

Globox Project

A/B Testing

1 Objectives

According to this Globox project, here are the objectives to achieve:

1. To demonstrate which version of the website users convert more.
2. To analyze other insights from the data
3. To visualize and communicate all insights to both technical and non-technical audiences.

2 Data

The database has three tables: users, groups, and activity table. Figure 1 shows how each table relates to each other.

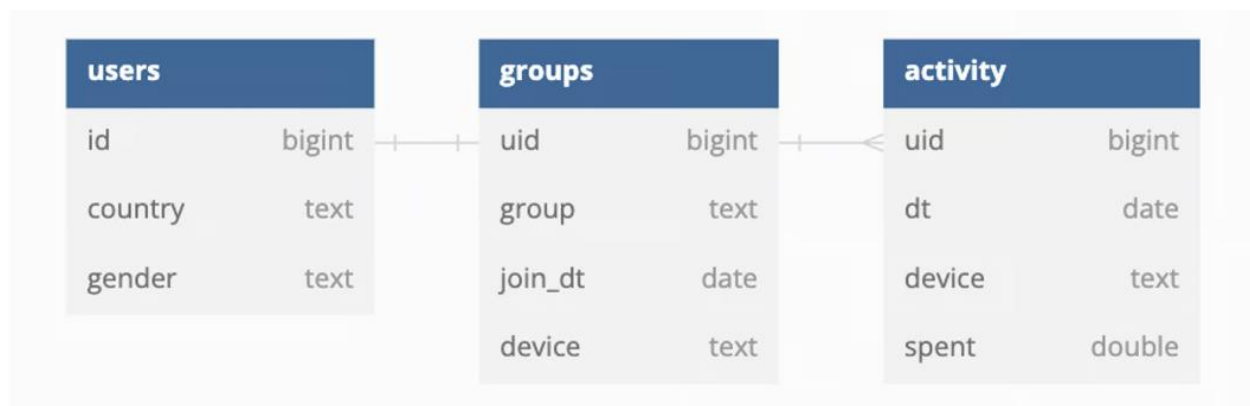


Figure 1 Data structure [1]

3 Methodology

3.1 Access the database using BeeKeeper Studio

To acquire the data, BeeKeeper was installed to access the database. After that, SQL language was implemented to query for required data.

Here is the list of SQL commands that are used to filter dataset:

- SELECT: using for showing the data from database column
- CASE: using for filtering groups into control and treatment groups
- SUM: using for add values
- COALESCE: using for replacing null data with a value (0)
- LEFT JOIN: using for joining tables
- GROUP BY: using for grouping data based on some columns

All SQL code is shown in Appendix A

3.2 Data analysis

3.2.1 Inferential statistic

1. Calculate a two-sample t-test with unpooled variance to hypothesis test the conversion rate between the control group and the treatment group

Hypotheses: $H_0: \mu_1 - \mu_2 = \mu_0$

$H_1: \mu_1 - \mu_2 \neq \mu_0$

Test statistic: $T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

p-value: $2 * P(t_{df} > |T|)$

$df = \min(n_1 - 1, n_2 - 1)$

2. Calculate a two-sample z-test with pooled proportion to hypothesis test the average spending between the control group and the treatment group

Hypotheses: $H_0: p_1 - p_2 = p_0$

$H_1: p_1 - p_2 \neq p_0$

Test statistic: $T = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$

p-value: $2 * P(Z > |T|)$

3. Plot the Confidence level graph to show the results from 1. And 2.

3.2.2 Novelty Effect

1. Plot the conversion and date of the dataset
2. Analyze the graph from 1. to see whether the user behavior changes over time or not

3.2.3 Power Analysis

1. Use Statig Calculator to calculate the sample size for conversion by two-tailed testing with 10% MDE and 4.0% Baseline Conversion rate [2]
2. Use Statulator Calculator to calculate the sample size for means by using Expected Difference between Means [3]

3.3 Data visualization using Tableau

To visualize data for other insights, using Tableau will be helpful to present them on one dashboard.

