 6 months we have $3.445 average spent for users in the treatment group and $3.316 average spent for users in the control group.

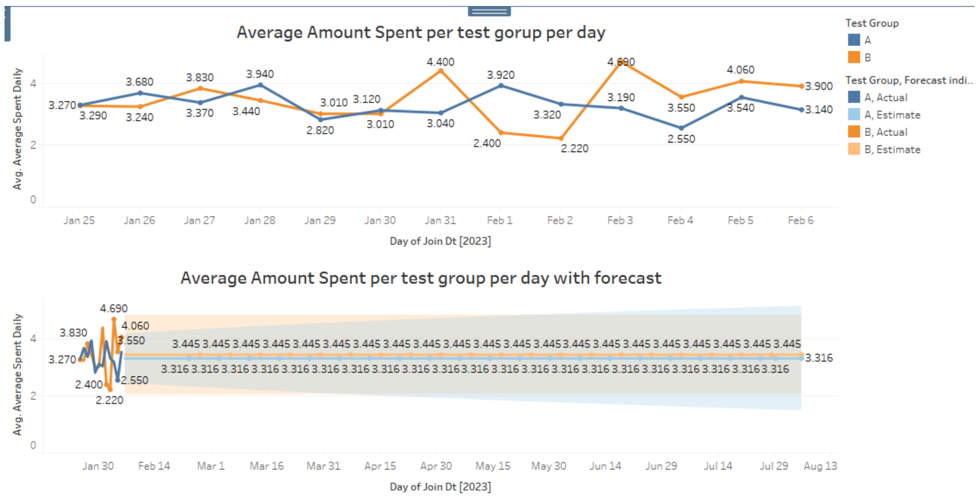

Fig 17. Showing average spending of users on a daily basis during the test period and in 6 months time.

6. POWER ANALYSIS

The sample size used to run this experiment is 48943 that is 24343 users in the control group(A) and 24600 users in the treatment group(B). With a minimum detectable effect (MDE) of 10%, a statistical power of 80%, an A/B split ratio of 0.5(control:treatment) and a significant level of 95%, a power analysis was conducted for both key metrics to detect the sample sizes needed to run this A/B test.

After conducting a power analysis with the baseline conversion rate of 3.92, the statsig calculator suggested a total sample size of 77k, and with a 50/50 split ratio, it suggests 38.5k users for the control group and 38.5k treatment group.

For the average spending, a statulator calculator was used with a standard deviation of 25.94, difference in mean of 0.337, statistical power of 80%, MDE of 10% and a significance level of 5%(two sided), a sample size of 186k was suggested.

Therefore, the power analysis suggests that the sample size used to run this A/B test is smaller than the recommended sample size required to achieve a 10% MDE with 80% statistical power at 95% significance level. It is thus advisable to use a larger sample size in the future to ensure that the statistical power is enough to detect the desired level of effect.

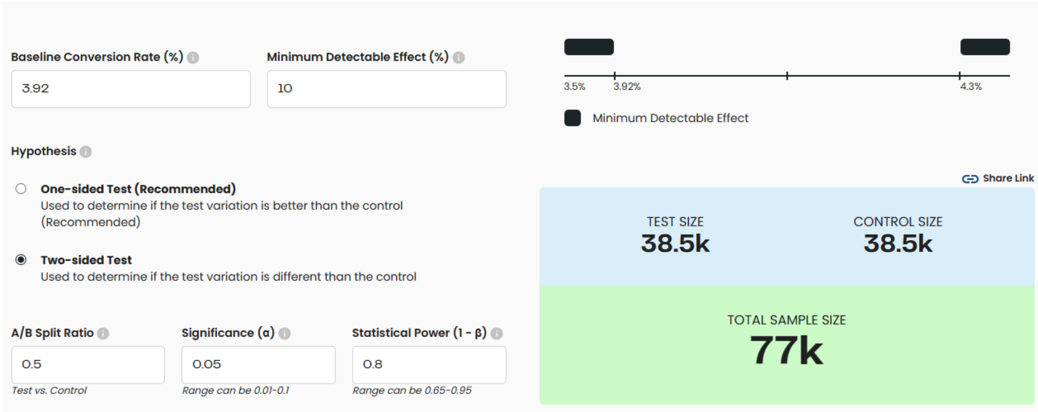

Fig 18. Power analysis of conversion rate using Statsig calculator to detect the sample size needed.

