A/B test of e-commerce company webpage

By: Manuk Mikayelyan

May 2023

Summary

An A/B test was conducted to determine whether GloBox should launch a food and drink banner on its website to increase revenue for mobile users. The primary metrics used were the conversion rate (CR) and the average revenue per user (ARPU). The analysis showed that the test group had a statistically significant increase in CR. However, the difference in average revenue per user ARPU was not statistically significant between the control and test groups. Therefore, it is recommended that GloBox does not launch the experience to all users, but iterates on the experience with variations to the banner design, placement, and messaging to increase the impact on revenue per user. Further testing can refine the impact of the banner on revenue and inform future decision-making regarding this feature.

Context

GloBox is an online marketplace that specializes in sourcing unique and high-quality products from around the world. GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, companies' food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue. The Growth team decides to run an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the test group sees it as shown below:
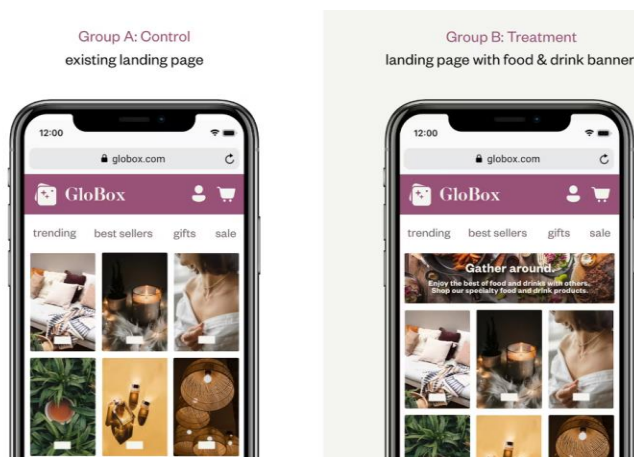
The setup of the A/B test

The setup of the A/B test is as follows:

1. The experiment is only being run on the mobile website.

2. A user visits the GloBox main page and is randomly assigned to either the control or test group. This is the join date for the user.

3. The page loads the banner if the user is assigned to the test group, and does not load the banner if the user is assigned to the control group.

4. The user subsequently may or may not purchase products from the website. It could be on the same day they join the experiment, or days later. If they do make one or more purchases, this is considered a "conversion".

Our task is to analyze the results of the A/B test and provide a recommendation to the stakeholders about whether GloBox should launch the experience to all users. The primary metric we would want to use is the conversion rate (CR), which is the percentage of users who made a purchase after visiting the

website. We would compare the CR for the test group (those who saw the food and drink banner) to the CR for the control group (those who did not see the banner) to determine if the banner had a positive impact on the CR. We would also consider the average revenue per user (ARPU) when making a recommendation to the stakeholders about whether to launch the experience to all users. We would analyze the difference in average spending between the test and control groups to see if there is a significant increase in spending among users who saw the banner. If the test group spends significantly more than the control group, it could be a good indication that launching the experience to all users would be a profitable decision.

The Dataset

GloBox stores its data in a relational <u>database</u>. We will be using the following datasets:

1. users - user demographic information

    - *id* - the user ID

    - *country* - ISO 3166 alpha-3 country code

    - *gender* - the user's gender (M = male, F = female, O = other)

2. groups - user A/B test group assignment

    - *uid* - the user ID

    - *group* - the user's test group

    - *join_dt* - the date the user joined the test (visited the page)

    - *device* - the device the user visited the page on (I = iOS, A = Android)

3. activity - user purchase activity, containing 1 row per day that a user made a purchase

    - *uid* - the user ID

    - *dt* - date of purchase activity

    - *device* - the device type the user purchased on (I = iOS, A = Android)

    - *spent* - the purchase amount in USD

We used SQL to get all the data from the relational database and then download it in CSV format. All the further data preprocessing and analysis have been done in Python (the detailed Analysis Plan and the link to the notebook are attached to the Annex).

