GloBox A/B Test Analysis:
Food & Beverage Banner

By: Olivia Hamilton

17 July 2023

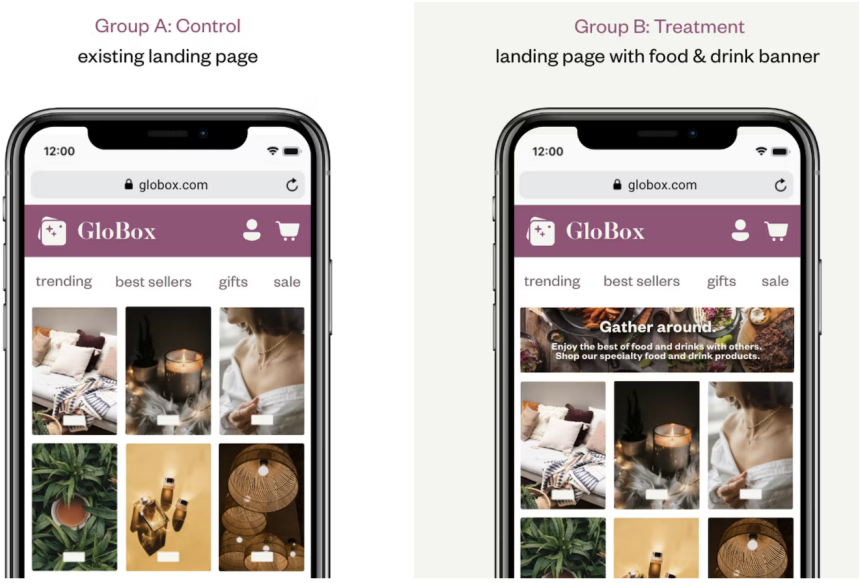
Summary

An A/B Test was performed to determine the effectiveness of a banner strategically placed at the top of the GloBox website highlighting the increased offerings in the food and drink category. The two main metrics being assessed for the overall objective of increasing revenue were "Average Amount Spent per User" and "Conversion Rate". While we did not see a significant difference in the Average Amount Spent per User across the two test groups, we did see a significant increase in the Conversion Rate for the Treatment Group who were assigned the banner. Based on my analysis I am confident that the increase in conversion rate is due to the banner. It's my recommendation that GloBox launch the banner for all users.

Context

GloBox, an international Online Marketplace primarily known for its boutique fashion items and high-end decor products, has recently seen tremendous growth in its food and drink offerings in the last few months, the company wanted to bring awareness to this product category with the primary objective of increasing revenue.

The Growth team launched an A/B test consisting of 48,943 users displaying a banner highlighting the food and beverage category at the top of the website. The Control Group (Group A) consisting of 24,343 users, sees the original website and is not shown the banner, while the Treatment Group (Group B) consisting of 24,600 users was shown the banner at the top of the website (for reference see the image below).



