A/B Test to Improve Homepage of GloBox (An Online Marketplace)

A Mastery Project Report Submitted to Masterschool

Submitted by: S M Asif Al - Samy

Submission Date: 8th July 2023

Abstract

This report is written based on an A/B testing of an online marketplace called “GloBox”. More precisely, GloBox wants to perform an experiment to see if they should launch a banner highlighting food and drink category products at the top of their website. GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. In case of A/B testing the customers are divided into two groups: Control group (Group A) who does not see the banner and the treatment group (Group B) who sees the banner.

Based on the spending and non-spending behaviors of the customers, the dataset has been customized and at last two metrics has been identified for the purpose of this A/B testing. From those two metrics, first one is categorized as “the difference in the conversion rate between two groups” where the conversion rate defines the rate of purchases in each group and the second one is categorized as “the difference in the average amount spent per user between two groups”.

After performing some statistical operations (hypothesis test and confidence interval) on both metrics, it is found that one of the metrics is statistically significant and another one is statistically insignificant.

Though only first metric (the difference in conversion rate) is staistically significant, it is recommended to launch the banner for everyone who is browsing their website. This is because, a banner is not typically an expensive feature to launch, in terms of engineering time or operational overhead. So if either metric has a significant increase, that’s enough justification to go ahead and launch it.

Table of Contents

1 Introduction......3
2 Context...... 3
2.1 Motivation of the experinment......3
2.2 A/B Test Setup...... 4
2.3 Overview of the Dataset...... 5
2.4 Extract the A/B Test Data......5
2.5 Calculate A/B Test Statistics...... 6
2.5.1 Hypothesis Testing......6
2.5.2 Confidence Intervals...... 7
3 Results......8
3.1 Results of A/B Test Statistics...... 8
3.2 Visualize the Results in Tableau...... 10
4 Recommendation......12
Appendix...... 13

List of Tables

Table 1: Values to perform the test statistics……8
Table 2: Values to perform the test statistics……9

List of Figures

Figure 1: Differentiation between the Control and Treatment groups……..4
Figure 2: Setup of the A/B Test……..5
Figure 3: ER (Entity Relationship) diagram of the database……5
Figure 4: Figure 4: Flowchart to choose the right statistical test……7
Figure 5: Overall conversion rate and average amount spent per group……10
Figure 6: Distribution based on device type, gender, and country……11
Figure 7: Visualization of confidence interval - Conversion Rate..……11
Figure 8: Visualization of confidence interval - Average Amount Spent.……11

1 Introduction

In the competitive landscape of e-commerce, businesses are constantly striving to optimize their online platforms to enhance customer experience and increase revenue. One widely used approach for this purpose is A/B testing, a powerful experimentation technique that allows businesses to compare two versions of a webpage, advertisement, or product feature to determine which one performs better in achieving specific goals.

In this paper, we explore a case study on GloBox, an online marketplace specializing in unique and high-quality products sourced from around the world. GloBox aimed to improve their homepage by conducting an A/B test focused on increasing awareness and revenue in their food and drink product category.

The test featured a prominent banner highlighting key products from the food and drink category on the mobile version of their website. The control group did not see the banner, while the test group was exposed to it. The main metrics of interest in this experiment were the user conversion rate (percentage of users who made a purchase) and the average amount spent per user in both the control and test groups.

Based on the metrics some sort of statistical analyses were performed such as hypothesis test, calculating confidence intervals in order to complete the A/B test. The dataset of GloBox including matrics of interest was also visualized to leverage the test. Based on the results and some important criterias, a solution has been recommended to GloBox if they include a banner for the food and drink categories.

2 Context

Throughout this section the whole context of the experiment with it's motivation, parameters, overview of the dataset, etc are discussed.

2.1 Motivation of the experinment

GloBox is an online marketplace that specializes in sourcing unique and high-quality products from around the world. They want to run an A/B test experiment to improve their homepage.

An A/B test is an experimentation technique used by businesses to compare two versions of a webpage, advertisement, or product feature to determine which one performs better. By randomly assigning customers or users to either the A or B version, the business can determine which version is more effective at achieving a particular goal.

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, their food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue.

