# GloBox A/B Test- Report and Codes

This A/B test was done in order to test whether we should include a banner at the top of the site that highlights key products in the food and drink category or not, in purpose to bring awareness to this product category to increase revenue. For this it was decided to run an A/B test, which is fit for such a case.

Some important highlights are that all the users were randomly assigned to either the control or treatment group when they visited the GloBox main page. This is the join date for the user. The control group (Group A) didn't see the banner, while the treatment group (Group B) saw it as described above. The test was running for 13 days, from 2023-01-25 to 2023-02-06, and if a user made one or more purchases during that time, this was considered a "conversion".

```
SELECT min as first_day,
    max as last_day,
    max-min +1 as total_test_time
FROM (
SELECT uid, join_dt,
     MAX(join_dt) over() as max,
     MIN(join_dt) over() as min
FROM groups
LIMIT 1) a
```

| first_day<br>DATE | last_day<br>DATE | total_test_time<br>INT4 |
|---|---|---|
| 2023-01-25 | 2023-02-06 | 13 |

Also, it is important to note that in order to minimize as much interference as possible, it was decided that the test is only being run on the mobile website. In total, this test had 24,343 users in the control group (group A), and 24,600 users in the treatment group (group B).

All calculations were done in a spreadsheet and SQL.

In this study, two aspects were examined.

**1. First of all, it was checked if there is a difference in the mean amount spent per user between the control and treatment.** The null hypothesis (H0) says that there is no difference in the mean between them, while the alternative hypothesis (H1) claims that there is a significant difference.

**2. In the second hypothesis, was tested whether there is a difference in the conversion rate between the control group and treatment group.** The null hypothesis (H0) in this experiment is that there is no significant difference between them. That is, any experimentally observed difference is due to chance alone, and causative relationship doesn't exist.

On the other hand, our alternative hypothesis (H1) is that there is a difference between the groups, and the treatment group will show more awareness to this product category (when this is our goal - to increase revenue). If a statistical significance is obtained it will mean that the results of the A/B test aren't due to chance, and we will reject the null hypothesis.

**Experiment 1:**

To test the first hypothesis, it was decided to test the confidence interval of the two groups in order to test whether there is a difference between them in terms of revenue, and it was also decided to do a hypothesis test to check whether this difference is not due to chance. In this experiment, an alpha of 5% (two-sided hypothesis) was decided in advance, which means that the confidence level would be 95%.

After the transition from SQL to spreadsheet I added another column that describes whether the user converted (1), or not converted (0). Therefore, before switching to the spreadsheet, I checked in SQL what was the average amount spent per user for the control and treatment.

It was found that the control group spent **$3.366**, and the treatment group: **$3.380**:

```sql
with t1 as (
SELECT g.uid as id, g.group as group, COALESCE(a.spent,0) as spent
FROM groups g
LEFT JOIN activity a
ON g.uid = a.uid
)

SELECT "group", AVG(spent) as avg_spent
FROM t1
GROUP BY "group"
```

After that, the confidence intervals of the two groups were also checked, to see the range of each of them, when we are 95% sure that the values of each group are in this range. The 95% confidence interval for the average amount spent per user in the control group is **(3.049, 3.700)**:

```sql
WITH t1 as (
SELECT g.uid as id, g.group as group, COALESCE(SUM(a.spent),0) as total_spent
FROM groups g
LEFT JOIN activity a
ON g.uid = a.uid
WHERE "group"='A'
GROUP BY id, "group"
)

SELECT AVG(total_spent) - 1.96 * STDDEV(total_spent) / SQRT(COUNT(id)) AS lower_bound,
    AVG(total_spent) + 1.96 * STDDEV(total_spent) / SQRT(COUNT(id)) AS upper_bound
FROM t1;
```

Query Results
1 ROWS

| lower_bound FLOAT8 | upper_bound FLOAT8 |
|---|---|
| 3.048697852929847 | 3.700339082927839 |

