# Comparison of revenue and conversion rate between two versions of the GloBox mobile website.

Antonio Caradonna, GloBox

August 21th, 2023

## SUMMARY

The hypothesis test conducted on the new version of the mobile website compared to the previous version revealed a statistically significant increase in the conversion rate. This suggests that the changes made to the new version positively impacted user engagement and interaction with the website, leading to a higher likelihood of desired user actions.

However, the analysis did not provide conclusive evidence regarding the revenue generated by the new version compared to the previous version. While this outcome leaves uncertainty about the direct revenue impact, it also offers an opportunity for further investigation and optimization.

Considering these results, **I recommend launching the new version of the mobile website** and monitoring its performance to run further analysis.

# 1. INTRODUCTION

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, given the exceptional growth of the company's food and drink offerings in the last few months, conducting an investigation into modifications aimed at raising awareness within this product category could potentially increase revenue.

This presentation provides a comparative analysis of two mobile website versions for the mobile website. In able to do this analysis, the team collected data for thirteen consecutive days where users who visited the mobile website were randomly assigned to either the control or test group and shown the corresponding landing page. That is, the page loads the banner if the user is assigned to the test group, and does not load the banner if the user is assigned to the control group (as shown in Figure 1). The user subsequently may or may not purchase products from the website. It could be on the same day they join the experiment, or days later. If they do make one or more purchases, this is considered a "conversion".

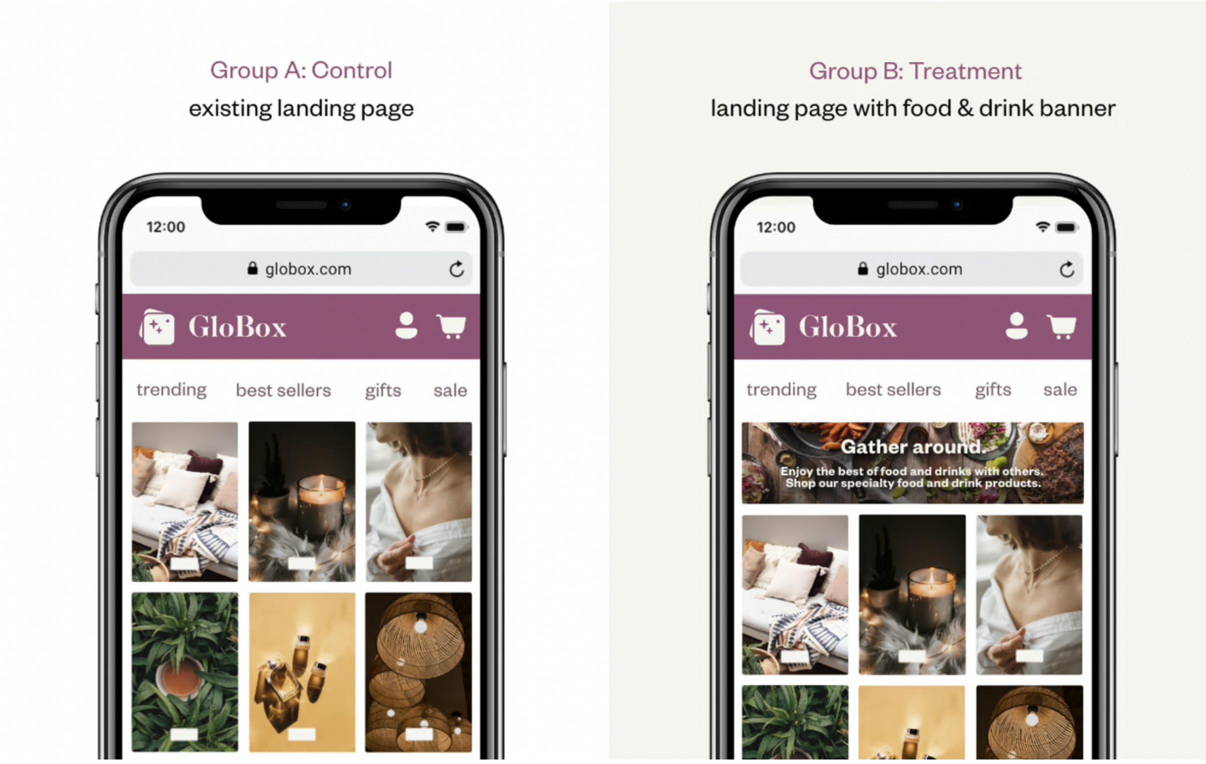


Figure 1: **Left:** Control group. **Right:** Treatment group. The control group does not see the banner, and the test group sees it as shown above.

