New Banner A/B Test Analysis

By: The Amazing Analyst at GloBox

Summary

This report aims to analyze the results of the A/B test and provide a recommendation to below stakeholders about whether GloBox should launch the new banner design to all users.

- Growth Product
- Engineering
- Marketing

Context

As an online e-commerce marketplace, we found the food and drink offerings have grown tremendously in the last few months. Management wants to bring awareness to this product category to wider customer bases in order to increase revenue. So the Growth Product 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 mobile website.

This A/B test compares two versions of the website (with/without the banner) to determine which one is better to increase revenue and conversion. The control group (A) does not see the banner, and the test group (B) sees it. By assigning users randomly to either the A or B version, management can determine which version is more effective at achieving the targets.

Below is the A/B test's setup:

1. A user will be assigned to either the control or test group randomly when visiting the GloBox main page. This is the join date for the user.
2. If the user is assigned to the test group, the banner is loaded; otherwise, the user cannot see the banner like those assigned to the control group.
3. Users may or may not purchase products on the day they join the experiment, or days later. When they make one or more purchases, it is considered a "**conversion**".

The following spells out the A/B test results and provides a recommendation whether we should launch the experience to all users. The intended audiences of this report are the following:

- Leila Al-Farsi, Product Manager, Growth: Leila is the product manager for the Growth product and engineering team.
- Alejandro Gonzalez, User Experience Designer, Growth: Alejandro is the designer for the Growth product and engineering team.
- Mei Kim, Head of Marketing: Mei oversees the Marketing team.

Leila, Alejandro, and Mei will decide together whether or not to launch the new banner feature based on the results.

Results

We are going to conduct hypothesis testing comparing the groups (A, B) for the metrics of revenue and conversions.

a) Revenue
   What is the 95% of confidence interval (CI) for the difference in the average revenue per user between the test and the control (test - control)?

   The Null hypothesis is that the mean revenue for the control is the same as the mean revenue for the test.

   H0: mean(test) == mean(control) or mean_diff = 0
   H1: mean(test) != mean(control) or mean_diff != 0

   CI can be calculated with this formula: mean +- critical_value * standard error

   We use Python (see References) for detailed calculations with these results:

   - 95% confidence intervals (-0.438638, 0.471335). In other words, we are 95% confident that the mean difference of A and B revenue is between (-0.438638, 0.471335).
   - p-value 0.9438

- As p-value 0.944 > alpha 0.05, it is not statistically significant and indicates strong evidence for the null hypothesis. We failed to reject the null hypothesis H0 and we rejected the alternative hypothesis H1.

b) Conversion

Is there any difference in conversion proportions between the test and the control? (significance level of α=0.05)

The Null hypothesis is that the conversion proportion for the control (Pa) is the same as the conversion proportion for the test (Pb).

H0 (Null hypothesis) : Pa = Pb
Ha (Alternative hypothesis): Pa <> Pb

- We are interested in testing the hypothesis that the conversion proportions from these two groups are unequal.
- The test statistic, Z, under the assumption of the null hypothesis, is

  Z = Pa_hat - Pb_hat / sqrt(P_hat(1-P_hat) * (1/Na + 1/Nb)) ~Normal(0,1)

  Where
  Pa_hat = proportion A (Xa) / sample size A (Na)
  Pb_hat = proportion B (Xb) / sample size B (Nb)
  P_hat = Xa + Xb / Na + Nb

- P_hat is the pooled sample proportion. Under H0, we assumed this test statistic is normally distributed with mean 0 and variance 1. The test is performed at a significance level of alpha and we will compute a p-value for the test statistic:
  p-value= 2(1 - ø(Z)),
  where ø is the cumulative distribution function for normal distribution.

Again, Python (see References) is used for detailed calculations with these results:

- p-value is 0.00011, because p-value <= alpha 0.05, it is statistically significant. It indicates strong evidence against the null hypothesis, as there is less than 5% probability the null is correct (and the results are random). Therefore, we reject the null hypothesis H0 and accept the alternative hypothesis. However, this does not mean that there is 95% probability that the alternative hypothesis is true.

- It is useful to recall the p-value is conditional upon the null hypothesis being true but is not related to the truth of the alternative hypothesis.

Recommendation

For revenue, we can say that the new banner experiment has no material effect on this metric. On the other hand, we found the experiment had an impact on the conversion metric but the degree was inconclusive.

The new banner does not show any tangible benefits. It takes resources and investment to roll out the experience to all users. On this basis, we recommend not to launch this feature.

References

Scripts to load data into Google Sheets:

```
-- Aggregate total amount spent per each user id
with t1 AS (
  select uid, sum(spent) as total_spent
  from activity
  group by uid
)
-- Match group to user id
select g.uid, g."group" AS gp,
      COALESCE(total_spent, 0) AS total_spent
from groups g
left outer join t1 a
on a.uid = g.uid
```

```
-- Aggregate total amount spent per each user id
with t1 AS (
  select uid, sum(spent) as total_spent
  from activity
  group by uid
)
-- Match group to user id
select g.uid, g."group" AS gp,
      COALESCE(total_spent, 0) AS total_spent
from groups g
```

```
left outer join t1 a
on a.uid = g.uid
-- Extract user id that made purchases (conversion)
where total_spent > 0
```

Link to Sheets

Python A/B Test