The 95% confidence interval for the average amount spent per user in the treatment group is **(3.073, 3.708)**:

```
WITH t1 as (
SELECT g.uid as id, g.group as group, COALESCE(SUM(a.spent),0) as total_spent
FROM groups g
LEFT JOIN activity a
ON g.uid = a.uid
WHERE "group"='B'
GROUP BY id, "group"
)

SELECT AVG(total_spent) - 1.96 * STDDEV(total_spent) / SQRT(COUNT(id)) AS lower_bound,
    AVG(total_spent) + 1.96 * STDDEV(total_spent) / SQRT(COUNT(id)) AS upper_bound
FROM t1;
```

Query Results
1 ROWS

| lower_bound FLOAT8 | upper_bound FLOAT8 |
|---|---|
| 3.07327943439715 | 3.708454457374415 |

Here you can see a bar chart illustrating the difference between the groups in terms of the average spent by a user in each group (chart 1):

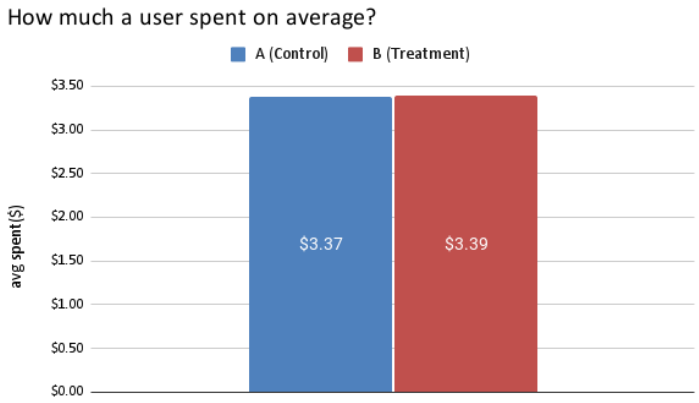


Chart 1

From looking at these data alone we can see that we are 95% sure that the ranges are different for the two groups. But for the purpose of testing and deepening, a test of the confidence level between the two groups was also done.

**To check this and the rest of the calculations in this report, I decided to continue the analysis in spreadsheet. I used this code in SQL to move forward:**

```
WITH cte AS (
SELECT uid, "group", SUM(spent) AS total_spent
FROM groups
left join activity
using(uid)
GROUP BY uid,"group"
),

cte_2 as
(select uid, "group", (COALESCE(total_spent, 0)) total_spent from cte)

select * from cte_2
```

**link to the spreadsheet:**
**https://docs.google.com/spreadsheets/d/1YDNZobgrf2E3r13VuW_WFSpZhfUgfrZ9bhDYYK3oCMk/edit?usp=share_link**

In order to calculate the CI I used a two-sample t-interval for a difference in means. In view of unequal variance, I used the unpooled standard error. It was found that the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control is **(-0.439, 0.471)**. **This result shows that we are 95% sure that there is a difference between the two groups**, this is duo the fact that the value 0 is not in the range.

To check whether this is a significant difference, I also conduct a hypothesis test between the two groups. With using a two-sided t-test for a difference in means, and again unpooled standard error, with a 5% significance level (predetermined) it was found that the **p-value is 0.944**. This value is higher than 0.05 and the meaning of this is that we got the result by chance. In other words**, we fail to reject the null hypothesis which says that there is no difference in the average amount spent per user between the control and treatment.**

**Experiment 2:**

In addition to examining whether the new banner does affect user awareness of the food and drink category and results in an increase in revenue, I also wanted to check what percentage of users in each group became converted users. This is the second hypothesis.

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". All calculations were done in a spreadsheet and SQL.

First, the conversion rate in each group was checked separately (percentage of converted users, divided by all users in the same group). It was found that in the control group the conversion rate was **3.92%**, and in the treatment group **4.63%**.