2.2 A/B Test Setup

The Growth team decides to run an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the treatment/test group sees it as shown below:

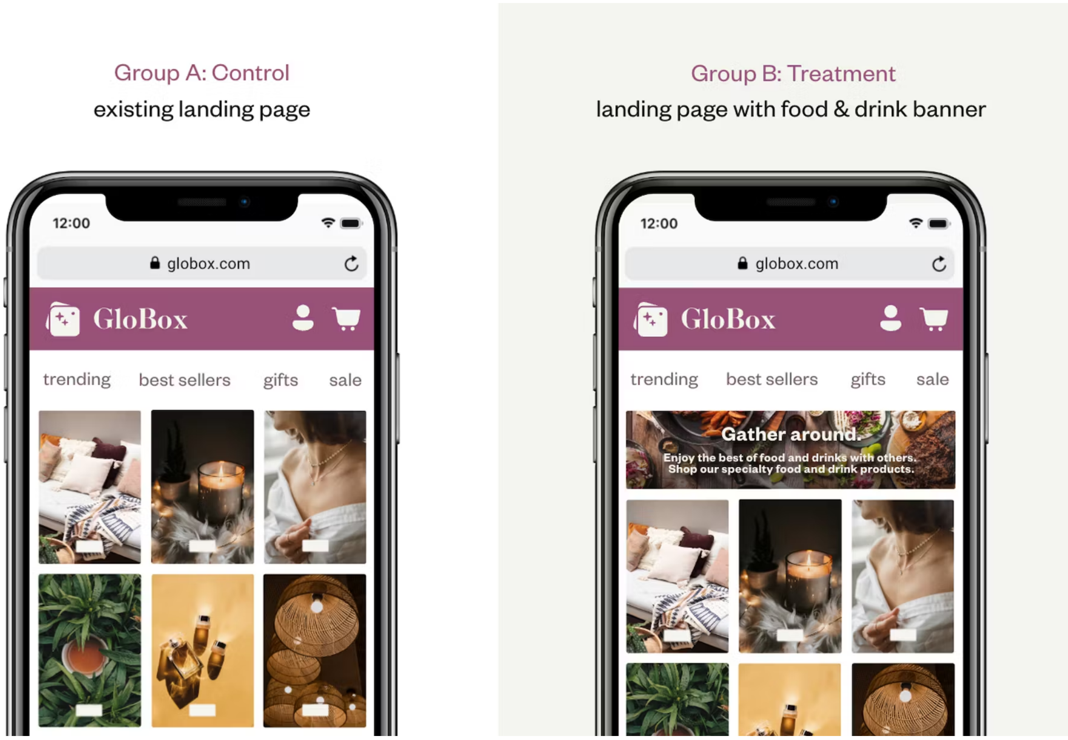

Figure 1: Differentiation between the Control and Treatment groups

The setup of the A/B test is as follows:

1. The experiment is only being run on the mobile website.
2. A user visits the GloBox main page and is randomly assigned to either the control or test group. This is the join date for the user.
3. 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.
4. 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".

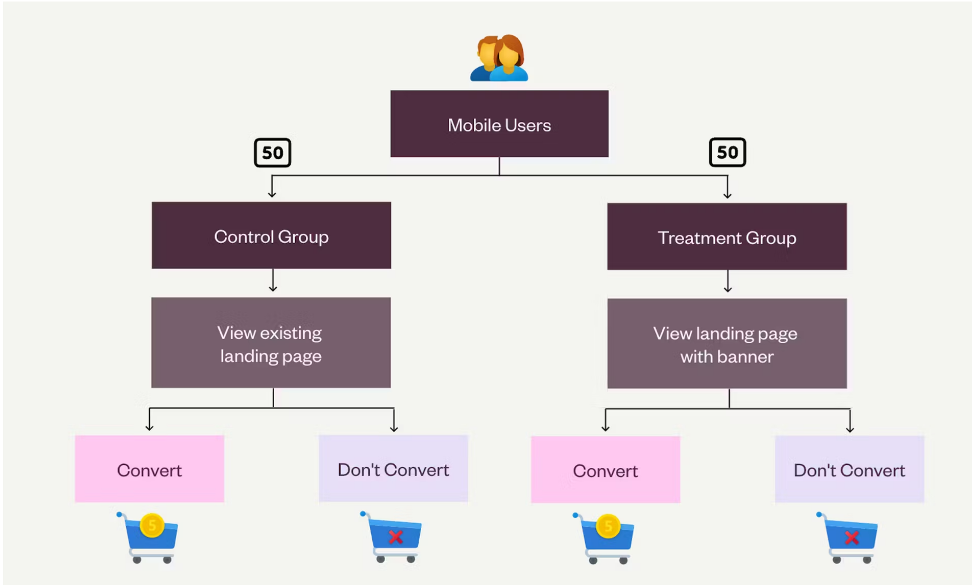

Figure 2: Setup of the A/B Test.

