GloBox Company

A/B TEST

Suna Jayyousi

Introduction

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, their 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:

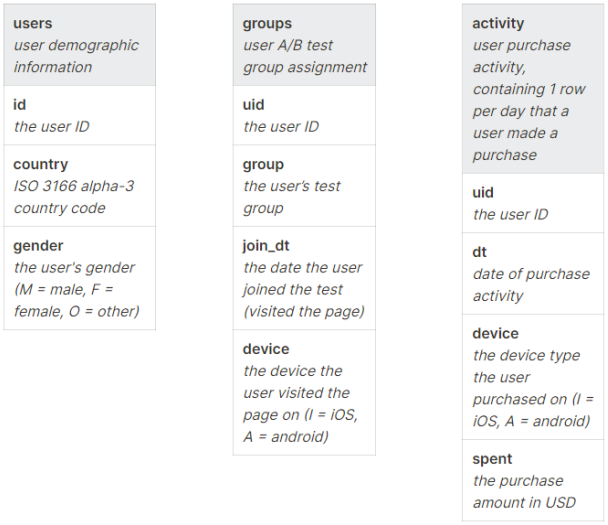
Group A: Control
existing landing page

Group B: Treatment
landing page with food & drink banner

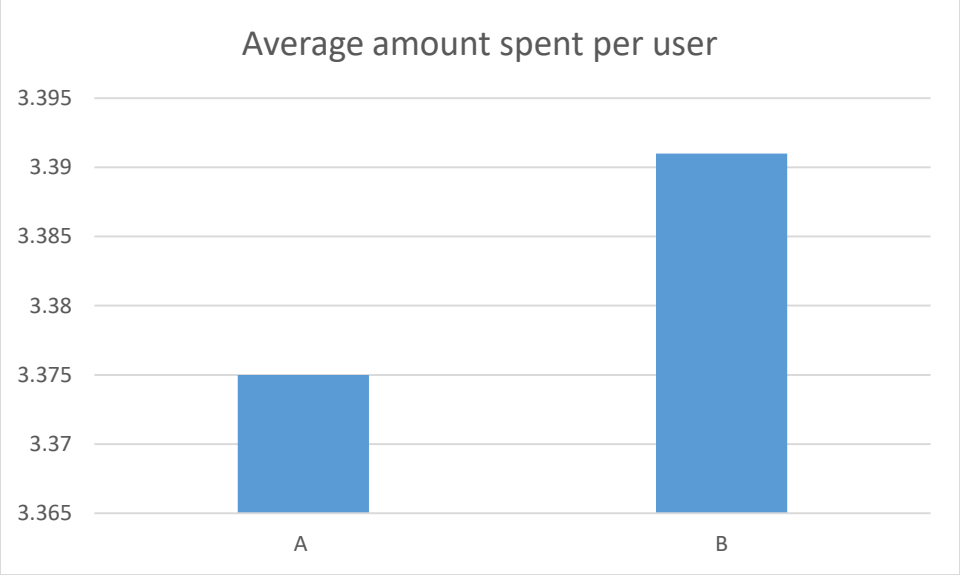
The Dataset

GloBox stores its data in a relational database, which you will access through bit.io.

You can find a description of each table and its columns below.



Average amount spent per user in for control and treatment



I write query in sql to find average amount spent per user for the control group

And 95% confidence interval for the average amount spent per user.

“””select distinct users.id ,sum(activity.spent)

from groups

join users

on groups.uid=users.id

left join activity

on users.id=activity.uid

where groups.group='A'

group by 1”””

then I saved the query result and apply t test by excel

mean	3.374518			level	0.95
std	25.93639			alpha	0.05
sample size	24343			t-score	1.960061
				EBM	0.325831
				mean-EBM	3.048688
				mean+EBM	3.700349

I write query in sql to find average amount spent per user for the treatment group

And 95% confidence interval for the average amount spent per user.

“””select distinct users.id ,sum(activity.spent)

from groups

join users

on groups.uid=users.id

left join activity

on users.id=activity.uid

where groups.group='B'

group by 1”””

Then I saved the query result and calculate the mean and confidence interval

mean	3.390867			level	0.95
std	25.41411			alpha	0.05
sample size	24600			t-score	1.96006
				EBM	0.317597
				mean-EBM	3.07327
				mean+EBM	3.708464

Hypothesis test between control and treatment groups base on average spent per user

1. Control group average spend: 3.375
2. Treatment group average spend: 3.391
3. 95% Confidence Interval for the mean total amount in the Control group: [3.049, 3.700]
4. 95% Confidence Interval for the mean total amount in the Treatment group: [3.073, 3.708]

By using p-value calculator the p-value =0.945

Statistically insignificant. We fail to reject the null hypothesis that there is no difference in the mean amount spent per user between the control and treatment.

So there is NO sufficient evidence to launch/no launch.

Conversion rate for users

I write this query to get conversion rate for control and treatment

“””with cte as(

select uid, sum(spent) total_spent,

case when sum(spent)>0 then 1 else 0 end as conversion

from activity

group by uid)

select g.group as experiment_group,

count(distinct(users.id)) sample_size,

avg(COALESCE(t.conversion,0)) as avg_conversion

FROM users

JOIN groups g ON users.id = g.uid

left JOIN cte t on t.uid=users.id

group by g.group”””

Query Results

2 ROWS

experim TEXT	sample_size INT8	avg_conversion NUMERIC
A	24343	0.03923099042845992688
B	24600	0.04630081300813008130

95% confidence interval for the conversion rate of users in the control and treatment groups

I used excel for calculate it

1-control group

conversion rate	0.0392
sample size	24343

1-conversion rate	0.9608
conversion rate(1-conversion rate)	0.03766336
(conversion rate(1-conversion rate))/sample size	1.54719E-06
standard error of the proportion(sqrt of I26)	0.001243863
Z*standard error	0.0024
mean-(z*standard error)	0.0368
mean+(z*standard error)	0.0416

2-treatment group

conversion rate	0.0463
sample size	24600
1-conversion rate	0.9537
conversion rate(1-conversion rate)	0.04415631
(conversion rate(1-conversion rate))/sample size	1.79497E-06
standard error of the proportion(sqrt of I26)	0.001339766
Z*standard error	0.002626022
mean-(z*standard error)	0.043673978
mean+(z*standard error)	0.048926022

Hypothesis test between control and treatment group

1. Conversion rate for users in control group: 0.0392
2. Conversion rate for users in treatment group: 0.0416
3. 95% confidence interval for the conversion rate of users in the control: [0.0368, 0.0416]
4. 95% confidence interval for the conversion rate of users in treatment: [0.0437, 0.0489]

Then by p-value calculator I got that the p-value would be <0.001, it is statistically significant (p-value<0.05), we reject the null hypothesis that there is no difference in the user conversion rate between control and treatment.

Conclusions:

1- We fail to reject the null hypothesis that there is no difference in the mean amount spent per user between the control and treatment.
2- We reject the null hypothesis that there is no difference in the user conversion rate between control and treatment.

   We have no enough evidence to prove that there is a no difference /difference in average amount spent per user between control and treatment groups

   I suggest to launch the banner of food and drink category.