## 2. CONVERSION RATE ANALYSIS

First, we define conversion rate as:

$$\text{Conversion rate} = \frac{\text{Purchases}}{\text{Visits}} \tag{1}$$

where we only take into account unique users. We begin by calculating the sample statistics to compare the two groups: the conversion rate for the **control** group is $\hat{p}_c = (3.92 \pm 0.24)\%$ and for the **treatment** group is $\hat{p}_t = (4.63 \pm 0.26)\%$. The $95\%$ confidence interval was calculated by:

$$\hat{p} \pm Z\frac{SE}{\sqrt{n}} \tag{2}$$

where $Z$ is the number of standard deviations from the mean that we want to take into account (for a confidence interval of $95\%$, the value of $Z$ is $1.96$), $SE$ is the standard deviation of the sampling distribution and $n$ is the sample size.

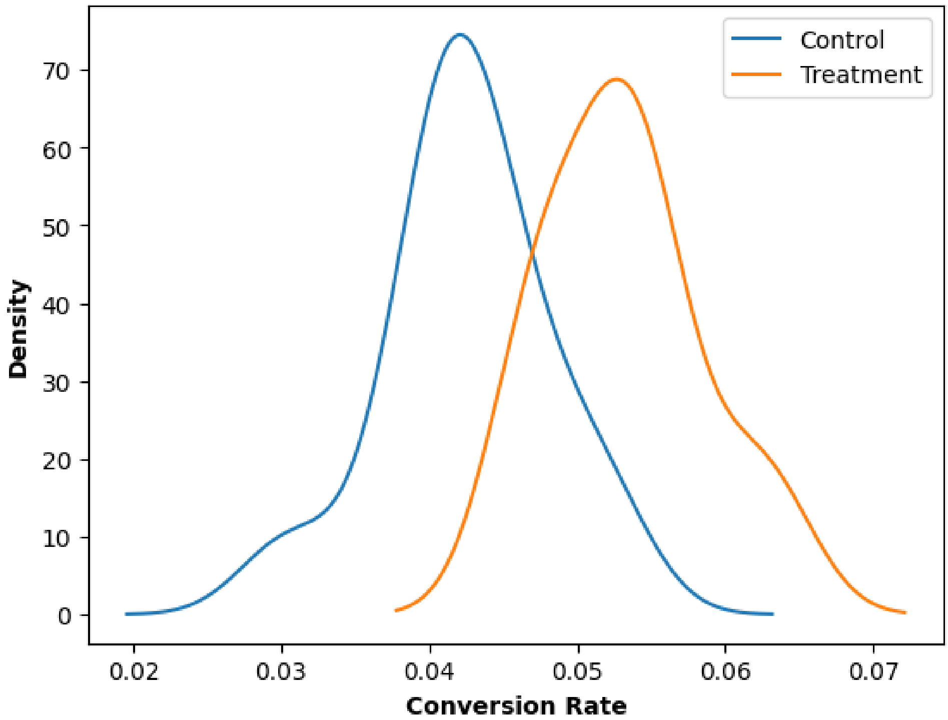


Figure 2: Sampling distributions for the conversion rate of the control and treatment groups.

From Figure 2, there is a clear difference between the two groups, where the treatment group has a higher conversion rate. That is true for this sample, but what is the probability that we would obtain the same result if we repeated the test multiple times?

Let's define two hypotheses for this experiment:

$$H_0 : p_c = p_t$$
$$H_1 : p_t > p_c$$

where $p_c$ is the proportion of people who converted in the control population and $p_t$ is the proportion of people who converted in the treatment population. Then, the null hypothesis $(H_0)$ tells us that the proportion of conversion rate is equal for both groups, and the alternative hypothesis $(H_1)$ is that the proportion of conversion rate is different between the groups.