A/B Test Fields and Setup of the Experiment

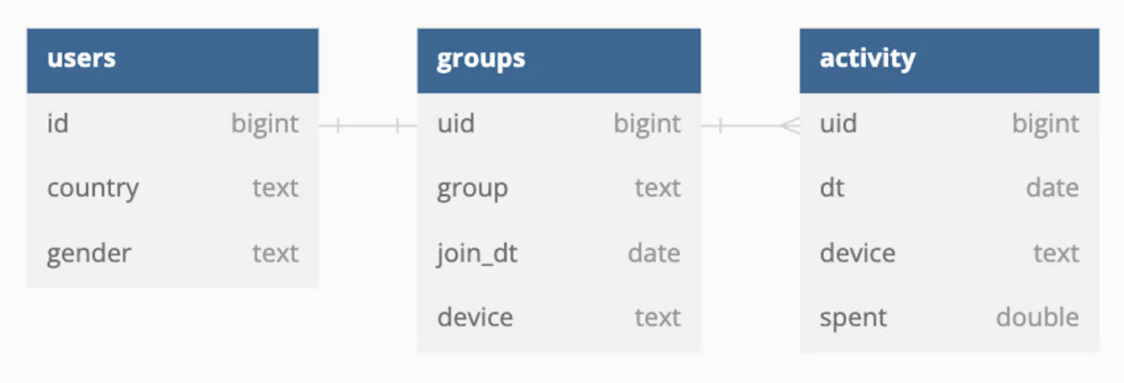
- The test was launched globally and focused solely on the GloBox mobile website.
- A user lands on the GloBox website and is then randomly assigned to either the Control Group showing the original website with no banner loaded or the Treatment Group showing the website displaying the banner.
- The user is assigned a unique User ID, and their Join Date, Device, Country, and Gender are logged.
- The user subsequently may or may not purchase products from the website. When a user makes one or more purchases the Amount Spent and Date of Purchase are recorded and considered a conversion.
  *** To Note:** Some users made more than one purchase which created duplicate user id's in these particular cases.

Dataset:

SQL queries were run to access, understand, modify, and extract the data from the provided PostgreSQL relational database, with tables and fields as below.

*** To Note**: There were null values in the "device" fields in both the "activity" and "groups" tables and nulls (n/a) in the Gender field in the "users" table. There were 139 duplicates due to these users having more than one purchase event.

Table Schema:



Below you'll find a description of each table and its columns:

- **users**: user demographic information
  - id: the user ID
  - country: ISO 3166 alpha-3 country code
  - gender: the user's gender (M = male, F = female, O = other)
    * Note: there are missing genders ("n/a")
- **groups**: user A/B test group assignment
  - 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 type the user purchased on (I = iOS, A = Android)
  - spent: the purchase amount in USD

Confidence Interval and Significance Level

- Confidence Intervals of 95%
- Significance Level 5%

Results

PHASE 1: Extract the A/B Test Data

* All queries available to be copied, pasted, and tested in the Google Docs link in the Appendix below

First, let's understand and extract the required data from the provided relational dataset using SQL queries:

1. **Let's determine the exact start and end dates of the experiment:**
   Confirmed start and end dates: **2023-01-25 to 2023-02-06**

```sql
SELECT MIN(join_dt)
FROM groups;
```

min
2023-01-25

```sql
SELECT MAX(join_dt)
FROM groups;
```

max
2023-02-06

2. **How many users were in the experiment and in each test group?**
   Total users in the experiment: **48943**

```sql
SELECT COUNT(DISTINCT id)
FROM users;
```

count
48943

GROUP A User Count:
Total users in Group A (Control Group): **24343**

```sql
SELECT COUNT(groups.group)
FROM groups
WHERE groups.group = 'A'
GROUP BY groups.group;
```

count
24343

GROUP B User Count:
Total users in Group B (Treatment Group): **24600**

```sql
SELECT COUNT(groups.group)
FROM groups
WHERE groups.group = 'B'
GROUP BY groups.group;
```

count
24600

3. **What is the conversion rate for the Control and Treatment Groups?**

   Control Group (A): CR = **3.92 %**

   Treatment Group (B): CR = **4.63%**

```sql
WITH conversion_stats AS (
    SELECT
  g.group AS test_group,
  COUNT(DISTINCT CASE WHEN a.spent IS NOT NULL THEN g.uid END) AS count_converted_users,
  COUNT(DISTINCT g.uid) AS count_total_users
    FROM groups AS g
    LEFT JOIN activity AS a ON g.uid = a.uid
    GROUP BY g.group
)
SELECT
    cs.test_group,
    cs.count_converted_users,
    cs.count_total_users,
    ROUND((cs.count_converted_users * 100.0 / cs.count_total_users), 2) AS conversion_rate
FROM conversion_stats AS cs;
```

test_group	count_converted_users	count_total_users	conversion_rate
A	955	24343	3.92
B	1139	24600	4.63