Below is a bar chart illustrating the conversion (in %) in each group (chart 2):

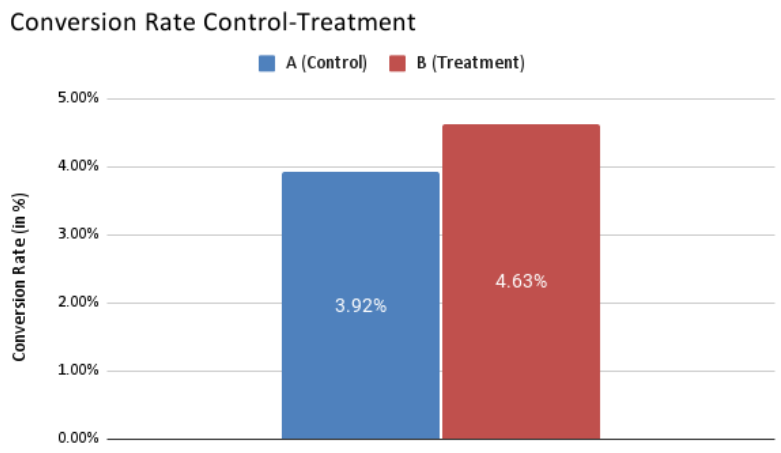


Chart 2

After that, the 95% confidence interval for the conversion rate of users in each group was measured here as well- It was found that the CI is **(0.0368, 0.0417) for the control group**, and **(0.0437, 0.0489) for the treatment group**.

As in the first experiment, here it also seems that there are no overlapping values, but for this purpose a 95% confidence interval for the difference in the conversion rate between the two groups was also tested. The unified group includes all the trial users, and the proportions was checked. Therefore, a two-sample z-interval for a difference in proportions was used. Since the proportions are the same we use the unpooled standard error. It was found that the CI is **(0.0035, 0.0107),** And we see that the value 0 is not included in the range. **This means that we are 95% sure that there is a difference between the two groups in terms of conversion rate.**

To see whether there is a significant difference in the conversion rate, a hypothesis test was performed for a normal distribution and a significance level of 5%. With a two-sample two-sided z-interval for a difference in proportions, and using a pooled standard error, it was found that the p value is **p = 0.0001**. **This result is statistically significant** because the p-value is less than an alpha of 5%, which means that we reject the null hypothesis that there is no difference in the user conversion rate between the control and treatment.

## conclusions and recommendations

The A/B test and the two hypotheses that were made following it, as well as the confidence intervals that were tested, came to check whether a banner at the top of the company app which highlight the food and drink category would encourage users to buy more and so increase the company's revenue. First it was checked whether the banner would cause a difference in the amount spent per user between the control and treatment group. From the confidence interval between the two groups, we could be 95% sure that there is a difference in the average amount

due to the fact that the interval is above 0, but the hypothesis test showed that this difference isn't significant for alpha of 5% (p = 0.944), and we failed to reject the null hypothesis.

After that it was checked whether the banner would cause a difference in the conversion rate between the control group and treatment group. The confidence interval did not include 0 here as well, and we could be sure at a 95% confidence level that there is a difference between the two groups (where the treatment group is higher than the control group). In addition, in this hypothesis test, the result was significant when alpha is 5% (p = 0.0001), which means that there is a significant difference in terms of conversion (in %) in each group, and the null hypothesis was rejected. To be more specific, from looking at the confidence interval of each group, as well as the confidence interval for the difference between them, it was found that in the treatment group more users made at least one or more purchases throughout the experiment.

**In light of all the above**, although we can say with 95% certainty that there are more users from the treatment group which made at least one purchase, than the control group (relatively), but in the end there was no significant difference between the two groups in terms of the average amount each user spent to be able to say that this difference was not due to a chance. Therefore, based on all of the results, my recommendation is to not launch the experiment for all users, since in the end the change doesn't yield us enough significant profit to justify this change. Although this is only the placement of a banner, the meaning and the aim are that we will eventually offer new and wider variety of products in the food and drink category, which probably doesn't have enough demand. I would suggest doing more market surveys, and several focus groups to understand if there is other category that worth to invest in (even a new category). Additionally, we can repeat the experiment with the banner for the new category, and see if the profit (the average spent) this time from each user will be significant higher and justify such a big change for the company.