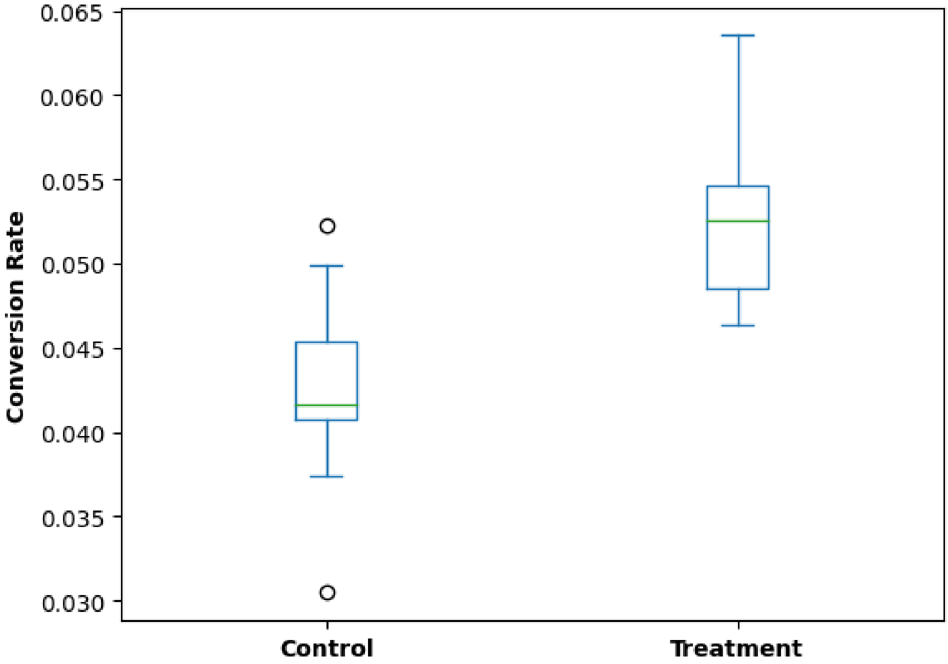


Figure 3: Summary of distribution data comparison for the conversion rate between control and treatment groups.

From Figure 2 we see that both distributions are approximately normal, people in each group were randomly selected and the value of one observation does not influence or affect the value of another observation. That being said, we can safely continue with our test.

Now, let's set the **significance level** to $\alpha = 0.05$. That is the maximum probability of rejecting the null hypothesis when it is actually true. This is often referred to as **Type I error**.

$$P(\textbf{Type I Error}) = \alpha \tag{3}$$

The next step is to assume the Null Hypothesis is true and calculate the probability of obtaining a difference between $\hat{p}_c$ and $\hat{p}_t$ that is at least as large as the one we got. To obtain this probability, we begin by calculating the $Z$ value.

$$Z = \frac{\hat{p}_t - \hat{p}_c}{\sigma_{\hat{p}_t - \hat{p}_c}} \tag{4}$$

where $Z$ indicates how many standard deviations a data point is above or below the mean, and $\sigma_{\hat{p}_t - \hat{p}_c}$ is the standard deviation of the sampling distribution of the difference between the sample proportions.

$$Z = \frac{\hat{p}_t - \hat{p}_c}{\sigma_{\hat{p}_t - \hat{p}_c}} \approx \frac{\hat{p}_t - \hat{p}_c}{\sqrt{\frac{\hat{p}_{comb}(1-\hat{p}_{comb})}{n_c} + \frac{\hat{p}_{comb}(1-\hat{p}_{comb})}{n_t}}} \tag{5}$$

where $\hat{p}_{comb}$ is the combined proportions:

$$\hat{p}_{comb} = \frac{\text{Purchases Control} + \text{Purchases Treatment}}{n_c + n_t} \tag{6}$$

and $n_c$, $n_t$ are the sample size or total website visits for the control and treatment groups, respectively. In the appendix section we can see the query utilized to find the amount of people that visited and converted for each group.

