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

Mitchell McCallum

2023-12-24

## A screenshot of a phone Description automatically generated

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 user engagement. However, amongst the users making purchases there was a considerably drop in spending leading to lackluster overall growth in revenue.

Based on these results, the recommendation is to retain the original site design but consider launching it to select locations. 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.

One point of note is in this study both average spending among the population and average spending of actual customers are analyzed where applicable we have tried to make the distinction but as a general point of reference average dollar values below $10 dollars will be values among the population but above $10 dollars will be the average spent amongst customers.

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

A diagram of a company

Description automatically generated

## **Summary**

## **Summary of Findings**

## In our recent A/B test conducted on the Globox mobile website from January 26 to February 6, we aimed to discern the influence of a newly introduced food and drink banner on user engagement. The hypothesis was that this banner would amplify user interaction and stimulate more purchases by accentuating available offers, potentially boosting GloBox's sales and revenue.

## During the test, users were randomly allocated to either the control or test group, ensuring comparable user characteristics. The control group experienced the original site interface, while the test group was presented with the new food and drink banner.

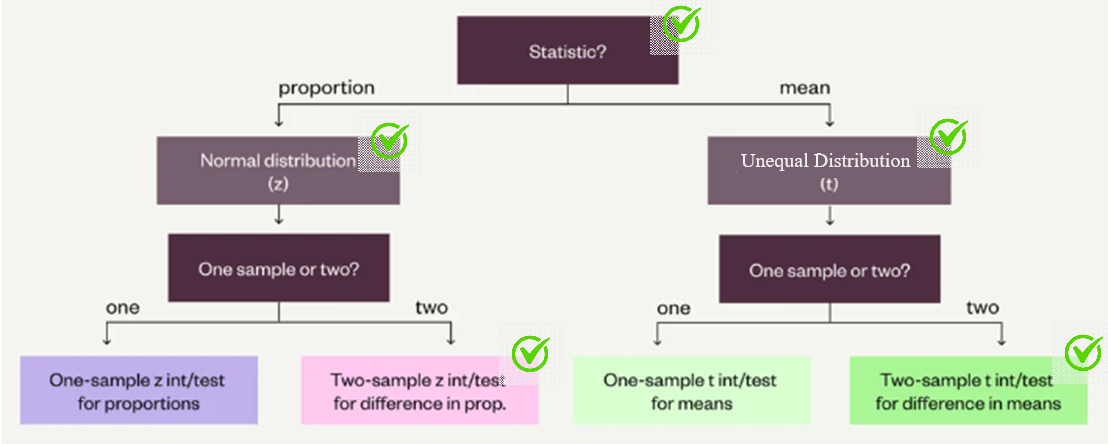
## Our primary metrics were the conversion rate and the average expenditure per user. The conversion rate represents the fraction of users who initiated a purchase during their site visit or subsequent visits. The average spending was deduced by dividing the total expenditure by the number of active users. Both metrics were separately computed for groups 'A' and 'B'. An uptick in either or both metrics would signify a revenue enhancement for GloBox.

## The test group demonstrated a heightened conversion rate, surging by approximately 12.47% in comparison to the control group. However, this was juxtaposed by a notable decline of around 15.53% in the average amount spent per customer between the two groups. This means that while more users in the test group were inclined to make a purchase, they spent less. This resulted in the treatment population only contributing to $1,269.43 (1.53%) of additional gross revenue.

## Given these findings, our counsel is to maintain the original banner for most users. We also advocate for additional testing either in select nations such as the UK where average spending was stronger for the test groups or generally expand the sample size the sample size to corroborate these observations.

## It's worth noting that our dataset, encompassing data on nearly 49,000 users, had several null values in fields like Country (643), Gender (6855), and device used (294). Despite this, the decision was made to retain these data rows, substituting the null values with placeholders like "Unknown" or "U" during the data extraction from the PostgreSQL database. The specifics of the SQL Query can be reviewed in Appendix III, Query 2.

