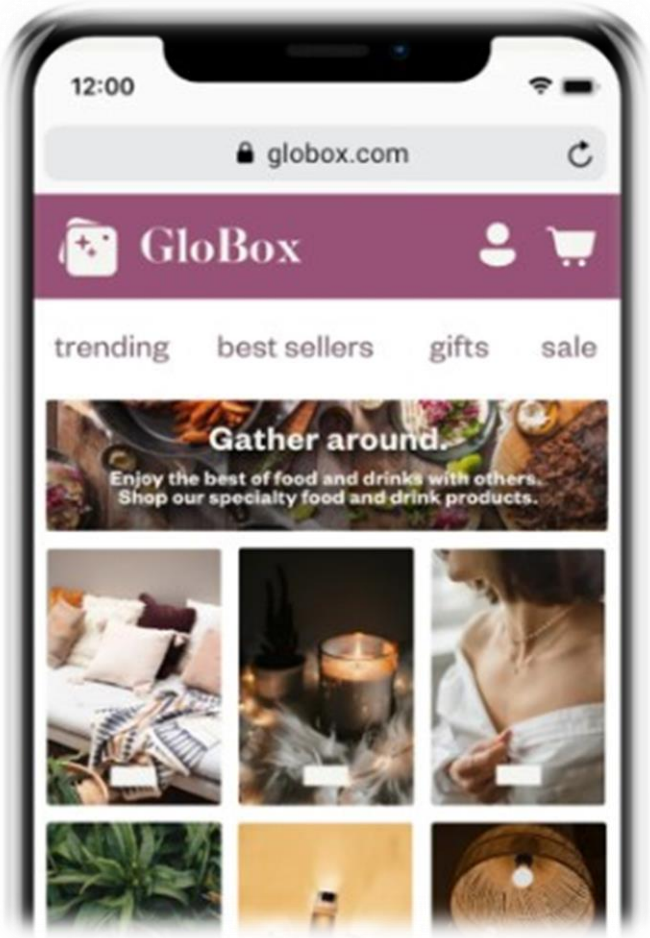


A/B TEST ANALYSIS REPORT

GloBox **FOOD AND DRINK BANNER EXPERIMENT**

Chukwuemeka Anyakwu | Master School Data Analysis Cohort | 2023

Table of Content

Summary

The main objective of conducting the A/B Testing for GLOBOX is to know whether there is a significant advantage in the Food and drink banner design on the landing page for their e-commerce mobile app.

Results on the following parameters:

- Increase in revenue.
- Increase in feasibility.
- Awareness of their food and drink offerings.

At the end of the experiment, **at this moment, we recommend the LAUNCHING of the Food and Drink Banner.**

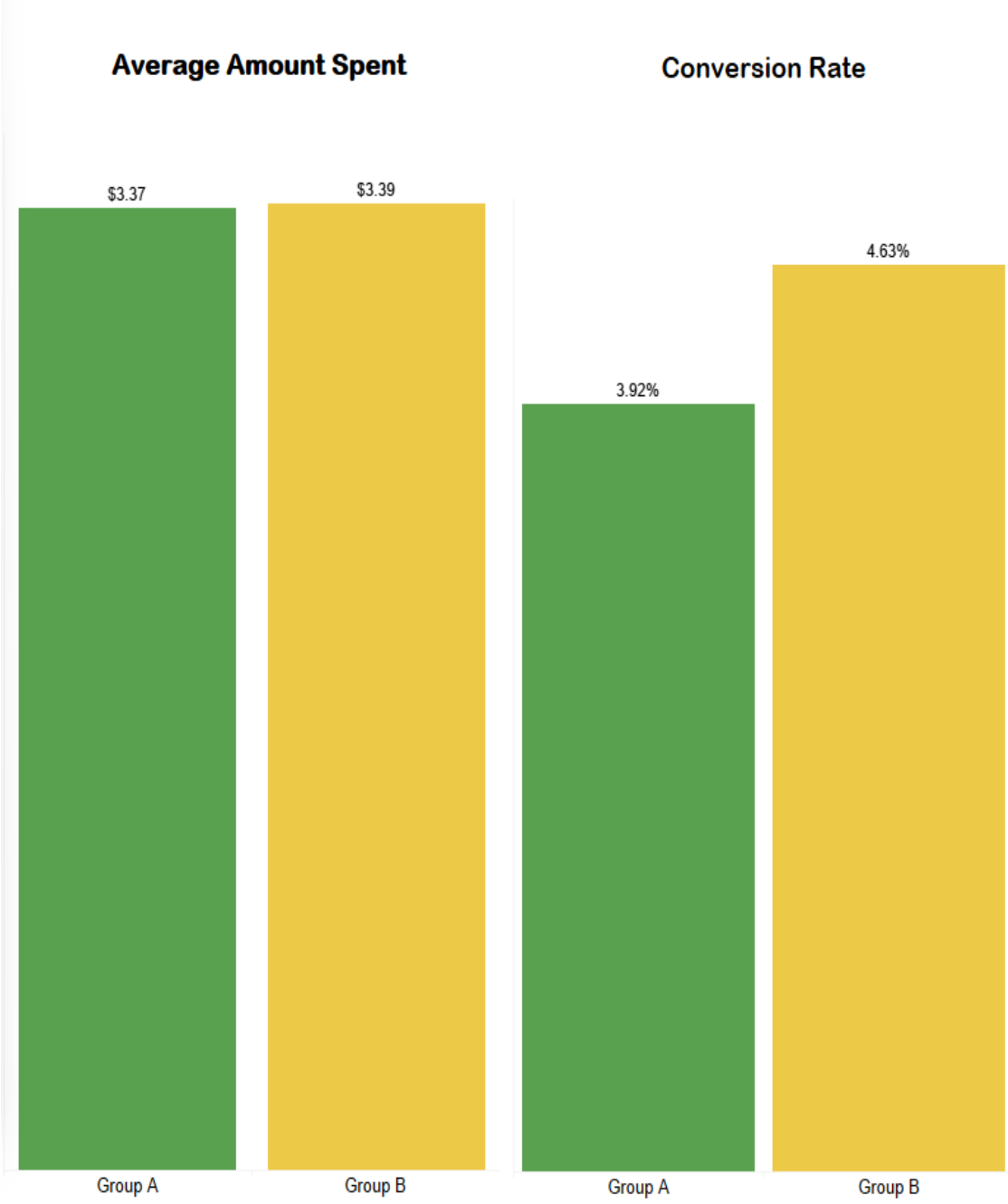
The test revealed a significant difference in conversion rate between Group A: Control (with no food and drink banner) and Group B: Treatment (with Food and drink banner on its landing page). More users were attracted, and purchases were made in Group B compared to Group A, indicating that the food and drink banner made the desired impact and was informative.