1. Visualize the gender and conversion analysis using bar charts between two test groups.
2. Visualize the device and conversion analysis using bar charts between two test groups.
3. Visualize the country and conversion analysis using bubbles between two test groups.
4. Visualize the time and conversion analysis using line charts between two test groups.

4 Results

4.1 Database Query

After querying that database using SQL code from Figure 9, the result as shown in Figure 2.

uid	cluste	spent	count	gend	devic
1000000	Treatment		0 CAN	M	
1000001	Control		0 BRA	M	
1000002	Control		0 FRA	M	
1000003	Treatment		0 BRA	M	
1000004	Control		0 DEU	F	
1000005	Treatment		0 GBR	F	
1000006	Treatment		0 ESP	M	
1000007	Control		0 BRA	F	
1000008	Control		0 BRA	F	
1000009	Control		0 USA		
1000010	Treatment		0 BRA	M	
1000012	Treatment		0 USA	M	
1000013	Control		0 GBR	F	
1000014	Control		0 USA	M	
1000015	Control		0 AUS	F	
1000016	Control		0 DEU	F	
1000017	Treatment		0 MEX	O	
1000018	Treatment		0 USA	M	
1000019	Treatment		0 MEX	F	
1000020	Control		0 AUS	M	
1000021	Treatment		0 GBR	M	
1000022	Treatment		0 DEU	F	
1000023	Treatment		0 BRA	M	
1000024	Control		0 CAN	M	
1000025	Control		0 USA	M	
1000026	Treatment		0 DEU	F	
1000027	Treatment		0 BRA		
1000028	Control		0 BRA	F	
1000029	Control		0 MEX	M	
1000030	Control		0 FRA	F	
1000031	Treatment		0 GBR	F	
1000032	Control		0 GBR		
1000033	Control		0 FRA		
1000034	Control		0 USA	M	
1000035	Treatment		0 USA	F	
1000036	Treatment		0 USA	F	
1000037	Treatment		0 TUR	M	
1000038	Control		0 BRA	M	
1000039	Treatment	36.65	GBR	F	A
1000040	Control		0 USA	F	
1000041	Control		0 CAN	M	
1000042	Control		0 BRA	F	
1000043	Treatment		0	F	
1000044	Treatment		0 USA		
1000045	Treatment	51.58	USA	F	I
1000046	Control		0 BRA	M	

Figure 2 The data after the query

4.2 Inferential Statistic Result

4.2.1 Two-sample t-test with unpooled variance result

Table 1 Two-sample t-test with unpooled variance result for average spending analysis

Parameters	Notation	Values
Sample size 1 (Control)	n_1	24343
Sample size 2 (Treatment)	n_2	24600
Sample mean (control)	\bar{x}_1	3.375
Sample mean (treatment)	\bar{x}_2	3.391
Sample std dev (control)	s_1	25.936
Sample std dev (treatment)	s_2	25.414
Standard Error	SE	0.232
Test statistic	T	0.070
Degree of freedom	df	24342
P-value	P-value	0.944
Confidence Level	α	0.05

