GLOBOX A/B TESTING

MASTERSCHOOL
MD TAUHIDUL HAQUE NAIME

GLOBOX A/B TESTING

Table of Contents

Executive summary

Globox, a well-established online marketplace renowned for its diverse and top-quality product selection sourced globally, is on the cusp of enhancing its website through the implementation of an A/B test experiment. In my capacity as a data analyst at Globox, I'm tasked with dissecting the results of this A/B test, which holds the primary objective of boosting the company's sales performance. This analytical journey seeks to distil valuable insights from data and ultimately offer data-driven recommendations based on these findings.

This analytical process unfolds in three crucial phases. Initially, we dig into the data trenches, employing SQL to extract user-level data, which is then brought to life through visualizations in Tableau. This initial step lays the foundation for understanding user behaviour patterns. Subsequently, we transition into the realm of statistical analysis, where we meticulously scrutinize the experiment results using spreadsheets. This stage is instrumental in revealing patterns and trends that might otherwise remain concealed. Finally, we wrap up the process by distilling our discoveries into a comprehensive report, encapsulating the experiment's outcomes and their wider implications.

The insights stemming from our analysis are profound. Notably, the introduction of a banner featuring food and drinks to mobile users resulted in the treatment group exhibiting a significantly higher rate of purchases compared to the control group. This striking observation underscores the banner's influential impact on customer behaviour, translating into increased conversions and sales.

In essence, our A/B test results paint a compelling picture. Conversion rates in the treatment group outshone those in the control group, and the average spending of users within the treatment group significantly surpassed that of the control group. These findings unequivocally validate the effectiveness of our A/B test in shaping customer behaviour and driving sales. Based on this data-driven evidence, launching the product is not just advisable but highly recommended to capitalize on the observed positive impact on the company's performance.

Context

Motivation

The primary motivation behind this experiment is to assess the impact of displaying a banner featuring key products to Globox's customers. The rationale behind implementing this experiment is for Globox to gain valuable insights into whether featuring such a banner on the landing page will result in increased sales for the company.

In the context of this experiment, two fundamental hypotheses come into play. The null hypothesis postulates that the banner showcasing key products will not bring about a significant change in customer purchases at Globox. Conversely, the alternative hypothesis suggests that the banner will indeed have a noteworthy and positive effect on customer purchasing behaviour. The aim is to gather data and conduct statistical analysis to determine whether there is sufficient evidence to reject the null hypothesis in favour of the alternative hypothesis, thus providing a robust conclusion.

The execution of an A/B test is pivotal in this process, as it serves as the conduit for data-driven insights that will inform the decision on whether to proceed with the product launch. This data-driven approach not only sheds light on user experience and conversion rates but also emphasizes the importance of recognizing that a boost in conversion rates doesn't necessarily equate to an immediate revenue increase for the company. It may, however, be indicative of heightened customer satisfaction and loyalty. Over the long term, this loyalty can be a driving force behind sustained revenue growth for the company. Therefore, this experiment goes beyond mere short-term sales impact and delves into the realm of cultivating lasting customer relationships and overall business growth.

Parameters of the test

In this experiment, two distinct test groups have been established: Group A, also referred to as the Control Group, and Group B, designated as the Treatment Group. To ensure the impartiality of the results, the assignment of individuals to these groups has been executed randomly. Group B, the Treatment Group, will experience the mobile website with a prominently displayed banner showcasing products, while Group A, the Control Group, will encounter the existing mobile website without this banner.

The experiment's timeline spans a duration of 13 days, commencing on January 25, 2023, and concluding on February 6, 2023. This carefully chosen timeframe allows for a comprehensive assessment of customer behaviour over a reasonable period.

The dataset under scrutiny encompasses a total of 48,943 customers, meticulously divided into two defined groups. Group A, the control group, includes 24,343 customers, while Group B, the treatment group, encompasses 24,600 customers. The careful division of the dataset ensures that the sample size for each group is sizable enough to draw meaningful conclusions from the experiment.

Two key metrics are at the heart of this experiment's evaluation. The first metric is the conversion rate of customers, which serves as a critical indicator of whether customers are making purchases because of the experimental banner. The second metric, the average purchase amount, is pivotal in assessing the overall revenue implications of the experiment. Together, these metrics provide a robust foundation for evaluating the experiment's impact on both customer behaviour and the company's financial performance.

Overview of the dataset

The dataset comprises three tables: the users table, groups table, and activity table. The users table contains demographic information about the users, while the groups table focuses on assigning users to two different groups for the A/B test. The activity table records user purchase activity.

Data types

Column	Data type
id	bigint
uid	bigint
country	text
gender	text
group	text
Join_dt	date
device	text
dt	date
spent	double

Table 1

The table reveals the presence of four distinct data types, namely bigint, text, date, and double.

- bigint is employed for handling larger-scale integer data or whole numbers.

- text is utilized for storing textual information, such as words or sentences.
- date data type serves the purpose of storing date-related information.
- double data type is designated for storing floating-point numbers or decimal values.

Summary of the dataset

	Control Group (Group A)	Treatment Group (Group B)	Total
Distinct users	24343	24600	48943
Distinct converted users	955	1139	2094
Conversion rate	3.92%	4.63%	4.28%
Average spent	$3.37	$3.39	$3.38

Table 2

Results

Understanding the GloBox Database

I extracted data from the dataset for statistical analysis using Beekeeper Studio in conjunction with PostgreSQL.

