

Goblox A/B Test by João Morgado

08/07/2022

Summary

Based on the analysis conducted, it is recommended to proceed with launching the experiment and implementing the banner on the website. The A/B test revealed a significant increase in the conversion rate by 0.7% and a slight increase in the average amount spent. These positive metrics indicate the potential for increased revenue and customer engagement. While the sample size used in the test was smaller than the recommended size, the observed effects suggest a positive impact of the banner. However, to gather more conclusive evidence and mitigate risks, further A/B tests with larger sample sizes should be conducted. Continuous monitoring and data collection will provide valuable insights for making informed decisions and optimizing the effectiveness of the banner implementation.

Context

The purpose of the report was to investigate the potential impact of implementing a banner on the Goblox website in order to increase revenue. To assess this, an A/B test was conducted. The experiment involved randomly assigning website visitors to two groups: Group A, which did not see the banner, and Group B, which was exposed to the banner. The test aimed to measure the conversion rate and the average amount spent by customers in each group. The dataset was obtained by tracking user interactions on the website during the test period. The data was cleaned and analyzed using SQL and spreadsheets, and the results were visualized using Tableau to facilitate a comprehensive understanding of the findings.

Results

The analysis led to several key discoveries, presented and visualized in the subsequent sections of this report. These findings offer a comprehensive understanding of user behavior, providing important insights for future strategies.

Approach:

The approach to this analysis involved multiple steps. It began with data extraction and cleaning using SQL, crafting a dataset that included vital variables such as user ID, country, gender, device, group assignment, and a column to signify whether a purchase was made. Hypothesis testing and calculating confidence intervals were then performed using Sheets. Finally, the findings were visually represented using Tableau, ensuring clear and intuitive comprehension of the results. The complete queries, Sheets, and Tableau dashboards are documented in the appendix.

The first visualization illustrates a significant increase in the conversion rate between Group A and Group B, attributable to the introduction of the banner. It's clear that the banner implementation positively influenced conversions, as indicated by the difference between the two groups in the visualization.

Null Hypothesis (H0)	There is no difference in conversion rate between Group A and Group B	pA = pB
Alternative Hypothesis (H1)	There is a difference in conversion rate between Group A and Group B	pA != pB

α	0,05
N° Users A (nA)	24343
N° Users B (nB)	24600
N° Converted A (xA)	955
N° Converted B (xB)	1139

Conversion Rate A (pA)	0,039231
Conversion Rate B (pB)	0,046301
Difference in Conversion Rate (Pdiff)	0,007070

pooled proportion (p^)	$(xA+xB) / (nA+nB)$	0,042784464
standard error (SE)	$\sqrt{p^ * (1-p^ * ((1 / nA)+ (1 / nB)))}$	0,001829526
test statistic (Z-score)	$(conv_rateB - conv_rateA) / SE$	3,86429177
critical value (crit)	Assumed	1,96
margin of error (ME)	Crit * SE	0,003585871
P-value	$(=2*(1-NORM.S.DIST(ABS(Z-score);1)))$	0,0001114
confidence level (CL)	1 - α	95%
confidence interval Lower Bound (LB)	Pdiff - ME	0,003484
confidence interval Upper Bound (UB)	Pdiff + ME	0,010656

Given that the p-value (0.000111412) is less than the significance level (0.05), we can conclude that there is sufficient evidence to reject the null hypothesis.
--

In contrast, the second visualization shows no significant difference in the average amount spent per user between the two groups. It's interesting that while the banner successfully drove conversions, it did not seem to affect how much users spent on average.

Null Hypothesis (H0)	There is no difference in average spent between Group A and Group B	$pA = pB$
Alternative Hypothesis (H1)	There is a difference in average spent between Group A and Group B	$pA \neq pB$

α	0,05
Nº Users A (nA)	24343
Nº Users B (nB)	24600
Average A (μA)	3,37
Average B (μB)	3,39
Average Diff ($\mu Diff$)	0,02