After the experiment ran from January 25th to February 6th, the control group consisted of 20,551 users, while the treatment group had 20,736 users, totaling 41,278 users.
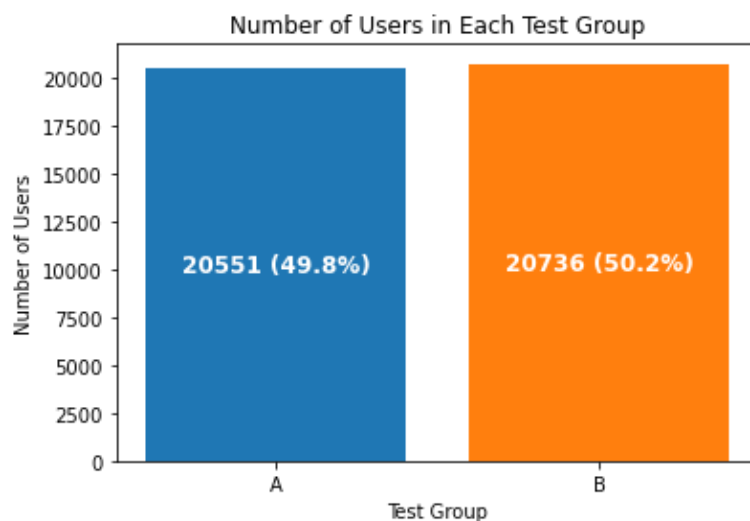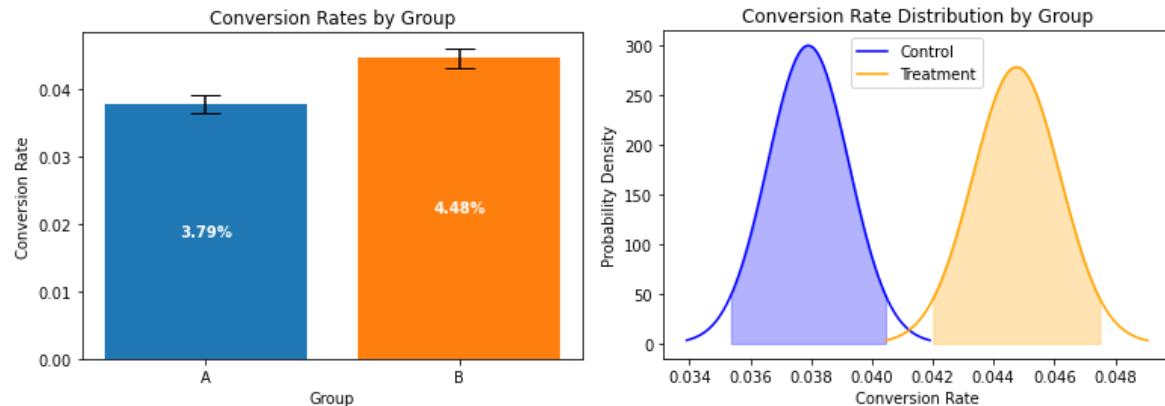
The control and treatment groups were well-balanced in terms of demographics, with the majority of users being from the USA and Brazil and using either iOS or Android devices.
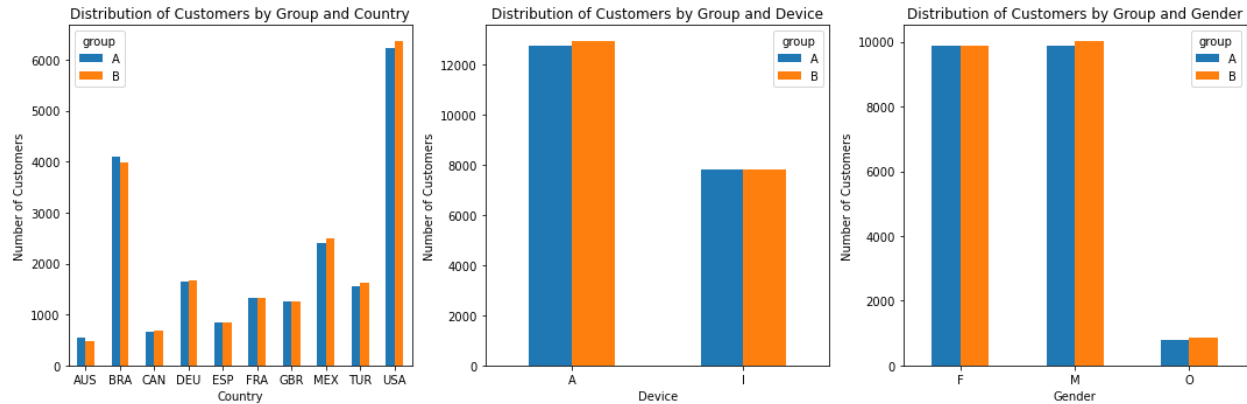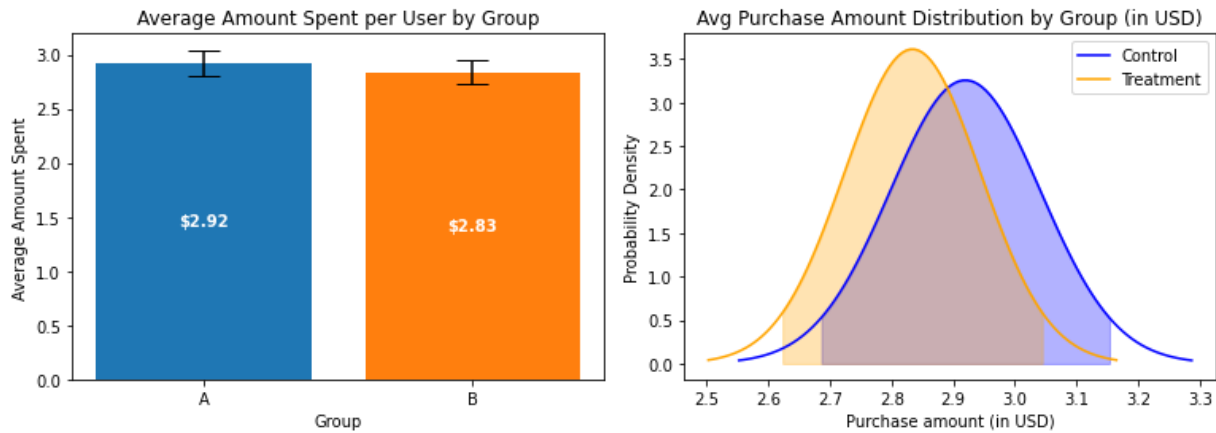