4. **What was the Average Amount Spent per User per Group?**

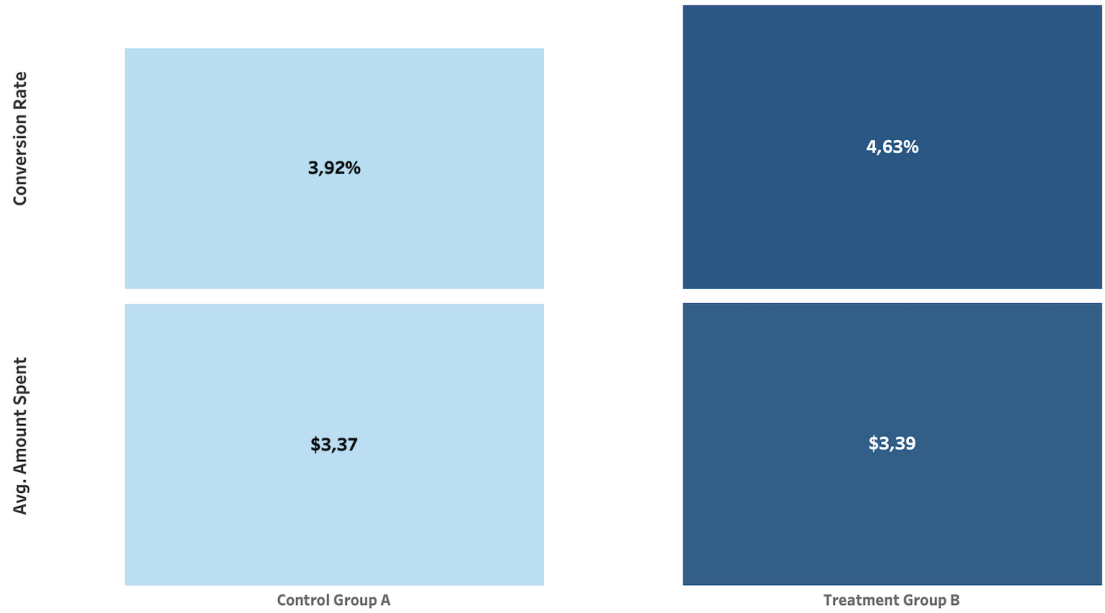
   Control Group A: $3.37

   Treatment Group B: $3.39

```sql
WITH conversion_stats AS (
  SELECT
    g.group AS test_group,
    COUNT(DISTINCT CASE WHEN a.spent IS NOT NULL THEN g.uid END) AS count_converted_users,
    COUNT(DISTINCT CASE WHEN a.spent IS NULL THEN g.uid END) AS not_converted_users,
    COUNT(DISTINCT g.uid) AS count_total_users,
    ROUND(COALESCE(SUM(a.spent), 0)::numeric, 3) AS total_spend,
    ROUND((COALESCE(SUM(a.spent), 0) / COUNT(DISTINCT g.uid))::numeric, 3) AS avg_spend_per_user
  FROM groups AS g
  LEFT JOIN activity AS a ON g.uid = a.uid
  GROUP BY g.group
)
SELECT
  cs.test_group,
  cs.count_converted_users,
  cs.not_converted_users,
  cs.count_total_users,
  ROUND((cs.count_converted_users * 100.0 / cs.count_total_users), 3) AS conversion_rate,
  cs.total_spend,
  cs.avg_spend_per_user
FROM conversion_stats AS cs;
```

test_group	count_converted_users	not_converted_users	count_total_users	conversion_rate	total_spend	avg_spend_per_user
A	955	23388	24343	3.923	82145.903	3.375
B	1139	23461	24600	4.630	83415.327	3.391