## **Process chart for statistical analysis**



## **Results**

### In our A/B test, we analyzed a total of 48,943 users, with 24,343 allocated to the Control group and 24,600 to the Test group. The conversion rates for the Control and Test groups stood at 3.923% and 4.630% respectively. When examining the average expenditure for active users, the Control group averaged $81.01, while the Test group averaged $68.43. Amongst the entire population we see average spending more on par with treatment group at roughly 3.39 and control group at 3.37. The statistical test conducted on these averages do not lead to a rejection of the null hypothesis thus the original site design. Subsequent sections will elucidate the statistical significance of these disparities in both conversion rate and average expenditure.

### Diving into user demographics, the majority hailed from the United States, trailed by Brazil and Mexico. Looking at users I will describe as uninvested with the brand due to nondisclosure of personal details of ‘gender’, ‘device’ or ‘country’. We see further indications amongst these users that the treatment group (average spent $65.29) has lower quality customers, having spent less than the control group (average spent $79.73). Furthermore we see a higher count of users among the treatment group (197) than the control group (155) being uninvested so the treatment group has 27% more uninvested users.

### For a detailed breakdown of the subsequent four computations and tests, please refer to the Excel document titled 'globox data.xlsx', which can be located in Appendix II Figure 2, Appendix III Fig 1-2.

### 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: p1 = p2 (The conversion rates in the control and treatment groups are equal)  
HA: p1 ≠ p2 (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: √(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 – NORM.S.DIST(3.86429177, 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: √((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 amongst sample population

We are testing to see if there is a statistically significant difference in the average spent between the two groups of the sample population. 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: μ1 = μ2 (The average amount spent per user in the control and treatment groups is equal) HA: μ1 ≠ μ2 (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: √((25.935850542 / 24343) + (25.413588722 / 24600)) = 0.232135762

T-Test p-value: 0.9524038 (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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Group A: Control** | **Group B:**  **Treatment** | **Grand Total** | **P Value** |
| **Total Number of Users** | **24,343** | **24,600** | **48,943** | **N/A** |
| **Conversion Rate** | **3.92%** | **4.63%** | **4.28%** | **0.0001** |
| **Average Amount Spend $** | **$3.37** | **$3.39** | **$3.38** | **0.9438** |

## 95% Confidence Interval for Average spent

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 on profit shows that:

* An overall increase in profitability of $1269.43 or 1.54%.
* A drop in average spending amongst spenders from $81.01 to $68.43 or -15.53%.
* We see the sample for control group at 1014 and the treatment group at 1219 so an increase of 20.21% in activity for an increase of only 1.54% in gross revenue.

For detailed charts on conversion rate and charts breaking down spend rate,, see Appendix II, Fig 1, Appendix III, Fig 1

The analysis for users displaying a level of disinterest in the brand, demonstrated parts of their personal information blank reveal the following results:

* Here we see a higher count of users in the treatment group (197) with the control group (155) an increase of 27% for customers failing to or refusing to disclose their personal details.
* The average spent amongst customers mirrors that of the overall sample $79.73 in the control group, $65.29 for the treatment group .
* 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 far lower in the test group ($72.59) compared to the control group ($320.37).

A graph of different sizes and colors

Description automatically generated with medium confidence

For detailed charts on conversion rate and average spend with NULL values, see Appendix II Fig 10,11 Appendix III, Fig 2

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Control Group Users** | **Test Group Users** | **Control Group Conversion Rate** | **Test Group Conversion Rate** | **Control Group Average Spend** | **Test Group Average Spend** |
| USA | 7,309 | 7,463 | 4.21% | 4.77% | $79.48 | $66.19 |
| Brazil | 4,805 | 4,629 | 3.92% | 4.63% | $80.43 | $73.16 |
| Mexico | 2,815 | 2,923 | 3.93% | 4.62% | $89.95 | $69.35 |
| Germany | 1,906 | 1,948 | 3.91% | 4.63% | $96.74 | $54.95 |

A graph of a bar chart

Description automatically generated with medium confidence

For detailed charts on the distribution by country, see Appendix II Fig 6, Appendix III Fig 3, 4

## 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: Fig 9) 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 9). 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 (p1) = 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 - β) = 0.8

The sample size should be approximately 124,132 in each group, or almost five times the size of the current test. Further data should be collected particularly for the activities table to ensure that these banners lead to quality conversions significantly improving revenue.