Results

The CR of all users is 4.13%. The CRs for the control and treatment groups are 3.79% and 4.48% respectively.

In order to determine whether there is a statistically significant difference in CR between the control and treatment groups, we ran a hypothesis test. We saw a statistically significant difference between the two groups at the 5% significance level (p=0.005). The 95% confidence interval for the difference in revenue per user between the two groups is (0.0030, 0.0107). Note that the interval does not include 0.

The ARPU for the control and treatment groups are 2.92 and 2.83 respectively.

In order to determine whether there is a statistically significant difference in ARPU between the control and treatment groups, we ran a hypothesis test. We did not see a statistically significant difference between the two groups at the 5% significance level (p=0.602). The 95% confidence interval for the difference in revenue per user between the two groups is (-0.409, 0.237). Note that the interval includes 0.

Recommendation

Given these results, it is recommended that GloBox doesn't launch the experience to all users - while the A/B test showed an increase in conversion rate, the impact on revenue per user is not significant, and there is a possibility that the banner may not lead to a significant increase in revenue.

It is also suggested to iterate on the experience and test it again with variations to the banner design, placement, and messaging to potentially increase the impact on revenue per user. Further testing can help to refine the impact of the banner on revenue and inform future decision-making regarding this feature.

To further improve revenue, we suggest conducting a segmentation analysis to identify specific segments that drove the increase in CR. This analysis can also identify segments where the test did not perform well and may require further testing or optimization. Additionally, we recommend conducting funnel analysis to identify any specific areas of the website that need improvement to increase CR.

Price analysis can also be conducted to identify if there are any specific price points that drive higher revenue. Personalization analysis can help identify the impact of product recommendations or messaging on CR and average spending. By conducting these analyses and tests, we can identify opportunities to increase revenue and average spending without negatively impacting CR.

Appendix

Analysis Plan

Here is a suggested analysis plan for this A/B test:

1. Data cleaning:
   a. Gather data, assess data both manually and programmatically, and store it in CSV format
   b. Check for missing values in all variables and handle them accordingly
   c. Check for duplicates and remove them if necessary
   d. Check for outliers and anomalous values in continuous variables and decide whether to keep them or remove them
   e. Check for inconsistencies and errors in the data, such as invalid dates or incorrect values, and correct them if possible

2. Descriptive analysis
   a. Calculate the number of users in each test group (control vs test)
   b. Check if the groups are balanced in terms of user demographics (country, gender, device)
   c. Calculate the CR for each group, defined as the percentage of users who made a purchase after joining the experiment

      d.   Calculate the ARPU for each group, defined as the total revenue divided by the number of users who made a purchase

3. Hypothesis testing

      a.   Perform a two-sample proportion test to determine if there is a statistically significant difference in CR between the control and test groups, with a significance level of 0.05.

      b.   Calculate the p-value and confidence interval for the test.

      c.   If the p-value is less than the significance level (0.05), we can reject the null hypothesis that there is no difference in CR between the groups, and conclude that the banner had a significant effect on CR.

      d.   Perform a two-sample t-test to determine if there is a statistically significant difference in ARPU between the control and test groups

      e.   Calculate the p-value and confidence interval for the test

      f.   If the p-value is less than the significance level, we can conclude that the banner had a significant effect on revenue as well.

4. Recommendation - based on the results of the analysis, provide a data-driven recommendation to the stakeholders about whether GloBox should launch the banner to all users

Code used for analysis

Attached as a separate file (Project_Code.ipynb)