Conversion Rate and Avg Spend by Group



5. Extracting the Data for further Analysis in Google Sheets and Tableau:

```sql
SELECT
    u.id AS user_id,
    u.country,
    u.gender,
    g.device AS device,
    g.group AS test_group,
    CASE WHEN SUM(COALESCE(a.spent, 0)) > 0 THEN 'converted' ELSE 'not converted' END AS conversion_status,
    ROUND(COALESCE(SUM(a.spent), 0)::numeric, 2) AS total_spent
FROM
    users AS u
LEFT JOIN
    groups AS g ON u.id = g.uid
LEFT JOIN
    activity AS a ON u.id = a.uid
GROUP BY
    u.id, u.country, u.gender, g.device, g.group;

```

Save Run

user_id	country	gender	device	test_group	conversion_status	total_spent
1000000	CAN	M	I	B	not converted	0.00
1000001	BRA	M	A	A	not converted	0.00
1000002	FRA	M	A	A	not converted	0.00
1000003	BRA	M	I	B	not converted	0.00
1000004	DEU	F	A	A	not converted	0.00
1000005	GBR	F	A	B	not converted	0.00

PHASE 2: Analyze A/B Test Results - Hypothesis Testing

In this phase of the analysis, we construct both a hypothesis test and confidence interval comparing the two test groups to establish if the changes that we are A/B testing are leading to a real and meaningful change in the metrics of interest, i.e. whether the new food and drink banner is leading to changes in the “User Conversion Rate” and the “Average Amount Spent per User”. The dataset we extracted in .csv format in the previous Phase is now imported into Google Sheets to calculate our formulas to achieve clear conclusions. All calculations can be found on the Google Sheet, and the Confidence Intervals visualizations can be viewed on the Tableau Dashboard, both links are below in the Appendix.

Conversion Rate: Hypothesis Test

First, a Hypothesis Test was conducted to see whether there is a statistically significant difference in the conversion rate between the two groups by finding the resulting p-value and

forming our conclusion.
"*A p-value is a statistical measurement used to validate a hypothesis against observed data.*" - Investopedia
We used the normal distribution and a 5% significance level, and the pooled proportion for the standard error.

Null Hypothesis: There is no difference in conversion rate between the two groups, therefore the conversion rate is equal in both groups.

Alternative Hypothesis: There is a difference in conversion rate between the two groups. Therefore, one group has a higher conversion rate than the other.

To find the resulting p-value for the conversion rate let's first clarify what type of test we are using.

Type of test we are using:
Proportion. A proportion can be represented as a percentage, therefore Conversion Rate is a proportion.

Number of samples we're evaluating, one or two samples:
We're comparing two samples.

One-sided or two-sided test?
Two-sided.

1. First, we calculate our **test statistic (the Z-score)** as:
   -3.8643.

   Formula to calculate the test statistic (Z-score):
   =(Conversion Rate Group A-Conversion Rate Group B)/Standard Error
   =(0.0392-0.0463)/0.0018

2. Next, we calculate the **resulting p-value of 0.0001**.

   Formula to calculate the resulting p-value:
   =2 * (1 - NORM.S.DIST(ABS(Z-Score)))
   =2 * (1 - NORM.S.DIST(ABS(-3.8643)))

We conclude that we **Reject the Null Hypothesis**. (Reject H0 in favour of H1) that there is no difference in conversion rate between the two groups.

GloBox

Conversion Rate: Confidence Intervals and Significance Levels

Next, we must establish the 95% confidence intervals for the difference in the conversion rate between the treatment and control (treatment-control). Confidence intervals help us understand the *magnitude* of the difference between the two groups. Here we used the normal distribution and unpooled proportions for the standard error. To calculate the Lower and Upper Bound Confidence Intervals for the Conversion Rate see the formulas below.

Type of interval we are computing:
Confidence Interval for a Difference in Proportion, and we are comparing two samples.