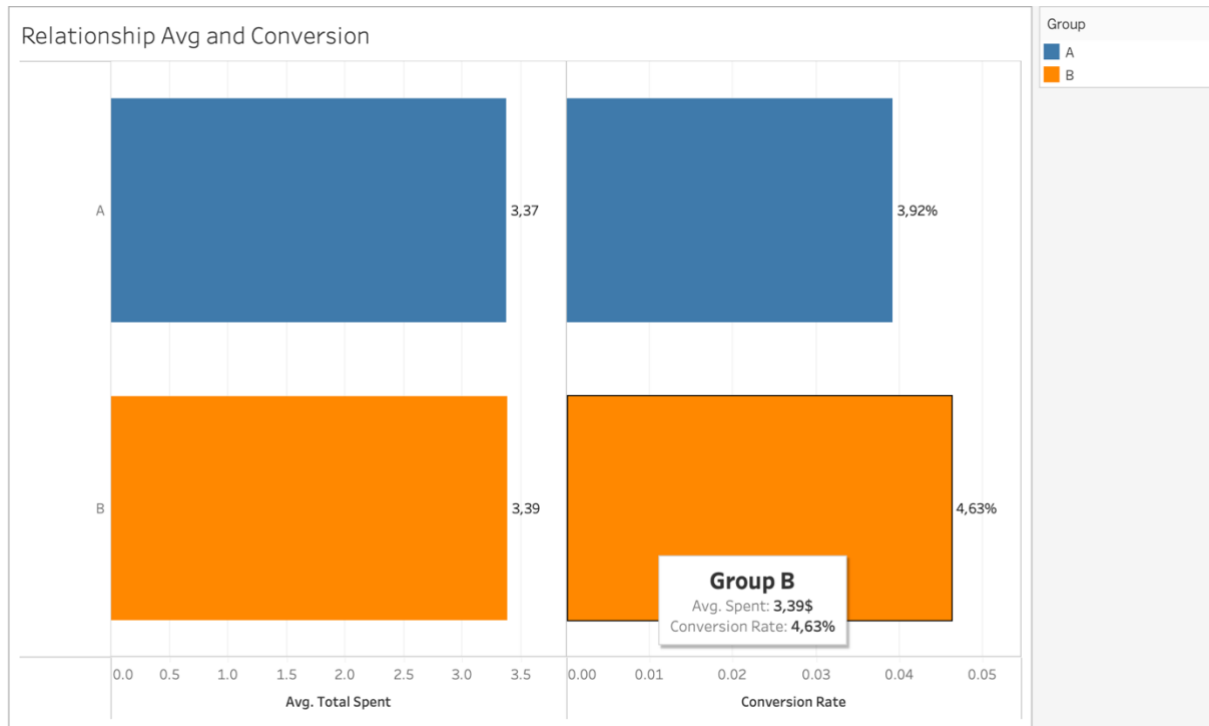
Standard Deviation A (sA)	25,936391
Standard Deviation (sB)	25,41411

standard error (SE)	$\sqrt{((sA^2 / nA) + (sB^2 / nB))}$	0,232140559
degrees of freedom (df)	$(nA-1) + (nB-1)$	48941
critical value (crit)	TINV(1- α /2, df)	1,96
margin of error (ME)	critical value * standard error	0,454998387
P-value	T.TEST, 2,3	0,998655951
confidence level (CL)	1 - α	95%
confidence interval Lower Bound (LB)	$(\mu Diff) - ME$	-0,44
confidence interval Upper Bound (UB)	$(\mu Diff) + ME$	0,47

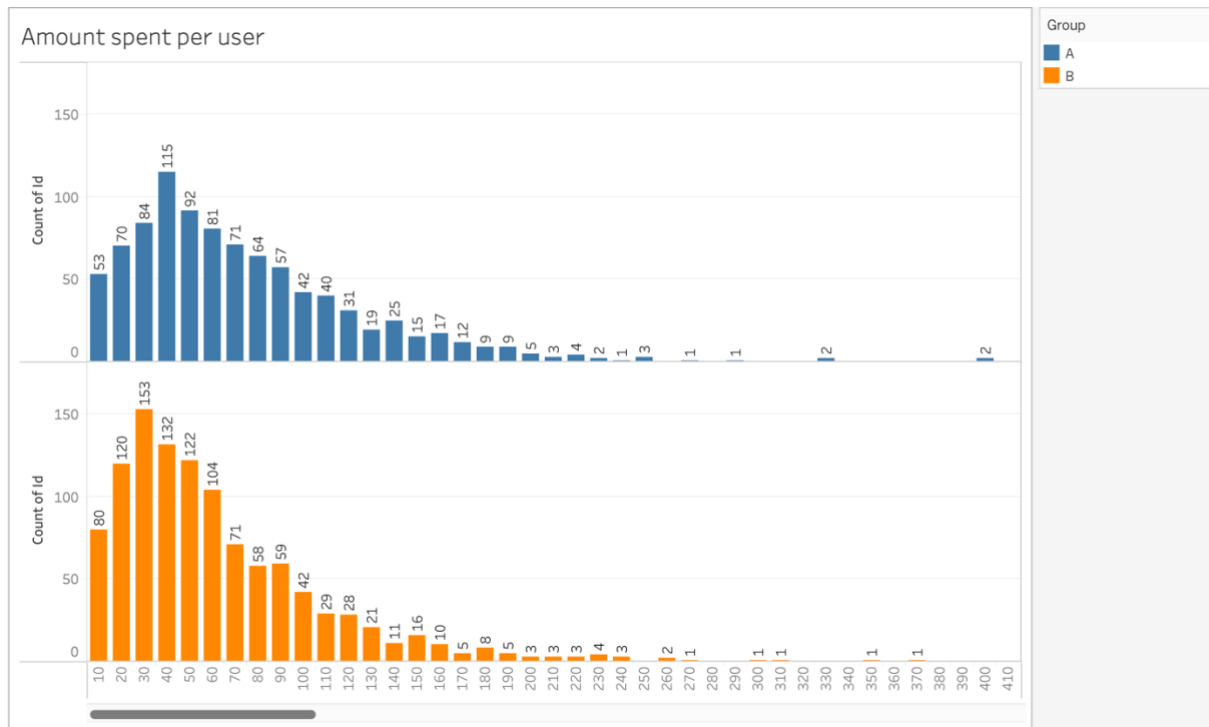
Given that the p-value (0.998655951) is greater than the significance level (alpha) of 0.05, we fail to reject the null hypothesis.		
--	--	--