Can a user show up more than once in the activity table?

Yes a user can show up more than once as they can make purchases multiple times.

Type of join should be used to join the user's table to the activity table

Left join should be used to join the users table to the activity table.

SQL function can be used to fill in NULL values

COALESCE()

Start and end date of the experiment

I employed an SQL query to extract the start and end dates of the experiment, focusing on the "groups" table. This query relies on aggregation functions to compute the earliest (minimum) and latest (maximum) join dates recorded in the "join_dt" column. Specifically, the MIN(join_dt) function identifies the initial date when the experiment commenced, while MAX(join_dt) identifies the concluding date of the experiment. Upon execution, the query produces a result set containing these two date values, thus furnishing essential insights into the temporal aspects of the experiment's duration.

Start date	25th January 2023
End Date	6th February 2023

[1]

Total distinct users

The SQL query is aimed at ascertaining the total count of unique users within the "groups" table. To accomplish this, it leverages the COUNT(DISTINCT(uid)) function, which quantifies the distinct user IDs (uid) and assigns the outcome the name "total_users." When executed, the query provides a single numerical value, serving as a valuable indicator of the comprehensive diversity and scale of user involvement within the "groups" table. This insight contributes to our understanding of the breadth and variety of user participation in this specific dataset.

Total unique users	48943

[2]

Users in control group and treatment group

The SQL query I ran was designed to retrieve the count of unique users in both the control group and treatment group from the "groups" table. Its objective was to break down the number of distinct user IDs (uid) in each of these groups. This was achieved by utilizing the COUNT(DISTINCT(uid)) function to count unique user IDs within each group, with results grouped by the "group" column.

When this query was executed, it generated a result set consisting of two columns: "group," which represents the different groups (control and treatment), and "total_users," which displays the count of unique users for each respective group. This information serves as a valuable tool for the assessment and comparison of user participation across the various groups listed in the dataset.

Control Group (Group A)	24343
Treatment Group (Group B)	24600

[3]

Conversion rate for all users

In the SQL query I executed to extract data on user conversion, non-conversion, conversion rates, and non-conversion rates among all users, the process begins with the creation of a Common Table Expression (CTE) named "temp_tab." Within this CTE, distinct user IDs are counted in the "users" table (aliased as "us") and the "activity" table (aliased as "ac") through a full join operation between the two tables. This calculation covers both converted and non-converted users.

The primary query then draws its data from the "temp_tab" CTE. It selects the counts for "converted" users and "Not_converted" users, while also computing the conversion rate (the percentage of converted users among the total user count) and the non-conversion rate (the percentage of users who have not converted among the total user count). The conversion and non-conversion rates are calculated as percentages by dividing the "converted" and "Not_converted" counts by the "total_users" count, and then multiplying the result by 100.

The output dataset provides valuable insights into user conversion performance by revealing the number of conversions, non-conversions, and the corresponding rates in percentage form, facilitating a comprehensive understanding of user engagement with the platform.

Conversion	count	Conversion rate
Converted	2094	4.28%
Not converted	46849	95.72%

[4]

user conversion rate for the control and treatment groups

In the SQL query, I executed to obtain the count of user conversions, non-conversions, and their respective rates for both the control and treatment groups, the process begins with the creation of a Common Table Expression (CTE) named "temp_tab." Within this CTE, the total number of users, converted users, and non-converted users are counted. This is achieved through full joins between the "users" and "activity" tables and between the "users" and "groups" tables. The results are grouped based on the group names extracted from the "groups" table.

The main query then retrieves data from the "temp_tab" CTE. It selects the "group_name," "converted," "Not_converted," and proceeds to calculate the conversion rates and non-conversion rates as percentages. The outcome is a result set that displays group names, the counts of converted and non-converted users, as well as their respective conversion and non-conversion rates. This dataset offers valuable insights into user engagement and conversion performance within distinct groups, shedding light on how users from the control and treatment groups are interacting with the platform.

Group Name	Conversion	Count	Conversion rate
Control Group (Group A)	Converted	955	3.92%
	Not converted	23388	96.08%
Treatment Group (Group B)	Converted	1139	4.63%
	Not converted	23461	95.35%

[5]

Average amount spent per user

The SQL query I executed to determine the average spending of all users calculates the average expenditure per user across all individuals present in the "users" table (aliased as "us"). These users are associated with activities recorded in the "activity" table (aliased as "ac") and are also members of groups as defined in the "groups" table (aliased as "gr"). To establish these associations, the query employs full joins that link users to their activities and group memberships based on user IDs.

The "spent_per_user" metric is derived by dividing the total amount spent in all activities (summed from the "spent" column in the "activity" table) by the count of distinct users. In essence, this computation offers insight into the average spending behaviour of all users, irrespective of their group affiliations, presenting a comprehensive view of overall user spending patterns.