1. **First, we calculate the Sample Statistic (Mean Difference):**
   Answer: **0.0071**

   Formula:
   =Conversion Rate Group B - Conversion Rate Group A
   =0.0463 - 0.0392

2. **Next, calculate the Standard Error:**
   Answer: **-0.0018**

   Formula:
   = SQRT((Conversion Rate Group A * (1 - Conversion Rate Group A) / Sample Size Group A) + (Conversion Rate Group B * (1 - Conversion Rate Group B) / Sample Size Group B))

   = SQRT((0.0392 * (1 - 0.0392) / 24343) + (0.0463 * (1 - 0.0463) / 24600))

3. **Now, calculate the Critical Value:**
   Answer: **1.96**

   Formula:
   =NORM.S.INV(1 - (1 - confidence level) / 2)
   =NORM.S.INV(1 - (1 - 0.95) / 2)

4. **Calculate the Margin of Error:**
   Answer: **0.0036**
   = Critical Value * Standard Error
   = 1.96 * 0.0018

This brings us to the **Confidence Intervals:**

Lower Bound: 0.0035
Formula to calculate the Lower Bound Confidence Interval:
= (Conversion Rate of Group B - Conversion Rate of Group A) - Margin of Error
= (0.0463 - 0.0392) - 0.0036

Upper Bound: 0.0107
Formula to calculate the Upper Bound Confidence Interval:
= (Conversion Rate of Group B - Conversion Rate of Group A) + Margin of Error
= (0.0463 - 0.0392) + 0.0036

** You can view the visual of Confidence Intervals including the Point Estimate on the Tableau Dashboard, and all the Formulas in the Google Sheet, links in the Appendix.*

Average Amount Spent per User: Hypothesis Test

We now need to establish our Hypothesis to see whether there is a statistically significant difference in the "Average Amount Spent per User" between the two groups by finding the resulting p-value and forming our conclusion. We used the t distribution and a 5% significance level and assumed unequal variance.

***Null Hypothesis*:** The average amount spent per user in Control Group A is equal to the average amount spent per user in Treatment Group B.

***Alternative Hypothesis*:** The average amount spent per user in Control Group A is different from the average amount spent per user in Treatment Group B.

To find the resulting p-value for the average amount spent per user let's first clarify what type of test we are using.

Type of test we are using:
Mean. A mean is an average of continuous values.

Number of samples we're evaluating, one or two samples:
We're comparing two samples.

One-sided or two-sided test?
Two-sided.

1. **First, we calculate the test statistic:**
   Answer: **-0.0704**

   Formula:
   = (Sample Mean Group A - Sample Mean Group B) / sqrt((Standard Deviation Group A^2 / Sample

Size Group A) + (Standard Deviation Group B^2 / Sample Size Group B))

= (3.375 - 3.391) / sqrt((25.94^2 / 24343) + (25.41^2 / 24600))

2. **Next, we calculate the p-value using the T.Test**
   Answer: **0.944**

   Formula:
   =T.TEST(range_total_spent_a, range_total_spent_b, 2, 2)
   =T.TEST(G2:G, O2:O, 2, 2)

We conclude that we **Fail to Reject the Null Hypothesis**. (Fail to Reject H0). There is not a significant difference in the average amount spent per user in control group A compared to the average amount spent per user in treatment group B.

Average Amount Spent per User: Confidence Intervals and Significance Levels

Recall that Confidence Intervals help us understand the *magnitude* of the difference between the two groups. Here, we used the t distribution and assumed unequal variance.
To calculate the Lower and Upper Bound Confidence Intervals for the Average Amount Spent per User see the formulas below.

Type of interval we are computing:

Confidence Interval for a Difference in Means, and comparing two samples.

