

A/B Test Analysis for Globox Food and Drink Banner Experience

Bradley Hazelton

2023-06-29

Summary

This report presents the results of an A/B test conducted to evaluate the impact of a new banner experience on user behavior on the Globox mobile website. The banner is to bring awareness to the food and drink offerings. The test divided users into two groups, a control group that saw the original site and the test group that saw the new banner. The primary metrics of interest were the conversion rate and the average amount spent per user. The test occurred between January 26 and February 6.

The analysis revealed a statistically significant increase in the conversion rate for the test group compared to the control group. This suggests that the banner led to a higher proportion of the users making a purchase. However, there was not a significant difference in the average amount spent per user between the two groups.

Based on these results, the recommendation is to launch the banner to all users. There is also data to suggest a delay the launch to expand the sample size to confirm the results and further investigate the impact on the average amount spent per user.

The following sections provide a detailed description of the test, the data analysis process, and the rationale behind these conclusions.

Context

The A/B test was conducted on the Globox mobile website over a period of 12 days, from January 26 to February 6. The purpose of the test was to evaluate the impact of a new banner promoting the food and drink offerings on user behavior.

The new food and drink banner was introduced with the expectation that it would enhance user engagement and encourage more purchases by highlighting the variety of offerings available. If successful, this could lead to increased sales and revenue for GloBox.

Users visiting the site during the test period were randomly assigned to the control or test group. The control group saw the original site and the test group saw the food and drink banner. The assignment of users was done randomly to ensure that the two groups were comparable in terms of user characteristics as described below.

The primary metrics of interest for the test were the conversion rate and the average amount spent per user. The conversion rate was defined as the proportion of users who made a purchase during their visit or subsequent visits to the site. The average amount spent was calculated as the total amount spent divided by the number of users. Both metrics were calculated for each group separately. These metrics are important as an increase in one or both would be an overall increase in revenue for GloBox.

The dataset used for the analysis includes data on approximately 49,000 users who visited the site during the test period. The data collected was the group they were assigned to, whether they made a purchase or not, the amount they spent, and other user characteristics such as gender, device, and country of residence.

During the data exploration process it was discovered that there were many null values in Country (643), Gender (6855), and device used (294). The decision to include these rows of data was made as this would have removed a large portion of the test. The null value was replaced with "Unknown" for country or "U" for gender and device to represent the missing data during the export from the PostgreSQL database. The SQL Query can be reviewed in Appendix III, Query 7.

Results

Data Overview

There were 48,943 total users in the A/B test with 24,343 in the Control group and 24,600 in the Test group. The Control group had a conversion rate of 3.923% while the Test group had a conversion rate of 4.630%. The average spent in the Control group was \$3.37 and was \$3.39 in the Test group.

In terms of user characteristics, most of the users were from the United States followed by Brazil and Mexico. The gender distribution was approximately equal with there being slightly more females in the Control group (10,069 vs 10,054) and slight more males in the Test group (10,235 vs 10,061). In addition there were 1,669 (808 in Control, 861 in Test) users of 'Other' gender (Transgender, Non-Binary, agender, genderfluid, and other gender identities). There were also 6,855 users with unknown gender split very evenly between the groups. There were approximately 5,900 more Android users than iOS users in each group. In every country we had users from Android holding approximately 62% of the devices.

The spreadsheet file for the following 4 calculations and tests are found in the Excel file, 'globox data.xlsx' referenced in Appendix IV.

Hypothesis Testing for Conversion Rate

We are testing to see if there is a statistically significant difference in the conversion rate of the two groups. Excel was used to complete the calculations. We are using a significance level of 0.05

H0: $p_1 = p_2$ (The conversion rates in the control and treatment groups are equal)

HA: $p_1 \neq p_2$ (The conversion rates in the control and treatment groups are not equal)

Control Group: 955 Conversions from 24,343 users for a conversion rate of 0.03923099

Test Group: 1139 Conversions from 24,600 users for a conversion rate of 0.046300813

Pooled Proportion: $(955 + 1139) / (24343 + 24600) = 0.042784464$

Standard Error: $\sqrt{(0.042784464 * (1 - 0.042784464) * (1 / 24343 + 1 / 24600))} = 0.001829526$

Test Statistic: $(0.03923099 - 0.046300813) / 0.001829526 = 3.86429177$

P-Value: $2 * (1 - \text{NORM.S.DIST}(3.86429177, \text{TRUE})) = 0.000111$

Since P is less than α we reject the null hypothesis, there is a statistically significant difference in the conversion rate between the groups.

95% Confidence Interval for Conversion Rate

The 95% confidence interval for the difference in the conversion rate between the treatment and control groups (treatment - control) is approximately [0.0035, 0.0107].

Standard Error Unpooled: $\sqrt{((0.03923099 * (1 - 0.03923099) / 24343) + (0.046300813 * (1 - 0.046300813) / 24600))} = 0.001828488$

Difference in Conversion Rates: 0.00706982

Lower Bound: $0.00706982 - 1.96 * 0.001828488 = 0.003485985$

Upper Bound: $0.00706982 + 1.96 * 0.001828488 = 0.010654$

Hypothesis Testing for Average Spent

We are testing to see if there is a statistically significant difference in the average spend between the two groups. Again, Excel was used to complete the calculations. We are using a significance level of 0.05 and assuming unequal variances and used a Welch's t-test.

H0: $\mu_1 = \mu_2$ (The average amount spent per user in the control and treatment groups is equal)

HA: $\mu_1 \neq \mu_2$ (The average amount spent per user in the control and treatment groups is not equal)

Control Group average spent is 3.37451752 with a Standard Deviation of 25.93585054

Test Group average spent is 3.390866667 with a Standard Deviation of 25.41358872

Standard Error: $\sqrt{((25.93585054^2 / 24343) + (25.41358872^2 / 24600))} = 0.232135762$

T-Test p-value: 0.943853 (Using T.TEST() function in Excel)

Since this P-value is greater than α we fail to reject the null hypothesis. This means that there is not a statistically significant difference in the average amount spent between the two groups.

