Saranya Lahiri Mukhopadhyay

**GLOBOX: A/B Test Analysis**

**Contents**

[**1.** **Summary** 2](#_Toc139670090)

[**2.** **About the Company: GLOBOX** 3](#_Toc139670091)

[**3.** **The Experiment** 4](#_Toc139670092)

[3a. Step 1 - SQL: 4](#_Toc139670093)

[3b. Step 2- Hypothesis Testing: 6](#_Toc139670094)

[3c. Visualization: 8](#_Toc139670095)

[3d. Result 11](#_Toc139670096)

[**4.** **Recommendations** 12](#_Toc139670097)

# **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 H₀*:* **The conversion rate is not dependent on showing the banner at the top of the website.**
* Alternate Hypothesis H₁*:* **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 (α): 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.

# **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.

# **The Experiment**

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

Used SQL SELECT query 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

Query 2) To get the number of users in Groups A and B and the number who converted ie, have spent after visiting the website:

SELECT "group", COUNT(DISTINCT(groups.uid)) as total\_group\_no, COUNT(DISTINCT(activity.uid)) as Converted\_no

FROM activity

RIGHT JOIN groups

on activity.uid = groups.uid

GROUP BY "group"

Query 3) To get the conversion rates of Groups A & B:

SELECT "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"

Query 4) To get the average spent for Groups A & B:

SELECT "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"

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"

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



*(Click on the excel icon to view the excel sheet)*

* Null Hypothesis H₀*:* The conversion rate is not dependent on showing the banner at the top of the website.
* Alternate Hypothesis H₁*:* 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 H₀ and accept H₁ due to sufficient statistical evidence in favour of H₁.
2. Fail to reject H₀ due to insufficient statistical evidence in favour of H₁.

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

Null Hypothesis, H₀ is p1 = p2

Alternative Hypothesis H₁, p1 ≠ p2

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

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

Rejection Region = α/2= 0.05/2 = **0.025**

Now when the Z-score 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 α = 0.025, the Z-score from the z-table can be found as: Z = +- 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 > α, we Accept the null hypothesis

And if p-value < α, 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 (α) | 0.05 |
| Confidence Interval | 95% |

After conducting the test, the following has been found:

|  |  |
| --- | --- |
| No. of users in A | 24343 |
| No. of users in B | 24600 |
| Conversion Rate A, p1 | 3.92% |
| Conversion Rate B, p2 | 4.63% |
| Z-Score | -3.867 |
| p-value | 0.000110415 |

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

## **3c. Visualization:**

3c.1. Comparing the Conversion Rate & Average Spent in Groups A & B:

A screenshot of a graph

Description automatically generated

3c.2. The Spent distribution of Groups A & B:

A screenshot of a graph

Description automatically generated

3c.3. Comparison between test Groups A & B as per Device the uses use:

A screenshot of a graph

Description automatically generated

3c.4. Comparison between test Groups A & B as per their Gender:

A screenshot of a graph

Description automatically generated

3c.5 Comparison of test Groups A & B as per their country:

A graph of different colored bars

Description automatically generated

Links to visuals:

1. 3c.1. <https://public.tableau.com/views/Project1Globox/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link>
2. 3c.2. <https://public.tableau.com/views/Project1Globox/Dashboard2?:language=en-US&:display_count=n&:origin=viz_share_link>
3. 3c.3. <https://public.tableau.com/views/Project1Globox/Dashboard3?:language=en-US&:display_count=n&:origin=viz_share_link>
4. 3c.4. <https://public.tableau.com/views/Project1Globox/TestGroupsABandGender?:language=en-US&:display_count=n&:origin=viz_share_link>
5. 3c.5. <https://public.tableau.com/views/Project1Globox/TestGroupsABandCountry?:language=en-US&:display_count=n&:origin=viz_share_link>

## **3d. Result**

1. After conducting the Hypothesis Test its found that Since p-value < 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.

1. From visual 3c.1, we find that although the conversion rate of Treatment Group > Control Group, the average spent of both groups are almost same which is further elaborated in the visual 3c.2.
2. From visual 3c.3, we see that the conversion rate of users using iOS is almost double with respect to users using Android or Unspecified users. However, users who have not specified their device type ie, Unspecified have higher average amount spent.
3. 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
4. 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 just touches the average bar.

# **Recommendations**

* Launch the banner at top of the website for the next quarter to observer 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.