Launching the food and drink banner would:

- Increase GLOBOX revenue.
- Increase its feasibility.
- Create greater awareness of their food and drink offerings.

However, the average amount spent in Group A and Group B is the same. We need to monitor this closely by increasing the sample size and conducting further experiments to obtain a more conclusive result regarding the difference in means.

However, the sample size is substantial and robust. However, for future tests, we recommend a more extended testing period for a more reliable result while also encouraging monitoring of the impact of different types of devices on future sales as we make necessary adjustments to maximize the user experience and revenue further.


Average Amount Spent
$3.37
$3.39
Group A
Group B
Conversion Rate
3.92%
4.63%
Group A
Group B

Context

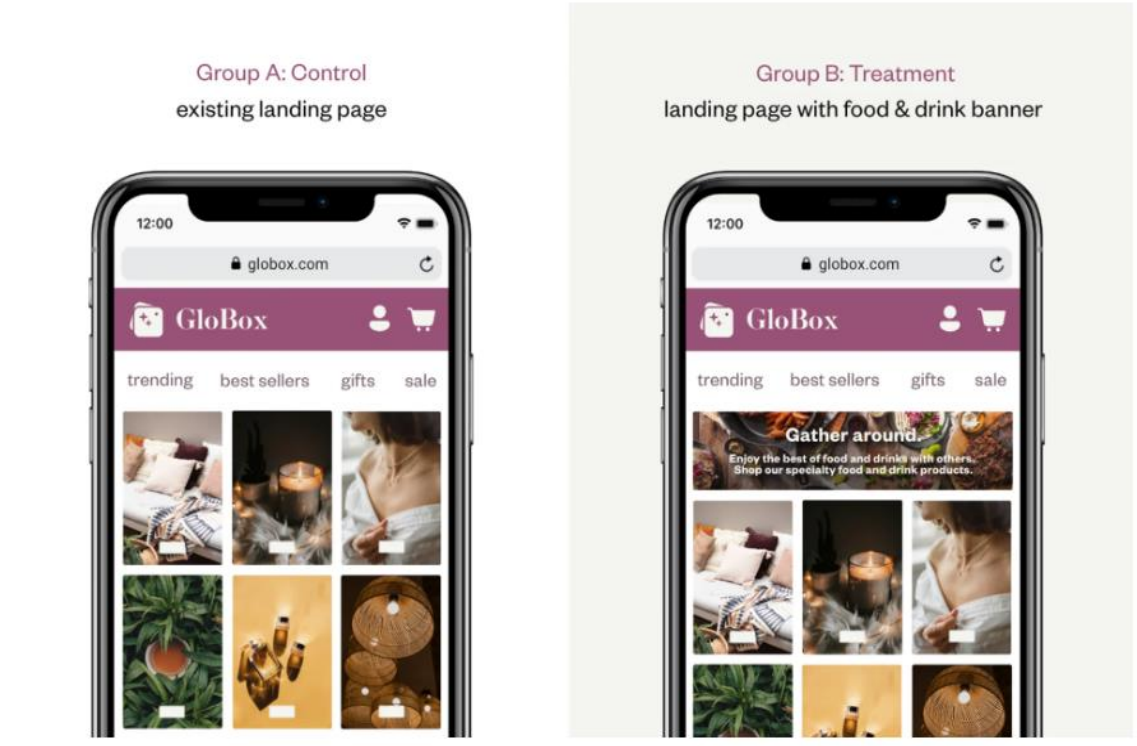
GloBox is primarily known for its boutique fashion items and high-end decor products; however, its food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue. Therefore, there is a need to introduce the Food and drink offering onto the organization's mobile app application to increase awareness of this new product category (i.e., the food and drink offering) and grow the company's revenue.

The Growth team decided to run an A/B test highlighting critical products in the food and drink category as a banner at the top of the website.

Group A: Control: This group uses the existing mobile app without the Food and drink banner.

Group B: Treatment: This group has a landing page with a food and drink banner.

Both tests were run simultaneously between January 25th and February 6th, 2023.



The experiment aims to verify if there is a significant impact made by the new design (Group B) over the existing design (Group A); the result of the experiment is going to be based on the patterns and behaviour of the consumer of the product and services based on the following parameters such as gender, device type, country or region, to assist us in better marketing better and reach.

Data and Approach

The methodology and approach for gathering and transforming the data are as follows:

Step 1: Data Extraction with SQL:

The SQL query aims to extract, manipulate, and transform the dataset provided in the 3 (three) tables below from the company's database.

Users: User demographic information, which includes the ID, country, and gender.

Groups: User A/B test group assignment, this has the following fields

- uid: the user ID
- group: The user's test group
- join-dt: the date the user joined the test (visited the page)
- device: the device the user visited the page on (I = iOS, A = Android)

Activity: user purchase activity, containing 1 row per day that a user made a purchase

- uid: the user ID
- dt: date of purchase activity
- device: the device types the user purchases on (I = iOS, A = Android)
- spent: the purchase amount in USD.

The next stage is **Data Modeling**, which involves building the Star Schema relationship between the three tables; find the relationship table below.

Which is a many-to-one relationship between the Fact table (activity table) and the dimension table (users, groups)

Step 2: Hypothesis Testing and Spreadsheets:

The result of the SQL query extraction was downloaded as an Excel file and uploaded in the Excel spreadsheet for further analysis. Hypothesis testing was performed on Groups A and B to compare the results of the Conversion Rate and Average amount spent.

1. A hypothesis test was conducted to see if there is a difference in the conversion rate between the two groups, using the **Null Hypothesis Ho** theory, which states that there is no difference in the conversion rate for both groups, while the **Alternate Hypothesis H1**: states that there is a difference in the conversion rate for both groups.

2. A hypothesis test was also conducted on the Average amount spent between both groups using the **Null Hypothesis Ho** theory, which states that there is no difference in the average amount spent for both groups. In contrast, the **Alternate Hypothesis H1** states that there is a difference in the average amount spent by both groups.