95% Confidence Interval for Conversion Rate

The 95% confidence interval for the difference in the average spent between the treatment and control groups (treatment - control) is approximately [-0.439, 0.471].

Lower Bound:

$(3.390866667 - 3.37451752) - T.INV(0.975, (24343 + 24600) - 2) * 0.232135762 = -0.43863984$

Upper Bound:

$(3.390866667 - 3.37451752) + T.INV(0.975, (24343 + 24600) - 2) * 0.232135762 = 0.471338$

Exploratory Data Analysis

The analysis by gender shows the following results:

- For females, the conversion rate in the test group (5.44%) was slightly higher than in the control group (5.14%), and the average amount spent was slightly lower in the test group (\$4.13) compared to the control group (\$4.46).
- For males, the conversion rate in the test group (3.79%) was higher than in the control group (2.63%), and the average amount spent was also higher in the test group (\$2.60) compared to the control group (\$2.25).
- For users of other genders, the conversion rate was slightly lower in the test group (3.02%) compared to the control group (3.22%), and the average amount spent was almost the same in both groups (\$2.77).

For detailed charts on conversion rate and average spend by gender, see Appendix II, Figures 3 and 5

The analysis by device shows the following results:

- For Android users, the conversion rate in the test group (3.52%) was higher than in the control group (2.77%), and the average amount spent was also slightly higher in the test group (\$2.47) compared to the control group (\$2.31).
- For iOS users, the conversion rate in the test group (6.47%) was higher than in the control group (5.85%), but the average amount spent was slightly lower in the test group (\$4.92) compared to the control group (\$5.07).
- For users with unknown devices, the conversion rate was higher in the test group (4.08%) compared to the control group (2.04%), but the average amount spent was lower in the test group (\$3.46) compared to the control group (\$6.54).

For detailed charts on conversion rate and average spend by gender, see Appendix II, Figures 4 and 6

The analysis by country shows the following results for the top 4 countries represented in this test:

- For USA there were 7,309 users in the control group and 7,463 users in the test group. The conversion rate in the test group (4.77%) was higher than the control group (4.21%), and the average spend was slightly higher in the test group (\$3.99) compared to the control group (\$3.93)

- For Brazil there were 4,805 users in the control group and 4,629 users in the test group. The conversion rate in the test group (4.63%) was higher than the control group (3.92%), and the average spend was almost the same with the test group (\$3.39) spending \$0.02 more than the control group (\$3.37)
- For Mexico there were 2,815 users in the control group and 2,923 users in the test group. The conversion rate in the test group (4.62%) was higher than the control group (3.93%), and the average spend was almost the same with the test group (\$3.39) spending \$0.01 more than the control group (\$3.38)
- For Germany there were 1,906 users in the control group and 1,948 users in the test group. The conversion rate in the test group (4.63%) was higher than the control group (3.91%), and the average spend was almost the same with the test group (\$3.39) spending \$0.03 more than the control group (\$3.36)

For detailed charts on the distribution by country, see Appendix II, Figures 7-10, 15

Further Analysis

Novelty Test

After observing a significant difference in the conversion rates between the control and test groups, we conducted a Novelty Effect Test. The novelty effect refers to the tendency of an individual to respond more positively to a new experience, object, or piece of information than to a familiar one. In the context of an A/B test, this could mean that users might change their behavior just because something is new, not necessarily better. Therefore, it's important to check if the observed effect was due to the novelty of the change.

In this test, we plotted the daily average conversion rate and observed a general downward trend. (Appendix II: Figure 12) We also plotted the daily visits to the mobile website, which showed the same general downward trend. The parallel trends in both the conversion rate and site visits suggest that there was no Novelty Effect in this test. The observed increase in conversion rate in the test group is likely due to the new banner, rather than just a response to something new.

Conversion Rate Over Time

To further understand the impact of the new banner on user behavior, we also analyzed the conversion rate of the control and test groups over the span of the test. This analysis helps to identify any trends or patterns in the conversion rate over time, which could provide additional insights into the effectiveness of the new banner.

We plotted the daily conversion rate for both the control and test groups (Appendix II, Figure 11). The plot showed that the conversion rate for the test group was consistently higher than the control group throughout the test period. This consistent difference further supports the conclusion that the new banner had a positive impact on the conversion rate.

There were two instances where the control group conversion rate was slightly higher than the test group, January 28 (4.76% vs 4.57%) and February 1 (4.27% vs 3.91%). As the groups were assigned randomly this can occur on occasion, but on the other 11 days of the test the test group had a higher conversion rate.

This analysis further confirms the results of the A/B test, showing that the new banner led to a higher conversion rate.

Sample Size Calculation

Using Python we calculated the required sample size for an accurate and reliable result. (Appendix I)

Baseline conversion rate (p_1) = 3.92% = 0.0392

Minimum detectable effect (d) = 5% of the baseline conversion rate = $0.05 * 0.0392 = 0.00196$

Desired significance level (α) = 0.05

Desired statistical power ($1 - \beta$) = 0.8

The sample size should be approximately 124,132 in each group, or almost five times the size of the current test.

Conclusions and Recommendations

In this analysis we found that there was a statistically significant difference in conversion rate between the control and test group, suggesting that the new banner had a positive impact on user behavior. While there was no significant difference in the average amount spent per user between the two groups, the increase in the number of users making a purchase in the test group implies a potential for increased revenue.

The test group out converted the control group by approximately 12.47%. This significant increase suggests that the new banner had a substantial positive impact on user conversion rates.

