Aviad Lange, 311525786 09.07.2023

GloBox A/B Testing Project

Summary

In this brief document, I will present my findings regarding the A/B testing launched by GloBox. The design team created a banner to promote the food category on the company's app. As a data analyst, my task was to assess whether it increases sales and conversions. The results did show a significantly higher conversion rate for the treatment group, but the average purchases were nearly equal between both groups. With that in mind, along with additional calculations and considerations that I'll disclose in this paper, my recommendation is to continue iterating and testing. I will elaborate on this further in the following sections.

Context

GloBox was searching for methods to boost its revenue, and one of the potential solutions explored was adding a banner to the app, which directs users to the food & beverages category. To assess the effectiveness of this banner, the data team randomly assigned new users to one of two groups – Group A (Control) and Group B (Treatment) – and analyzed the differences in user behavior between the two groups[1]. The data analysis focused on two primary aspects: the percentage of conversions and the amount of money spent by these users via the app.

The data set was composed of three tables: 'activity', 'groups', and 'users', each of which was associated with a unique user ID. The 'users' table contained information about each user's gender and country of residence. The 'groups' table included details about the user's device type (Android or iOS), the date they joined the experiment, and, crucially, their assigned group (A/B).

[1] Appendix N. 1.

On the other hand, the 'activity' table differed slightly as it only included information about users who had made a purchase. Each purchase was recorded separately, which meant that a user could appear multiple times in this table if they made multiple purchases. The 'activity' table included data such as the purchase date, the user's device type, and the price they paid for the purchase.

Data Processing

First of all, I had to manipulate the database to create new fields that would facilitate extracting clear insights. I crafted a query that joined the necessary data into one table, which included each user's unique ID, country, gender, device type, assigned group (A/B), whether they converted or not, and the total amount of money they spent[2]. Then, I wrote two additional queries: one groups the number of conversions and the average spending by day[3], while the other sums up the number of users and adds it to the cumulative total from the previous days[4].

Once I had the data organized, I used a spreadsheet in order to perform some statistical calculations. I performed a 't-test', and found different 'p-value's, the margin of 95% and 99% error, and the 'Confidence Intervals' for all sort of values[5]. Finally, I used Tableau to visualize my findings, including change over time (in order to check for a possible 'Novelty Effects', and growth forecast in terms of users[6].

[2] Appendix N. 2.
[3] Appendix N. 3.
[4] Appendix N. 4.
[5] Appendix N. 5.
[6] Appendix N. 6.

Findings

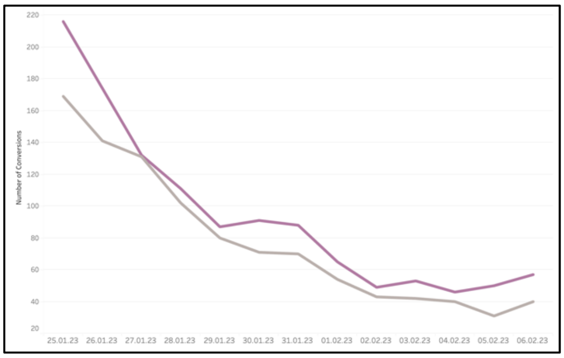
Conversion Rate

First and foremost, when considering the conversion rate, there was a noticeable increase for the treatment group, rising from 3.92% to 4.63%. If we analyze the data by gender, it appears that women tend to convert more frequently than men, but they were less affected by the banner, with their conversion rate going up from 5.14% to 5.44%. On the other hand, men's purchases increased from 2.63% to 3.79%, which is an increase of more than a whole percent!

Moreover, when comparing users based on their device type, IOS users exhibited a significantly higher conversion rate (increasing from 5.85% to 6.47%) compared to Android owners (who had a conversion rate of 2.77% to 3.52%).

As for the country division analysis, I won't delve too deeply into it, but my visualization clearly shows the country with the largest change (Canada), the one with the lowest conversion rate (Austria), and the only one with a negative change (Turkey).

Regarding the concern about 'Novelty Effect', the visualization indicates a gradual decrease in conversion rates over time, but this trend is evident for both the control and treatment groups.