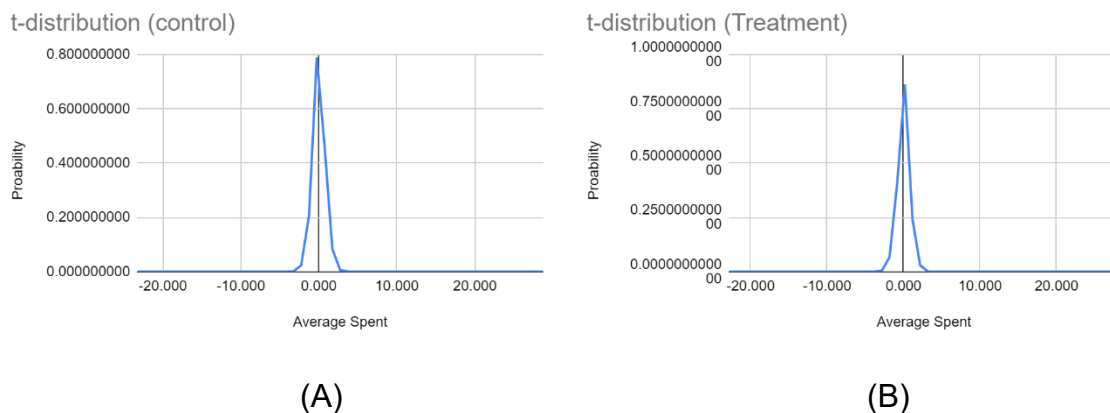


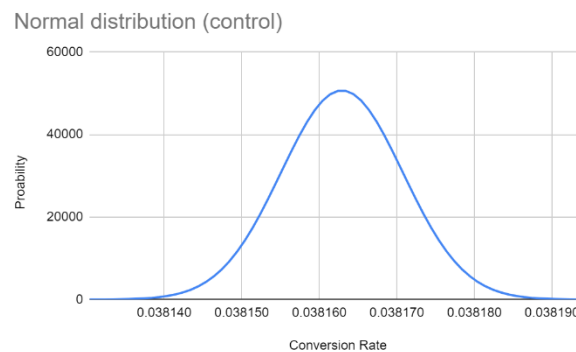
Figure 3 (A) The t-distribution of average spending on the control group
(B) The t-distribution of average spending on the treatment group

Conclusion: Since the p-value exceeds the significant level (α), the null hypothesis (H_0) is accepted, which means there is no significantly difference on average spending between the two test groups. The estimated average spending of most users in the control group is between \$3.049 and \$3.700, while the treatment group is between \$3.073 and \$3.708.

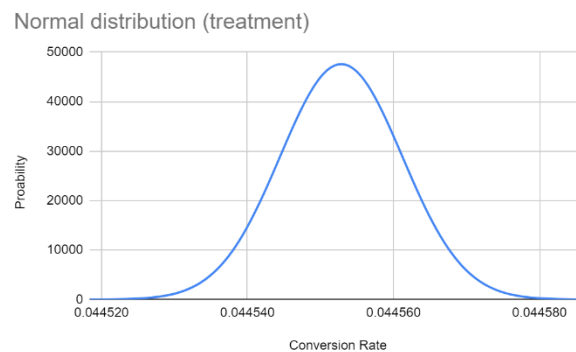
4.2.2 Two-sample z-test with pooled proportion result

Table 2 Two-sample t-test with pooled proportion result for conversion rate analysis

Parameters	Notation	Values
Sample size 1 (Control)	n_1	24343
Sample size 2 (Treatment)	n_2	24600
Proportion 1 (Control)	\hat{P}_1	0.038
Proportion 2 (Treatment)	\hat{P}_2	0.045
Proportion mean	\hat{P}	0.041
Standard Error	SE	0.002
Test statistic	T	-3.551
P-value	P-value	0.0004
Confidence Level	α	0.05



(A)



(B)

Figure 4 (A) The normal distribution of conversion rate on the control group
 (B) The normal distribution of conversion rate on the treatment group

Conclusion: Since the p-value is less than the significant level (α), the null hypothesis (H_0) is failed to accept, which means there is a significant difference on conversion rate between control and treatment. The estimated conversion rate of most users in the control group is between 3.815% and 3.818%, while the treatment group is between 4.454% and 4.457%.

4.3 Novelty effect result

For novelty effect analysis, the result is shown in Figure 5. Focusing on the treatment group (Orange line) shows that the conversion has fluctuated over time, which means this dataset has a novelty effect. Therefore, the user behavior might change because of the new feature addition rather than its functionality.