Based on these findings, we would recommend the banner be launched for all users as there appears to be an overall positive effect on conversions and by extension, revenue. Even though the sample size is smaller than the ideal calculated size there was an immediate impact on the conversion rate that persisted during the test. If time and resources permit, it would be beneficial to conduct further testing with a larger sample size to confirm these findings.

Additionally, our analysis revealed a higher conversion rate among female users compared to male and other gender users. Understanding the factors behind this difference could provide valuable insights for improving targeting strategies and further increasing conversion rates. Future research could focus on exploring these gender differences in more detail.

Appendix I: Python Code for Sample Size Calculation

Code can be found in a Notable notebook at

https://app.noteable.io/f/6277107e-40d9-4290-bd08-8dfe060d5a7e/sample_size_calculation.ipynb

```
import math
from scipy.stats import norm
from scipy.special import erfinv

# Given values
p1 = 0.0392 # Baseline conversion rate
d = 0.00196 # Minimum detectable effect
alpha = 0.05 # Significance level
beta = 0.2 # 1 - Statistical power

# Calculating critical values
z_alpha_2 = abs(math.sqrt(2) * erfinv(2 * alpha - 1))
z_beta = abs(math.sqrt(2) * erfinv(2 * beta - 1))

# Calculating pooled probability
p_bar = p1 + d / 2

# Calculating sample size
n = ((z_alpha_2 * math.sqrt(2 * p_bar * (1 - p_bar)) + z_beta * math.sqrt(p1 * (1 - p1) + (p1 + d) * (1 - (p1 + d)))) / d) ** 2
n = math.ceil(n) # Rounding up to the nearest whole number because you can't have a fraction of a sample

print(f"The required sample size is approximately {n} in each group.")
```

