

# **GLOBOX: A/B Test Analysis**

Saranya Lahiri Mukhopadhyay

## Contents

|  |           |
|--|-----------|
| <b>1. Summary.....</b>                       | <b>2</b>  |
| <b>2. About the Company: GLOBOX .....</b>    | <b>3</b>  |
| <b>3. The Experiment .....</b>               | <b>4</b>  |
| <b>3a. Step 1 - SQL: .....</b>               | <b>4</b>  |
| <b>3b. Step 2- Hypothesis Testing: .....</b> | <b>7</b>  |
| <b>3c. Visualization: .....</b>              | <b>9</b>  |
| <b>3d. Result.....</b>                       | <b>14</b> |
| <b>4. Recommendations .....</b>              | <b>15</b> |

## 1. Summary

GloBox is an e-commerce company in the high-end décor and food & drinks categories. Since in recent times there is a huge growth in the Food & Drinks category, the company has decided to conduct an A/B Test to determine if a banner highlighting the key products in the food and drink category at the top of the website will increase the company's revenue.

Tools used to reach a solution to the above question:

- **SQL**: to extract the data required to conduct the test from the dataset provided
- **Hypothesis Testing**: to show if there is a statistically significant difference between the two groups, namely
  - Group A or the **Control Group**: the existing landing page.
  - Group B or the **Treatment Group**: adding the food and drink banner in the landing page.
- **Tableau**: To create visualisations for clear and easy understanding.

While conducting the A/B testing the following was considered:

- Null Hypothesis  $H_0$ : **The conversion rate is not dependent on showing the banner at the top of the website.**
- Alternate Hypothesis  $H_1$ : **The conversion rate is directly dependent on showing the banner on the website** ie the conversion rate increases or decreases with displaying the banner at the top of the website.
- Significance Level ( $\alpha$ ): 0.05

The p-value was found to be 0.0001104 which is much below the significance level and hence the Null Hypothesis was REJECTED.

Thus implying that showing the banner at the top of the landing page has an impact on the conversion rate. Hence the banner should be launched at the top of the website to increase the revenue.

## 2. About the Company: GLOBOX

As an e-commerce company, GloBox specializes in unique and high-quality products from around the world. Their products range from handmade jewellery and textiles in the boutique fashion items and high-end decor products category to exotic spices and rare teas in the food and drink category.

In the last few months, the company had a huge growth in the food and drinks division. As a result, the company wants to launch a banner containing the company's highlighted products in the Food & Drinks category with the expectation of increasing the revenue. To achieve this purpose, the company is conducting an A/B test.

### Stakeholders related to A/B Testing:

- **Growth Product & Engineering Team:** This is the team that you work with at GloBox. The team is made up of a product manager, a user experience designer, an engineering manager and several software engineers, and you, the data analyst. The team develops features for the GloBox website that drive growth in users and revenue.
- **Leila Al-Farsi, Product Manager, Growth:** Leila is the product manager for the Growth product and engineering team. Alongside Alejandro, she leads the Growth team by deciding their goals and projects, measuring their success against defined KPIs, and communicating results to other company leaders like Mei.
- **Alejandro Gonzalez, User Experience Designer, Growth:** Alejandro is the designer for the Growth product and engineering team. He conducts user research and designed the experience that the A/B test is evaluating.
- **Mei Kim, Head of Marketing:** Mei oversees the Marketing team, which works on targeting audiences with effective marketing campaigns to drive customers to the GloBox website. She collaborates frequently with Leila and Alejandro to design website experiences that will align well with the current marketing efforts.

### 3. The Experiment

#### 3a. Step 1- SQL:

Used SQL SELECT queries to understand the dataset and also extract data from the dataset into a CSV file for further analysis through A/B Testing.