In conclusion, the visualizations offer insightful revelations. The banner was effective in boosting the conversion rate, but it did not influence the average spending per user. This suggests that while current promotional efforts have been successful in increasing purchases, further strategies may need to be considered to increase the average spending per user.

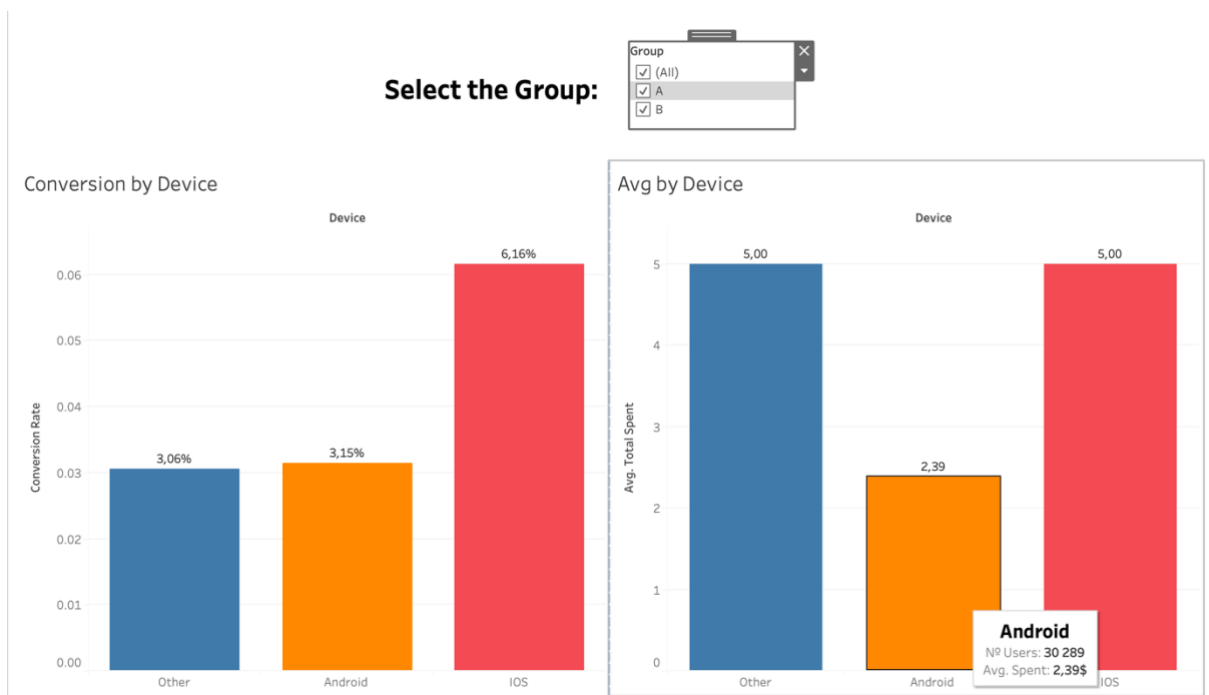
In Tableau, several visualizations were created to explore the relationships between different variables. The first visualization displayed the differences in average amount spent and conversion rate for each group. Group A had an average amount spent of \$3.37 and a conversion rate of 3.92%, while Group B had an average amount spent of \$3.39 and a conversion rate of 4.63%.