Appendix II: Charts from Tableau

Figure 1: Conversion Rate with Confidence Range

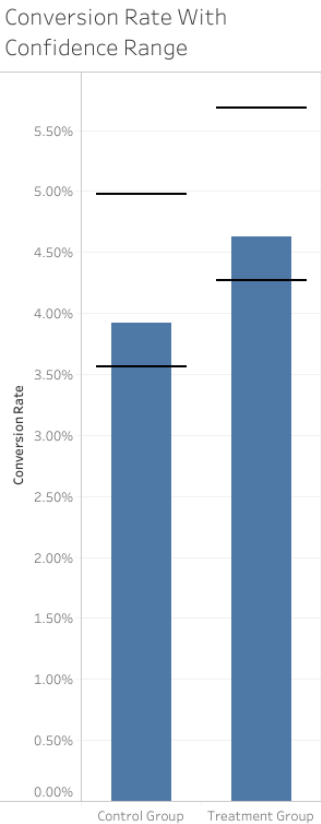


Figure 2: Average Spend with Confidence Range

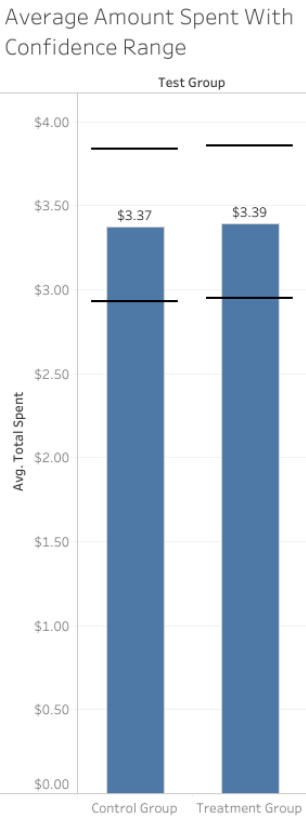


Figure 3: Conversion Rate by Gender

Conversion Rate By Gender

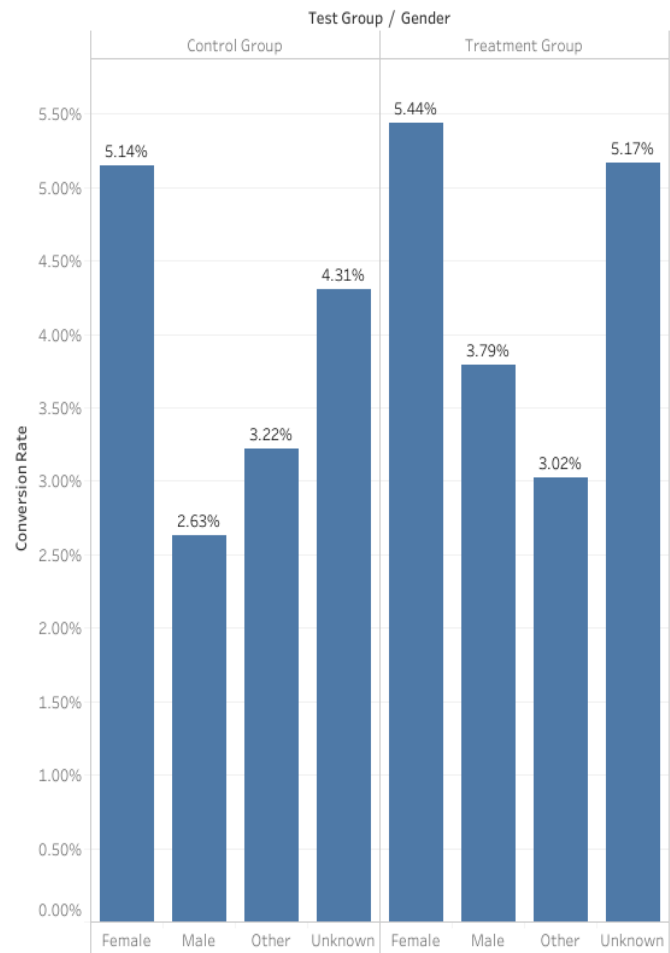


Figure 4: Conversion Rate by Device

Conversion Rate By Device

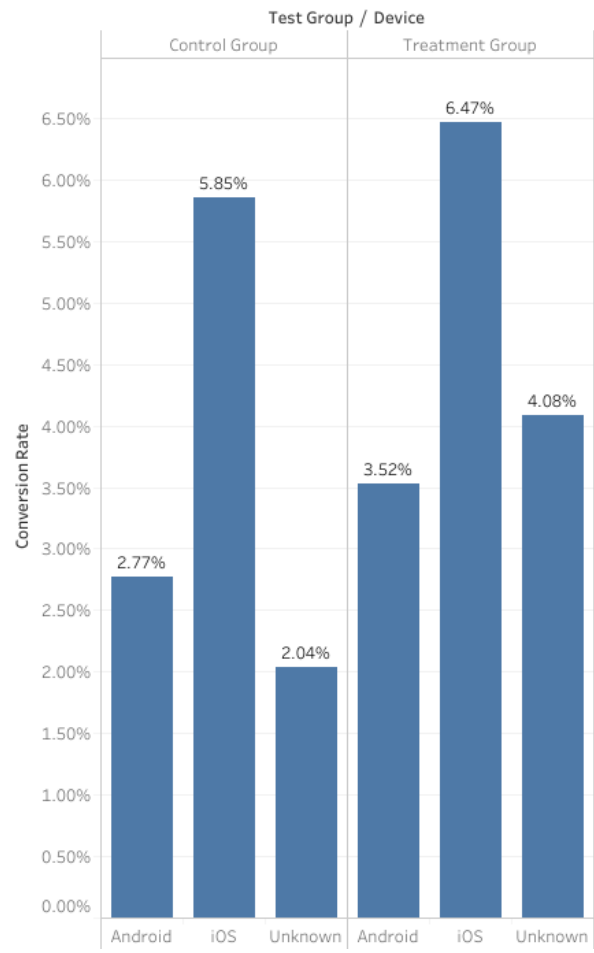


Figure 5: Average Spent by Gender
Average Amount Spent by Gender

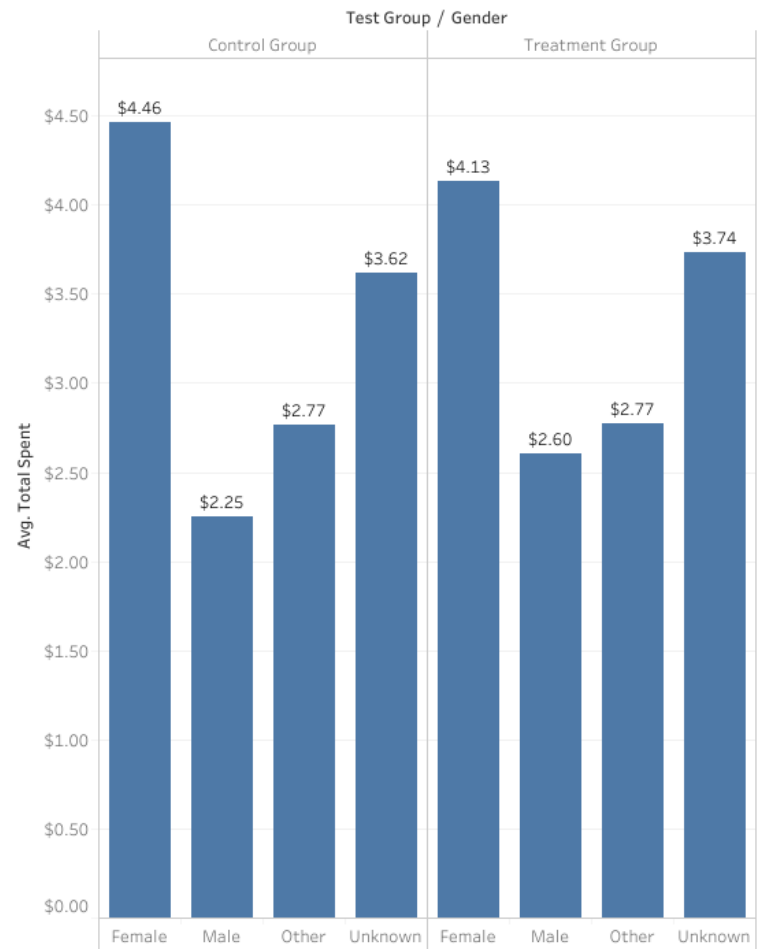


Figure 6: Average Spent by Device
Average Amount Spent by Device



Figure 7: Users by Country

Users By Country

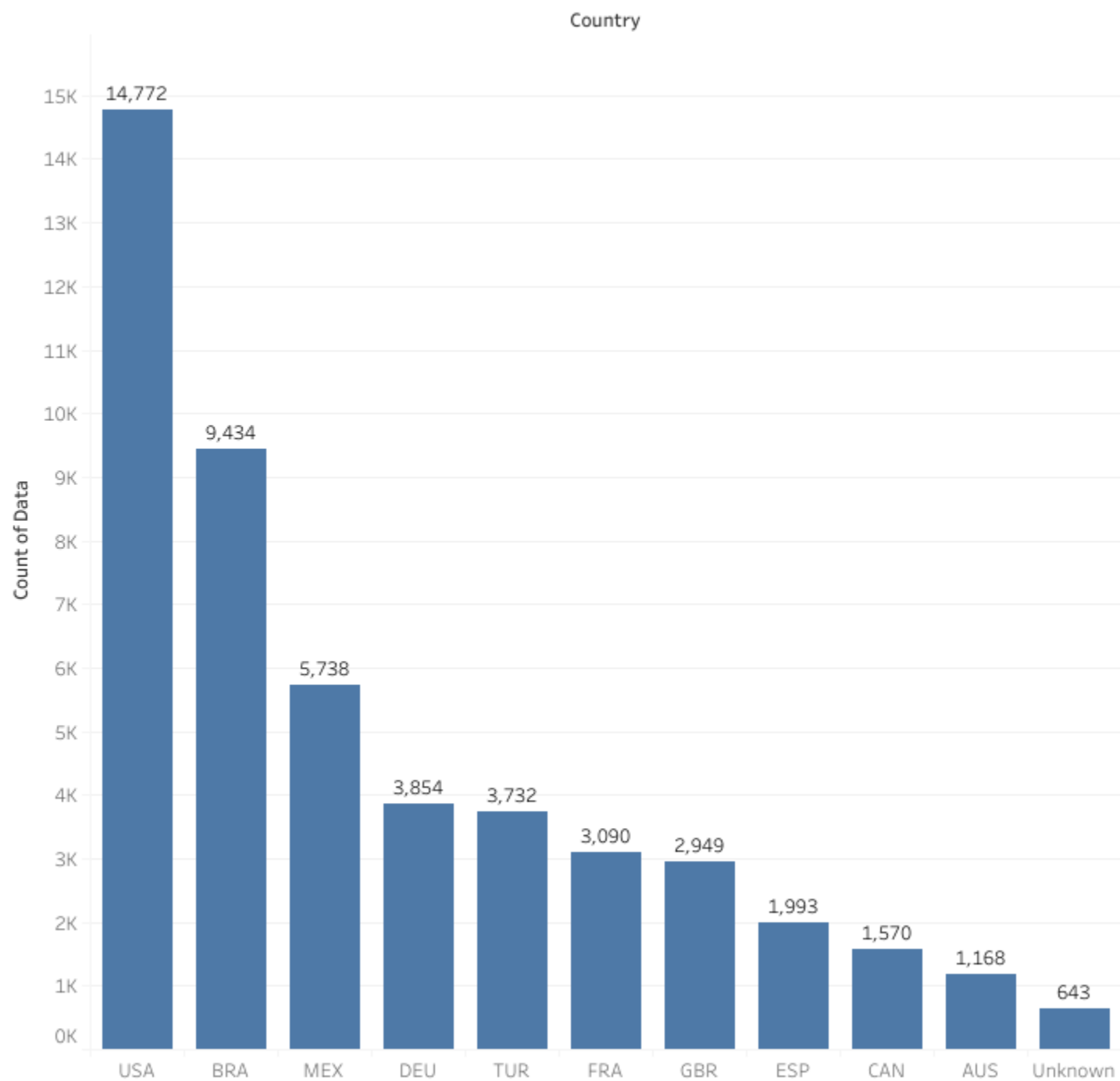


Figure 8: Users by Country in Test Groups

Users By Country by Test Group

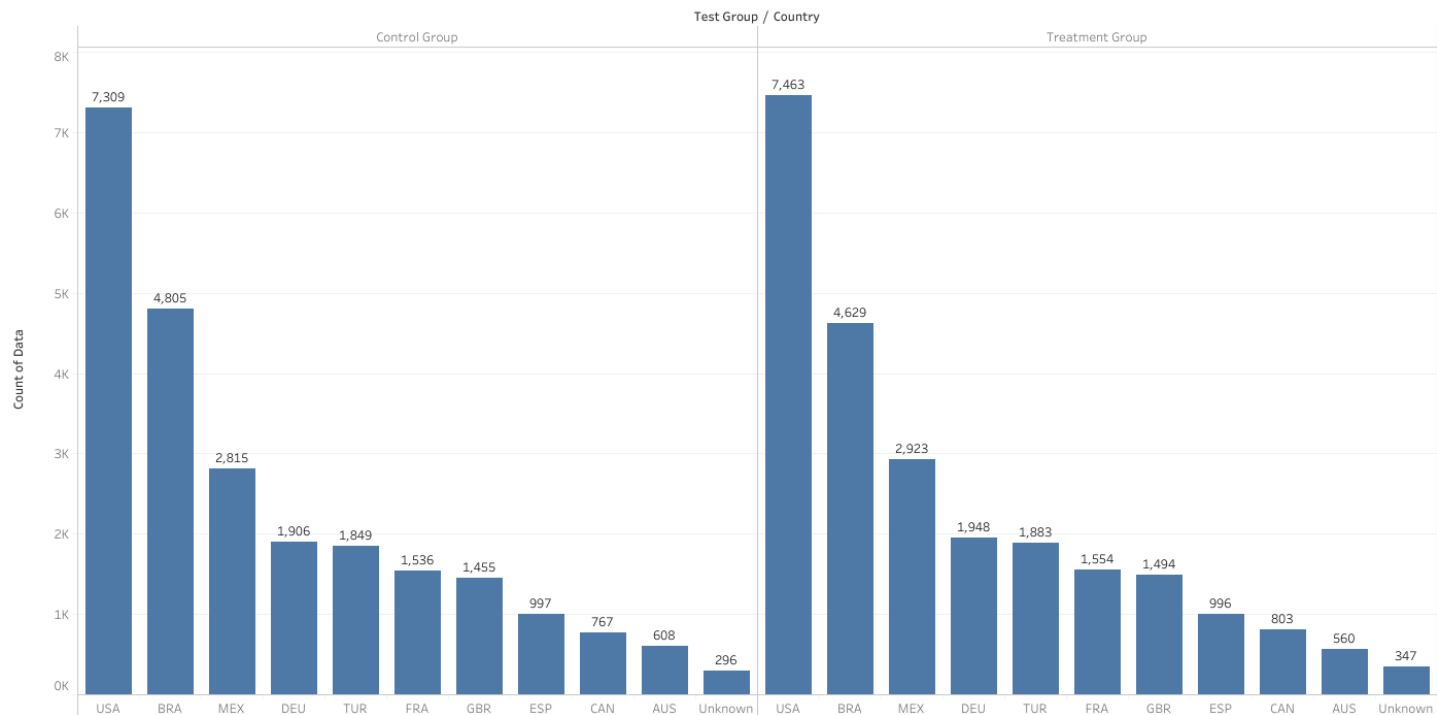


Figure 9: Conversion Rate by Country