**Query 1** To get the total sample size for the experiment:

```
SELECT COUNT(DISTINCT(uid))  
FROM groups
```

Result

| count ▲ |
|---------|
| 48943   |

**Query 2** To get the number of users in Control Group and Treatment Group and the number who converted i.e., have spent after visiting the website:

```
SELECT CASE WHEN "group" = 'A' THEN 'Control'  
            WHEN "group" = 'B' THEN 'Treatment'  
            ELSE 'Error' END AS Group,  
       COUNT(DISTINCT(groups.uid)) AS total_group_count,  
       COUNT(DISTINCT(activity.uid)) AS no_of_conversions  
FROM activity  
RIGHT JOIN groups  
ON activity.uid = groups.uid  
GROUP BY "group"
```

Result

| group ▲   | total_group_count ▲ | no_of_conversions ▲ |
|-----------|---------------------|---------------------|
| Control   | 24343               | 955                 |
| Treatment | 24600               | 1139                |

**Query 3** To get the conversion rates of Control Group & Treatment Group:

```
SELECT CASE WHEN "group" = 'A' THEN 'Control'
           WHEN "group" = 'B' THEN 'Treatment'
           ELSE 'Error' END AS Group,
       CAST(COUNT(DISTINCT activity.uid ) AS FLOAT8)/
       CAST(COUNT(DISTINCT groups.uid) AS FLOAT8)*100) as
       Conversion_Rate
FROM activity
RIGHT JOIN groups ON activity.uid = groups.uid
GROUP BY "group"
```

Result

| group     | conversion_rate    |
|-----------|--------------------|
| Control   | 3.9230990428459926 |
| Treatment | 4.630081300813008  |

**Query 4** To get the average spent for Control Group & Treatment Group:

```
SELECT CASE WHEN "group" = 'A' THEN 'Control'
           WHEN "group" = 'B' THEN 'Treatment'
           ELSE 'Error' END AS Group,
       SUM(COALESCE(spent, 0))/COUNT(DISTINCT(Joined_Table.uid)) as
       Average_spent
FROM (select groups.uid, "group", spent
      FROM activity
      RIGHT JOIN "groups"
            activity.uid = groups.uid) as Joined_Table
GROUP BY "group"
```

Result

| group     | average_spent       |
|-----------|---------------------|
| Control   | 3.3745184679288412  |
| Treatment | 3.39086694588578326 |

**Query 5** Query that 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+):

```
SELECT users.id,  
       country,  
       gender,  
       groups.device,  
       "group",  
       SUM(COALESCE(spent, 0)) as spent_all  
FROM users  
INNER JOIN "groups" on groups.uid = users.id  
LEFT JOIN activity on users.id = activity.uid  
GROUP BY users.id, country, gender, groups.device, "group"
```

### Result

| id      | country | gender | device | group | spent_all |
|---------|---------|--------|--------|-------|-----------|
| 1000000 | CAN     | M      | I      | B     | 0         |
| 1000001 | BRA     | M      | A      | A     | 0         |
| 1000002 | FRA     | M      | A      | A     | 0         |
| 1000003 | BRA     | M      | I      | B     | 0         |
| 1000004 | DEU     | F      | A      | A     | 0         |
| 1000005 | GBR     | F      | A      | B     | 0         |
| 1000006 | ESP     | M      | A      | B     | 0         |
| 1000007 | BRA     | F      | A      | A     | 0         |
| 1000008 | BRA     | F      | A      | A     | 0         |

48943 0 affected 1.604 s

### 3b. Step 2- Hypothesis Testing:

Access the excel sheet project-1.xlsx to view the Hypothesis testing done from the SQL extract using the SQL **Query 5** in section **3a. Step 1**.

project-1.xlsx file link:

<https://1drv.ms/f/s!AqPopfURvUT6grp4OWoBFjDpN78Y2Q?e=wAVwbr>

- Null Hypothesis  $H_0$ : The conversion rate is not dependent on showing the banner at the top of the website.
- Alternate Hypothesis  $H_1$ : The conversion rate is directly dependent on showing the banner on the website ie, the conversion rate is directly dependent on showing the banner on the website
- Working with Proportions
- Comparing between two samples
- two-sided test

We can have two outcomes from the test:

1. Reject  $H_0$  and accept  $H_1$  due to sufficient statistical evidence in favour of  $H_1$ .
2. Fail to reject  $H_0$  due to insufficient statistical evidence in favour of  $H_1$ .