$$\hat{p}_{comb} = \frac{955 + 1139}{24343 + 24600} = 0.043 \tag{7}$$

Then,

$$\sigma_{\hat{p}_t - \hat{p}_c} = \sqrt{\frac{0.043(1-0.043)}{24343} + \frac{0.043(1-0.043)}{24600}} = \sqrt{1.69 \times 10^{-6} + 1.67 \times 10^{-6}} = 1.83 \times 10^{-3} \tag{8}$$

Finally, the $Z$ score is:

$$Z \approx \frac{0.0463 - 0.0392}{0.00183} \approx 3.88 \tag{9}$$

The probability of obtaining a result like the one we got in the data, assuming that the Null Hypothesis is true would be:

$$\text{p-value} : P(Z \geq 3.88) = 0.0001 \tag{10}$$

This result comes from looking at a $Z$ score table and looking at the probability of obtaining a value that is $3.88$ standard deviations away from the control proportion mean. Assuming a $95\%$ confidence level (two-tailed test), the critical Z-values are approximately $\pm 1.96$. Since $3.88$ is greater than $1.96$ (meaning that $3.88$ is even further away from the mean), we reject the null hypothesis.

Based on the data and the test, **we have sufficient evidence to conclude that there is a significant difference in the proportions of conversions between the control and treatment groups**.

# 3. REVENUE ANALYSIS

In this section, we do a similar analysis as previously shown, but this time we focus on the revenue generated by each group. The calculation of the $95\%$ confidence interval for the sample statistics is done using equation 2, but using the standard deviation of the sample statistics.

$$\overline{\mu} \pm t \frac{S}{\sqrt{n}} \tag{11}$$

This time we obtain that the control group generates an average revenue of $\overline{\mu}_c = (3.366 \pm 0.317)\$$, while the treatment group produces an average revenue of $\overline{\mu}_t = (3.380 \pm 0.307)\$$.

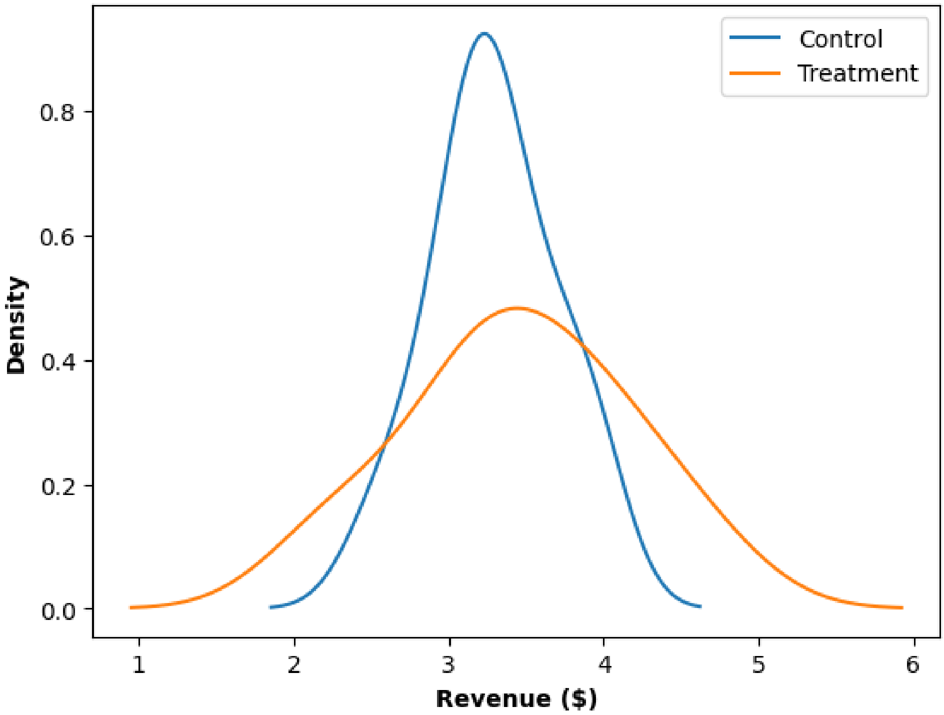


Figure 4: Sampling distributions for the average revenue of the control and treatment groups.

The distributions for the average revenue generated by group are presented in Figure 4. Moreover, Figure 5 shows a box plot for the average revenue between groups. It is noticeable that the treatment group has a wider distribution.

