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 came up with a banner that promotes the food category on the company's app, and as a data analyst I had to determine whether it increases sells and conversions. The results did show a significant higher conversion rate for the treatment group, but the average purchases were almost equal between both groups. With that in mind, and additional calculations and consideration which I'll reveal in this paper, my recommendation is to continue iterating and testing, and I'll about that elaborate later on.

Context

GloBox was looking for a way to increase its income, and one of the solutions explored was adding a banner to the app, a banner that directs the user to the food&beverages category. In order to examine the effectiveness of this banner, the data team decided to randomly assign new users into one of two groups – A-Control, B-Treatment[1] – and see if there is any difference in the users' habits between the groups. The data examined was focused on two main aspects, the percentage of conversions and the amount of money these users spent via the app.

The data set was assembled of three tables, 'activity', 'groups' and 'users', and first of all, each one referred to each user with a unique id. The 'users' table holds each user's gender and country of residence, and 'groups' has each user device (Android or IOS), the date they joined the experiment and of course their group (A/B). 'activity' is a bit different because it has information only about those who actually made a

[1] Appendix N. 1

purchase, and refers to each purchase separately, meaning a user can appear there more than once. It basically holds the date of the purchase, the device type, and the price the user paid.

Data Processing

First of all, I had to play around with the database in order to create some new fields that will help me extract clear insights. I wrote a query that joins the necessary data into one table that includes each user and his user id, country, gender, device, group, whether he converted or not, and the total amount of money he spent. I had to put my mind into a few important things, first, upon joining the tables I had to make sure I include the not converted users with 'LEFT JOIN'. Second, I needed to sum every purchase made by the same user, so I wrote a sub-query which does exactly that. Finally, in order to include the non-spending users in the average spent calculations, I used 'coalesce()' to replace the 'NULL' values with 0[2].

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[3]. 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[4].

Findings

First of all, in terms of conversion rate, there was a clear increase for the treatment group, from 3.92% to 4.63%. If we look by gender, women tend to convert more than

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

men, but were less affected from the banner - they went up from 5.14% to 5.44%, as men' purchases increased from 2.63% to 3.79% - more than a whole percent! In addition, IOS users' conversion rate (5.85% to 6.47%) is significantly higher than android owners' (2.77% to 3.52%). Finally, I won't go too deep into the country division analysis, 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 'Novelty Effects' concern, the visualization clearly shows gradual decrease of conversion per day, but it's true for both groups[5].

Regarding the average amount spent, there was a very slight change, from 3.37$ to 3.39$ per user. If we examine the gender division, women from group B even spent less somehow – 4.46$ to 4.13$, while men again were tempted by the banner and spent 2.60$, oppose to the 2.50$ spent by their fellows in group A. In terms of device, IOS users spend more in general, but went down from 5.07$ to 4.92$, while there was an increase from 2.31$ to 2.47$ among android owners. In addition, country wise, it's easy to see the one with the largest change (GB), the less spending one (Austria), and that a negative change is common (5/10 countries). Regarding 'Novelty Effects' concern, the visualization doesn't reveal any detectable pattern, and again, it's true for both groups. Lastly, a very interesting result is the average spending for only converted users, where there is a huge difference in favor of the control group – 86.46$, compared to the treatment group – 72.78$.

As for my confidence in the results for the conversion rate, first of all, I got a very low 'p-value' - 0.0001 – meaning I can quite calmly reject the null hypothesis that there is no difference between the groups. Second of all, I checked the 'confidence intervals' for both 95% and 99% certainty, and the minimum difference between the

[5] Appendix N. 5

groups' conversion rate is 0.2352% (99%)[6]. Furthermore, I even visualized the CI for each group, to get a deeper view of the clear difference between them. Finally, a power analysis tells me it would have been better to keep the testing on for four more days in order to get enough users (60.6k[7]) for the commonly accepted 0.8 'statistical power'[8].

On the other side of the coin, the amount-spent analysis is really not conclusive. First, the 'p-value' is super high - 0.944 – so I failed to reject the null hypothesis that there is no difference between the groups. Second, the maximum difference in cents is 0.615$ (99%), which in terms of spending is not at all high. Furthermore, I again visualized the CI for each group, and the fact there is a lot of overlap - suggesting there is no much of clear difference - stands out[9]. Finally, a power analysis shows it would be very long and costly to have 0.8 'statistical power', as it will require no less than 38,716,244 users[10].

Recommendations

After reviewing and analyzing the data, I think the right way to go is to continue iterating with a couple of parameters in mind. First, obviously the banner's exposure made an impact in terms of conversion rate, so it's clearly a good idea. However, business wise, there was no significant increase of income per user, and that is the final goal. In my opinion, the huge drop I mentioned in converted users' spending can definitely imply that the banner focuses on the wrong category. The users in the control group weren't tempted to go specifically to the food&beverages, and therefore it's possible the ones who did purchase did so in a more expensive category. On the

[6] Appendix N. 6
[7] Appendix N. 7
[8] Appendix N. 8
[9] Appendix N. 9
[10] Appendix N. 10