Average spent for all users	$3.38

[6]

Average amount spent per user for the control and treatment groups

The SQL query I executed to obtain the average spending in the control group and treatment group commences with full joins between the "users," "activity," and "groups" tables, aligning users with their respective activities and group affiliations using user IDs as the key. Subsequently, it calculates the "spent_per_user" metric, which signifies the average spending per user within each group. This is achieved by dividing the total amount spent in each group (aggregated from the "spent" column in the "activity" table) by the count of unique users within that group. Finally, the query organizes the results by the "group" column, presenting the average spending per user in each specific group. This facilitates insights into user spending patterns across different groups, allowing for a comparative analysis of spending behaviour.

Control Group (Group A)	$3.37
Treatment Group (Group B)	$3.39

[7]

Extracting the Analysis Dataset

For data extraction to facilitate visualization, I exported the data in .CSV format from the dataset using PostgreSQL. The query executed for this purpose amalgamates information from the "users," "activity," and "groups" tables, offering a comprehensive snapshot of user attributes and their conversion status across various groups.

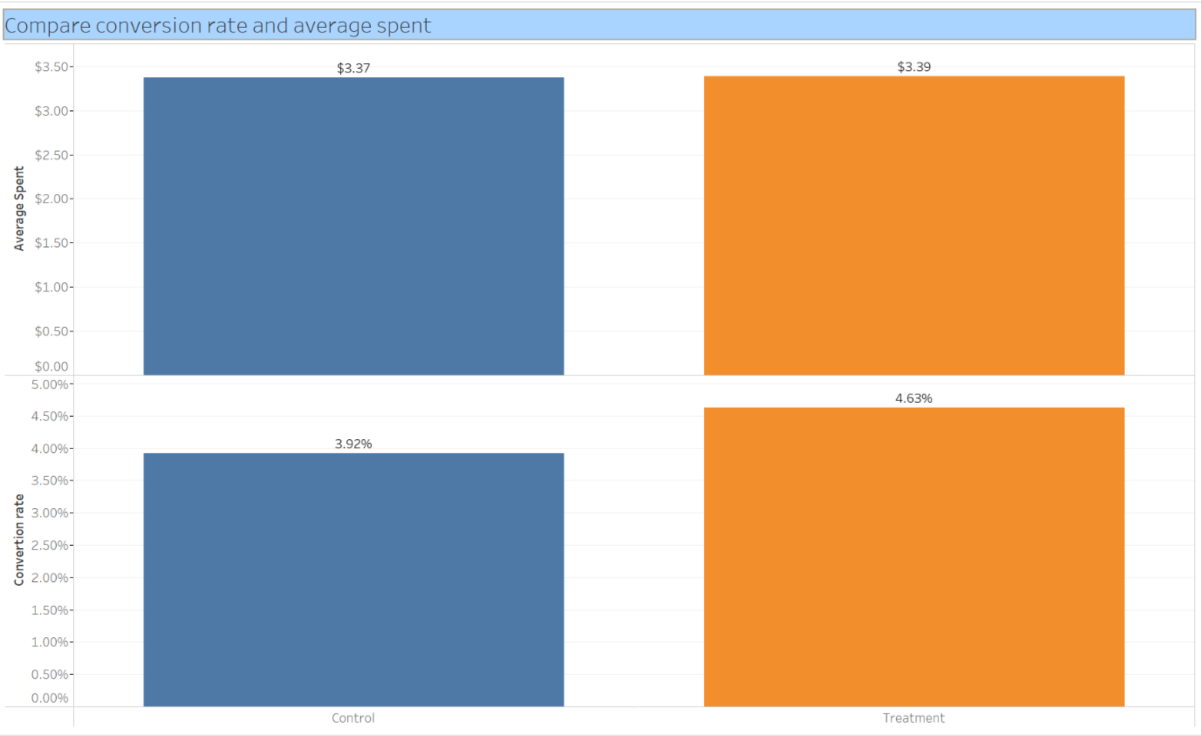
The query employs left joins to correlate user IDs with device details and group affiliations. It calculates the total spending per user, applying the COALESCE function to handle instances where there is no spending data. Additionally, users are categorized into "Converted" if they have incurred any spending, "Not converted" if they haven't spent at all, and "Error" to account for any unusual cases.

The results are organized by user ID, device, and group, presenting a granular breakdown of user information, including spending behaviour, group associations, and conversion status. This dataset serves as a valuable resource for generating visual representations that can provide deeper insights into user characteristics and their engagement across various groups.

Visualize the Results in Tableau

Visualization to compare the conversion rate and average amount spent between the test groups

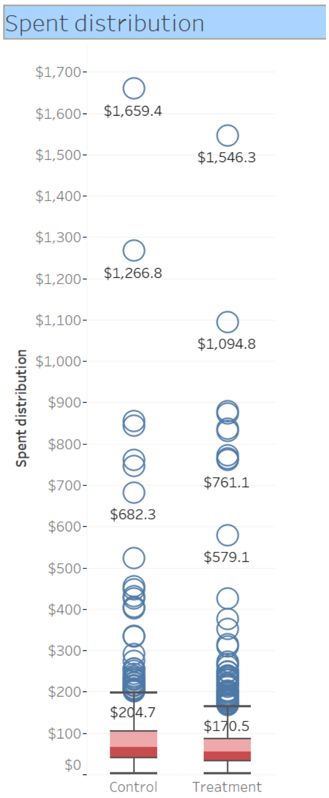
The chart reveals that the control group's average spending is $3.37, while the treatment group's is $3.39. The conversion rate for the control group stands at 3.92%, whereas the treatment group records 4.63%. These findings indicate that the treatment group exhibits higher average spending and a greater conversion rate compared to the control group, underscoring the effectiveness of displaying the banner as depicted in the chart.