These findings provide valuable insights into the impact of the banner on the app's conversion rates across different user segments, indicating areas where the banner was more effective and areas where further investigation may be warranted.

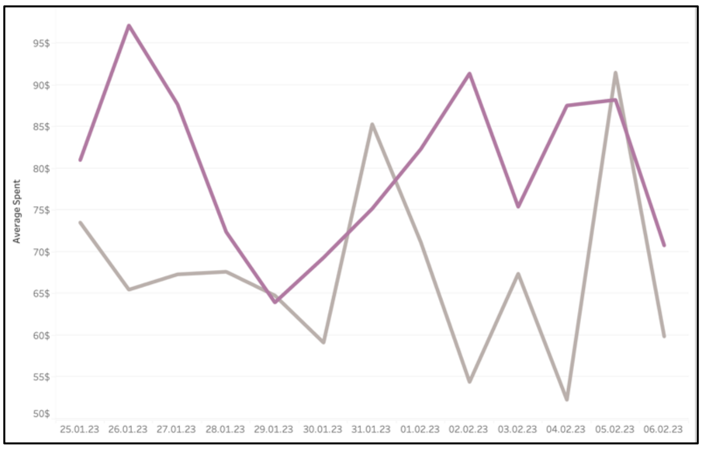
Average Spending

Regarding the average amount spent, there was only a very slight change, from 3.37$ to 3.39$ per user. If we analyze the data by gender, it appears that women from group B even spent less somehow – from 4.46$ to 4.13$, while men, once again, were tempted by the banner and spent 2.60$, compared to the 2.25$ spent by their counterparts in group A.

When considering device type, IOS users generally spend more, but their average spending went down from 5.07$ to 4.92$, while there was an increase from 2.31$ to 2.47$ among Android owners.

In addition, based on the country-wise analysis, it's evident that one country had the largest change in average spending (GB), one country had the lowest average spending (Austria), and negative changes were observed in five out of ten countries.

Regarding the concern about 'Novelty Effect', the visualization does not reveal any detectable pattern, and this observation holds true for both the control and treatment groups.



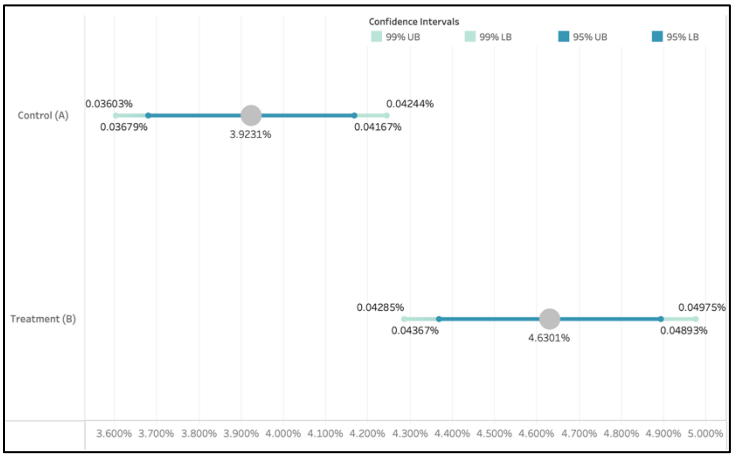
Lastly, a very interesting result is the average spending for only converted users, where there is a substantial difference in favor of the control group – 86.02$, compared to the treatment group's 73.24$.

These findings shed light on the impact of the banner on average spending across different user segments, suggesting varying responses based on gender, device type, and country. The significant difference in average spending for converted users between the control and treatment groups warrants further investigation and consideration when making decisions about the banner's effectiveness.