2.3 Overview of the Dataset

GloBox stores its data in a relational database, which can be accessed in Beekeeper Studio, a popular open-source SQL editor that makes it easy to query databases.

Figure 3: ER (Entity Relationship) diagram of the database.

Three key things to note:

1. All users should be assigned to one A/B test group
2. Not all users make a purchase
3. Purchase activity is for all product categories, not just food and drink

2.4 Extract the A/B Test Data

In order to become familiar with the dataset and extract the data to use for the future statistical analysis, a bunch of SQL queries has been written and applied. The queries are available in the Appendix section.

Two important queries, one is to find the user **conversion rate** for the control and treatment groups and another one is to find the **average amount spent** per user for the control and treatment groups, including users who did not convert, are very useful to perform hypothesis testing.

After extracting the final dataset as CSV, it represents the columns such as the user ID, the user's country, the user's gender, the user's device type, the user's test group, whether or not they converted (spent > $0), and how much they spent in total ($0+).

2.5 Calculate A/B Test Statistics

The dataset has been exported to a spreadsheet and performed **Hypothesis Testing** and calculated **Confidence Intervals** to complete the test. Thre are two columns which are used as the test metrics such as **conversion** and **total_spent**. From this two columns we have found our **metrics of interest**: the user conversion rate for the control and treatment groups and the average amount spent per user for the control and treatment groups.

2.5.1 Hypothesis Testing

Hypothesis test helps us determine whether or not there is a statistically significant difference between the two test groups by using **Statistical Tests**. According to Rebecca Bevans (2020), Statistical Tests can be used to[1]:

- determine whether a predictor variable has a statistically significant relationship with an outcome variable
- estimate the difference between two or more groups

Statistical Tests assume a null hypothesis of no difference (for conversion rate / average amount spent) between groups and an alternative hypothesis of difference between groups. Then the tests determine whether the observed data fall outside of the range of values predicted by the null hypothesis. If yes, then we can reject the null hypothesis and accept the alternative hypothesis and vice versa. As we already knew the types of variables we are dealing with, we have used the following flowchart (Figure 4) to choose the right statistical test for our data.[1]

[1] https://www.scribbr.com/statistics/statistical-tests/ [Accessed on 15th of June 2023]