3. Confidence interval tests were also conducted on the conversion rate and the average amount spent.

Step 3: Data visualization in Tableau

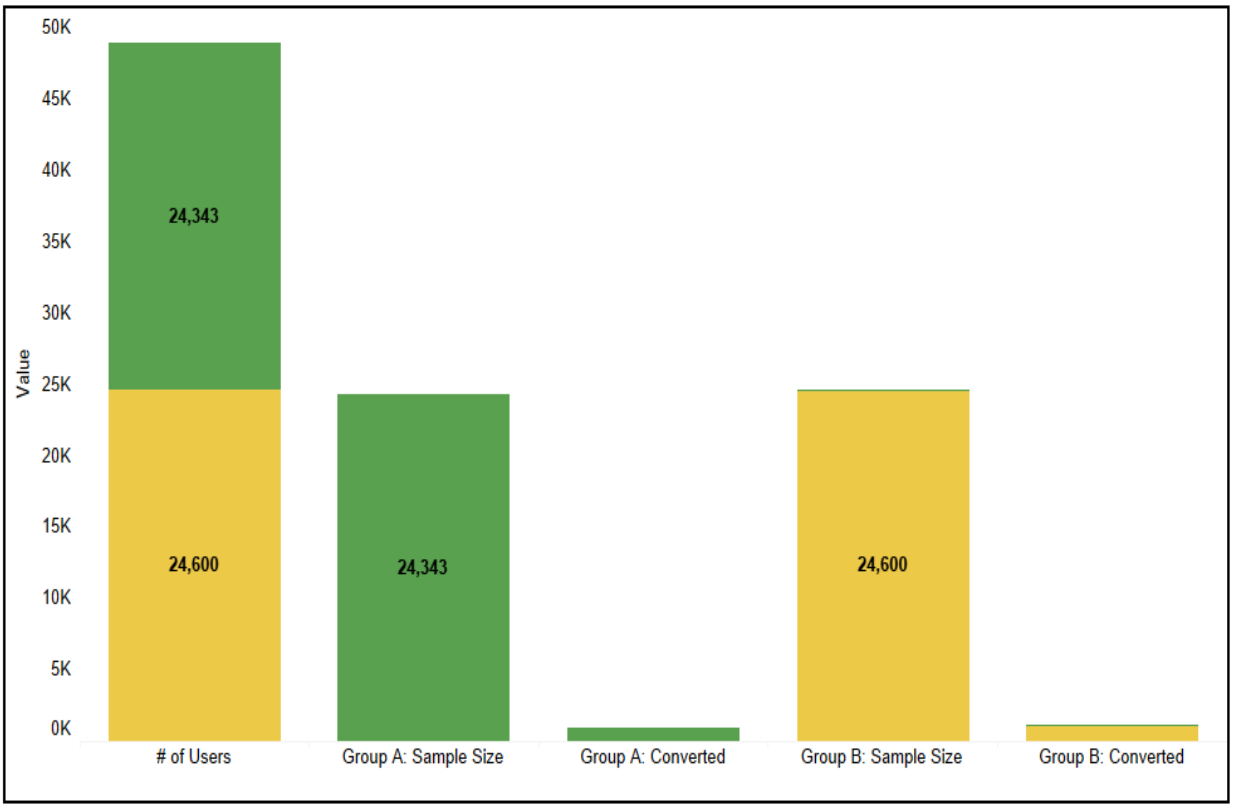
Tableau was used as our primary ETL tool to enhance our storytelling and provide further analysis and insight into the dataset. With the ETL Tool, it was easy to build various visualizations for the conversion rate, average amount spent, average mount distribution, gender vs test metrics distribution, user's device vs test metrics distribution and country vs test metrics distribution. The experiment's conclusion was based on the results from the above parameters. The experiment's result helped us make reasonable conclusions based on the results obtained after the analysis.

Find below various visualizations and results of the insight.

Results

Inferential Statistics

# of Users	Group A: Sample Size	Group A: Converted	Group B: Converted	Group B: Sample Size	Avg Amount Spent	Conversion rate
48,943	24,343	955	1,139	24,600	$3.38	4.28%



Conversion Rate

Purpose: To conduct a Hypothesis test on the conversion rate.

Null Hypothesis H_0: states that there is no difference in conversion rate between the two groups (control vs treatment)

$$H_0: p_1 - p_2 = p_0$$

Alternative Hypothesis H_1: The conversion rate in the two groups is different.

$$H_1: p_1 - p_2 <> p_0$$

The hypothesis test used was a two-sample z-test with pooled proportions.

Test statistic

$$T = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\hat{p} = \frac{\hat{p}_1 * n_1 + \hat{p}_2 * n_2}{n_1 + n_2}$$

P-value:

$$2 * P(Z > |T|)$$

Significance level (α): 0.05

The methodology and approach for gathering and transforming the data are as follows:

Group	Population	Count of Conversion	The standard deviation of conversion	Average of conversion
A (Control)	24343	955	0.194	0.039
B (Treatment)	24600	1139	0.210	0.046
Grand Total	**48943**	**2094**	**0.202**	**0.043**

Calculation	Notation	Value
Sample Size (Control Group A)	n1	24343
Sample Size (Treatment Group B)	n2	24600
Sample mean (Control Group A)	x1 bar	0.039
Samples mean. (Treatment Group B)	x2 bar	0.046
Sample std (Control Group A)	s1	0.194
Sample std (Treatment Group B)	s2	0.210
Standard error	SE	0.002
Test statistic	T	-3.866
Degrees of freedom	df	24342
p-value	pval	0.0001

Test statistic = -3.866
p-value = 0.0001

a) Our **signifiance level** **α** = 0.05
Our p-value cut of or threshold = 0.05
If p-value <= 0.05, we REJECT the Null Hypothesis
If p-value > 0.05, we FAIL TO REJECT the Null Hypothesis

Group	Population	Count of Convert	StdDev of convert	Average of convert
A (Control Group)	24343	955	0.194	0.039
B (Treatment Group)	24600	1139	0.210	0.046
Grand Total	**48943**	**2094**	**0.202**	**0.043**

Conclusion:

With a p-value of 0.0001, statistically significant. We **REJECT the null hypothesis** that there is no difference in the user conversion rate between the control and treatment groups.

Computing the Confidence Interval ~ 95% CI

95% Confidence interval for the difference in the conversion rate between Group A (Control) and Group B (Treatment)

Methodology:

Two-sample z-interval with unpooled proportions

sample statistic	$\hat{p}_1 - \hat{p}_2$
critical value	$z_{1-\frac{\alpha}{2}}$
standard error	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

▼ Notation

- $\hat{p}_1$, $\hat{p}_2$ = sample proportions for samples 1 and 2, respectively
- n_1, n_2 = sample size for samples 1 and 2, respectively
- $z_{1-\frac{\alpha}{2}}$ = z-value, the value of the standard normal distribution at the $(1-\frac{\alpha}{2})th$ percentile

Calculation	Notation	Value
Group A convert		955
Group A sample	n2	24343
Group B convert		1139
Group b sample	n2	24600
Sample proportion for A	p1	0.03923099
Sample proportion for B	p2	0.046300813
Difference	df	0.007069823
Standard Error	SE	0.001828488
Z-critical	Z1	1.959963985
Margin of Error		0.003583771
Lower CI		0.0035
Upper CI		0.0107

Conclusion:

The confidence level of the difference between the conversion rate falls between 0.35% and 1% at a 95% confidence level. This value is less than α: 0.05; **hence, we REJECT the null hypothesis**. The experiment group (i.e., the Control and Treatment Group) has a conversion rate difference.