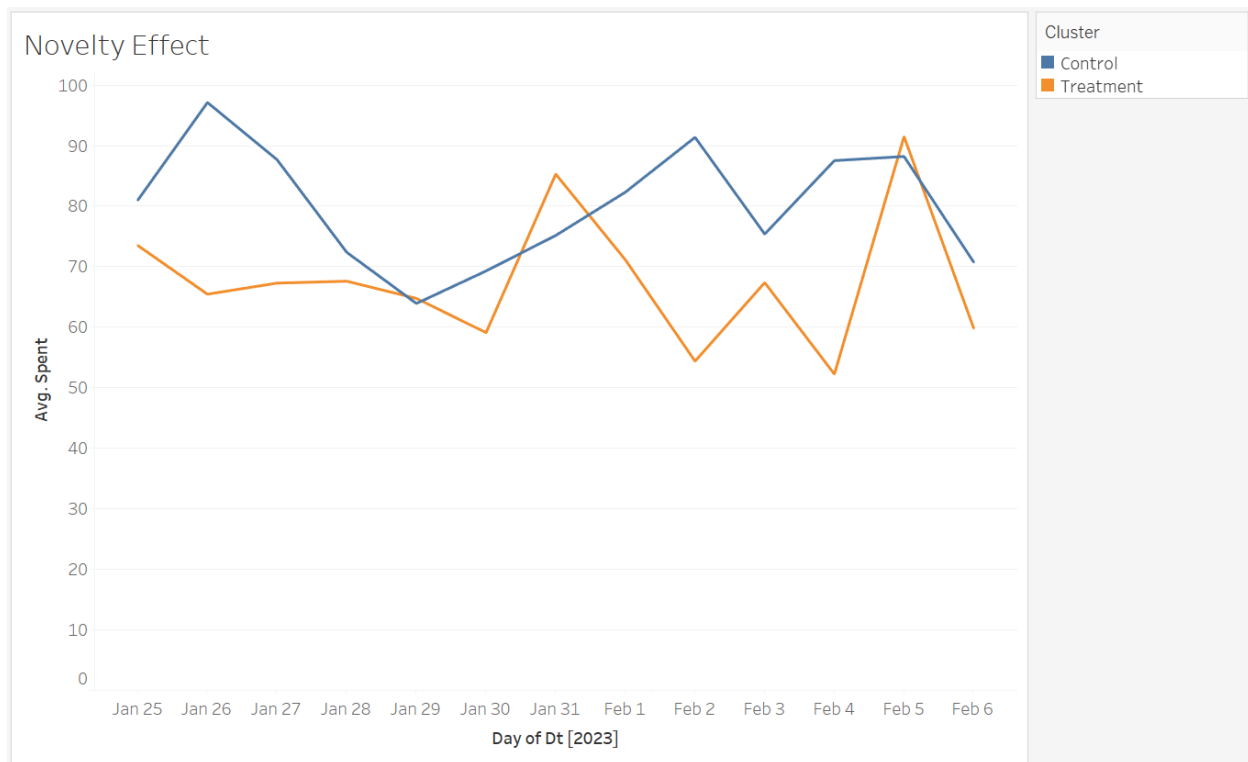


Figure 5 The novelty effect analysis result

4.4 Power analysis result

4.4.1 Statsig sample size calculator results for Conversions

According to the conversion, Statsig suggests the sample size to be 37,700 to achieve two-tailed testing with a power of 80% and a significance level of 5%.