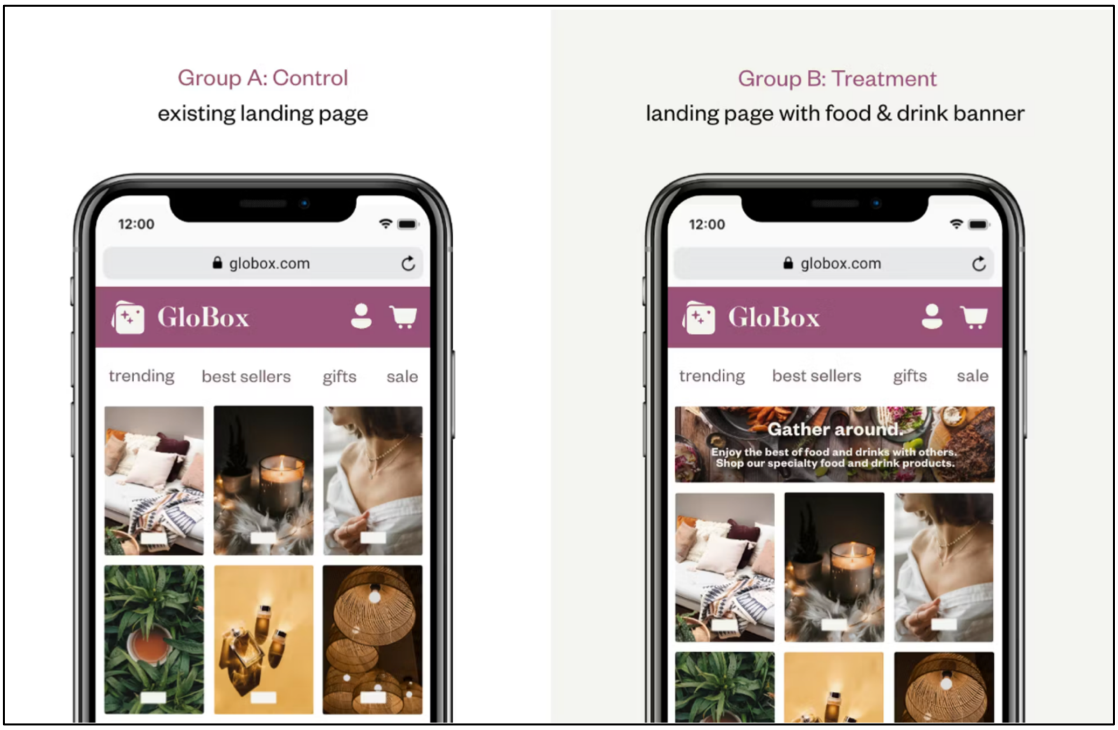
contrary, the users in the control group were indeed drawn to the food&beverages, but their purchases there weren't as expensive, because feed tends to have relatively low costs.

Therefore, I believe the food&beverages category is not as profitable as others, and a combination of an effective banner and a different target category can translate the conversion rate's increase to income going up as well. Although the space on the app is expensive, the banner development is quite cheap and easy, so I think a couple of two-weeks-periods – each focusing on different category – could be a great indicator on how to proceed further.

In addition, IOS users convert more than twice than Android owners (in both groups), so I highly recommend to make sure the Android's interface is user friendly, and has no major issues. Lastly, because women were much less affected by the banner, as I mentioned, I assume it would be wise to try and design the banner differently for the feminine users.

Appendix

1.



2. Related code - 'Main Code.sql'/'Main Code.png', files attached.

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

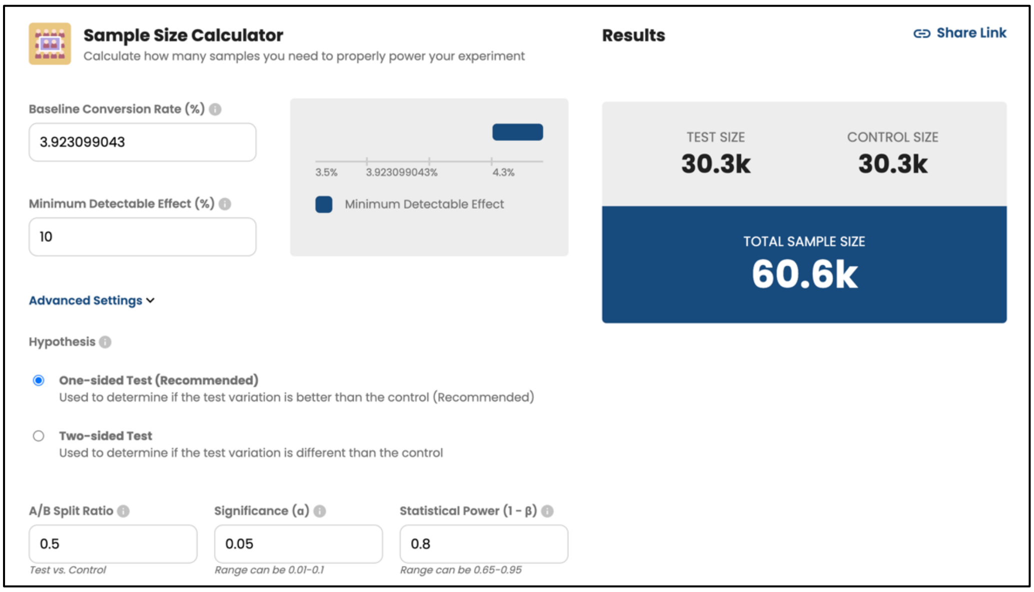
'Cleaned Data.xslx', file attached.

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

5. Related code - 'Add Date.sql'/'Add Date.png', files attached.

6. The data comes from 'Confidence Intervals%.csv', file attached.

7.



https://www.statsig.com/calculator

8. Related code - 'Running Total.sql'/'Running Total.png', files attached.

9. The data comes from 'Confidence Intervals AS.csv', file attached.

10. https://statulator.com/SampleSize/ss2M.html,

'SampleSize-TwoIndependentMeans.pdf', file attached.