#### 3b.1. Explaining a hypothesis testing:

Null Hypothesis,  $H_0$  is  $p_1 = p_2$

Alternative Hypothesis  $H_1$ ,  $p_1 \neq p_2$

Significance Level ( $\alpha$ ): The probability of Rejecting the Null Hypothesis if it is true

T-value and Z-value are examples of test statistics that is calculated from a sample data during a hypothesis test.

Z-value tells where the data lies in the data distribution whereas a T-value indicates the difference between two groups. If the T-value is large, it indicates that the groups are different and if the T-value is small, it indicates that the groups are similar.

Since we are considering a two-sided or two-tailed test with a significance level ( $\alpha$ ) of 0.05, we will get two rejection region depending on the significance level, where the rejection regions can be found as:

$$\text{Rejection Region} = \alpha/2 = 0.05/2 = \mathbf{0.025}$$

Now when the Z-value is calculated, if it falls to be in the rejection region then the Null Hypothesis will be rejected and if it does not then the Null Hypothesis will be accepted.

When  $\alpha = 0.025$ , the Z-score from the z-table can be found as:  $Z = \pm 1.96$ .

Hence if the Z-score from the test is lower than -1.96 or greater than 1.96, the Null Hypothesis is **rejected**, else **accepted**.



Whereas, p-value: Is the probability of falsely rejecting the Null Hypothesis.

Hence if **p-value** >  $\alpha$ , we **Accept** the null hypothesis

And if **p-value** <  $\alpha$ , we **Reject** the null hypothesis.

### 3b.2. Results from hypothesis testing:

To conduct the test, the following has been taken into consideration:

|                                 |         |
|---------------------------------|---------|
| Control Group                   | Group A |
| Treatment Group                 | Group B |
| Significance Level ( $\alpha$ ) | 0.05    |
| Confidence Interval             | 95%     |

After conducting the test, the following has been found:

|  |             |
|--|-------------|
| No. of users in Control group                              | 24343       |
| No. of users in Treatment group                            | 24600       |
| Conversion Rate of Control group, p1                       | 3.92%       |
| Conversion Rate of Treatment group, p2                     | 4.63%       |
| Z-value  | -3.867      |
| p-value  | 0.000110415 |
| T-Value of average amount spent by users of the two groups | -0.07042491 |

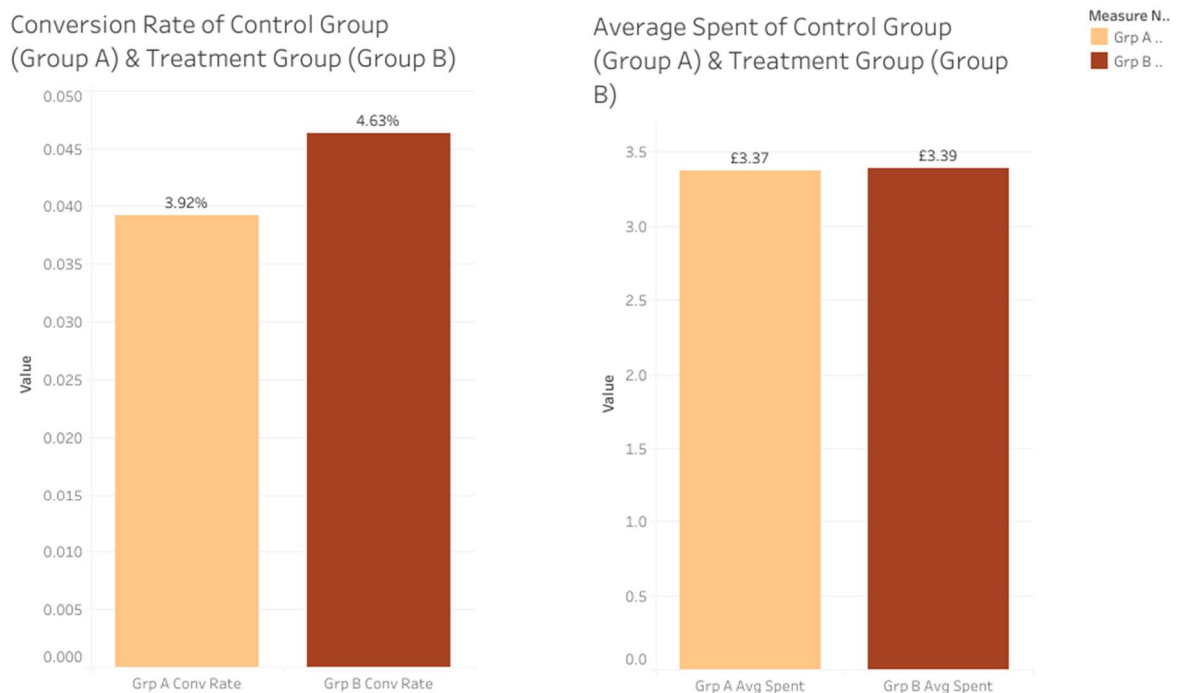
As can be seen from the above table that the p-value is **MUCH SMALLER** than the significance level ( $\alpha$ ), hence the Null Hypothesis is **REJECTED** in this experiment.