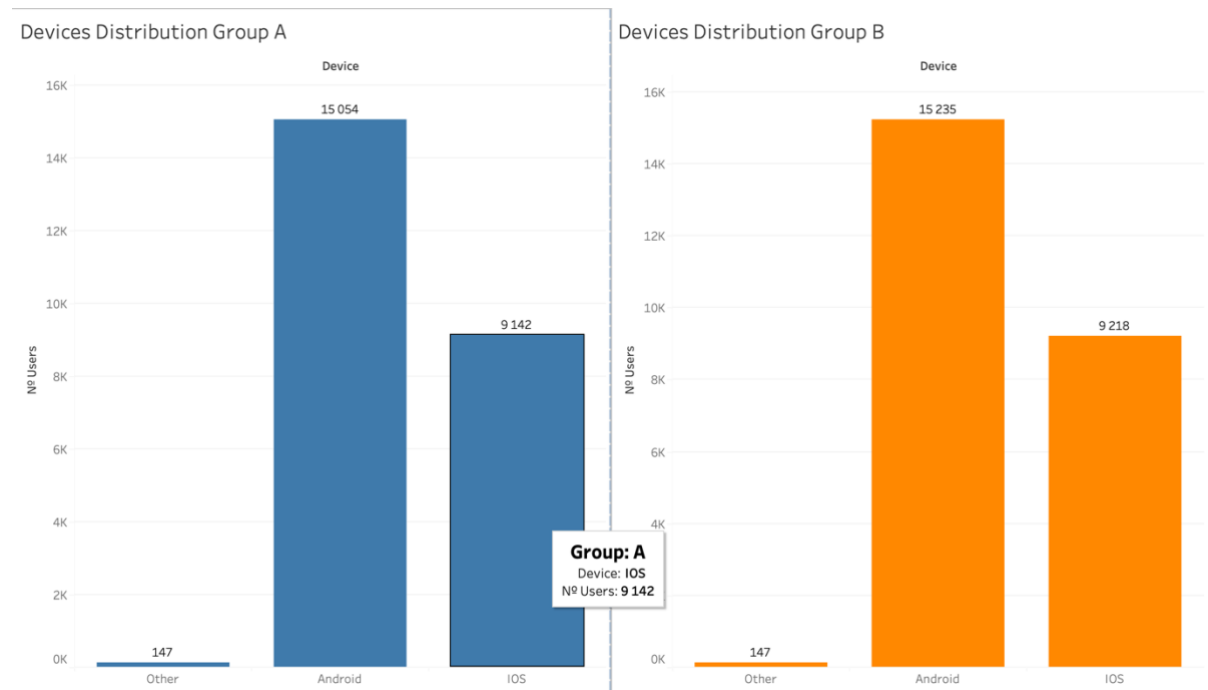
The second visualization focused on the amount spent per user. It revealed that even though Group B had a higher volume of products sold, Group A had a higher proportion of higher-priced items being sold. This suggests that while Group B may have had more transactions, Group A's sales were skewed towards higher-priced items.



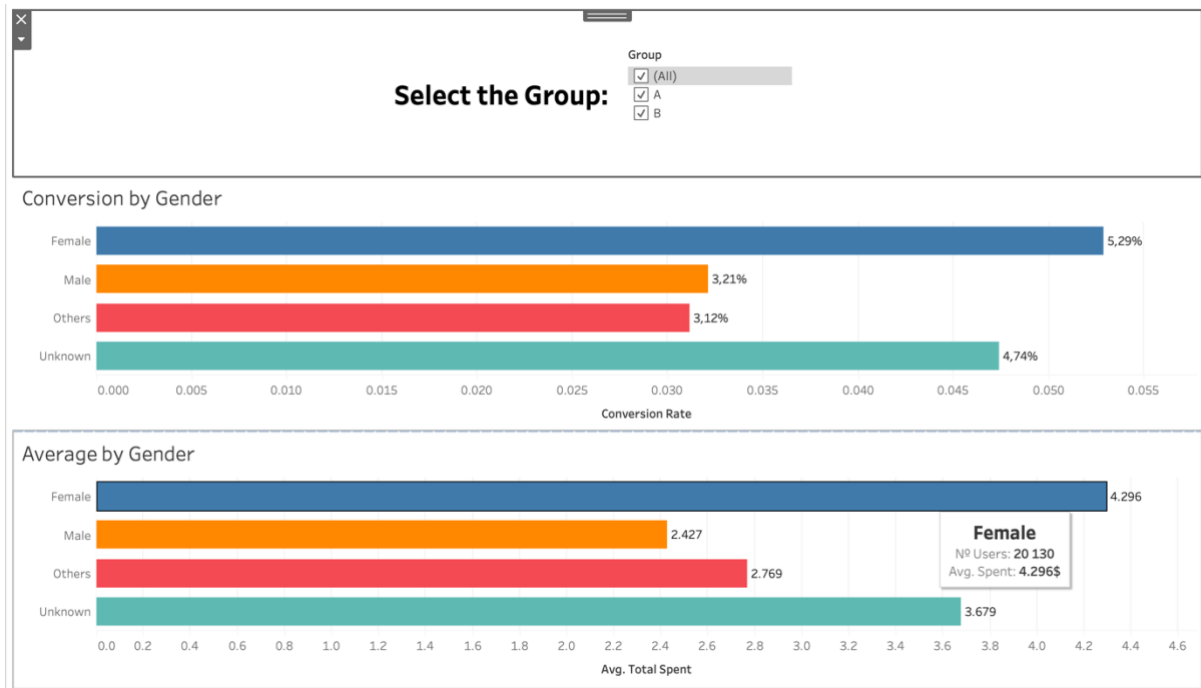
The third visualization explored the relationship between device type and the metrics. It showed that iOS users had both a higher conversion rate and a higher average amount spent compared to Android users. However, it was observed that Android users constituted a larger user base compared to iOS users. This finding suggests that the higher purchasing power of iOS users, potentially driven by the higher prices of iOS devices, contributed to the higher conversion rate and average amount spent for this group.