Calculation	Notation	Value
Group A convert		955
Group A sample	n2	24343
Group B convert		1139
Group b sample	n2	24600
Sample proportion for A	p1	0.03923099
Sample proportion for B	p2	0.046300813
Difference	df	0.007069823
Standard Error	SE	0.001828488
Z-critical	Z1	1.959963985
Margin of Error		0.003583771
Lower CI		0.0035
Upper CI		0.0107

Conclusion:

The confidence level of the difference between the conversion rate falls between 0.35% and 1% at a 95% confidence level. This value is less than α: 0.05; **hence, we REJECT the null hypothesis**. The experiment group (i.e., the Control and Treatment Group) has a conversion rate difference.

Average Amount Spent.

The next stage is to conduct a hypothesis test to see whether there is a difference in the average amount spent between Group A (Control) and Group B (Treatment).

a) **Null Hypothesis H_0** states no difference in the average amount spent between the Control and Treatment groups.

$$H_0 : \mu_1 - \mu_2 = \mu_0$$

b) **Alternative Hypothesis H_1**: States that there is a difference in the average amount spent between the control and treatment groups.

$$H_1 : \mu_1 - \mu_2 <> \mu_0$$

c) Methodology used: Two-sample t-test for a difference in means**.**

d) Test statistic:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

e) P-value:

$$2 * P(t_{df} > |T|)$$

$$df = min(n_1 - 1, n_2 - 1)$$

f) Significance level (α): 0.05

Group	Population	Average of the amount spent	StdDev of the amount spent
A (Control Group)	24343	3.375	25.936
B (Treatment Group)	24600	3.391	25.414
Grand Total	**48943**	**3.383**	**25.675**

Description	*Group A*	*Group B*
Mean	3.37451847	3.390866946
Variance	672.696355	645.8769667
Observations	24343	24600
Hypothesized Mean Difference	0	
df	48894	
t Stat	-0.0704249	
P(T<=t) one-tail	0.47192788	
t Critical one-tail	1.64488479	
P(T<=t) two-tail	0.94385575	
t Critical two-tail	1.9600125	

Conclusion:

The p-value is 0.944; this is statistically insignificant; **we fail to reject the null hypothesis** that there is no difference in the mean amount spent per user between the control and treatment groups.

What is the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control)

The methodology used: Two-sample t-interval for a difference in means.

sample statistic	$\bar{x}_1 - \bar{x}_2$
critical value	$t_{(1-\frac{\alpha}{2},\, df)}$ $df = min(n_1 - 1, n_2 - 1)$
standard error	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Group	Population	Average of the amount spent	StdDev of the amount spent
A	24343	3.375	25.936
B	24600	3.391	25.414
Grand Total	**48943**	**3.383**	**25.675**

Calculation	Notation	Value
Stdev A		**25.936**
Stdev B		**25.414**
Sample Statistic		0.016349146
Sample Population N	N	48943
n1	n1	24343
n2	n2	24600
Critical value	Z	1.959963985
Standard Error	SE	0.232140506
Margin of Error		0.454987031
Lower CI		-0.438637885
Upper CI		0.471336177

Conclusion:

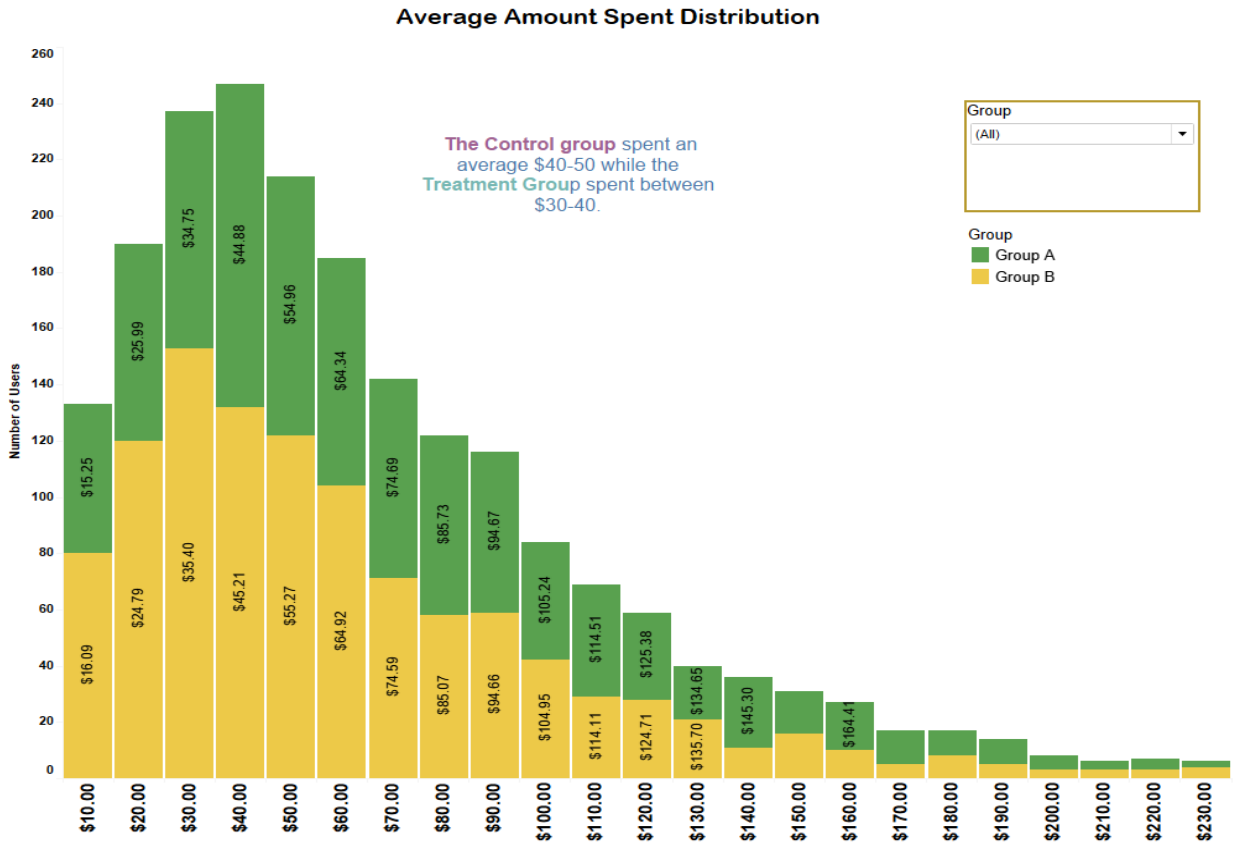
The confidence interval is between -0.439 and 0.471. based on the statistical analysis, **we fail to reject the null hypothesis**. The evidence suggests no significant difference in the amount spent per user in both groups.