### 3c. Visualization:

Following Section will show the visualizations created as part of the experiment.

#### 3c.1. Comparing the Conversion Rate & Average Spent for Control Group (A) & Treatment Group (B):

##### Comparison of Control Group & Treatment Group

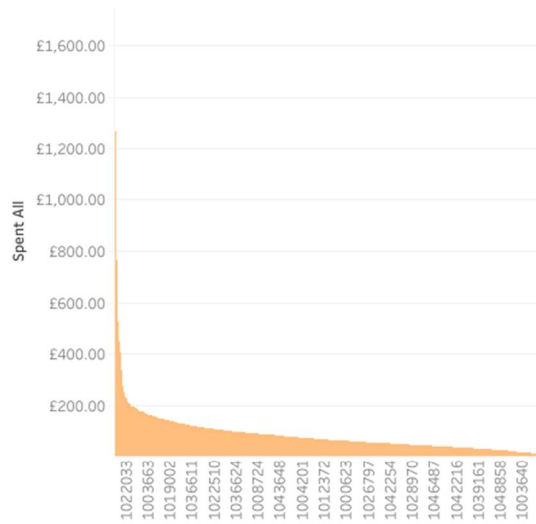


As can be seen from the above visual 3c.1, that although the conversion rate of Treatment Group is greater than the Control Group, the average spent of both groups are almost same which is further elaborated in visual 3c.2.

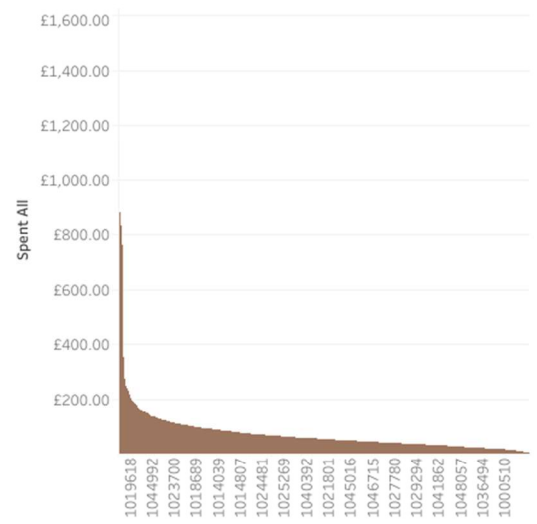
### 3c.2. The Spent distribution of Control Group & Treatment Group:

Spent Distribution for Control Group & Treatment Group

Group A: Spent Distribution per User



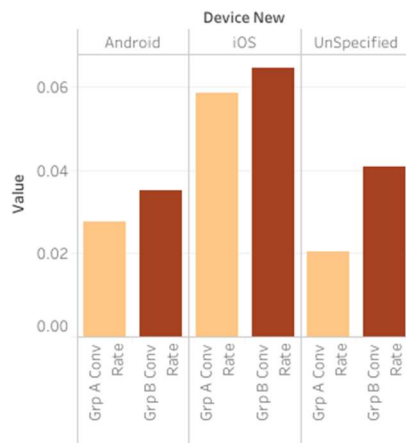
Group B: Spent Distribution pe User



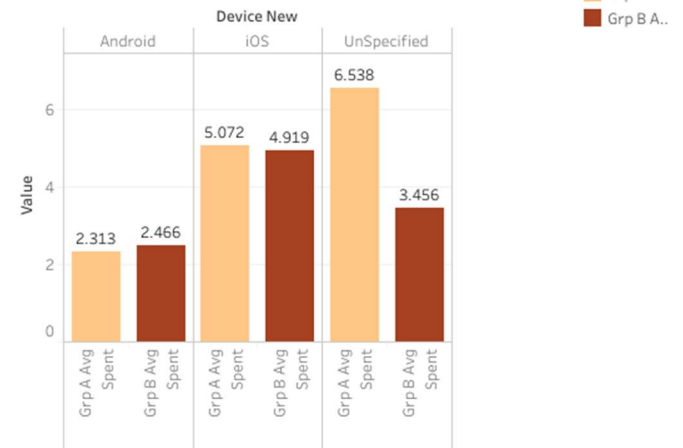
### 3c.3. Comparison between Control Group & Treatment Group as per Device the users use:

#### Comparison of Control Group & Treatment Group as per Device

Conversion Rate as per Device



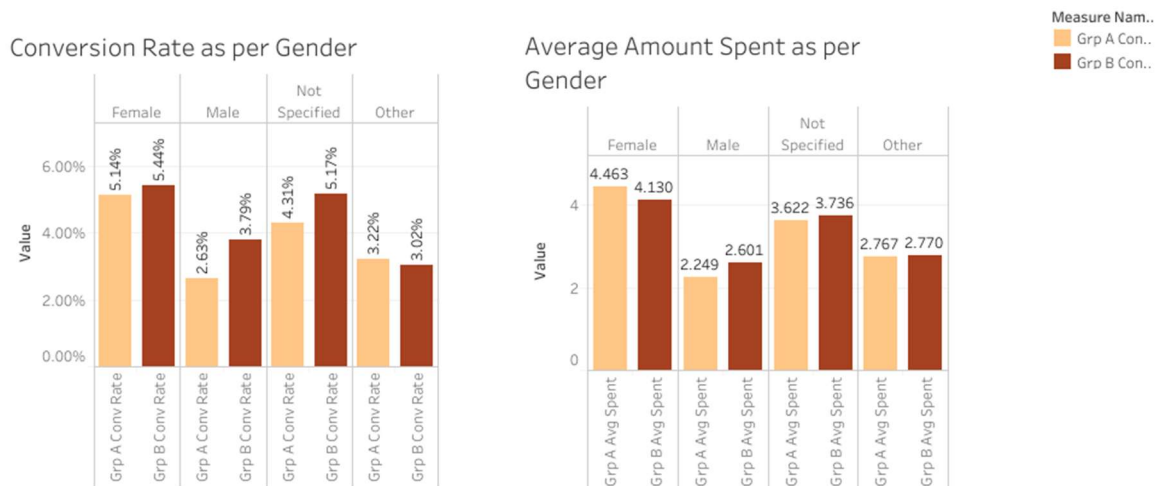
Average amount Spent as per Device



From visual 3c.3, we see that the conversion rate of the Treatment Group is higher than the Control Group irrespective of which device they are using. However, when the Average Amount Spent is considered, it seems that the Unspecified Users of the Control Group have a big impact on the revenue. This means that although the conversion rate of Control Group is around half of that of the Treatment Group for the Unspecified Users, the revenue generated by the Control Group is almost Double than that of the Treatment group.

### 3c.4. Comparison between Control Group & Treatment Group as per their Gender:

#### Comparison of Control Group & Treatment Group as per Gender

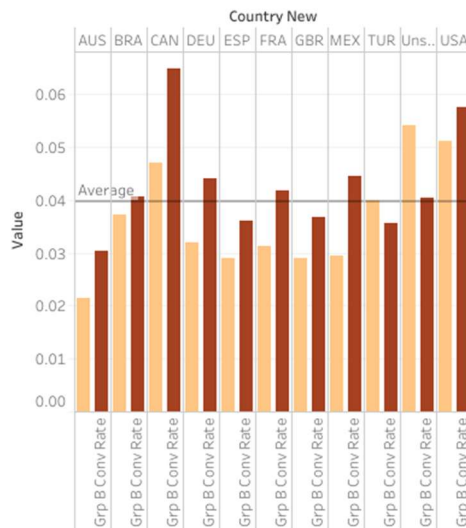


Similarly, the conversion rate and spent amount of female is much higher than the rest of the gender types. Which suggests that there is scope to increase revenue within the other gender types with proper marketing strategy. Refer visual 3c.4