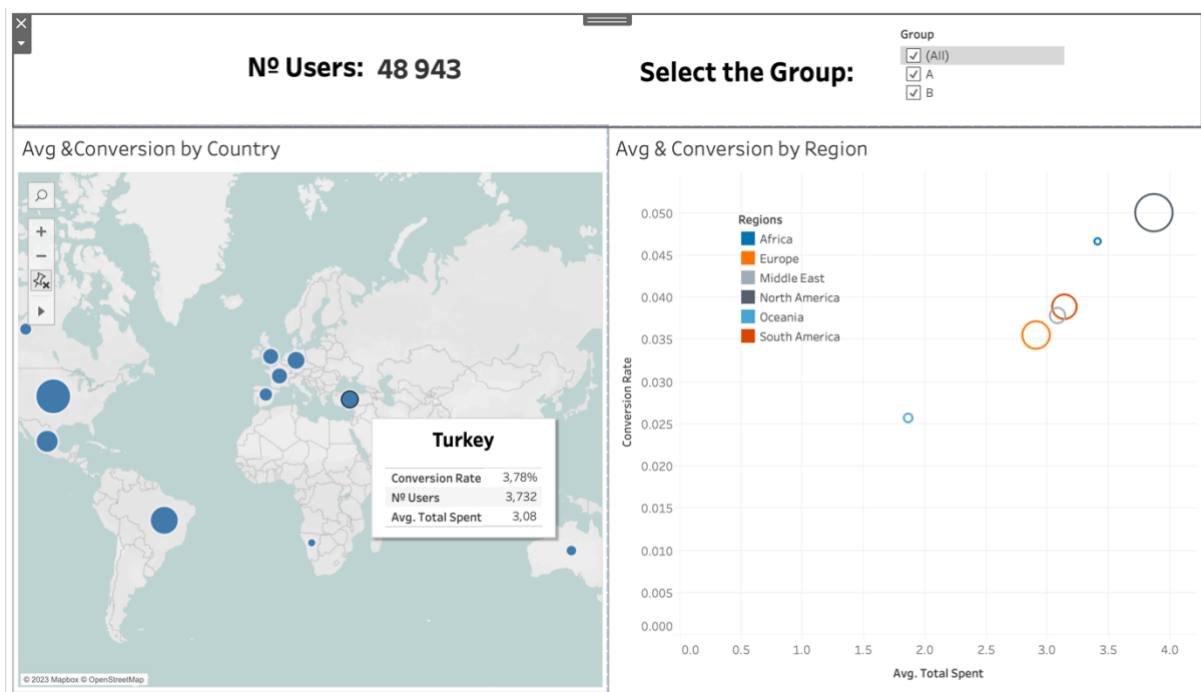
Upon further exploration, a detailed comparison was conducted to examine the distribution of device types across Group A and Group B. This examination was pivotal to ensure that the observed differences in conversion rates were not influenced by an imbalance in the device distribution between the groups. The newly generated visualization reveals a nearly identical distribution of device types in both groups, confirming that the type of device used by the participants did not contribute to the higher conversion rate observed in Group B. This further strengthens the conclusion that the banner implementation was the key driver of the increased conversions in Group B.



In the fourth visualization, the relationship between gender and the metrics was explored. It was observed that females accounted for a higher number of users, with a conversion rate of 5.29% and an average spend of \$4.30. This finding may be influenced by the website's focus on boutique fashion items, high-end decor products, and food and drink offerings, which tend to resonate more with female users.