Conversion Rate by Country

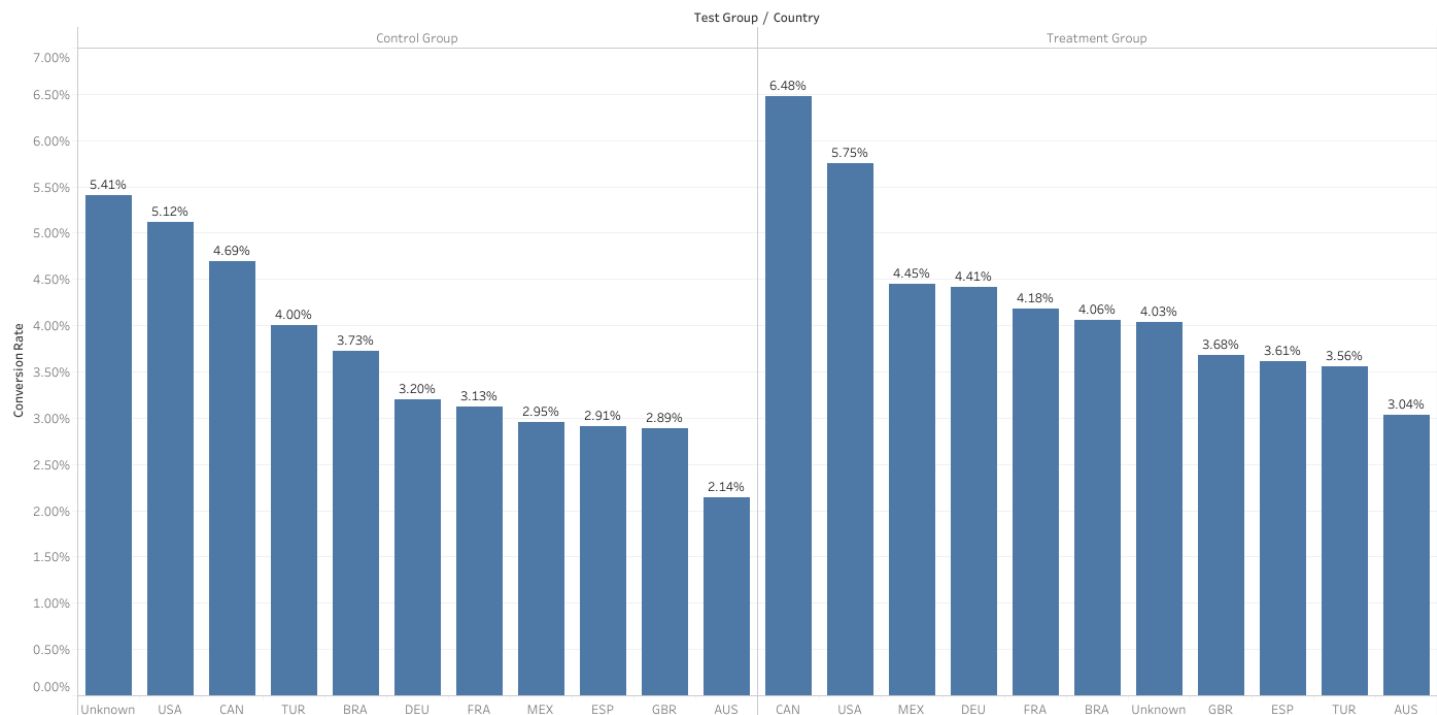


Figure 10: Average Spent by Country

Average Spend by Country

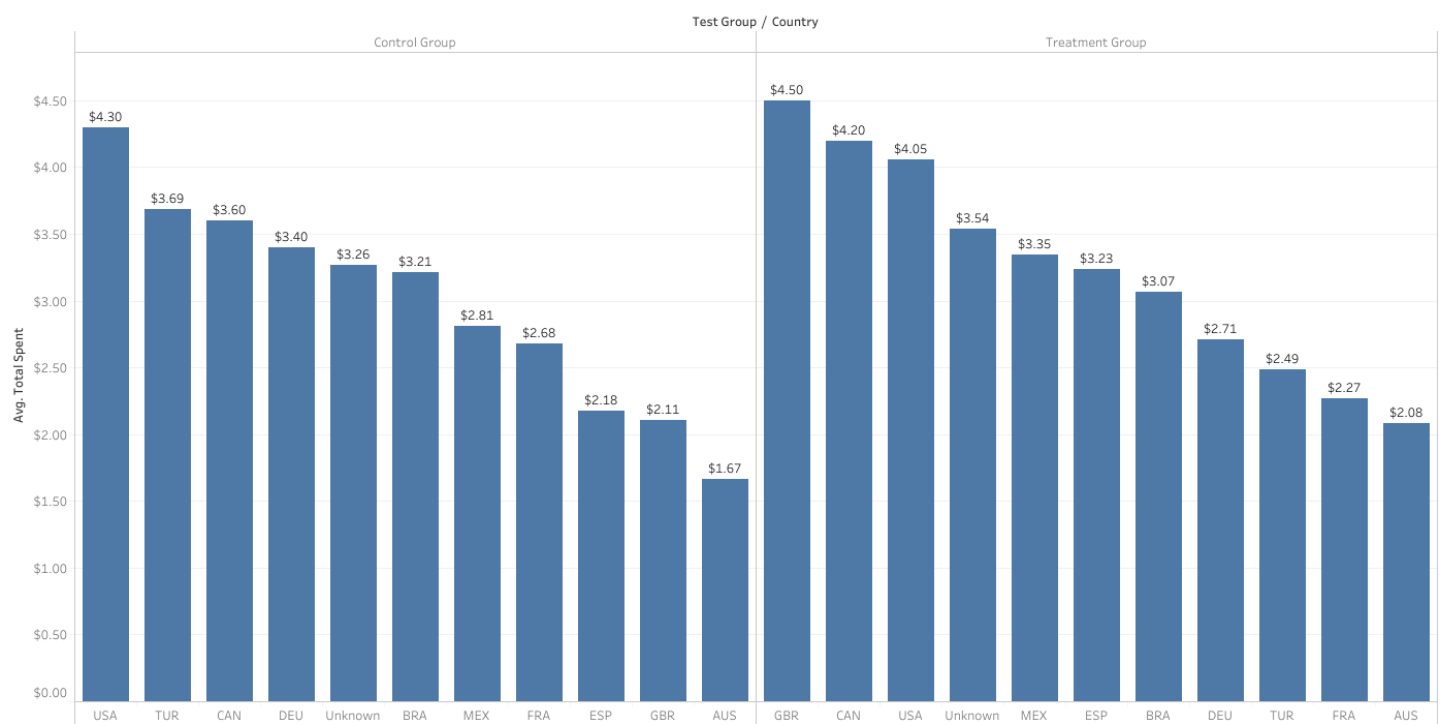


Figure 11: Conversion Trend

Conversion Trend

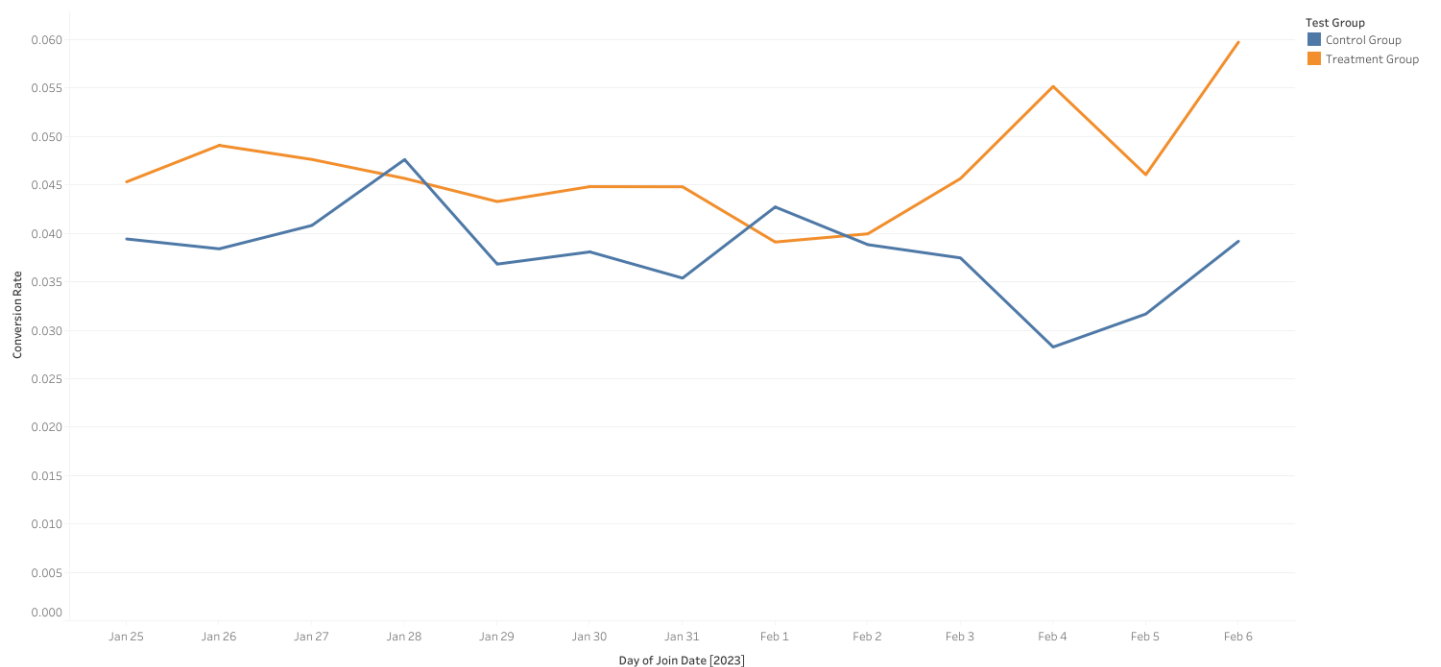


Figure 12: Novelty Test

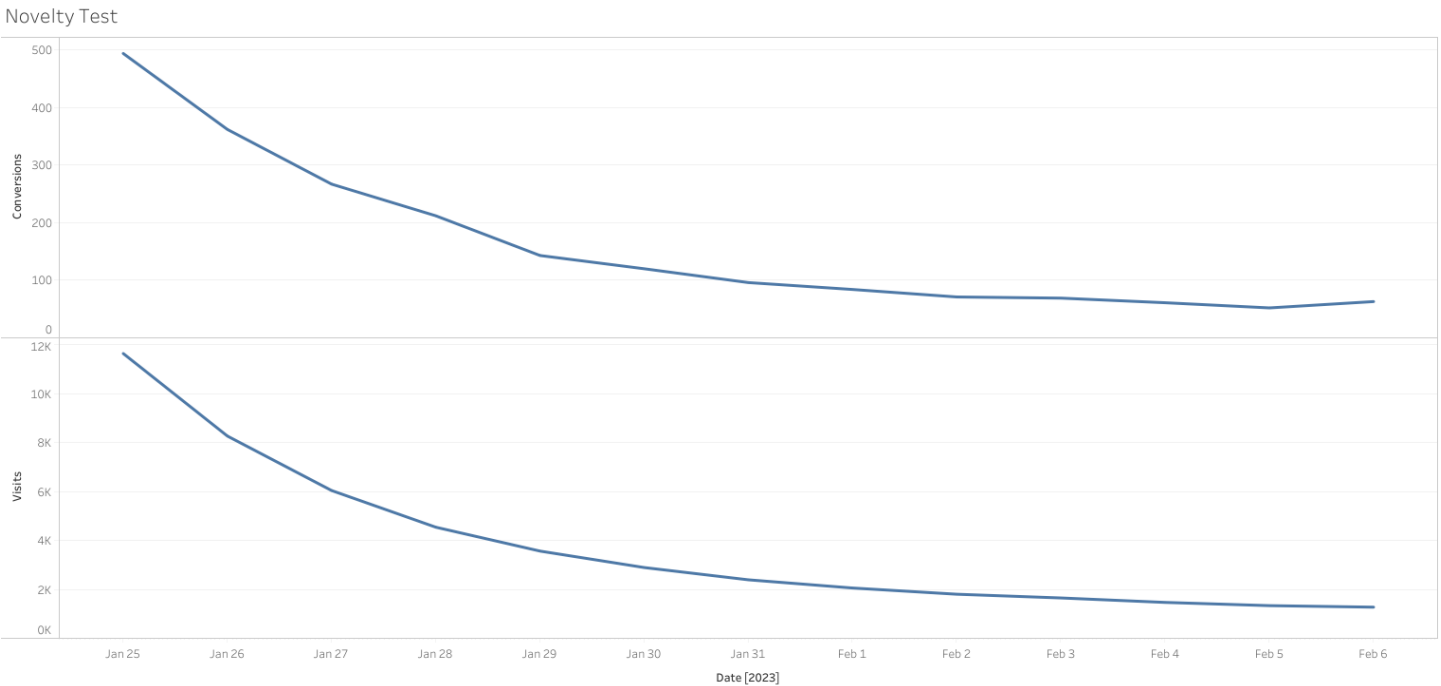


Figure 13: Device by Group

Device By Group

Device	Test Group	
	Control Group	Treatme nt Group
Android	15,054	15,235
iOS	9,142	9,218
Unknown	147	147

Figure 14: Gender by Group

Gender By Group

Gender	Test Group	
	Control Group	Treatme nt Group
Female	10,069	10,061
Male	10,054	10,235
Other	808	861
Unknown	3,412	3,443

Figure 15: Device by Country

Device By Country

	Country (group) / Country										
	Developed Countries							Emerging Countries			Unkno..
Device	AUS	CAN	DEU	ESP	FRA	GBR	USA	BRA	MEX	TUR	Unkno..