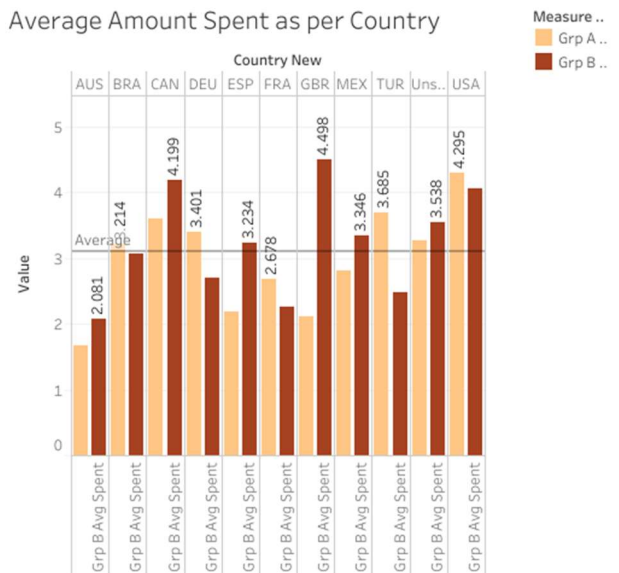
### 3c.5 Comparison of Control Group & Treatment Group as per their country:

#### Comparison of Control Group & Treatment Group as per Country

Conversion Rate as per Country



Average Amount Spent as per Country



From visual 3c.5, we can say that considering the average conversion rate, North American countries like Canada, USA & Mexico have above average conversion rate. Whereas, European countries like Germany, France & Spain barely touch the average bar.

#### Links to visuals

- 3c.1.  
[https://public.tableau.com/views/Project1Globox/Dashboard1?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Project1Globox/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)
- 3c.2.  
[https://public.tableau.com/views/Project1Globox/Dashboard2?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Project1Globox/Dashboard2?:language=en-US&:display_count=n&:origin=viz_share_link)
- 3c.3.  
[https://public.tableau.com/views/Project1Globox/Dashboard3?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Project1Globox/Dashboard3?:language=en-US&:display_count=n&:origin=viz_share_link)
- 3c.4.  
[https://public.tableau.com/views/Project1Globox/TestGroupsABandGender?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Project1Globox/TestGroupsABandGender?:language=en-US&:display_count=n&:origin=viz_share_link)
- 3c.5.  
[https://public.tableau.com/views/Project1Globox/TestGroupsABandCountry?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Project1Globox/TestGroupsABandCountry?:language=en-US&:display_count=n&:origin=viz_share_link)

### 3d. Result

Below are the results achieved out of the entire experiment.

1. After conducting the Hypothesis Test its found that Since  $p\text{-value} < \text{significance level}$  ie,  $0.0001104 < 0.05$ , Hence **REJECT** the null hypothesis.

This means that we accept the Alternate Hypothesis ie, the conversion rate is directly dependent on showing the banner on the website.

2. While determining if there is a difference in the average amount spent per user between the control & treatment group, it is found that the T-value is -0.0704 which is very small. Hence implying that although the average spent per user for the treatment group is marginally higher, but the two groups are similar with respect to the average amount spent per user.

## 4. Recommendations

So, at the end of performing the entire Data Analysis process for GloBox to answer the initial questions following are the recommendations.

- Launch the banner at top of the website for the next quarter to observe the actual conversion rate and do a reconsideration at the end of the quarter to continue with the banner or not.
- Further research to understand the Unspecified device segment, as the average spent amount is the highest, so there is a scope to increase revenue with that segment if proper marketing strategies can be applied after knowing their preferences.
- Derive marketing strategies to understand user behaviour the rest of the gender besides Female as there are lot of scope to increase revenue with them.