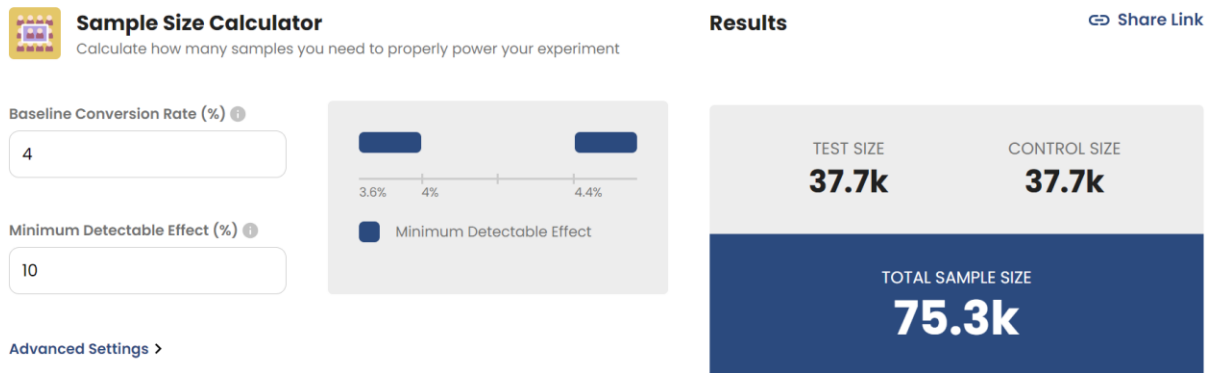


Figure 6 The result from Statsig

4.4.2 Statulator sample size calculator result for Means

According to the difference mean between control and treatment groups, the Statulator suggests the sample size to be 2,090 to achieve a power of 80% and a significance level of 5%.

Sample Sizes for Comparing Two Means

Influence of Changing Input values on Sample Size Estimates

Figure created by Statulator beta: www.statulator.com/SampleSize/ss2M.html

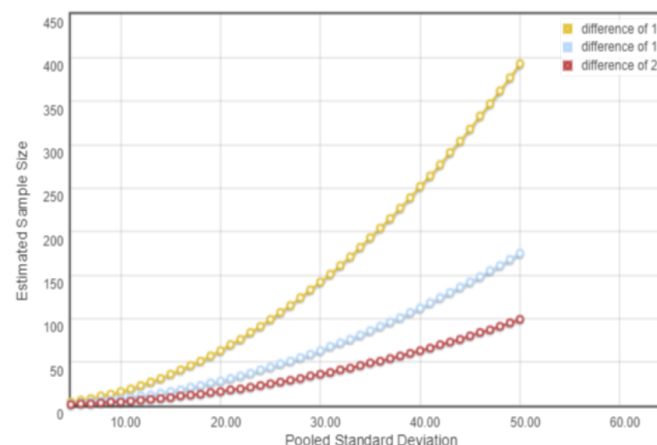


Figure 7 The result from Statulator

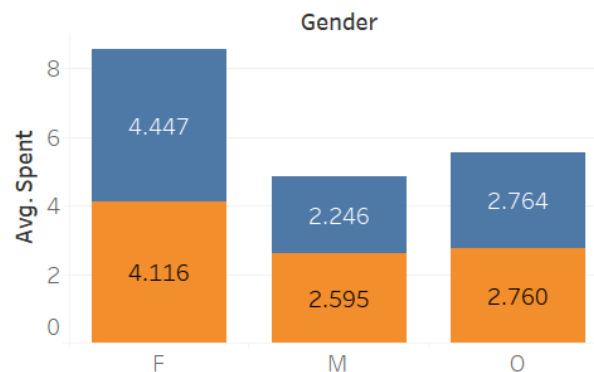
4.5 Other Insights Visualization

Speaking about data visualization using Tableau, here are vital takeaways further the A/B testing analysis:

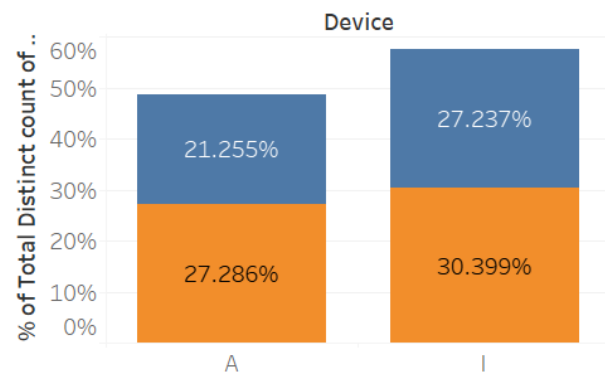
- Females spent more than other genders; the least was males at \$4.447, \$2.764, and \$2.246 on average for control group and at \$4.116, \$2.760, and \$2.595 on average for treatment group respectively.
- iOS users have a higher conversion rate than Android users at 57.629% and 48.541%, respectively.
- The nationality that has the most conversion is the USA.
- The days that users spent the most conversions were on Wednesday.