The fifth visualization analyzed the average amount spent and conversion rate by country and region. The data revealed that the United States (USA) had the highest conversion rate at 5.44%, along with the highest average amount spent per user at \$4.17. Additionally, the USA had the highest number of users, with 14,772 recorded. This significant presence of users in the USA heavily influenced the North American region, positioning it at the top in terms of both conversion rate and average spend.

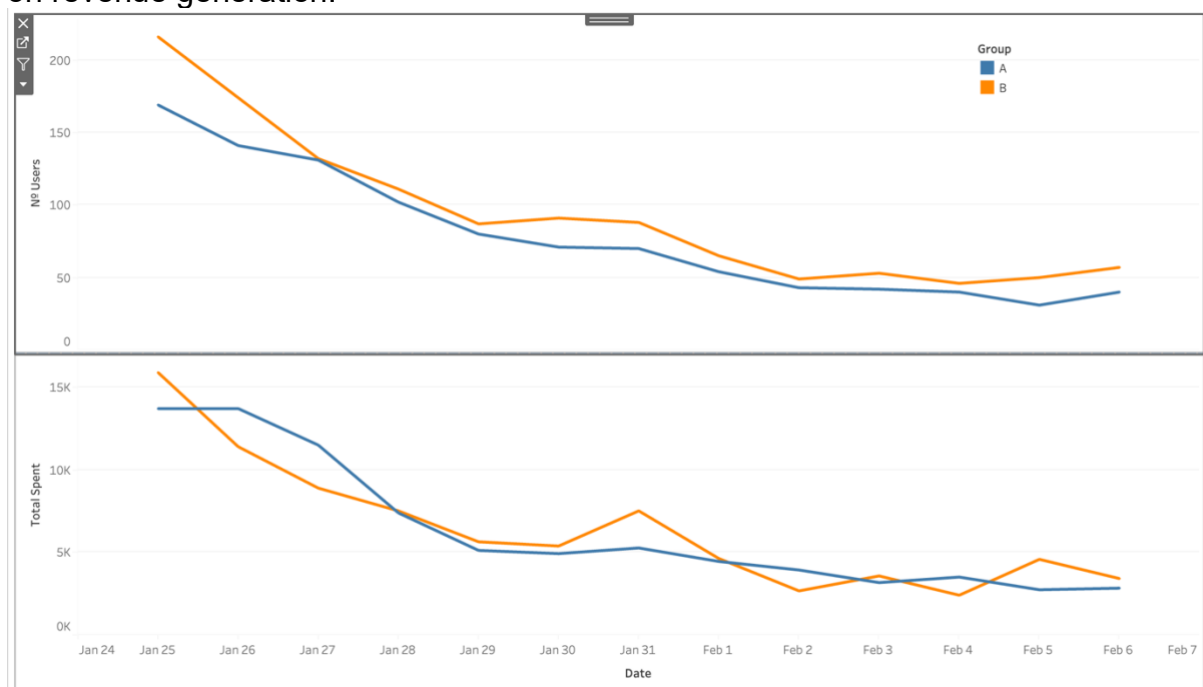


An additional part of the analysis involved exploring the possibility of a novelty effect in the A/B test. However, the visualization clearly indicated that Group B consistently outperformed Group A in terms of the number of users who made purchases and the

total amount of sales throughout the testing period. This consistent performance suggests there was no novelty effect; instead, the banner had a lasting positive influence.

Temporal patterns and trends were also considered in the analysis to give a more detailed view of user behavior over time. The fact that Group B consistently outperformed Group A throughout the test period underlines the success of the banner implementation.

In conclusion, considering the consistent superiority of Group B and the insights gained from the time-based analysis, it is recommended to continue with the banner implementation on the website. Regular performance monitoring and further analyses should be conducted to keep refining strategies and maximizing the banner's impact on revenue generation.