## 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. Yet with the minimal increase in revenue (1.54%) derived from the decrease in average spent per customer (-15.53%), the increase in the number of users making a purchase in the test group implies that when factoring for the increased costs to handle the traffic, the banner is likely a net loss for the company.

Based on these findings, we would not recommend the banner be launched for all users as there appears to be a poor effect on the quality of conversions and by extension, net profit. 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, most importantly further data for the users we convert. This would prove particularly valuable to validate quality of conversions data as we hope to ensure more than a mere increase in activity with our findings. Additional information on the costs for handling the increased traffic is recommended as then we can do a CVP analysis to better understand the banner’s impact on the company EBITDA (earnings before interest tax depreciation and amortization).

We would recommend considering checkout suggestions or suggestions on some page for users with items in cart in the marketing pipeline would be a worthwhile feature set as this could increase the average value of customers along with overall products rather than having a few additional sales of low value through the banner pipeline.

Additionally, our analysis revealed both a higher conversion rate as well as higher amount spent among select countries (such as the UK). 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 geographic 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/published/38b1a434-4707-4013-88ad-103a1e51ee93/Sample-size-calculation

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) # Sample size limitations necessitate rounding up to whole numbers

print(f"Necessary sample size in each group is approximately {n}")

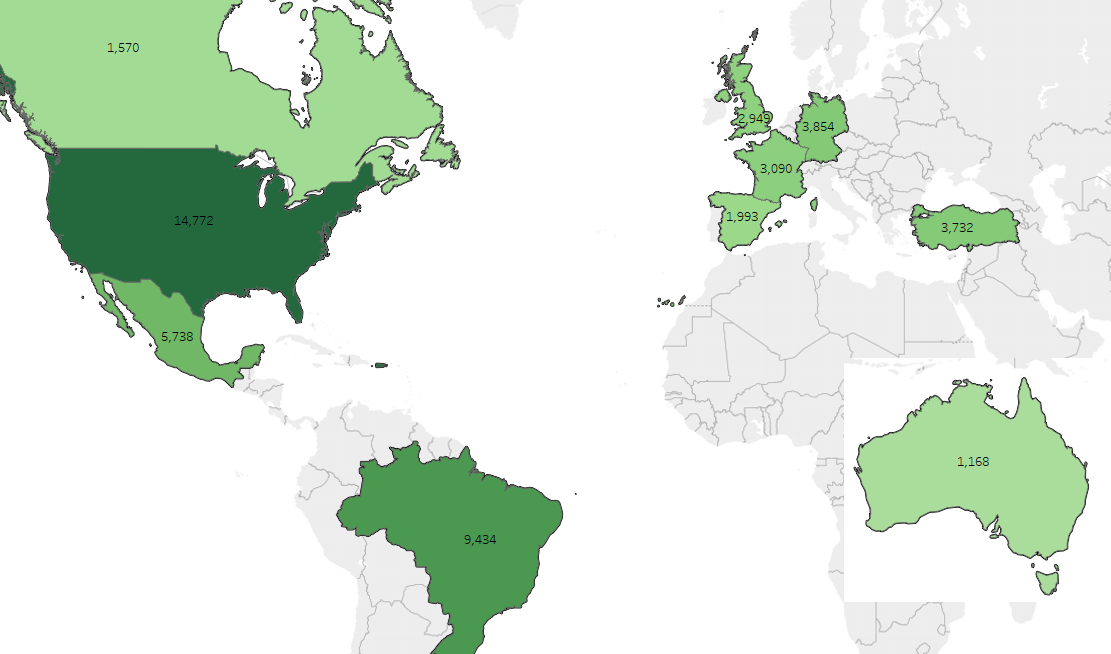
## Appendix II: Charts from Tableau

Tableau Link: [https://prod-ca-a.online.tableau.com/#/site/mitchellportfolio/workbooks/226142?:origin=card\_share\_link](https://prod-ca-a.online.tableau.com/" \l "/site/mitchellportfolio/workbooks/226142?:origin=card_share_link)