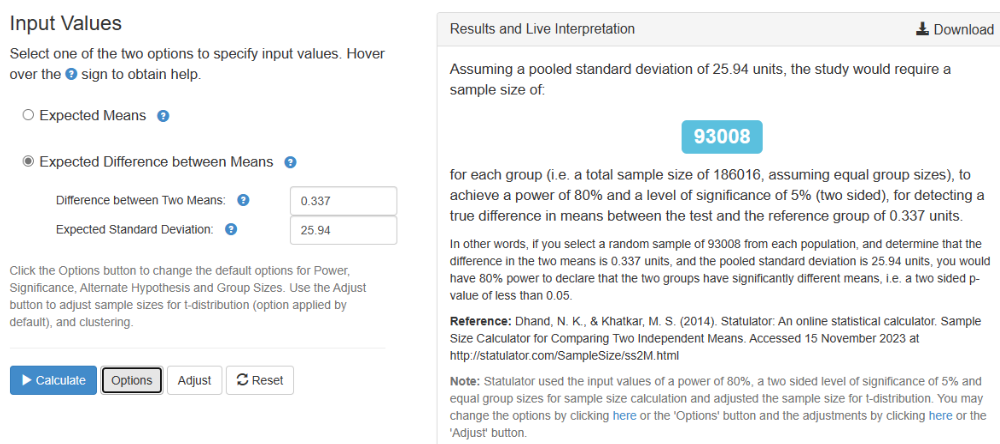

Fig 19. Power analysis of average spent using Statulator calculator to detect the total sample size needed.

7. Recommendations - DO NOT LAUNCH

7.1 Introduction:

The objective of this A/B test analysis was to evaluate the effectiveness of the treatment in increasing revenue per user, our primary experiment goal. The analysis aimed to provide data-driven insights to guide the decision-making process regarding the potential launch of the treatment. After analyzing the data using various tools, I recommend that we do not launch the food and drinks banner on the homepage.

7.2 Reasons for recommendation

- **Lack of Significant Increase in Revenue**:

The primary objective of our A/B test was to enhance revenue. However, the analysis reveals that the treatment did not lead to a noticeable improvement in this key metric. Without a clear positive impact on our primary goal, launching the treatment could lead to unproductive allocation of resources.

- **Small Sample Sizes and Limited Statistical Significance:**

A critical factor contributing to the inconclusive results is the small sample sizes used in the experiment. The power analysis indicates that the sample sizes were insufficient to detect a true effect reliably. As a result, the results lack statistical significance and might be influenced by random variation rather than a genuine treatment effect.

- **Recommendation - Do Not Launch:**

Considering the absence of a substantial increase in revenue per user and the limitations posed by small sample sizes, it is strongly advised not to proceed with the launch of the treatment. Implementing a feature without clear evidence of its positive impact could potentially harm user experience and revenue streams. It is essential to focus on initiatives that have demonstrated positive effects on our primary metrics.

7.3 **Conclusion**

While the treatment did not yield the desired outcome overall, there's an opportunity to optimize the use of our resources. If in future experiments, the treatment significantly boosts revenue per user, deploying the banner exclusively on the main page could be a strategic decision. Its prime placement offers a unique opportunity to maximize user engagement and revenue. However, this deployment should only be considered if substantial, consistent positive effects are observed in subsequent, well-powered experiments.

In conclusion, refraining from the launch of the treatment at this stage is the most prudent decision based on the available data. Moving forward, it is essential to conduct experiments with adequate sample sizes and rigorous methodology to ensure reliable results. Strategic deployment of features, especially in prime spaces, should be grounded in robust, data-backed insights to maximize their impact on our revenue and user experience.

8. APPENDIX

8.1 Analysis Files

[Tableau visualization]

[Deck]

Spreadsheet analysis

8.2 SQL Queries

Below is the code used to extract the data used for further analysis;

```
SELECT u.id                          AS user_id,
       u.country,
       u.gender,
       g.device,
       g.group                       AS test_group,
       CASE
         WHEN a.spent > 0 THEN 1
         ELSE 0
       END                           AS conversion,
       COALESCE(Sum(a.spent), 0) AS total_spent
FROM   users u
       LEFT JOIN groups g
              ON u.id = g.uid
       LEFT JOIN activity a
              ON u.id = a.uid
GROUP  BY 1,
          2,
          3,
          4,
          5,
          6
```

The code below is the code I used to calculate my novelty effect

```
SELECT    a.uid                        AS activity_id,
          g.group                      AS test_group,
          g.join_dt                    AS join_date,
          a.dt                         AS conversion_dt,
          Sum(COALESCE(a.spent,0))     AS total_spent,
          a.dt - g.join_dt             AS difference_dt
FROM      groups g
LEFT JOIN activity a
using     (uid)
```

```
GROUP BY  1,
          2,
          3,
          4
          -------------------------------------------NOVELTY EFFECT
DATA---------------------------------------------------------
          WITH data AS
          (
                    SELECT    g.join_dt                      AS join_dt,
                              g.GROUP                        AS test_group,
                              count(DISTINCT g.uid)          AS total_users,
                              count(DISTINCT a.uid)          AS
total_converted_users,
                              sum(COALESCE(a.spent, 0)) AS total_spent
                    FROM      groups g
                    LEFT JOIN activity a
                    using     (uid)
                    GROUP BY  1,
                              2 )
SELECT   data.join_dt,
         data.test_group,
         Round(Cast(Sum(data.total_converted_users)/ Max(data.total_users)
* 100 AS DECIMAL(10,2)), 2) AS conversion_rate_daily,
         Round(Cast(Sum(data.total_spent)
/Max(data.total_users)AS          DECIMAL(10,2)), 2) AS average_spent_daily
FROM     data
GROUP BY 1,
         2
```