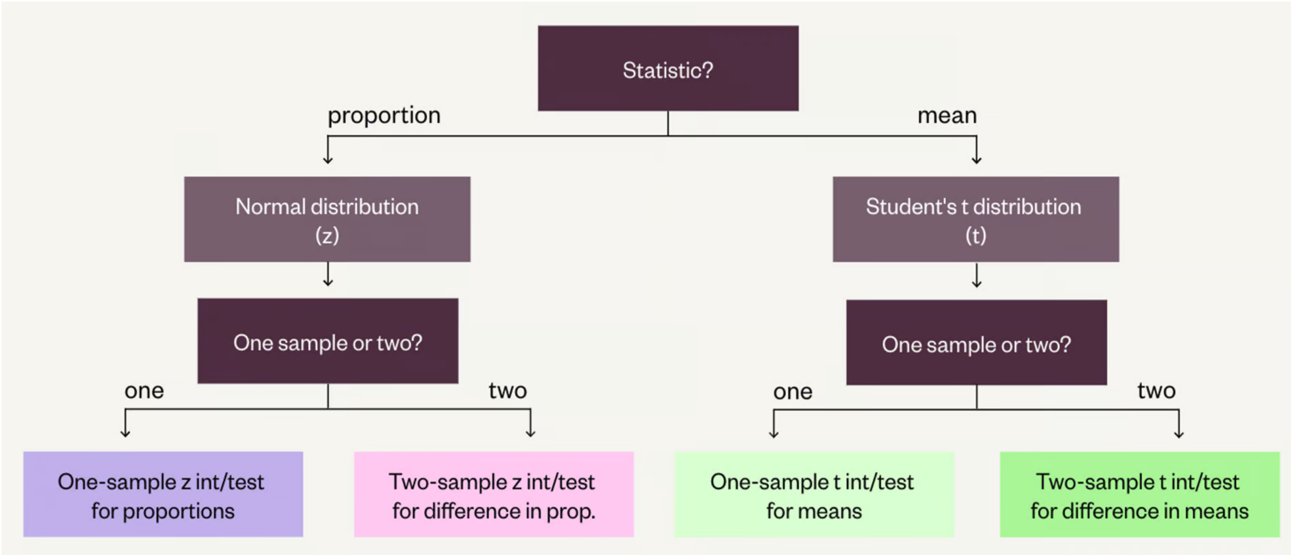

Figure 4: Flowchart to choose the right statistical test.

Both types of tests (t-test and z-test) provide a **p-value** or probability value that tells us the statistical significance of our findings by comparing with a significance level that is 5% (denoted as α) in our test. However, this threshold can also be set higher or lower. According to Pritha Bhandari (202[2]1), in a hypothesis test, the p-value is compared to the significance level to decide whether to reject the null hypothesis[2]. If the p-value is higher than the α (0.05), the null hypothesis is not refuted, and the results (differences in our test) are not statistically significant. Alternatively, if the p-value is lower than the αl, the results are interpreted as refuting the null hypothesis and reported as statistically significant.

Therefore, a hypothesis test has been conducted to see whether there is a difference in the conversion rate between the two groups using the normal distribution and a 5% significance level. Another hypothesis test is also conducted to see whether there is a difference in the average amount spent per user between the two groups using the t distribution and a 5% significance level.

2.5.2 Confidence Intervals

Confidence intervals help us understand the magnitude of the difference between the two groups. According to Rebecca Bevans (2020), the confidence interval is the range of values that we expect our estimate (difference in conversion rate and difference in average amount spent between 2 groups) to fall between a certain percentage (95% in our test) of the time if we run our experiment again or re-sample the population in the same way[3].

The **confidence level** is the percentage of times we expect to reproduce the estimate between the **upper** and **lower** bounds of the confidence interval, and is set by the α value.

---

[2] https://www.scribbr.com/statistics/statistical-significance/ [Accessed on 15th of June 2023]
[3] https://www.scribbr.com/statistics/confidence-interval/ [Accessed on 15th of June 2023]

Confidence, in statistics, is another way to describe probability. For example, if we construct a confidence interval with a 95% confidence level, we are confident that 95 out of 100 times the estimate will fall between the upper and lower values specified by the confidence interval.

If we want to calculate the confidence intervals for our estimates (for both the metrics), we need to know:

1. The point estimate we are constructing the confidence interval for
2. The critical values for the test statistic
3. The standard deviation of the sample
4. The sample size

Once we know each of these components, we can calculate the confidence interval for our estimate by plugging them into the confidence interval formula that corresponds to our data.

N.B. all the formulas to estimate each of the terms described above are written in the appendix section.

3 Results

All the results associated with A/B testing as well as their calculations are presented in this section.

3.1 Results of A/B Test Statistics

1. A hypothesis test to see whether there is a difference in the conversion rate between the two groups:

Null Hypothesis (**H_0**): There is no difference in the conversion rate between 2 groups.
If conversion rate for **group A** is denoted by **$p\hat{}_A$** and conversion rate of **group B** is denoted by **$p\hat{}_B$** then according to Null Hypothesis, **H_0: $p\hat{}_A - p\hat{}_B = 0$.**

Alternative Hypothesis (**H_1**): There is a difference in the conversion rate between 2 groups.
H_1: $p\hat{}_A - p\hat{}_B \neq 0$.

Table 1: Values to perform the test statistics.

Group A	Group B
$p\hat{}_A = 0.0392$	$p\hat{}_B = 0.0463$
$S_A = 0.1942$	$S_B = 0.2102$
$n_A = 24343$	$n_B = 24600$

Test Statistics (**z**): -3.88

p(-3.88>|z|) ≈ NORMDIST(-3.88, 0, 1, TRUE) ≈ 0.00005236572431 X 2 = 0.0001047314599

p-value < 0.05

Therefore, we reject the Null Hypothesis. There is a significant difference in the conversion rate between 2 groups.