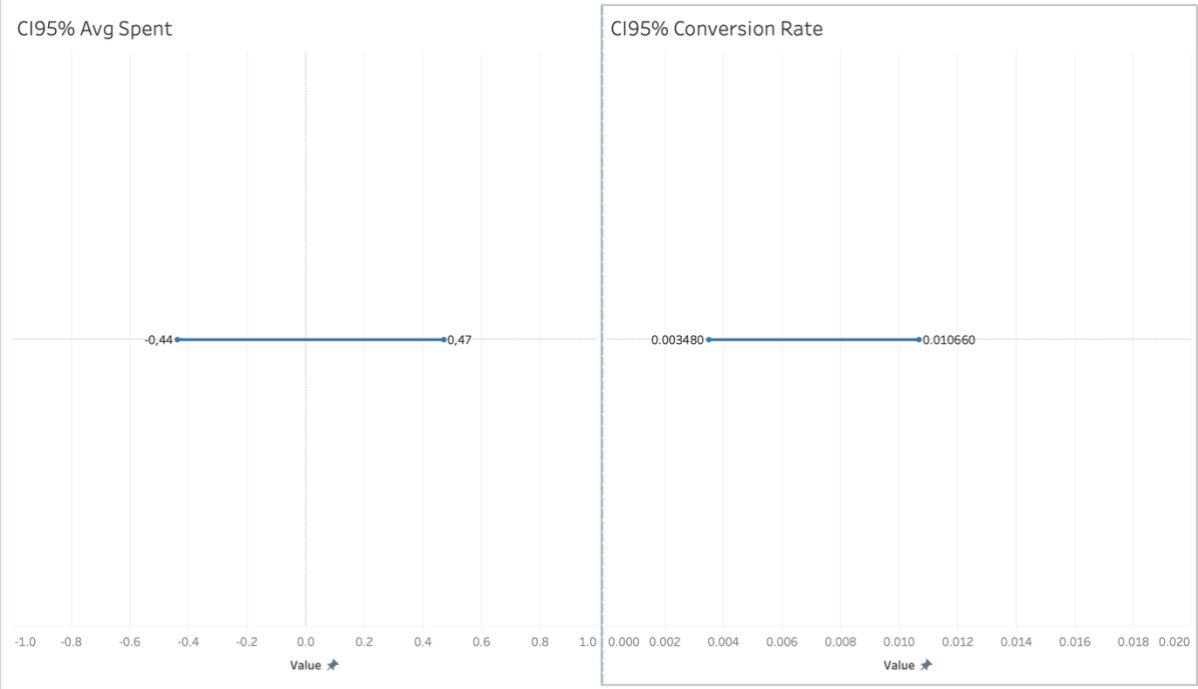


In the final visualization, the confidence intervals for the average amount spent and the conversion rate were displayed.

For the 95% confidence interval of the average amount spent, the lower bound was -0.44, and the upper bound was 0.47. This indicates that, with 95% confidence, the true average amount spent per user falls within this range. The confidence interval crossing zero suggests that there is some uncertainty regarding the impact of the banner on the average amount spent.


Regarding the 95% confidence interval of the conversion rate, the lower bound was 0.003480, and the upper bound was 0.10660. This range represents the estimated true conversion rate with 95% confidence. The fact that the interval does not cross zero indicates that the banner had a statistically significant positive effect on the conversion rate.

By visualizing the confidence intervals, the uncertainty surrounding the estimated values is made evident. It highlights the need to consider the range of potential values within which the true population values may lie. These confidence intervals provide a measure of the reliability and precision of the estimated effects, aiding in the interpretation and decision-making process based on the A/B test results.



In addition to the analysis, a power analysis was conducted to evaluate the sample size used in the A/B test. The analysis suggested larger sample sizes (77,400 for conversions and 245,278 for means) than the actual sample size of 48,943. The smaller sample size may limit the statistical power of the test to detect smaller effects accurately. Therefore, while the observed increase in conversion rate and average amount spent are promising, it is important to interpret the results cautiously due to the sample size limitations. Further experiments or data collection with larger sample sizes are recommended to validate the findings and make more confident decisions about implementing the banner on the website. Continuous monitoring and future A/B tests can provide more robust results and ensure that observed effects are not due to chance or sample variability.

Sample Size Calculator for Conversions

**Sample Size Calculator**
Calculate how many samples you need to properly power your experiment

Baseline Conversion Rate (%) ⓘ

3,9

Minimum Detectable Effect (%) ⓘ

10

Advanced Settings >

3.5%

3.9%

4.3%

Minimum Detectable Effect

Results

[Share Link](#)

TEST SIZE

38.7k

CONTROL SIZE

38.7k

TOTAL SAMPLE SIZE

77.4k

Sample Size Calculator for Means

Input Values

Select one of the two options to specify input values. Hover over the ? sign to obtain help.

☒ Expected Means ?

Mean of the Reference Group: ?

3.7

Mean of the Test Group: ?

3.9

Standard Deviation: ?

25

☐ Expected Difference between Means ?

Click the Options button to change the default options for Power, Significance, Alternate Hypothesis and Group Sizes. Use the Adjust

Results and Live Interpretation

[Download](#)

Assuming a pooled standard deviation of 25 units, the study would require a sample size of:

245278

for each group (i.e. a total sample size of 490556, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 0.19999999999999973 (i.e. 3.9 - 3.7) units.