Distribution of average amount spent.

Observation:

- ❖ The control group spent an average of $40-$50 with about 115 users.

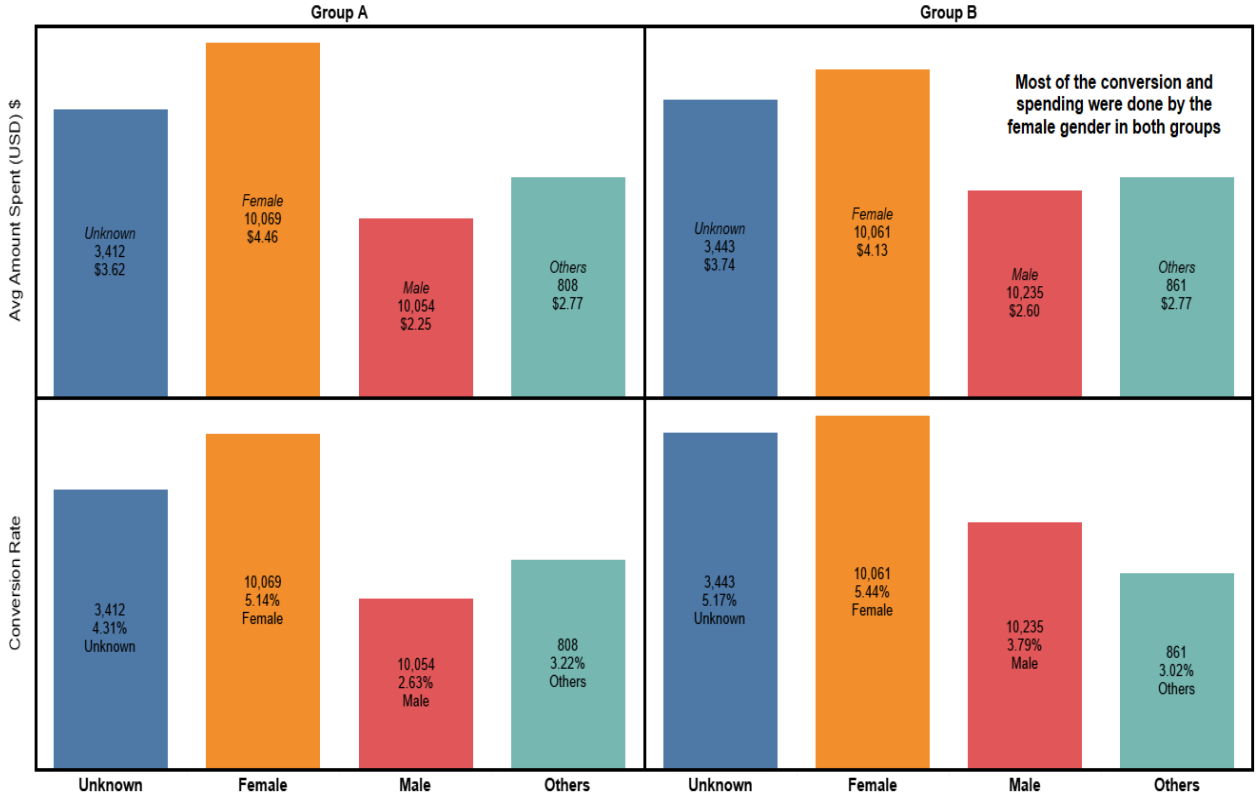
The treatment group spent an average of $30-$40 with about 153 users.



The outliers were removed to get a proper distribution and remove noise from the dataset.

- ❖ We need to check the reason for the spending patterns to help us with our pricing strategies, promotions, and product features within these price ranges.
- ❖ Tailor marketing messages and campaigns to target and engage users based on spending patterns.
- ❖ Monitor spending patterns over time and adjust strategies accordingly.

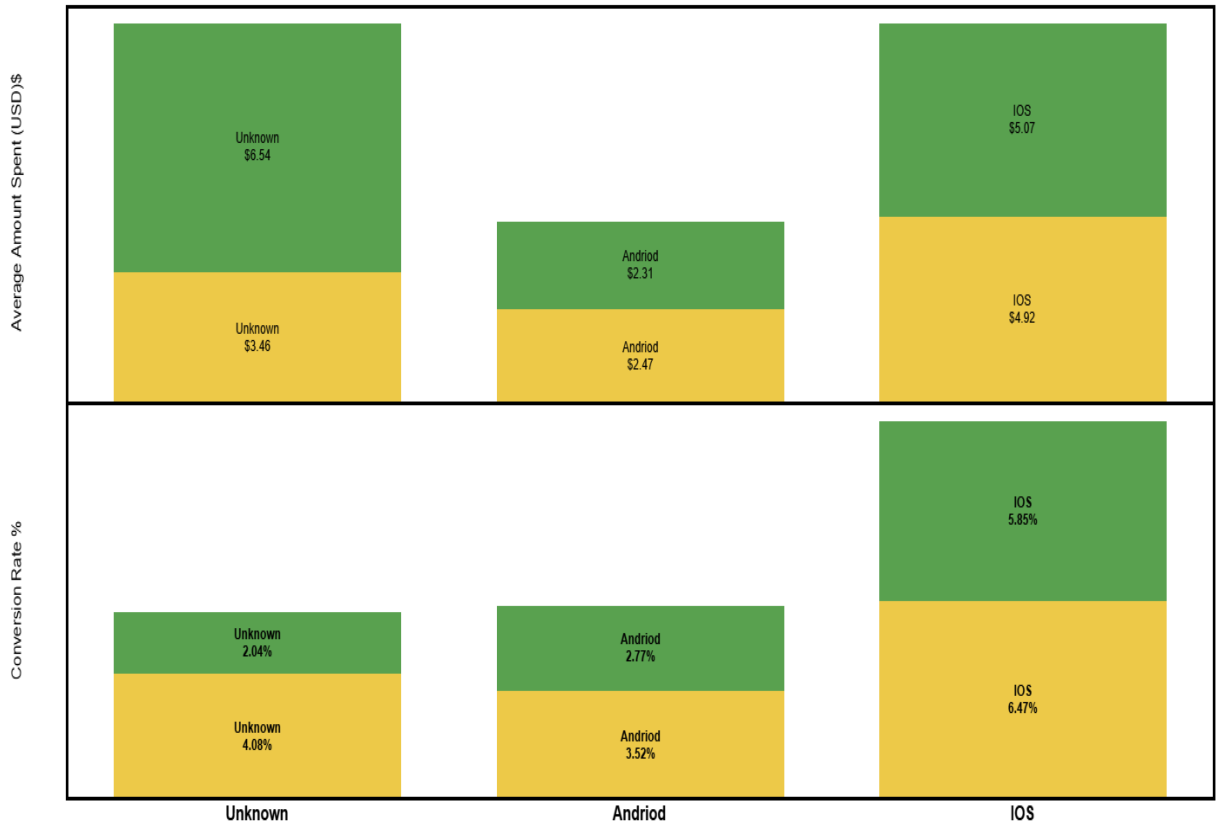
Relationship between the Test Metrics and Gender