2. 95% confidence interval for the difference in the conversion rate between the treatment and control (treatment-control):

Confidence interval = sample statistic ± critical value ∗ standard error

sample statistic = $p\^{}_A - p\^{}_B$, Critical value = $Z_{(1 - \alpha/2)}$ = 1.960, standard error = 0.00183

margin of error = critical value ∗ standard error = 0.00359

Lower bound = sample statistic - margin of error = 0.0035
Upper bound = sample statistic + margin of error = 0.0107

3. A hypothesis test to see whether there is a difference in the average amount spent per user between the two groups using the t distribution and a 5% significance level:

Null Hypothesis (**H_0**): There is no difference in the average amount spent per user between 2 groups.

If average amount spent per user for **group A** is denoted by **X_A** and average amount spent per user of **group B** is denoted by **X_B** then according to Null Hypothesis, **H_0: $X_A - X_B = 0$.**

Alternative Hypothesis (**H_1**): There is a difference in the average amount spent per user between 2 groups. **H_1: $X_A - X_B \neq 0$.**

Table 2: Values to perform the test statistics.

Group A	Group B
X_A = 3.374518468	X_B = 3.390866946
S_A = 25.93639056	S_B = 25.4141096
N_A = 24343	N_B = 24600

Test Statistics (**t**): -0.0706

$p(|t| \geq 0.0706)$ ≈ T.DIST(-0.0706, 24342, TRUE) ≈ 0.4718583452X 2 = 0.9437166904

p-value > 0.05

Therefore, we failed to reject the Null Hypothesis. There is insignificant difference in the average amount spent per user between 2 groups.

4. 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control):

sample statistic = $X_A - X_B$

critical value = $Z^* = t_{(1 - \alpha/2,\ df)} = 1.960$

margin of error = critical value $*$ standard error = 0.4551

Lower bound = sample statistic - margin of error = - 0.4390
Upper bound = sample statistic + margin of error = 0.4710

3.2 Visualize the Results in Tableau

The dataset has been uploaded into tableau to visualize the results. The following figure (Figure 4) represents the overall conversion rate and average amount spent per group from which it is