The next step is to do a hypothesis test on the sample statistics we have calculated in this section. The null hypothesis is going to state that there is no difference between the average revenue generated by the control and treatment groups. On the other hand, the alternative hypothesis is going to be that there is a difference between the average revenue of the groups.

$$H_0 : \overline{\mu}_c = \overline{\mu}_t$$
$$H_1 : \overline{\mu}_c \neq \overline{\mu}_t$$

Let's set the significance level to $\alpha = 0.05$. That is, if the experiment is repeated many times, the confidence level is the percent of the time each sample's mean will fall within the confidence interval. Furthermore, this time it is more convenient to use a t-test because the z-test tends to underestimate when not dealing with proportions. We assume the null hypothesis and calculate the t-statistics based on the sample data. The t-statistics is equal to:

$$t = \frac{\overline{\mu}_c - \overline{\mu}_t}{\sqrt{\frac{S_c^2}{n_c} + \frac{S_t^2}{n_t}}} \tag{12}$$

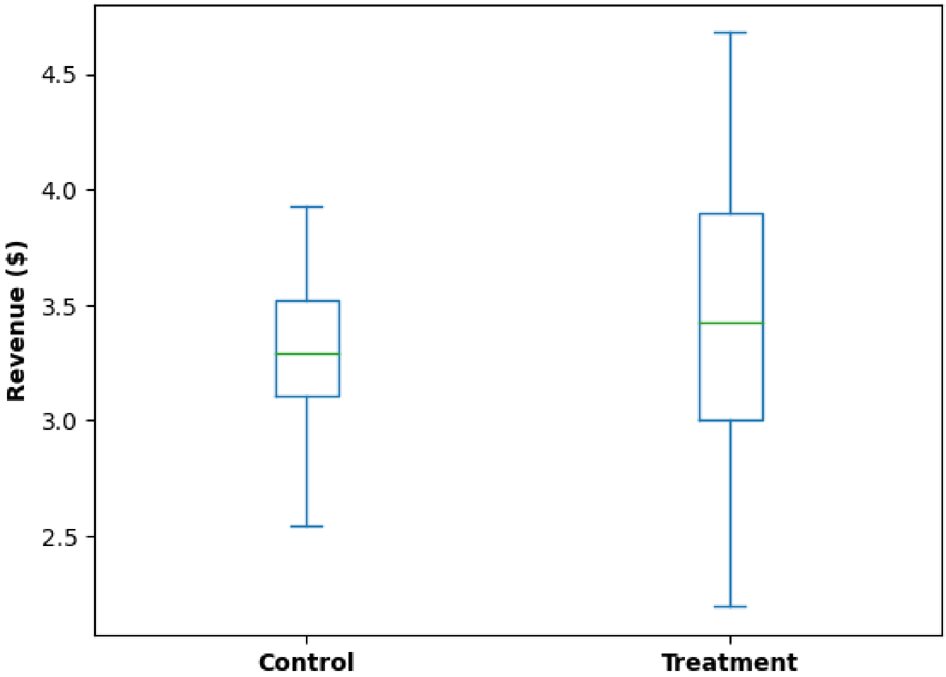


Figure 5: Summary of distribution data comparison for the average revenue between control and treatment groups.

where the denominator represents an estimation of the standard deviation of the sampling distribution of the difference of the sample means. Also, $S_c$ and $S_t$ are the

$$t = \frac{3.366 - 3.380}{\sqrt{\frac{(25.37)^2}{24343} + \frac{(24.79)^2}{24600}}} = \frac{-0.014}{0.23} \approx -0.06 \tag{13}$$

The negative value means that the result is lower than our sample mean. Then,

$$\text{p-value} : P(|t| \geq 0.06) \approx 0.944 \tag{14}$$

Since the p-value is greater than the chosen significance level, **we fail to reject the null hypothesis.** This means that **we do not have enough evidence to conclude that there is a significant difference in the average revenue between the control and treatment groups**.

# 4. CONCLUSION AND RECOMMENDATIONS

The analysis done on the new feature of the mobile website for GloBox indicates that the new version has led to a statistically significant increase in the conversion rate compared to the previous version. This outcome suggests that the changes implemented in the new version have positively impacted user engagement and interaction with the website.