Android	720	958	2,400	1,241	1,915	1,841	9,159	5,827	3,509	2,297	422
	61.64%	61.02%	62.27%	62.27%	61.97%	62.43%	62.00%	61.77%	61.15%	61.55%	65.63%
iOS	440	603	1,432	746	1,159	1,087	5,519	3,547	2,194	1,416	217
	37.67%	38.41%	37.16%	37.43%	37.51%	36.86%	37.36%	37.60%	38.24%	37.94%	33.75%
Unknown	8	9	22	6	16	21	94	60	35	19	4
	0.68%	0.57%	0.57%	0.30%	0.52%	0.71%	0.64%	0.64%	0.61%	0.51%	0.62%

Appendix III: SQL Queries

Query 1: Obtain Start and End Dates

```
SELECT MIN(join_dt) AS start_date, MAX(join_dt) AS end_date
FROM groups;
```

Query 2: Total User Count

```
SELECT COUNT(DISTINCT id) AS total_users
FROM users;
```

Query 3: User Count By Group

```
SELECT "group", COUNT(DISTINCT uid) AS num_users
FROM groups
GROUP BY "group";
```

Query 4: Conversion Rate for All Users

```
SELECT
    COUNT(DISTINCT CASE WHEN a.uid IS NOT NULL THEN a.uid END) AS converted_users,