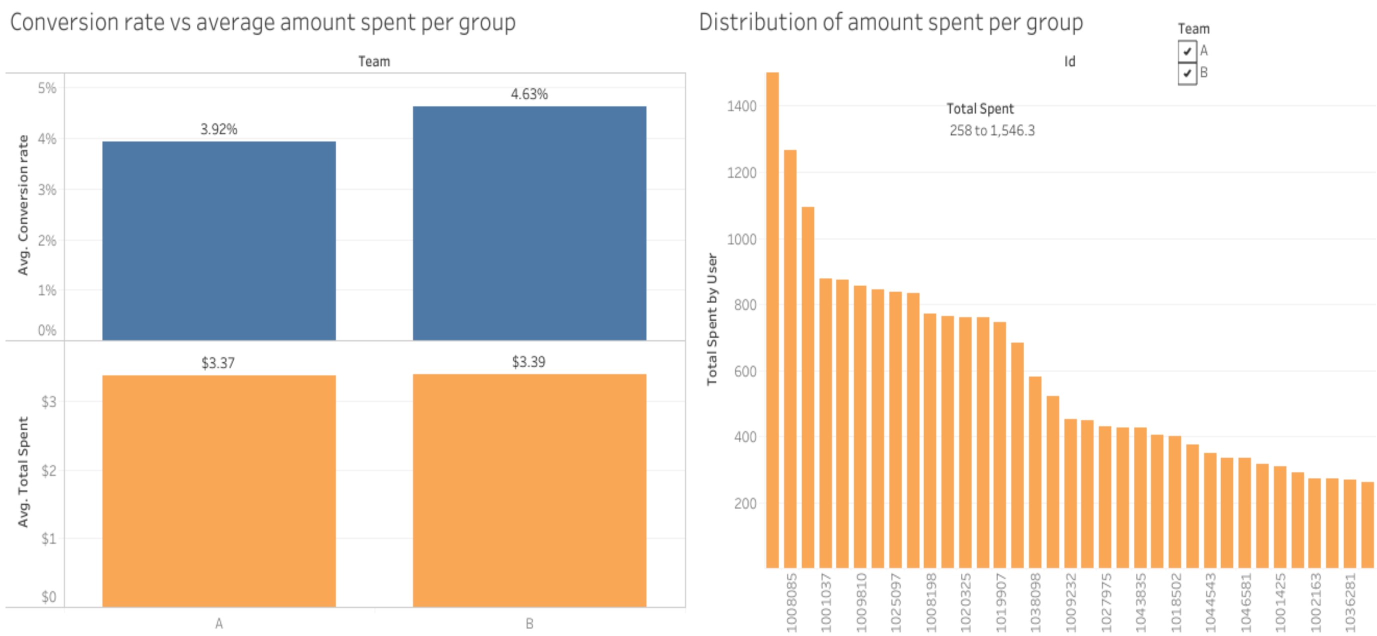

Figure 5: Overall conversion rate and average amount spent per group.

pretty much clear that the difference in conversion rate between two group is significant whereas the difference in average amount spent between two group is insignificant.

The following figure (Figure 5) shows the distribution of conversion rate with respect to average amount spent per group based on device usage, gender, and country. From the figure, though most users use android, ios has most purchased users. As the products are related to fashion & jewelry, it obvious that female users purchased more than male users.

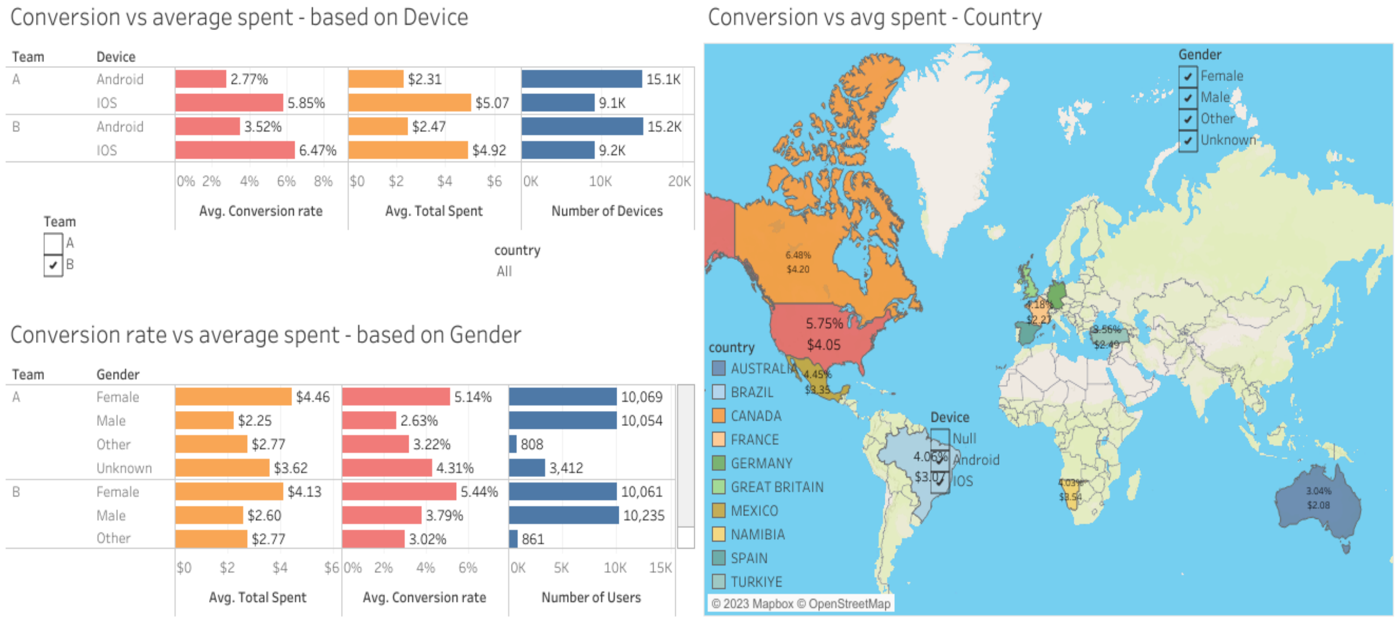

Figure 6: Distribution based on device type, gender, and country.