|  |  |
| --- | --- |
| Figure 1: Conversion Rate with Confidence Range | Figure 2: Average Spend with Confidence Range |
| A picture containing text, screenshot, diagram, parallel  Description automatically generated | A picture containing text, screenshot, diagram, line  Description automatically generated |

|  |  |
| --- | --- |
| Figure 3: Conversion Rate by Gender | Figure 4: Conversion Rate by Device |
|  |  |

|  |
| --- |
| Figure 5: Users by Country |
|  |

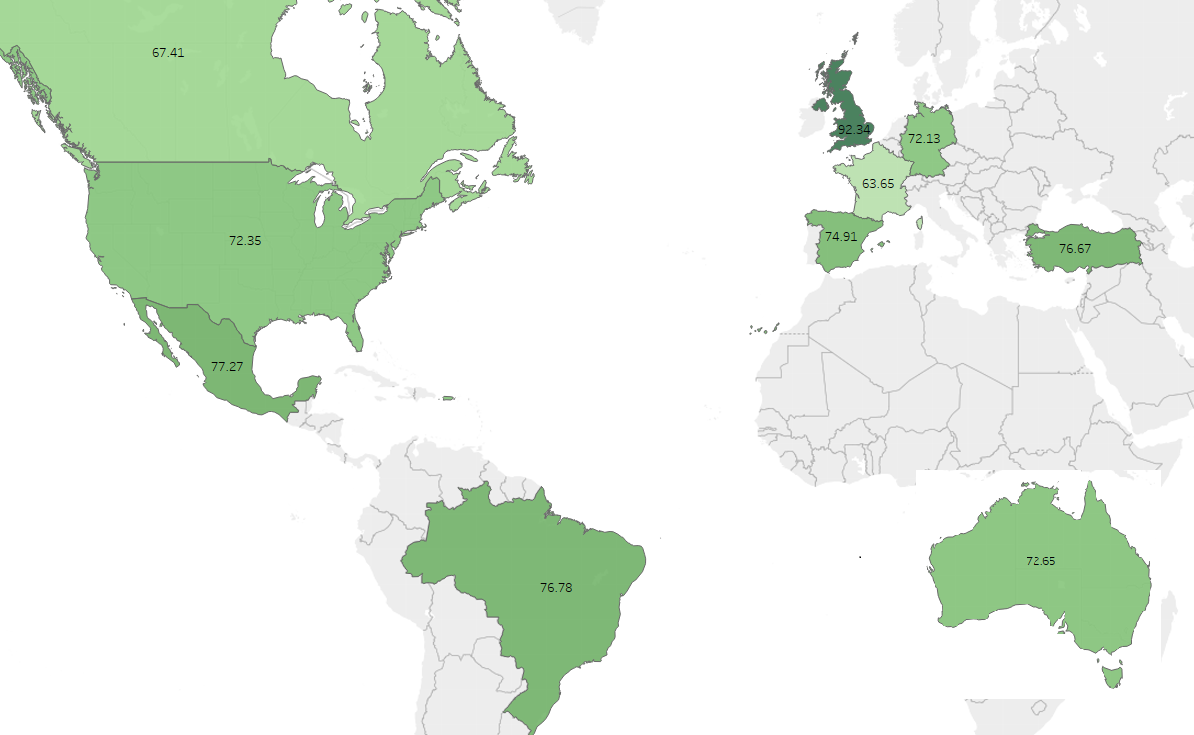


### Figure 6: Users by Country in Test Groups

A graph of a number of blue bars

Description automatically generated with medium confidence

### Figure 7: Average Spent by Country



|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| Figure 8: Conversion Trend |
| A graph of blue and orange lines  Description automatically generated |

|  |
| --- |
| Figure 9: Novelty Test |
| A graph of a person with a beard  Description automatically generated with medium confidence |

|  |  |
| --- | --- |
| Figure 10: Device by Group |  |
| A screenshot of a device  Description automatically generated |  |

|  |
| --- |
| Figure 11: Device by Country |
| A screenshot of a graph  Description automatically generated |