    COUNT(DISTINCT u.id) AS total_users,
    CAST(COUNT(DISTINCT CASE WHEN a.uid IS NOT NULL THEN a.uid END) AS DECIMAL) /
COUNT(DISTINCT u.id) AS conversion_rate
FROM
    users u
LEFT JOIN
    activity a ON u.id = a.uid;
```

Query 5: Conversion Rate for Each Group

```
SELECT
    g.group,
    COUNT(DISTINCT CASE WHEN a.uid IS NOT NULL THEN a.uid END) AS converted_users,
    COUNT(DISTINCT u.id) AS total_users,
    CAST(COUNT(DISTINCT CASE WHEN a.uid IS NOT NULL THEN a.uid END) AS DECIMAL) /
COUNT(DISTINCT u.id) AS conversion_rate
FROM
    users u
LEFT JOIN
    activity a ON u.id = a.uid
JOIN
    groups g ON u.id = g.uid
GROUP BY
    g.group;
```

Query 6: Average Spent for All Users

```
SELECT
    g.group,
    AVG(COALESCE(a.spent, 0)) AS average_amount_spent
FROM
    users u
LEFT JOIN
    groups g ON u.id = g.uid
LEFT JOIN
    activity a ON u.id = a.uid
GROUP BY
    g.group;
```

Query 7: Join All Tables For .CSV Export

```
SELECT
    u.id AS user_id,
    COALESCE(u.country, 'Unknown') as country,
    COALESCE(u.gender, 'U') AS gender,
    COALESCE(g.device, 'U') AS device,
    g.group AS test_group,
    g.join_dt AS join_date,
    CASE WHEN a.spent > 0 THEN 'Yes' ELSE 'No' END AS converted,
    ROUND(COALESCE(SUM(a.spent), 0)::numeric, 2) 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
    u.id,
    country,
    gender,
    g.device,
    g.group,
    g.join_dt,
    converted;
```

Appendix IV: Excel Data File

‘globox data.xlsx’ includes the raw data used in the analysis on the ‘Data’ worksheet, as well as detailed calculations for conversion rate and Average Spent. It serves as a reference for anyone who wishes to understand the data manipulation and analysis processes in more depth.