[8]

The distribution of the amount spent per user for each group

In this chart depicting the distribution of spending for both groups, it's evident that the control group's spending distribution exhibits a broader spread compared to the treatment group. The highest recorded purchase amount in the control group is $1659.4, whereas in the treatment group, it reaches $1546.3. Notably, there are numerous purchases concentrated in the lower spending range for both groups.



[8]

visualizations to explore the relationship between the test metrics (conversion rate and average amount spent) and the user's device

[8]

This chart illustrates the correlation between test metrics (conversion rate and average spending) and user devices. Notably, the highest conversion rate is observed among Treatment group users who have not specified their device type, whether Android or iPhone.

Additionally, across all device usage groups, the Treatment group consistently exhibits a higher conversion rate than the Control group.

Conversely, when considering average spending, the chart indicates that the Control group's users tend to have a higher average spending than the Treatment group, regardless of their device usage.

visualizations to explore the relationship between the test metrics (conversion rate and average amount spent) and the user's gender

This chart illustrates the connection between test metrics (conversion rate and average spending) and user gender. Notably, male users in the treatment group exhibit the highest conversion rate, while females in the control group have the highest average spending. There is a substantial difference between male and female average spending, but the difference in conversion rates between the genders is less pronounced.

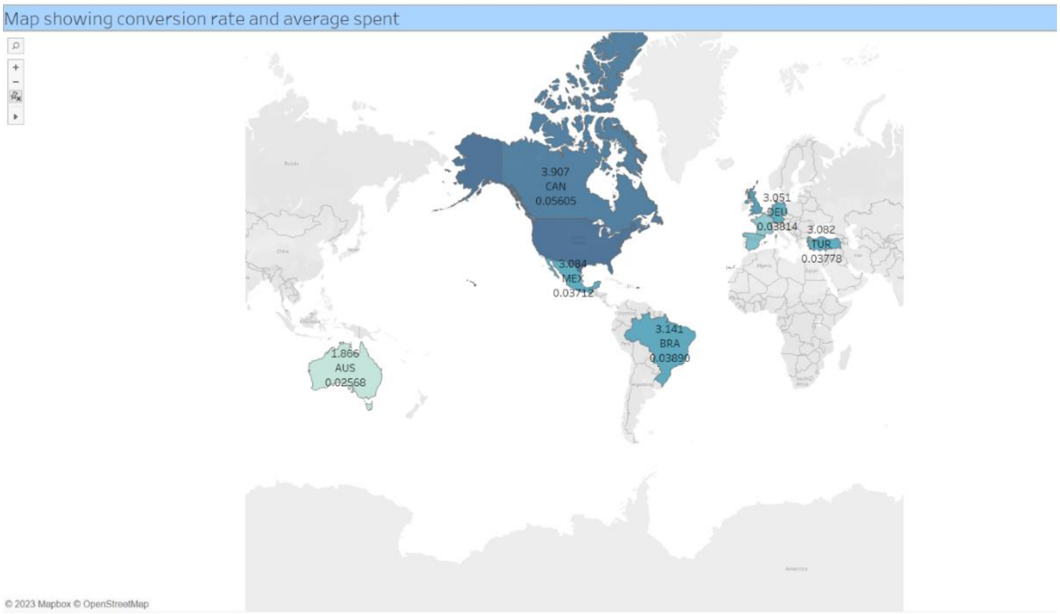
[8]

Visualizations to explore the relationship between the test metrics (conversion rate and average amount spent) and the user's country

Conversion rate and average spend demonstrate variations from one country to another. Canada boasts a high conversion rate of 5.605%, while Australia lags at 2.568%. In terms of average spend, USA users lead the pack with $4.173, whereas Canada's users spend an average of $3.907.