While exploring map in tableau public ([explore map in tableua](#)), it is identified that Cannada and Mexico have major changes in coversion rate as well as in average amount spent and other countries have minor mixed type changes. Therefore, it is suggested that put more emphasis on these countries such as offering more varieties in food and drinks categories to these areas which makes more sale and increase revenue.

The following figure (Figure 6) represents the 95% confidence interval of both the metrics of interest. 95% confidence interval for the difference in the conversion rate between the treatment and control group is (0.0035, 0.0107).

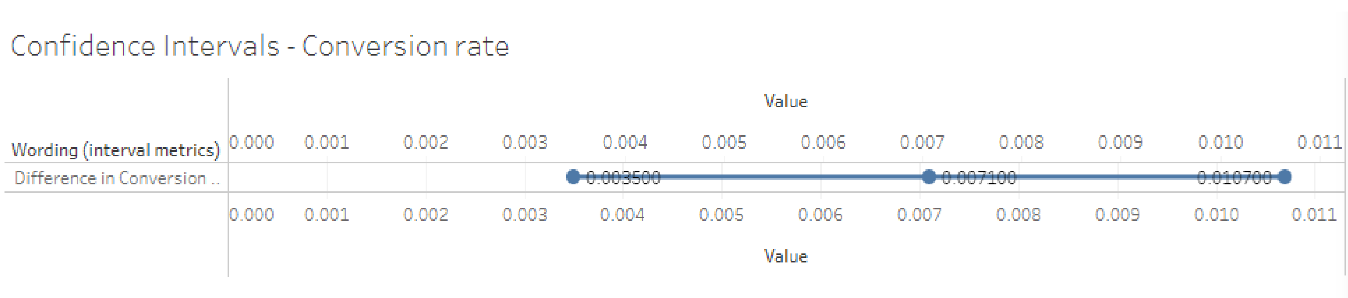

Figure 7: Visualization of confidence interval - Conversion Rate.

On the other hand, 95% confidence interval for the difference in the average amount spent per user between the treatment and the control is (- 0.4390, 0.4710) which is quite bigger interval than the difference in the conversion rate. Section 2.5 is the reference of this CI calcululation.

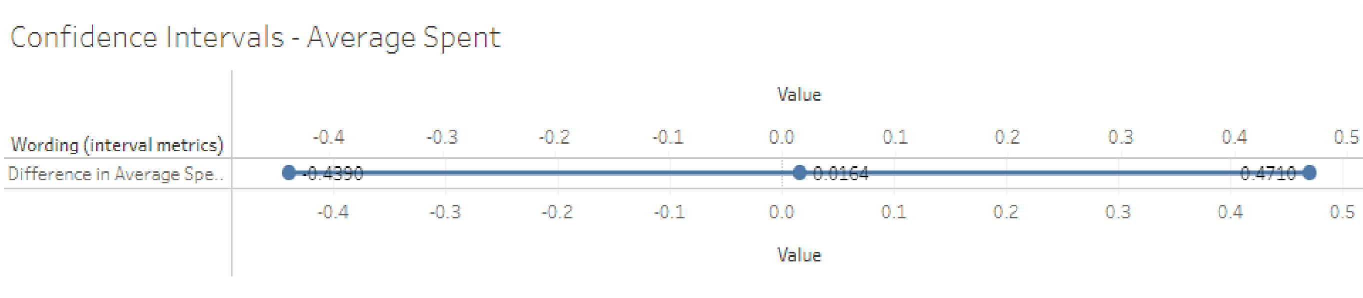

Figure 8: Visualization of confidence interval - Average Amount Spent.

4 Recommendation

Only one metric (the difference in conversion rate) is staistically significant, and another one is insignificant. We don't have enough evidence to take a decision. However, a banner is not typically an expensive feature to launch, in terms of engineering time or operational overhead. So if either metric has a significant increase, that's enough justification to go ahead and it is recommended to launch the banner for everyone who is browsing their website. The increase in the conversion rate might be a indication that in future more users can buy products that results in more revenue.