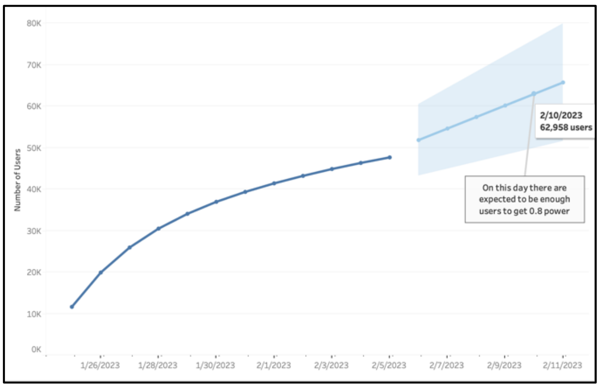
Data's Reliability

Based on the information provided, it seems that the data's reliability for the conversion rate analysis is quite high. The very low 'p-value' of 0.0001 indicates strong evidence against the null hypothesis, supporting the conclusion that there is a significant difference in conversion rates between the control and treatment groups. Additionally, the confidence intervals calculated at 95% and 99% certainty further emphasize the clear distinction between the two groups' conversion rates, with a minimum difference of 0.2352%.

The visualization of the confidence intervals for each group reinforces the finding of a substantial difference in conversion rates, providing a deeper understanding of the results.

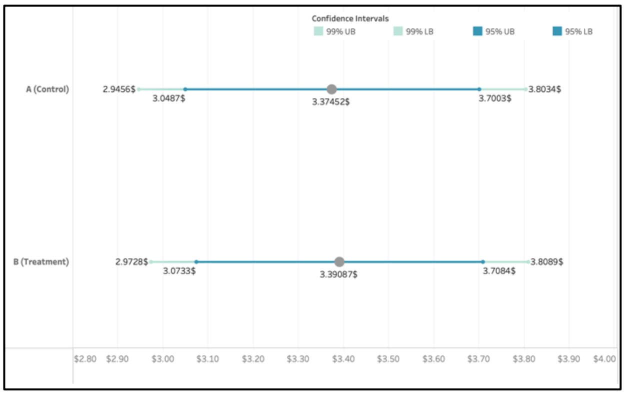


The only parameter that lacks the certainty stamp is the power analysis. In order to have a power of 0.8 (a commonly accepted threshold), the experiment should have lasted a few more days and achieve the 60,600 users required[7].



However, the analysis of the amount spent is less conclusive. The high 'p-value' of 0.944 indicates that there is not enough evidence to reject the null hypothesis, suggesting that there may not be a significant difference in the average amount spent between the control and treatment groups.

The maximum difference in cents (0.615$ at 99% certainty) is relatively small in terms of spending, and the visualization of the confidence intervals indicates considerable overlap, which further supports the notion that there might not be a clear difference in average spending between the two groups.



[7] Appendix N. 7.

Moreover, the power analysis reveals that achieving a statistical power of 0.8 for the amount spent analysis would require an extremely large number of users, more than 38 million, making it impractical and costly to obtain a meaningful result[8].

In conclusion, the data's reliability for the conversion rate analysis appears to be strong, showing a significant difference between the control and treatment groups. However, the amount spent analysis lacks conclusive evidence due to high 'p-values' and substantial overlap in confidence intervals, making it difficult to draw definitive conclusions about the impact of the banner on average spending.

Recommendations

Upon careful review and analysis of the data, I believe the optimal approach is not to launch, and continue iterating while considering specific parameters. It is evident that the banner's exposure had a positive impact on the conversion rate, which is indeed a positive sign. However, from a business perspective, the ultimate goal of increasing income per user has not been significantly achieved.

In my view, the substantial drop observed in spending among converted users may suggest that the banner is focusing on the wrong category. The control group users were not particularly enticed to explore the food & beverages category, and thus those who made purchases likely opted for more expensive product categories. Conversely, the treatment group users were drawn to the food & beverages category, but their purchases in this category did not amount to significant spending due to relatively lower costs associated with food-related items.

Hence, I am of the opinion that the food & beverages category might not be as profitable as other alternatives. To effectively capitalize on the increased conversion

[8] Appendix N. 8.

rate, it would be wise to combine an effective banner with a different target category that promises higher income potential. Although the space on the app is precious, considering that banner development is relatively inexpensive and straightforward, conducting several two-week periods, each focusing on a distinct category, would serve as a valuable indicator to chart our future course.

Furthermore, it is crucial to address the disparity in metrics between IOS and Android users. IOS users are converting and spending at approximately twice the rate of Android users in both groups. Therefore, it is highly recommended to ensure that the Android interface is user-friendly and free of any major issues to enhance performance among Android users.