In other words, if you select a random sample of 245278 from each population, and determine that the means of the test and the reference groups are 3.9 and 3.7 units, respectively, and the standard deviation is 25 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

Recommendation

Based on the analysis conducted, it is recommended to proceed with launching the experiment and implementing the banner on the website. The A/B test revealed a significant increase in the conversion rate by 0.7%, and a slight increase in the average amount spent. These positive metrics indicate the potential for increased revenue and customer engagement.

The primary purpose of the banner is not only to improve the conversion rate but also to draw attention to the foods & beverages products offered by Globox. This strategic emphasis on food and beverages is likely to have contributed to the observed improvement in the average spend histogram. By highlighting these products, customers may be more enticed to explore and purchase a wider range of items, resulting in higher average spending per transaction. This aligns with the company's goal of maximizing sales and optimizing the customer shopping experience.

Taking into account the encouraging results from the A/B test, implementing the banner with a focus on the foods & beverages products is expected to have a positive impact on key performance indicators, such as the average spend per customer and overall revenue.

While the sample size used in the test was smaller than the recommended sample sizes from the power analysis, the observed effects suggest that the banner implementation had a positive impact. However, it is important to approach the results with caution due to the limitations of the sample size.

To gather more conclusive evidence and mitigate potential risks, it is advisable to iterate on this experience and conduct further A/B tests with larger sample sizes. By doing so, it will help validate the observed effects and provide more confidence in the decision to implement the banner permanently.

Continuous monitoring of the implemented banner's performance is essential, along with collecting additional data to assess its long-term effects on key metrics. This iterative approach will allow for data-driven decision-making and provide insights into the banner's effectiveness in the long run.

In summary, launch the experiment based on the positive results observed, but also iterate on this experience by conducting future A/B tests with larger sample sizes to validate and strengthen the findings. Continuous monitoring and data collection will provide valuable insights for making informed decisions and optimizing the effectiveness of the banner implementation.

Appendix

First SQL code:

```
CREATE VIEW project_view AS
SELECT
    users.id,
    users.country,
    users.gender,
    groups.device,
    groups."group",
    COALESCE(SUM(activity.spent), 0) as total_spent,
    CASE WHEN SUM(activity.spent) > 0 THEN 'Yes' ELSE 'No' END AS
did_spend
FROM
    users
LEFT JOIN
    activity ON users.id = activity.uid
JOIN
    groups ON users.id = groups.uid
GROUP BY
    users.id, users.country, users.gender, groups.device, groups."group";
```

The **excel** file with it's analysis is provided inside the .zip file.

Visualizations:

Relationship of the avg and conversion by group

<https://public.tableau.com/app/profile/joao.morgado7999/viz/Relationshipoftheavgandconversionbygroup/RelationshipAvgandConversion?publish=yes>

Amount Spent per User

<https://public.tableau.com/app/profile/joao.morgado7999/viz/AmountSpentperUser/Amountspentperuser?publish=yes>

Relationship by Device

<https://public.tableau.com/app/profile/joao.morgado7999/viz/RelationshipbyDevice/RelationshipbyDevice?publish=yes>

Devices Distribution

<https://public.tableau.com/app/profile/joao.morgado7999/viz/DevicesDistribution/Dashboard6>

Relationship by Gender

<https://public.tableau.com/app/profile/joao.morgado7999/viz/RelationshipbyGender/RelationshipbyGender?publish=yes>

Avg and Conversion by Country and Region

<https://public.tableau.com/app/profile/joao.morgado7999/viz/AvgandConversionbyCountryandRegion/AvgConversionbyCountryRegion?publish=yes>

Novelty Effect

https://public.tableau.com/app/profile/joao.morgado7999/viz/NoveltyEffect_16888547246850/NoveltyEffect?publish=yes

SQL Query with dates

```
SELECT
    users.id,
    users.country,
    users.gender,
    groups.device,
    groups."group",
    COALESCE(SUM(activity.spent), 0) as total_spent,
    CASE WHEN SUM(activity.spent) > 0 THEN 'Yes' ELSE 'No' END AS did_spend,
    groups.join_dt,
    activity.dt as dt_purchase

FROM
    users
LEFT JOIN
    activity ON users.id = activity.uid
JOIN
    groups ON users.id = groups.uid
GROUP BY
    users.id, users.country, users.gender, groups.device, groups."group",
    groups.join_dt,
    activity.dt
```

CI95% for Avg and Conversion

<https://public.tableau.com/app/profile/joao.morgado7999/viz/CI95forAvgandConversion/CI95forAvgConversion?publish=yes>

Links for the Power Analysis

For Conversion

<https://www.statsig.com/calculator>

For Means

<https://statulator.com/SampleSize/ss2M.html#>