Globox Visualization

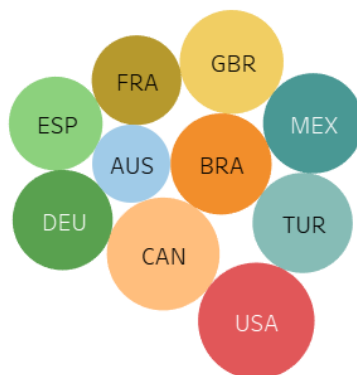
Gender Analysis



Device Analysis



Country Analysis



Time Analysis

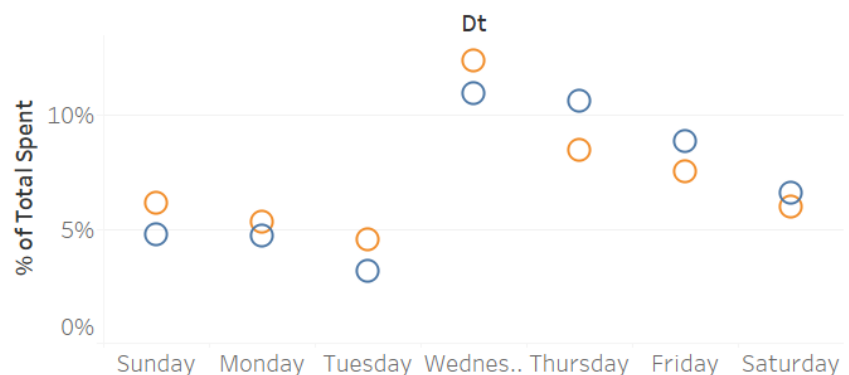


Figure 8 The result for other insights presentation using Tableau

5 Conclusion and Discussion

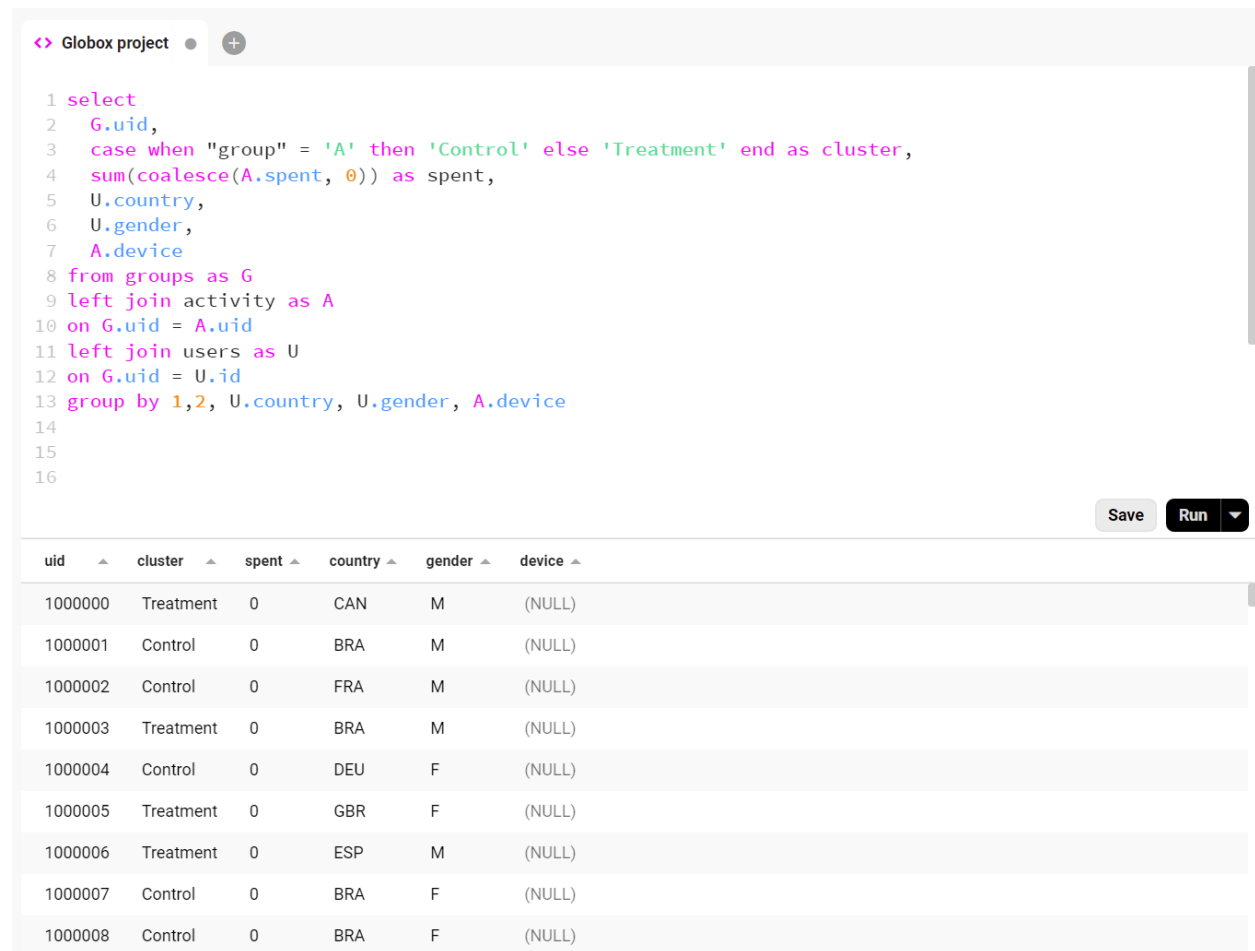
A/B testing has been implemented to find out how the new feature has affected users' conversion. After experimentation, data has been queried for data analysis.