Observation:

- The male has the lowest conversion rate and the average amount spent across all genders. In the treatment group, male users experienced a remarkable 44% increase in conversion rate.
- Female users had the highest conversion rate across all the genders, with a 5% growth in the treatment group.
- The result points to the effectiveness of the Food and drinks banners due to an increase in the conversion rate of the male gender.
- Further study is needed to know why the female gender conversion rate has stagnated.

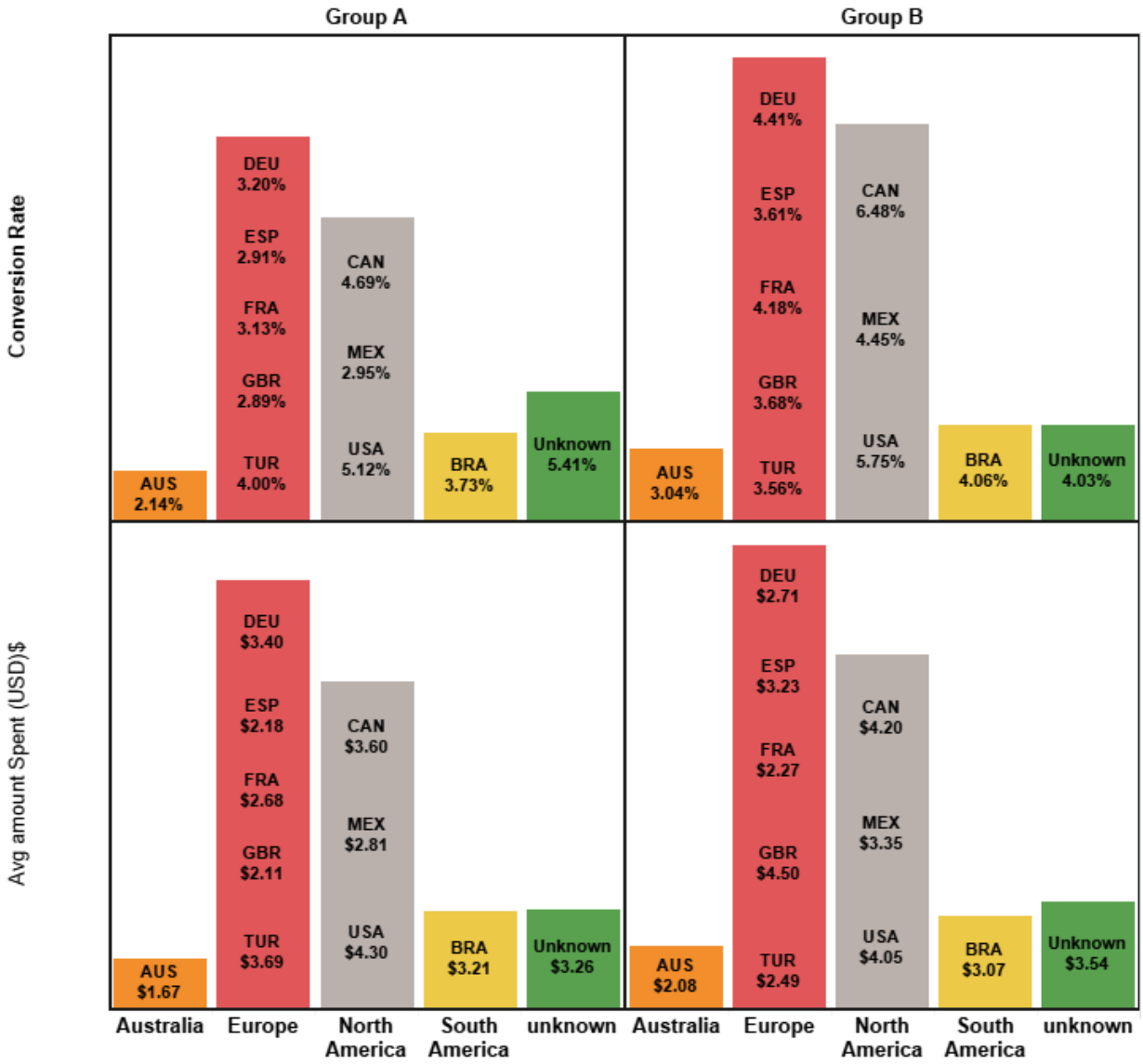
Relationship between the Test Metrics and User Device.



Observation:

- The iOS device users have the highest conversion rate of 6.47% for Group B and 5.85% for Group A, which shows an increase in the Conversion rate of the IOS device. There is a decline in the average amount IOS users spend, from $5.07 for group A to $4.92 for group B.
- Android users have the lowest conversion rate and the average amount spent.
- Finally, we have users who did not record their devices, so we converted them to unknown devices. They greatly impacted the average amount spent in Group A at $6.54 and $3.46 for Group B.