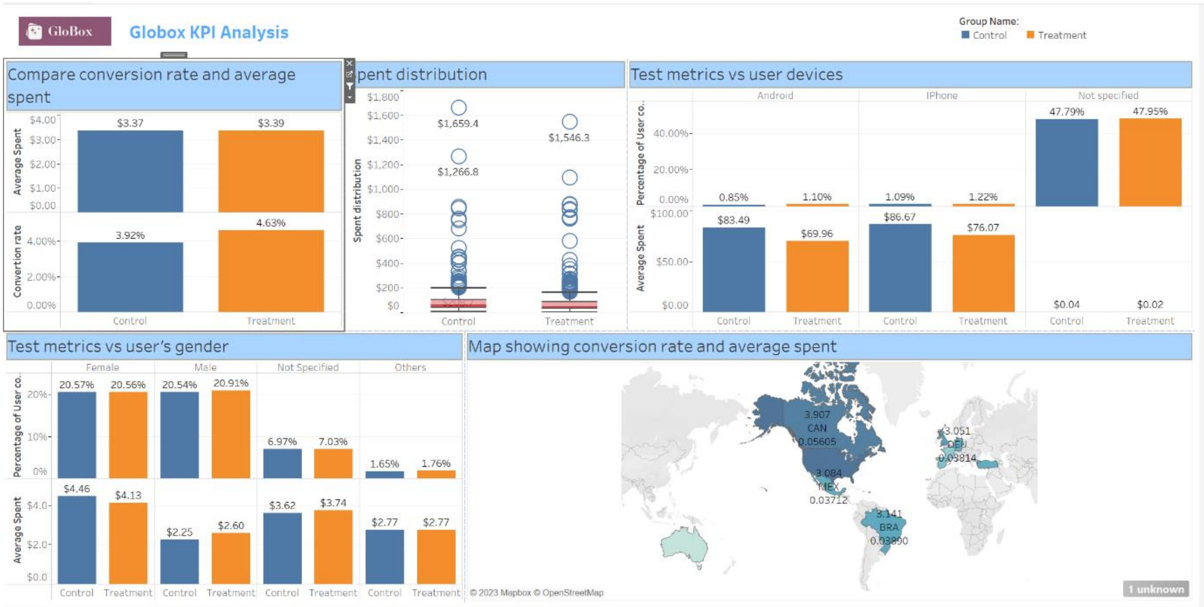
The highest conversion rate is found in Canada at 5.436%, while the lowest is in Australia at 2.568%. Similarly, the highest average spending is attributed to USA users at $4.173, while the lowest is reported in Australia at $1.866.

Meanwhile, Mexico, Brazil, Turkey, and Germany show nearly identical conversion rates and average spending figures.



[8]

Dashboard



[8]

From the dashboard, we can summarise that:

Female and male users exhibit nearly identical conversion rates across non-specified and other user groups. However, female users outpace males in terms of average spending across male, other, and non-specified user groups.

iOS users tend to spend more than users on other devices and enjoy a higher conversion rate compared to Android users. Most users have a non-specified device type but still manage to achieve a considerable conversion rate.

Comparing the treatment and control groups, the treatment group features more users in lower spending categories but fewer in higher spending brackets in contrast to the control group. Nonetheless, the treatment group outperforms the control group in both conversion rate and average spending, particularly attracting users in higher spending categories.

In summary, variations in conversion rates and spending behaviour are influenced by factors such as country, gender, and user devices.

Calculate A/B Test Statistics using Spreadsheets

Hypothesis test to see whether there is a difference in the conversion rate between the two groups

To test whether there's a difference in conversion rates between the two groups, a two-sample z-test for the difference in proportions was performed using the same dataset utilized for data visualization.

The Null Hypothesis was set as 'there is no difference in the conversion rate between the control and treatment groups,' while the Alternative Hypothesis assumed 'there is a difference in the conversion rate between the control and treatment groups.' The calculated z-score was -3.86429177, and the corresponding p-value was determined to be 0.000111412.

With a pre-defined significance level (α) of 0.05 or 5%, the p-value was compared. Since the p-value is smaller than the significance level, it was concluded that there is a significant difference in conversion rates between the control and treatment groups. In this context, the Null Hypothesis was rejected, indicating that there is sufficient evidence to reject it in favour of the Alternative Hypothesis. Or it can be concluded that there is enough evidence to reject the null hypothesis.

Null Hypothesis – H0	There is no difference in the conversion rate in control and treatment group.
Alternative Hypothesis – H1	There is difference in the conversion rate in control and treatment group.
H0 – Notation	$H0 : p1 = p2$
H1 - Notation	$H1 : p1 \neq p2$

	Control Group (Group A)	Treatment Group (Group B)
Converted	955	1139
Not Converted	23388	23461
Total	24343	24600
Conversion rate (Proportion)	0.03923099	0.046300813
Conversion rate (percentage)	3.92%	4.63%

Total converted	2094
Total not converted	46849
sample size	48943
Sample proportion	4.28%
Z-score	-3.86429177
p-value	0.000111412

[9]

95% confidence interval for the difference in the conversion rate between the treatment and control

To calculate the 95% confidence interval for the difference in the conversion rate between the treatment and control groups, several steps were taken:

- Standard Error: The standard error was computed, which involves assessing the variability and spread in the data.
- Critical Value: A critical value was determined based on the chosen confidence level (in this case, 1.959963985).
- Margin of Error: The standard error was multiplied by the critical value to calculate the margin of error. This represents the range within which the true difference in conversion rates is likely to fall.
- Lower Bound: The lower bound of the confidence interval was found by subtracting the margin of error from the sample statistic.
- Upper Bound: The upper bound was determined by adding the margin of error to the sample statistic.

As a result of these calculations, the 95% confidence interval for the difference in conversion rates between the treatment and control groups was found to be (0.0035, 0.0107). This interval suggests that we can be 95% confident that the true difference in conversion rates lies within this range.

Sample statistics	0.00706982
Std error Group A	0.001244334
Std error Group B	0.001339777
Std error	0.001828488
Critical value	1.959963985
Margin of error	0.003583771
Lower bound	0.00348605
Upper bound	0.01065359
Confidence Interval	(0.0035,0.0107)

[9]

Hypothesis test to see whether there is a difference in the average amount spent between the two groups

To test whether there is a difference in the average amount spent per user between the control and treatment groups, a two-sample t-test for the difference in means was conducted based on the dataset's characteristics.