According to all analyses shown in this report, there is a significant difference in conversion rate, while no significant difference on average spending between the two test groups. For the conversion rate analysis, this claim is unreliable by insufficient data supported from power analysis. Even though there is enough data for average spending analysis, the novelty effect should be considered since it shows the fluctuation over time. Therefore, the suggestion is to continue the experiment to gather more data and re-analyze after the sample size meets the requirement.

6 References

- [1] "Project Overview," *Project Overview*. <https://cms.master.school/project-overview> (accessed Aug. 25, 2023).
- [2] "A/B Test Sample Size Calculator - Statsig." <https://www.statsig.com/calculator> (accessed Aug. 25, 2023).
- [3] "Sample Size Calculator for Comparing Two Independent Means." <https://statulator.com/SampleSize/ss2M.html#> (accessed Aug. 25, 2023).

Appendix A



The screenshot shows a SQL query editor interface for a project named "Globox project". The query is as follows:

```
1 select
2   G.uid,
3   case when "group" = 'A' then 'Control' else 'Treatment' end as cluster,
4   sum(coalesce(A.spent, 0)) as spent,
5   U.country,
6   U.gender,
7   A.device
8 from groups as G
9 left join activity as A
10 on G.uid = A.uid
11 left join users as U
12 on G.uid = U.id
13 group by 1,2, U.country, U.gender, A.device
14
15
16
```

Below the query editor, there are "Save" and "Run" buttons. The "Run" button is active, and the results of the query are displayed in a table below.

uid	cluster	spent	country	gender	device
1000000	Treatment	0	CAN	M	(NULL)
1000001	Control	0	BRA	M	(NULL)
1000002	Control	0	FRA	M	(NULL)
1000003	Treatment	0	BRA	M	(NULL)
1000004	Control	0	DEU	F	(NULL)
1000005	Treatment	0	GBR	F	(NULL)
1000006	Treatment	0	ESP	M	(NULL)
1000007	Control	0	BRA	F	(NULL)
1000008	Control	0	BRA	F	(NULL)

Figure 9 The SQL query that used in this Globox project