Globox A/B Testing & Analysis Report

15 August 2023

Contents:

I. Extracting the Table Using SQL - Process Explanation Page 2-4

II. A/B Testing & Recommendation Report Page 5-7

III. Link to the Visualization in Tableau

By,

Deepthi Binu

dipkrishna@gmail.com

Extracting the Table Using SQL - Process Explanation

1. Objective:

The primary goal of this task was to extract data from the activity, groups and users table within the given database. This table extraction was performed to gather valuable data for A/B testing, analysis and decision-making purposes.

2. SQL Query:

The SQL query used for extracting the table data was as follows:

```
SELECT g.uid, g.group, g.join_dt, COALESCE(g.device, 'no info') as device,
COALESCE(u.country, 'no info') as country,
COALESCE(u.gender, 'no info') as gender, COALESCE(SUM(a.spent),0) as total_spent,
CASE WHEN COALESCE(SUM(a.spent),0) > 0 THEN 1 ELSE 0 END as converted
FROM groups as g
LEFT JOIN activity as a
ON g.uid = a.uid
LEFT JOIN users as u
ON g.uid = u.id
GROUP BY 1, 2, 3, 4, 5,6;
```

3. Process Steps:

Here are the steps I followed to successfully extract the table data:

- Accessing the Database: I connected to the relevant database using the appropriate credentials and tools, such as Beekeeper Studio , a cross-platform SQL editor and database manager available for Linux, Mac, and Windows.

- Writing the SQL Query: I crafted the SQL query mentioned above, ensuring that it fetches all the necessary columns and rows for our analysis.

- Executing the Query: After verifying the query for accuracy, I executed it within the SQL environment.The queries used during the process are :

```
// Can a user show up more than once in the activity table?
SELECT COUNT(uid) - COUNT(DISTINCT(uid)) AS repeated_uid FROM activity;

// the start and end dates of the experiment
SELECT MIN(dt) AS start_date , MAX(dt) AS end_date FROM activity;

// total users were in the experiment
SELECT COUNT(DISTINCT(id)) as total_users FROM users;

// users in the control and treatment groups
SELECT g.group, COUNT(DISTINCT(uid)) FROM groups as g GROUP BY 1;
```

```
// conversion rate of all users
WITH S1 AS(SELECT (SELECT COUNT(DISTINCT(uid))
                                FROM activity a
                               WHERE a.spent>0)AS active_users,
        COUNT(DISTINCT (id)) AS total_users
                                FROM users)
SELECT ROUND((CAST(active_users AS NUMERIC(10,2))/total_users)*100,2) AS conversion_rate
FROM S1;

// user conversion rate for the control and treatment groups
WITH S1 AS(SELECT groups.group, COUNT(DISTINCT(a.uid)) AS group_users
         FROM groups
         JOIN activity a
         ON groups.uid=a.uid
        GROUP BY 1),
     S2 AS(SELECT g.group, COUNT(DISTINCT(uid)) AS total_users
                                                FROM groups as g
                                                GROUP BY 1)

SELECT S1.group,ROUND((CAST(S1.group_users AS NUMERIC(10,2))/S2.total_users)*100,2) AS
conversion_rate
FROM S1
JOIN S2
ON S1.group=S2.group;

// average amount spent per user for the control and treatment groups, including users who did
not convert

WITH S1 AS
    (SELECT g.group, SUM(a.spent) AS amount_spent, COUNT(DISTINCT(g.uid)) AS user_count
                        FROM groups as g
     LEFT JOIN activity a
     ON a.uid=g.uid
                        GROUP BY 1)

SELECT S1.group, ROUND(S1.amount_spent/S1.user_count,3) AS conversion_rate
FROM S1;
```

- Exporting Results: The results of the query were exported in csv format for further analysis, which was done by using Excel and google sheets and Tableau.

4. Quality Control:

To ensure the integrity of the extracted data, I performed the following quality control measures:

- ✓ Data Validation:

  Validate data types and lengths to ensure they match the expected format.
  Use regular expressions to verify that text data follows a specified pattern.

- ✓ Data Cleansing:
  - Remove or correct any invalid or irrelevant data.
  - Convert data to a consistent case (e.g., uppercase or lowercase) for uniformity.
  - Handle missing values using appropriate methods like default values or data imputation.
- ✓ Cross-Reference Checks:
  - Verify that references to other tables (foreign keys) actually exist in those tables.
  - Check that any linked data is consistent and accurate.
- ✓ Aggregation Checks:
  - Validate aggregated data against known values or calculations to ensure correctness.
  - Compare aggregated data against source data to ensure data wasn't lost during transformation.
- ✓ Documentation:
  - Document your SQL scripts, including the purpose of each script, expected inputs, outputs, and any assumptions made.