Appendix

1. SQL query to extract the dataset where the query returns: the user ID, the user's country, the user's gender, the user's device type, the user's test group, whether or not they converted (spent > $0), and how much they spent in total ($0+).

```
WITH active_users
     AS (SELECT uid,
                Sum(spent) AS total_spent
         FROM   activity
         GROUP  BY uid)
SELECT u.*,
       g.device,
       g.team,
       CASE
         WHEN u.id = a.uid THEN 'converted'
         ELSE 'not converted'
       END AS conversion,
       CASE
         WHEN u.id = a.uid THEN a.total_spent
         ELSE 0
       END AS total_spent
FROM   users AS u
       LEFT JOIN active_users AS a
              ON u.id = a.uid
       JOIN groups AS g
         ON g.uid = u.id
```

2. Formalas of A/B Test Statistics:

$$T = \frac{sample\ statistics - null\ value}{statdard\ error}$$

a) For difference in the conversion rate: (Two-sample z-test with pooled proportion)

$$sample\ statistics = (\hat{p}_A - \hat{p}_B)$$

$$null\ value = p_0$$

$$standarde\ error = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$$T = \frac{(\hat{p}_A - \hat{p}_B) - p_0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

$$\hat{p} = \frac{(\hat{p}_A * n_A) + (\hat{p}_B * n_B)}{n_A + n_B}$$

b) For difference in the average amount spent: (Two-sample t-test with unpooled variance)

$$sample\ statistics = (\hat{x}_A - \hat{x}_B)$$

$$null\ value = \mu_0$$

$$standarde\ error = \sqrt{(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B})}$$

$$T = \frac{(\hat{x}_A - \hat{x}_B) - \mu_0}{\sqrt{(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B})}}$$

3. Link to the spreadsheet where the test statistics are performed:
Click here to explore

4. Link to the Tableau visualization dashboard:
Click here to explore

5. Link to the confidence interval (for conversion rate) visualization in Tableau:
Click here to explore

6. Link to the confidence interval (for average amount spent) visualization in Tableau:
Click here to explore

7. A list of additional SQL queries to explore the dataset:
    1. Can a user show up more than once in the activity table?

```
SELECT uid,
       Count(uid) count_id
FROM   activity
GROUP  BY uid
HAVING Count(uid) > 1
```

    2. What type of join should we use to join the users table to the activity table?

```
SELECT u.*,
       a.dt AS date,
       a.spent
FROM   users AS u
       LEFT JOIN activity AS a
              ON u.id = a.uid
```

3. What SQL function can we use to fill in NULL values?
```
SELECT u.*,
       a.dt AS date,
       COALESCE(a.spent, 0) AS spent
FROM   users AS u
       LEFT JOIN activity AS a
              ON u.id = a.uid
```

4. What are the start and end dates of the experiment?
```
SELECT Min(join_dt) AS start_date,
       Max(join_dt) AS end_date
FROM   groups
```

5. How many total users were in the experiment?
```
SELECT Count(DISTINCT uid) total_users
FROM   groups
```

6. How many users were in the control and treatment groups?
```
SELECT team,
       Count(DISTINCT uid) total_users
FROM   groups
GROUP  BY team
```

7. What was the conversion rate of all users?
```
WITH active_users
     AS (SELECT uid,
                Sum(spent) AS total_spent
         FROM   activity
         GROUP  BY uid)
SELECT Round(Avg(CASE
                   WHEN g.uid = au.uid THEN 1
                   ELSE 0
                 END), 4) AS pct_conversion
FROM   groups AS g
       LEFT JOIN active_users AS au
              ON g.uid = au.uid
```

8. What is the user conversion rate for the control and treatment groups?
```
WITH active_users
     AS (SELECT uid,
                Sum(spent) AS total_spent
         FROM   activity
         GROUP  BY uid)
SELECT g.team,
       Round(Avg(CASE
                   WHEN g.uid = au.uid THEN 1
```

```
                        ELSE 0
                      END), 4) AS pct_conversion
FROM    groups AS g
        LEFT JOIN active_users AS au
               ON g.uid = au.uid
GROUP  BY g.team
```

9. What is the average amount spent per user for the control and treatment groups, including users who did not convert?

```
WITH active_users
     AS (SELECT uid,
                 Sum(spent) AS total_spent
         FROM   activity
         GROUP  BY uid)
SELECT g.team,
       Avg(CASE
             WHEN g.uid = au.uid THEN au.total_spent
             ELSE 0
           END) AS avg_spent
FROM    groups AS g
        LEFT JOIN active_users AS au
               ON g.uid = au.uid
GROUP  BY g.team
```