## Appendix III: Charts from Google Sheets

|  |  |
| --- | --- |
| Figure 1: Average Spent and Total Spent among control and treatment group |  |
|  |  |

### Figure 2: Average Spent and user count with Null gender, device or country values

A graph of a bar chart

Description automatically generated with medium confidence

|  |
| --- |
| Figure 3: Average Spent by Country Treatment Group |
|  |

### Figure 3: Average Spent by Country Control Group

A graph with blue and red lines

Description automatically generated

## Appendix III: SQL Queries

### Query 1: Average spent, conversion rate & count of total users BY COUNTRY

SELECT

U.country AS Country,

AVG(A.spent) AS Average\_Spent,

(COUNT(DISTINCT A.uid) \* 100.0 / CAST(COUNT(DISTINCT U.id) AS DECIMAL)) AS Conversion\_Rate,

COUNT(U.id)

FROM

Users U

LEFT JOIN

Activity A ON U.id = A.uid

LEFT JOIN

Groups G ON U.id = G.uid

GROUP BY

U.country

ORDER BY

Average\_Spent DESC;

### Query 2: Obtain Average spent for customers, user count & conversion rate by group

SELECT

G.group AS group,

COUNT(DISTINCT U.id) AS User\_Count,

AVG(A.spent) AS Average\_Spent,

(COUNT(DISTINCT A.uid) \* 100.0 / CAST(COUNT(DISTINCT U.id) AS DECIMAL)) AS Conversion\_Rate

FROM

Users U

JOIN

Groups G ON U.id = G.uid

LEFT JOIN

Activity A ON U.id = A.uid

WHERE

U.country IS NULL OR U.gender IS NULL OR G.device IS NULL

GROUP BY

G.group;

### Query 3: Obtain Total spent for sample population, average spent among customers, along with standard deviation, total users for population sample.

SELECT

G.group AS group,

SUM(A.spent) AS Total\_Spent,

AVG(A.spent) AS Average\_Spent,

STDDEV(A.spent) AS stddev\_spent,

COUNT(DISTINCT U.id) AS total\_users

FROM

Users U

JOIN

Groups G ON U.id = G.uid

LEFT JOIN

Activity A ON U.id = A.uid

GROUP BY

G.group;

### Query 4: Obtain both spend rows and binary 1 or 0 rows with 1 representing conversion to perform advanced statistics and used to determine overall conversion rates.

SELECT

g.uid,

CASE

WHEN a.uid IS NOT NULL THEN 1

ELSE 0

END AS in\_both\_tables,

COALESCE(a.spent, 0) AS spent

FROM

Groups g

LEFT JOIN

Activity a ON g.uid = a.uid;

### Query 5: Obtain conversion count, total users and conversion ratio by gender.

SELECT

U.gender,

COUNT(DISTINCT C.uid) AS Conversions,

COUNT(DISTINCT U.id) AS Total\_Users,

(COUNT(DISTINCT C.uid) \* 100.0 / CAST(COUNT(DISTINCT U.id) AS DECIMAL)) AS Conversion\_Rate

FROM

Users U

LEFT JOIN

(SELECT G.uid FROM Groups G INNER JOIN Activity A ON G.uid = A.uid) C ON U.id = C.uid

GROUP BY

U.gender;

### Query 6: Obtain conversion count, total users and conversion ratio by device.

SELECT

U.device,

COUNT(DISTINCT C.uid) AS Conversions,

COUNT(DISTINCT U.id) AS Total\_Users,

(COUNT(DISTINCT C.uid) \* 100.0 / CAST(COUNT(DISTINCT U.id) AS DECIMAL)) AS Conversion\_Rate

FROM

Users U

LEFT JOIN

(SELECT G.uid FROM Groups G INNER JOIN Activity A ON G.uid = A.uid) C ON U.id = C.uid

GROUP BY

U.device;

## Appendix IV: Google Sheets Data File

Can be found at <https://docs.google.com/spreadsheets/d/1XoIvEKfch3UWNTMCqyjU8XGlQsCDEWNuiVrO_7j3LiA/edit?usp=sharing>

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