5. Conclusion:

The data extraction procedure has been accomplished, and downloaded the essential data required for the A/B testing and analysis.

A/B Testing & Recommendation Report

Executive Summary:

In this report, I present the results of an A/B test conducted to evaluate the impact of implementing a Banner with the key products in the food and drink category, at the top of the website of GloBox, .

The test was conducted between 25th Jan 2023 and 6th Feb 2023, with two groups: Group A (control) and Group B (treatment). The Group A does not see the banner, and the Group B sees it.

The purpose of the test was to provide a recommendation to the stakeholders about whether GloBox should launch the experience of the banner on their homepage to all the users.

**Hypothesis test1: **

- Null Hypothesis (H0): There is No requirement to improve the existing homepage. Pa-Pb=0
- Alternative Hypothesis (Ha): There is requirement to improve the existing homepage with food and drinks bar. Pa-Pb < > 0

Test Design:

- Total Users: 48943
- Group A (Control): 24343
- Group B (Treatment): 24600
- Metric being evaluated: Conversion rate

Results:	
Conversion Rate - Group A:	3.92%
Conversion Rate - Group B:	4.63%

Statistical Analysis:	
Calculated p-value:	0.0001
Confidence Level:	95%
Confidence interval	0.0035 to 0.0107

Conclusion:

Based on the A/B test results, we can draw the following conclusions:

- Since P-value(0.0001)<significance level (0.05), there is statistically significant evidence to reject the null hypothesis in favour of the alternative hypothesis. So it can be concluded that the findings favours the requirement to improve the existing homepage with food and drinks bar.

- The 95% confidence interval for the difference in conversion rate between the two groups is 0.0035 and 0.0107. Since the confidence interval does not includes zero, it suggests that there may be a statistically significant difference between the two group's conversion rates.

****Hypothesis test2 using T-distribution test: ****

Null Hypothesis (H0): There is no difference in the average amount spent per user between the two groups (μA=μB).

Alternative Hypothesis (H1): There is a difference in the average amount spent per user between the two groups (μA !=μB).

****Test Design:****

- Total Users: 48943
- Group A (Control): 24343
- Group B (Treatment): 24600
- Metric being evaluated: Average amount spent.

****Results:****	
Average Amount Spent - Group A:	$3.375
Average Amount Spent - Group B:	$3.391

****Statistical Analysis:****	
Calculated p-value:	0.946
Confidence Level:	95%
Confidence interval	(-)$0.471 to $0.439

****Conclusion:****

Based on the T-distribution test results, we can draw the following conclusions:

Since the p-value(0.946) is greater than the significant value, α(0.05), we would fail to reject the null hypothesis that there is no difference in the average amount spent per user between the two groups.

The 95% confidence interval for the difference in average amount spent per user between the treatment (Group B) and control (Group A) groups is from approximately -$0.471 to $0.439.

This means that we don't have strong statistical evidence to conclude that there is a significant difference in the average amount spent per user between the two groups. It's important to note that the confidence interval includes zero, which supports the idea that the difference could be negligible or non-existent.

****Recommendation:****

Based on the A/B test results and the T-test statistical analysis, I advise introducing a banner featuring key products from the food and drink category at the top of the GloBox website. This recommendation stems from the statistically significant enhancement observed in conversion rates. Nevertheless, I also propose ongoing monitoring and in-depth analysis to validate the sustained impact of this change.

****Limitations:****

It is observed that our dataset contains a significant number of missing values and areas where essential information is absent. Approximately a major % of the data has missing values or incomplete information. While we took steps to mitigate this issue through imputation methods, it's essential to acknowledge that these gaps could potentially introduce bias and affect the reliability of our analysis.

****Future Steps:****

- I believe that future research efforts should focus on improving data collection procedures to enhance the robustness of our findings. Despite these limitations, I've worked diligently to ensure the validity of the conclusions within the constraints of the data we have available."

****Acknowledgments:****

I would like to acknowledge the efforts of the Support team in guiding me at the designing and conducting of this A/B testing process.

The spreadsheet link:
https://docs.google.com/spreadsheets/d/1gH7MNivTpAm77E1xByxKSn19MJv6TjhPB_ulyKjOt60/edit?usp=sharing

Link to the visualization in Tableau :
https://public.tableau.com/app/profile/deepthi.binu/viz/GloboxABTestingAnalysis/GloboxABTesting

For inquiries or further information, please contact dipkrishna@gmail.com

---

Deepthi Binu
dipkrishna@gmail.com