While the analysis couldn't conclusively establish a direct relationship between the new version and revenue increase, the potential for a positive impact still exists. The fact that the new version has led to a statistically significant increase in the conversion rate is a positive indicator. Given that one of the primary goals is to improve user engagement and interaction with the mobile website, **I would recommend launching the new version** of the mobile website and keep monitoring its performance, gathering additional data, and implementing iterative improvements based on both quantitative and qualitative insights.

# 5. APPENDIX: SQL QUERIES

Listing 1: Conversion Rate calculation for the control group (A).

```
WITH full_data AS(
SELECT u.id AS user_id,
    u.country AS user_country,
  u.gender AS gender,
  g.device AS device,
  g.group AS test_group,
  CASE WHEN a.spent > 0 THEN 1 ELSE 0 END AS has_converted,
  CASE WHEN a.spent IS NULL THEN 0 ELSE a.spent END AS spent
FROM users u
LEFT JOIN groups g
    ON u.id = g.uid
LEFT JOIN activity a
    ON u.id = a.uid)

SELECT COUNT(DISTINCT user_id) AS total_users,
    COUNT(DISTINCT CASE WHEN spent > 0 THEN user_id END) AS ↩
        converting_users,
  ROUND(COUNT(DISTINCT CASE WHEN spent > 0 THEN user_id END) * 100.0 /↩
      (COUNT(DISTINCT user_id)), 2) AS conversion_rate
FROM full_data
WHERE test_group = 'A'
```

Listing 2: Conversion Rate calculation for the treatment group (B).

```
WITH full_data AS(
SELECT u.id AS user_id,
    u.country AS user_country,
  u.gender AS gender,
  g.device AS device,
  g.group AS test_group,
  CASE WHEN a.spent > 0 THEN 1 ELSE 0 END AS has_converted,
  CASE WHEN a.spent IS NULL THEN 0 ELSE a.spent END AS spent
FROM users u
LEFT JOIN groups g
    ON u.id = g.uid
LEFT JOIN activity a
    ON u.id = a.uid)

SELECT COUNT(DISTINCT user_id) AS total_users,
    COUNT(DISTINCT CASE WHEN spent > 0 THEN user_id END) AS ↩
       converting_users,
  ROUND(COUNT(DISTINCT CASE WHEN spent > 0 THEN user_id END) * 100.0 /↩
      (COUNT(DISTINCT user_id)), 2) AS conversion_rate
FROM full_data
WHERE test_group = 'B'
```

Listing 3: Average revenue calculation for the control group (A).

```
WITH full_data AS(
SELECT u.id AS user_id,
    u.country AS user_country,
  u.gender AS gender,
  g.device AS device,
  g.group AS test_group,
  CASE WHEN a.spent > 0 THEN 1 ELSE 0 END AS has_converted,
  CASE WHEN a.spent IS NULL THEN 0 ELSE a.spent END AS spent
FROM users u
LEFT JOIN groups g
    ON u.id = g.uid
LEFT JOIN activity a
    ON u.id = a.uid)

-- Average spent by group
SELECT AVG(total_amount) AS average_spent,
    COUNT(user_id)
FROM (
    SELECT user_id,
    SUM(spent) AS total_amount,
    test_group
    FROM full_data
    GROUP BY user_id, test_group
) user_totals
WHERE test_group = 'A';
```

Listing 4: Average revenue calculation for the treatment group (B).

```sql
WITH full_data AS(
SELECT u.id AS user_id,
    u.country AS user_country,
  u.gender AS gender,
  g.device AS device,
  g.group AS test_group,
  CASE WHEN a.spent > 0 THEN 1 ELSE 0 END AS has_converted,
  CASE WHEN a.spent IS NULL THEN 0 ELSE a.spent END AS spent
FROM users u
LEFT JOIN groups g
    ON u.id = g.uid
LEFT JOIN activity a
    ON u.id = a.uid)

-- Average spent by group
SELECT AVG(total_amount) AS average_spent,
    COUNT(user_id)
FROM (
    SELECT user_id,
    SUM(spent) AS total_amount,
    test_group
    FROM full_data
    GROUP BY user_id, test_group
) user_totals
WHERE test_group = 'B';
```