1. **First, we calculate the Sample Statistic (Point Estimate):**
Answer: **0.0163**

Formula:
=Sample Mean Group B - Sample Mean Group A
=3.391 - 3.375

2. Next, calculate the Standard Error:
Answer: **-0.2321**

Formula:
=SQRT((Standard Deviation Group B^2 / Sample Size Group B) + (Standard Deviation Group A^2 / Sample Size Group A))

=SQRT((25.41^2 / 24600) + (25.94^2 / 24343))

3. Now, calculate the Critical Value:

Answer: **1.96**

Formula:
=T.INV(1 - 0.05 / 2, degrees of freedom)
=T.INV(1 - 0.05 / 2, 48894.50)

This brings us to the **Confidence Intervals:**

Lower Bound: -0.439
Formula to calculate the Lower Bound Confidence Interval:
= Sample Statistic *Mean Difference* - (Critical Value * Standard Error)
= 0.0163 - (1.96 * 0.2321)

Upper Bound: 0.471
Formula to calculate the Upper Bound Confidence Interval:
= Sample Statistic *Mean Difference* + (Critical Value * Standard Error)
= 0.0163 + (1.96 * 0.2321)

** You can view the visual of Confidence Intervals including the Point Estimate on the Tableau Dashboard, and all the Formulas in the Google Sheet, links in the Appendix.*

Novelty Effect

When we look at the Treatment Group we do see what appears to be evidence of a Novelty Effect in action, at the 3-day mark on the 29th of January we see signs of stabilizing, however, conversions and revenue further drop off for another 4 days until the 2nd of February and then we see more consistent stabilization. However, when we look at the Control Group (Group A) alongside it, we see a similar trend effect in the Control Group, and therefore we do not attribute this downward trend in spend and conversions to the banner feature added.

Power Analysis

A Power Analysis was performed and the results show that the sample size was more than adequate for our test, and the duration of the experiment was sufficient to accumulate this sample size.
Minimum Sample Size: 12895 unique visitors per test variation
AB test duration: Minimum test duration 1.05 weeks *Round up to an A/B Test period of 2 weeks (discrete number of business cycles).

Minimum Detectable Effect

A Minimum Conversion Rate for Group B of 4.39% was needed for the duration of the 13-day A/B Test, with 24472 unique visitors per week on the website.
This amounts to a **Minimum Detectable Effect of 12% needed** to run the test at the required levels of power and confidence.
We achieved an 18% improvement on the 3.92% Conversion Rate in the Control Group, with a Conversion Rate in the Treatment Group of 4.63%

Calculations:

Control Group Conversion Rate:	**3.92%**
Improvement seen over Control Group:	**18%** (4.63% CR)
Average Unique Visitors on the Website per week:	**24472**
Maximum Number of Weeks of A/B Test:	**1 week 6 days (2 weeks)**
Two-sided Test	**Yes**
Statistical Power:	**80%**
Confidence Level:	**95%**

Recommendation

I recommend that we launch the banner, as we saw at least one of the metrics required, "User Conversion Rate", showed a statistically and practically significant increase. We, therefore, feel confident in releasing the banner to all users. The perceived cost and overall difficulty of launching and maintaining a banner is minimal and therefore worth the benefits we saw in the A/B test.

Note: As we have found the difference in conversion rate is statistically significant, this improvement is sufficient, as GloBox gains more paying customers and can be remarketed to so as to expand their purchasing habits over time and therefore increase the "Average Amount Spent per User".

APPENDIX *With Hyperlinks*:

Google Docs with SQL Queries for Phase 1:
All SQL Queries for Phase 1 of this analysis can be found on the above Google Doc hyperlink.

Google Sheets with Calculations in Column 'Q' on the first two sheets:
All calculations can be found on the above Google Sheet hyperlink, take note of the tab names.

Tableau Story (x3 Dashboards):
All Visualisations can be found at the Tableau hyperlink above.
Please click the "Full Screen" icon in the bottom right corner to ensure the full experience.

Calculators used to confirm Power Analysis results:
A/B Test Power and Significance Test Calculator
A/B Test Calculator by CXL
A/B Test Sample Size Calculator
Sample Size Calculator for Comparing Two Independent Means