Relationship between the Test Metrics and Countries



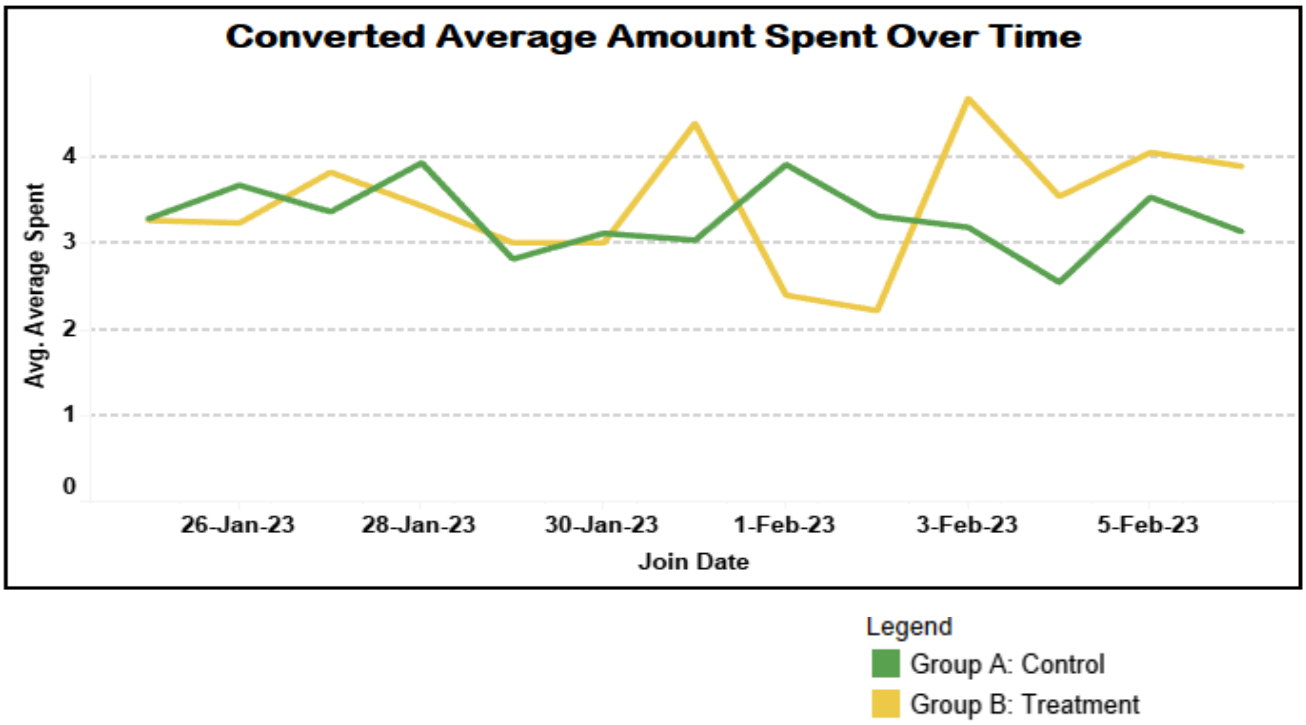
Observation:

- Europe, In the treatment group, shows a significant 32% improvement in conversion rate, followed by South America, where the treatment group also had an increase of 22%
- We equally have null countries, which are unknown.

Novelty Effect

Distribution of Average Amount Spent.

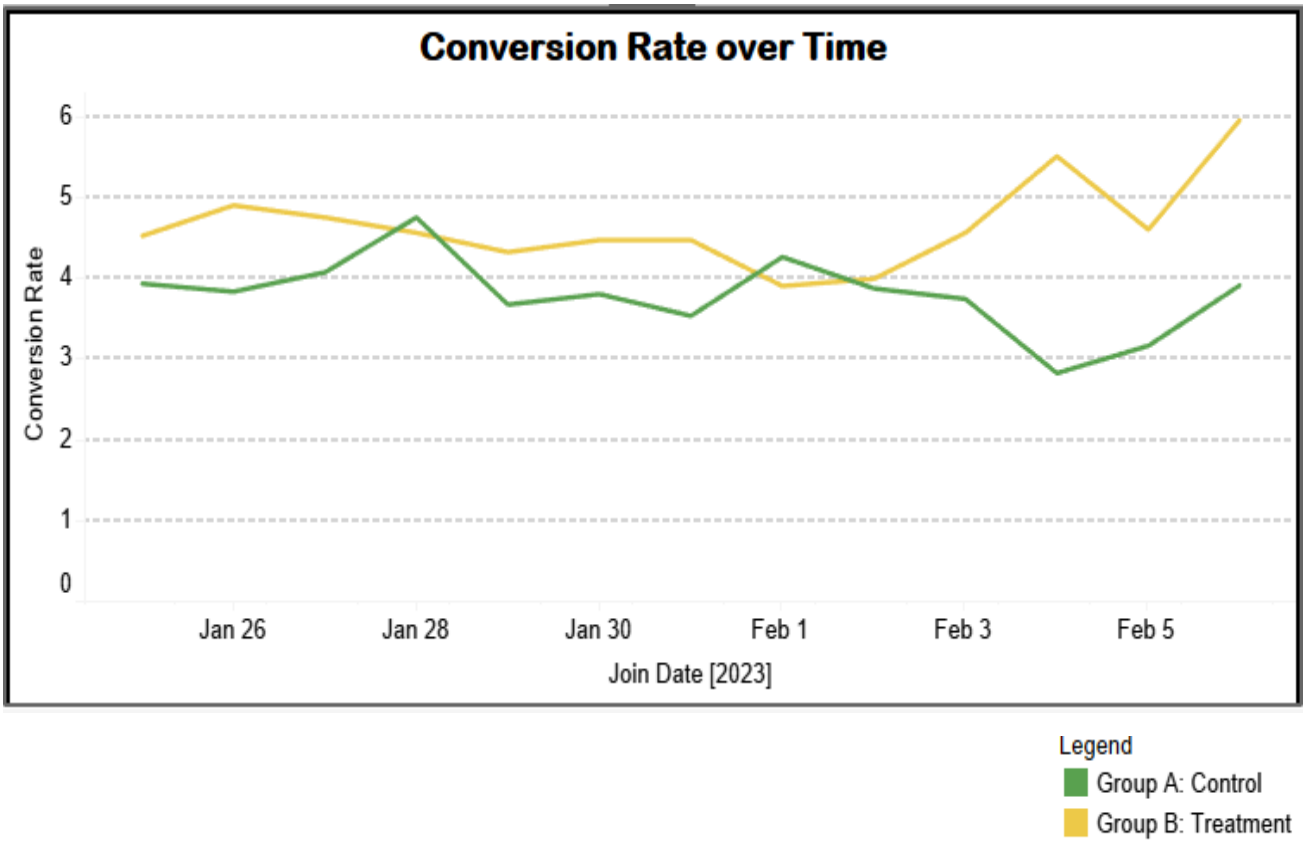
The novelty effect helps us study and understand users' behaviour and patterns in both test groups.



Observation:

- We can observe from the visuals that the Control group has a higher average amount spent over time than the treatment group.
- We can also observe no clear, consistent trend in a significant novelty effect. While there are some fluctuations, the differences between the Groups are not consistently significant enough to suggest a novelty effect.