The Null Hypothesis was formulated as 'there is no difference in the average amount spent between the control and treatment groups,' while the Alternative Hypothesis stated, 'there is a difference in the average amount between the control and treatment groups.' To determine whether to reject or fail to reject the Null Hypothesis, a p-value was calculated.

The p-value was obtained after calculating the standard deviations for both groups and subsequently computing the test statistic using the standard deviation and average of both groups. The p-value was determined to be 0.943856044.

A p-value of 0.943856044 is relatively high, and when compared to the chosen significance level (α) of 0.05 or 5%, it is notably higher. After this comparison, it can be concluded that the Null Hypothesis is fail to be rejected. In other words, there is insufficient evidence to reject the Null Hypothesis, suggesting that there is no significant difference in the average amount spent between the control and treatment groups.

Null Hypothesis – H0	There is no difference in the average amount spent in control and treatment group
Alternative Hypothesis – H1	There is difference in the average amount spent in control and treatment group
H0 – Notation	H0 : p1 = p2
H1 - Notation	H1 : p1 ≠ p2

Control Group (Group A) average amount spent	3.374518468
Treatment Group (Group B) average amount spent	3.390866946
Std deviation Control Group (Group A)	25.93639056
Std deviation Treatment Group (Group B)	25.4141096
Test statistics	-0.07042491
Degree of freedom	24342
P-value	0.943856044

[9]

95% confidence interval for the difference in the average amount spent between the treatment and the control

To establish a 95% confidence interval for the difference in the average amount spent per user between the treatment and control groups, a series of steps were meticulously executed. It began with the computation of sample statistics and the subsequent determination of the standard deviation for both groups. The degrees of freedom were ascertained to facilitate the calculation of the standard error, which in turn played a crucial role in determining the critical value associated with the chosen confidence level. By multiplying the standard error and the critical value, the margin of error was pinpointed. To find the lower bound of the confidence interval, the margin of error was subtracted from the sample statistics, while the upper bound was calculated by adding the margin of error to the sample statistics. Ultimately, the resulting 95% confidence interval for the difference in average spending between the treatment and control groups was established and found to be (-0.4387, 0.4714). This interval provides a reliable estimate of where the true difference in average spending between the groups is likely to lie with a 95% level of confidence.

Sample statistics	0.016348478
Std deviation Group A	25.93639056
Std deviation Group B	25.4141096
Degree of freedom	24342
Standard error	0.232140559
critical value	1.960061445
Margin of error	0.455009759
Lower bound	-0.43866128
Upper bound	0.47135824
Confidence interval	(-0.4387, 0.4714)

[9]