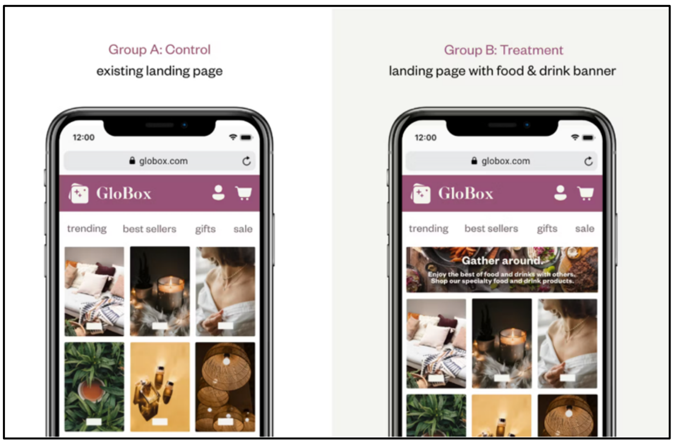
Lastly, as mentioned earlier, it appears that women were less influenced by the banner. Therefore, it would be prudent to explore tailored designs for the banner, specifically catering to the preferences and needs of female users.

In conclusion, by leveraging these insights and employing a data-driven approach, we can refine our strategies and continuously improve the banner's impact to achieve our ultimate goal of increased income per user.

Aviad Lange, 311525786 09.07.2023

Appendix

1.



2.

```sql
-- Gets a table with all the needed data
SELECT id, country, gender, g.device, "group",
    CASE WHEN total_per_cus IS NULL THEN FALSE
    ELSE TRUE
    END AS converted,
coalesce(total_per_cus, 0) AS total_spent
FROM "users" AS u
LEFT JOIN "groups" AS g
ON g.uid = u.id
LEFT JOIN
    (SELECT uid, SUM(spent) AS total_per_cus
    FROM "activity"
    GROUP BY uid) AS "new_activity"
ON "new_activity".uid = u.id
ORDER BY total_spent DESC
```

3.

```sql
--This query groups the number of conversions and the average spendig by day.
SELECT "group", dt AS date,
    COUNT(converted) AS num_of_conv,
    AVG(spent) AS avg_spent
FROM "Cleaned Data" AS c
RIGHT JOIN "activity" AS a
ON c.id = a.uid
WHERE converted = 't'
GROUP BY dt, "group"
ORDER BY date
```

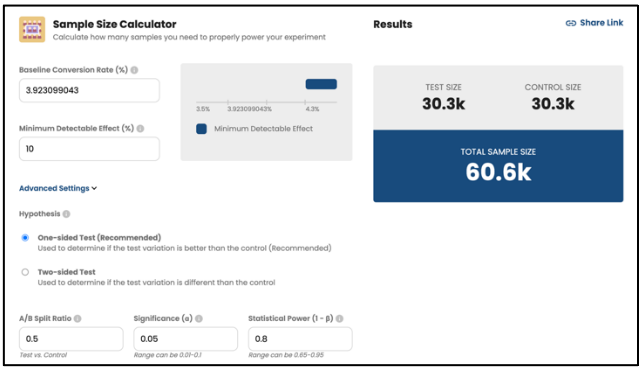
4.

```sql
--This query  counts the number of new users joined each day and sums it with the count of the days before.
SELECT join_dt,
  SUM(new_users) OVER
  (ORDER BY join_dt ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)
  AS RunningSum
FROM(
    SELECT join_dt,
        COUNT(converted) AS new_users
    FROM "Cleaned Data" AS c
    RIGHT JOIN "groups" AS g
    ON c.id = g.uid
    GROUP BY join_dt) AS count_users;
```

5.https://docs.google.com/spreadsheets/d/1egJtiNO6bDa-FH-As9HNk1HLlWF857dOOtv6MxdDq4s/edit#gid=1737822075,

'Cleaned Data.xslx', file attached.

6. https://public.tableau.com/authoring/GloBoxProject/GloBoxPresentation#1

7. 

https://www.statsig.com/calculator

8. https://statulator.com/SampleSize/ss2M.html, file attached.