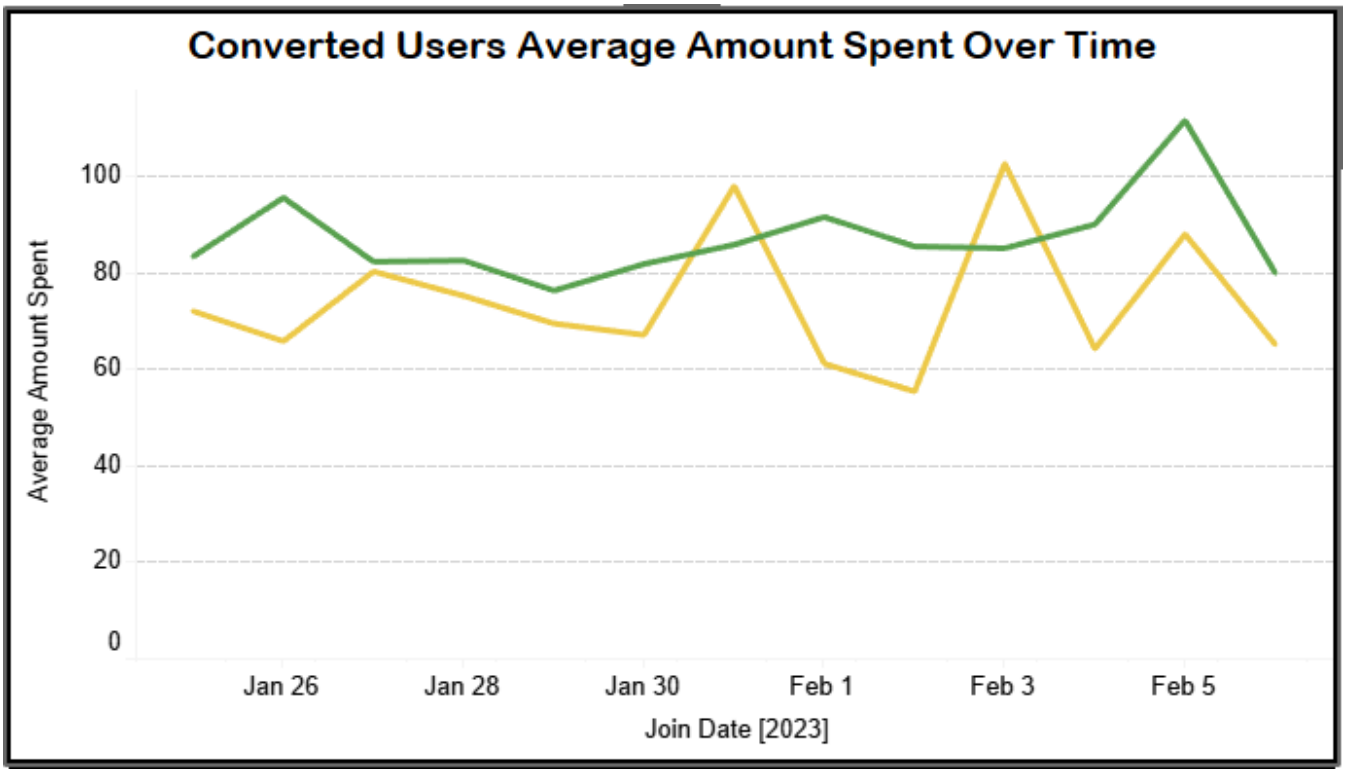
Conversion Rate for All Users



Observation:

The Conversion rates did not exhibit consistency and significant differences between the control and treatment groups. The fluctuations indicate that there is no novelty effect on the conversion rate.

Average Amount Spent by Converted Users



Observation:

- The visuals show variations in the average amount converted users spend, but this pattern does not indicate a novelty effect. There is no significant difference between the Control and Treatment groups.

Power Analysis

We implemented power analysis to prevent p-hacking. We used the power analysis online calculator to determine the appropriate sample size for the experiment to ensure a high probability that we can correctly REJECT or fail to Reject the Null Analysis and that there is no statistically significant difference between the Control and Treatment groups.

The study compared the conversion rates of a treatment group to a control group, with the control group conversion rate of $3.9 considered the baseline and a minimum detectable effect of 10%.

The Power analysis yielded a recommended total sample size of 60,600. This sample size is expected to provide adequate statistical power for detecting the desired effect size.

We also compared two independent means. For this analysis, the difference between the means was set at 10% of the control group's mean, resulting in a value of 0.337. The expected standard deviation used for this analysis was 24.67 units.

Based on these inputs, the power analysis determined that a total size of 182,164 would be required. This sample size was calculated to achieve a power of 80% and a significance level of 5%(two-sided) for detecting an actual difference in means of 0.337 units, assuming a pooled standard deviation of 25.67 units.

The power analysis results indicate that we had a sufficient sample size to evaluate the impact on conversion rates confidently. The effect size for the conversion rate was much larger (18%) than the minimum detectable effect of 10%, and the power analysis confirmed that the sample size used in the A/B test was adequate to detect this effect.

The effect size for the average spend was much smaller (0.5%) than the minimum detectable effect of 10%. The power analysis showed that a much larger sample size would be required to detect such a small effect with the desired statistical power. However, even with a larger sample size, the difference in average spending between the control and treatment groups may remain negligible.

The Power analysis confirms that the sample size used in the A/B test was sufficient to detect a substantial effect on conversion rate but not practically meaningful for detecting

the small effect on the average amount spent, making the focus on strategies to improve conversion rate.

Recommendation

Based on the result of the A/B experiment and further insights, there is sufficient statistical information and evidence to make a confident recommendation to **LAUNCH the Food and drink banner on our webpage**.

The A/B test result demonstrated a statistically significant difference in conversion rates between the control and treatment groups, indicating that the banner positively impacts driving conversions, leading to high revenue and business growth for the organization.

Although there is no statistically significant difference in the average amount spent between Group A and Group B, at this moment, we recommend that we increase the sample size of the dataset for an extendable amount of time and check if there would be a significant difference in the average amount spent between both groups.

Finally, the banner is a relatively inexpensive feature to launch regarding engineering and operational overhead, so we can always make alterations to the banner design to meet the market needs without affecting the overall project cost.

We recommend launching the food and drink category banner while conducting a more detailed analysis of user spending patterns to tailor the marketing strategies accordingly.

Appendix

1. Video Presentation: Video Presentation
2. Link to PowerPoint Presentation: Presentation Slide
3. Link to SQL Data Extraction: SQL Extraction
4. Link to Excel Spreadsheet File: Excel Spreadsheet
5. Link to Tableau: Tableau file