Advanced Tasks

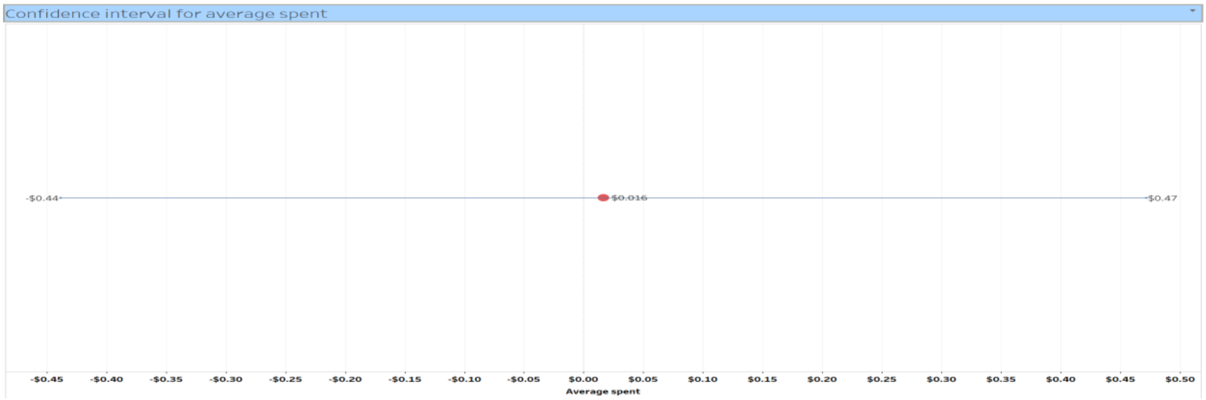
Visualize the Confidence Intervals

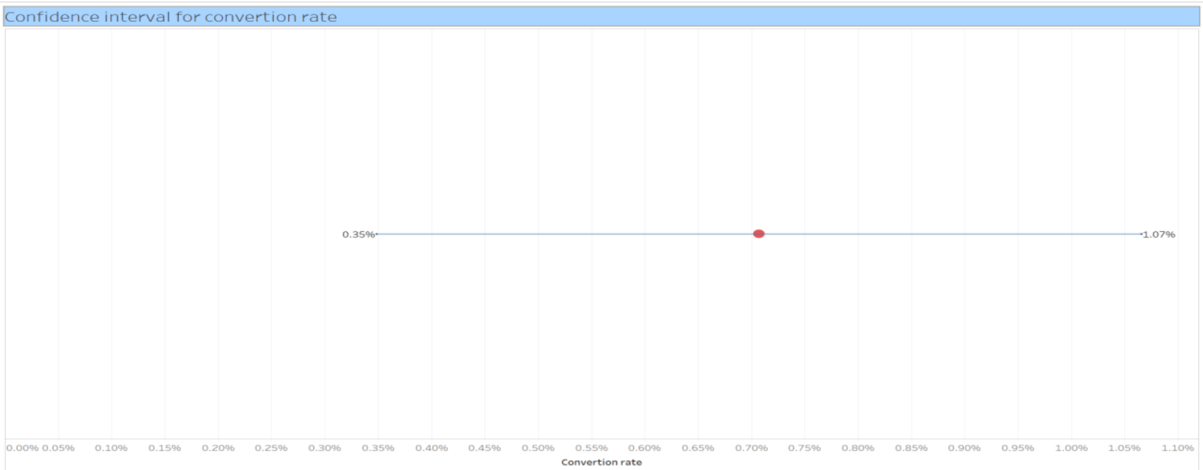
The confidence intervals serve to delineate the likely range of values within which the true population parameter can be found with a specified level of confidence. In this analysis, we have established confidence intervals for two key metrics: average amount spent per user and conversion rate.

Average amount spent:

For average amount spent per user, the 95% confidence interval, as derived from the data, spans from -0.44 to 0.47. This signifies that we can express 95% confidence that the true population's average spending falls within this range.

Additionally, we observed a difference of 0.02 in the conversion rate between the groups. These confidence intervals are essential for understanding the precision and reliability of our estimates, enabling us to make informed decisions and draw conclusions about the population parameters.





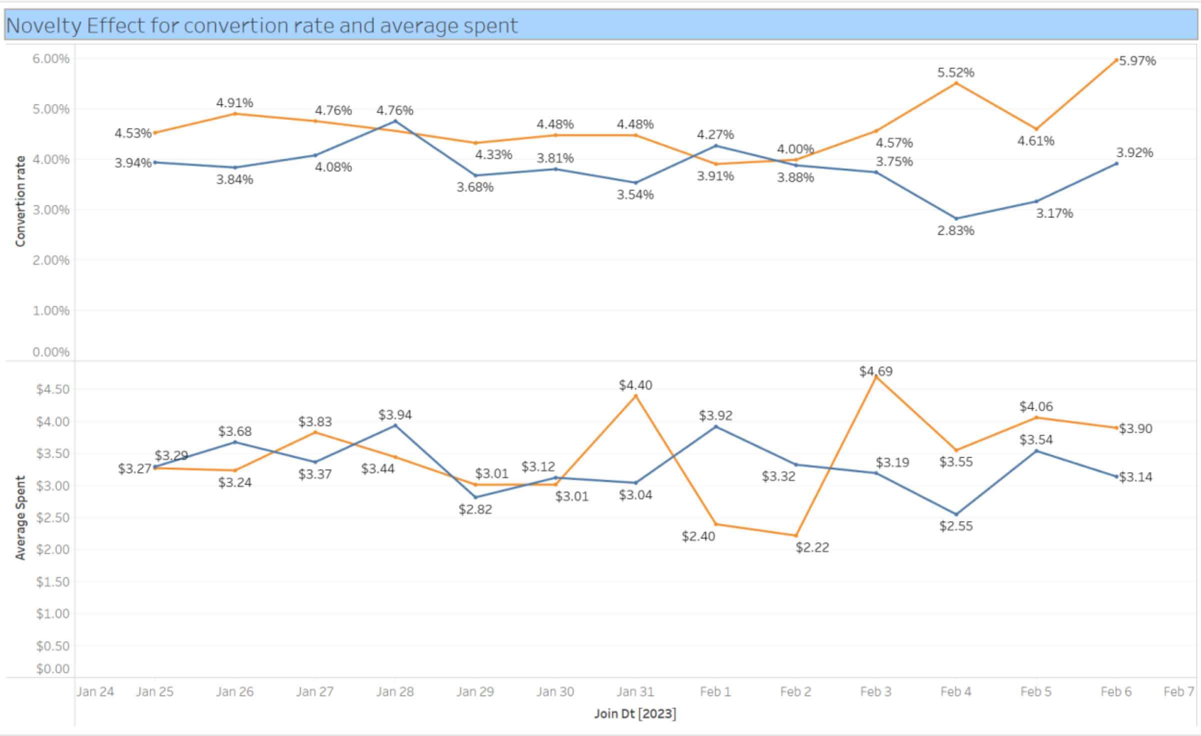
[8]

Conversion rate:

The 95% confidence interval for the conversion rate, as determined from the chart, is (0.0035, 0.0107). This indicates that we can assert with 95% confidence that the true population's conversion rate falls within this interval.

Moreover, we observed a difference of 0.0071 in the conversion rate between the groups. These confidence intervals provide a valuable range within which we can reasonably expect the actual population parameter to be situated, allowing us to make well-informed assessments and conclusions regarding conversion rates.

Check for Novelty Effects



[8]

The introduction of the new banner on Globox's platform has brought forth a noticeable "novelty effect." This is evident from the graph, where we observe an increase in both the conversion rate and the average amount spent. It appears that users are embracing the new banner enthusiastically, as the treatment group displays a more significant increase in these metrics compared to the control group.

This positive response suggests that users are genuinely interested in having key products highlighted on the landing page. The novelty and emphasis on key items seem to resonate well with users, driving improved conversion rates and higher spending, ultimately benefiting the company.

Power analysis

Metrics		Inputs	Sample size (Control group)	Sample size (Treatment group)	Sample Size total
Conversion rate	Baseline conversion rate (%)	3.92	38500	38500	77000
	minimum Detectable Effect (MDE) (%)	10			
	A/B Split ratio	0.5			
	Significance level (α)	0.05			
	Statistical Power (1-β)	0.08			
Means	Difference between Two Means	0.0164	39261696	39261696	78523392
	Expected Standard Deviation	25.93639056			

[10]

Conversion rate:

Based on the information provided, the calculated required sample size for each group is 38,500, totalling 77,000 for the entire experiment. This calculation considers several factors: the baseline conversion rate, a minimum detectable effect of 10%, an A/B split ratio of 0.05, a statistical level of 0.05, and a statistical power of 0.08.

However, it's worth noting that the actual sample size used for the experiment was 48,943, which is significantly smaller than the sample size indicated by the power analysis. This suggests that while the experiment was conducted with a sample size that exceeded the calculated requirement, there may still be room for increasing the sample size further to enhance the statistical power of the experiment and improve the reliability of the results.

Average amount spent:

Based on the provided information, the required sample size for each group is approximately 39,261,696 (assuming equal group sizes), resulting in a total sample size of 78,523,392. This sample size is necessary to achieve a statistical power of 80% and maintain a level of significance of 5% (two-sided) for detecting a true difference in means between the test and reference groups of 0.0164 units.

In practical terms, if you were to randomly select a sample of 39,261,696 individuals from each population and find that the difference in the means of the two groups is 0.0164 units, with a pooled standard deviation of 25.93639056 units, you would have an 80% probability of correctly identifying that the two groups indeed have significantly different means. This would be reflected in a two-sided p-value of less than 0.05, signifying statistical significance.

Recommendation:

Considering the observed improvements in our success metrics, I recommend moving forward with the launch of the experiment. However, it's essential to recognize that our power analysis has indicated potential limitations in our sample size, both in terms of conversion rate and average amount spent.

While these findings suggest that our data may not be as statistically robust as desired, it's worth considering the relative cost-effectiveness of launching the experiment. Given that the launch is not prohibitively expensive, the benefits of gathering more data need to be weighed against the potential delay in the product's release.

In this context, my recommendation is to proceed with the product launch based on our current sample size. This decision reflects a practical compromise that balances the desire for more data with the need for timely implementation. As the experiment progresses, we can continuously evaluate its impact and consider future refinements based on additional data and insights.

Appendix

1. SELECT MIN(join_dt), MAX(join_dt)
   FROM group

2. SELECT COUNT(DISTINCT(uid)) AS toal_users
   FROM groups

3. SELECT groups.group, COUNT(DISTINCT(uid)) AS toal_users
   FROM groups
   GROUP BY groups.group

4. WITH temp_tab AS(
   SELECT COUNT(DISTINCT(us.id)) AS total_users,
   COUNT(DISTINCT(ac.uid)) AS converted,
   (COUNT(DISTINCT(us.id))-COUNT(DISTINCT(ac.uid))) AS Not_converted
   FROM users AS us
   FULL JOIN activity AS ac ON ac.uid = us.id)
   SELECT converted, Not_converted,
   (CAST(converted AS FLOAT) /CAST(total_users As FLOAT))*100.0 AS convertion_rate,
   (CAST(Not_converted AS FLOAT) /CAST(total_users As FLOAT))*100.0 AS not_convertion_rate
   FROM temp_tab

5. WITH temp_tab AS(
   SELECT gr.group AS group_name,COUNT(DISTINCT(us.id)) AS total_users, COUNT(DISTINCT(ac.uid)) AS converted,
   (COUNT(DISTINCT(us.id))-COUNT(DISTINCT(ac.uid))) AS Not_converted
   FROM users AS us
   FULL JOIN activity AS ac ON ac.uid = us.id
   FULL join groups AS gr on gr.uid = us.id
   GROUP BY gr.group)
   SELECT group_name,converted, Not_converted, (CAST(converted AS FLOAT) /CAST(total_users As FLOAT))*100.0 AS convertion_rate,
   (CAST(Not_converted AS FLOAT) /CAST(total_users As FLOAT))*100.0 AS not_convertion_rate
   FROM temp_tab
   GROUP BY total_users,group_name, converted, Not_converted

6. SELECT SUM(ac.spent)/COUNT(DISTINCT(us.id)) AS spent_per_user
   FROM users AS us
   FULL JOIN activity AS ac ON ac.uid = us.id
   FULL join groups AS gr on gr.uid = us.id

7. SELECT gr.group, SUM(ac.spent)/COUNT(DISTINCT(us.id)) AS spent_per_user
   FROM users AS us
   FULL JOIN activity AS ac ON ac.uid = us.id
   FULL join groups AS gr on gr.uid = us.id
   GROUP BY gr.group

8. tableau-dashboard-link
9. Spreadsheet-link
10. Statsig-calculator-link
11. Statulator-link