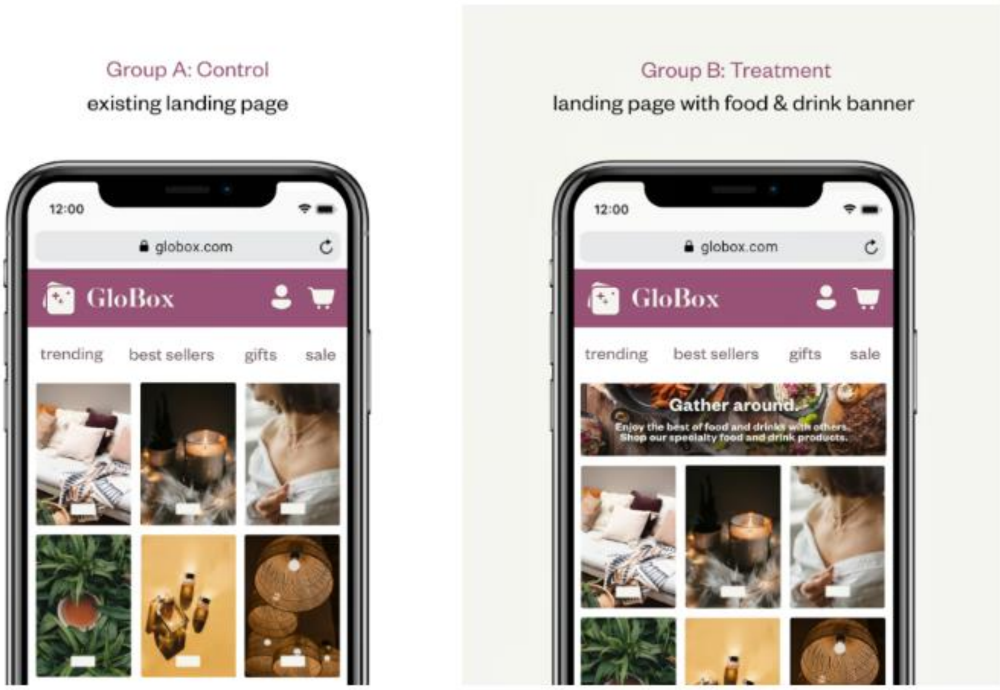
Mastery Project

Globox Homepage

A/B test experiment to improve homepage.

DA106 Mastery project



September 25, 2023

Yuran Belane

Summary

In this present report, we’ll conduct an A/B test overview of the goals and parameters for GloBox - An online marketplace specializing in sourcing unique and high-quality products from around the world.

This experiment has the purpose of improving their homepage, whether to compare two versions of a webpage with product features to determine which one performs better. By randomly assigning customers or users to either the group A (control) – existing landing page or B (treatment) – landing page with food and drinks banner, the Growth team can determine which version is more effective at achieving a particular goal which is increase the revenues.

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 which the control group does not see the banner, and the test group sees it. Although The A/B test was conducted over a period of approximately two weeks from January 25th, 2023, to February 6th, 2023, which is precisely 13 days.

This test will follow setup’s which the experiment is only being run on the mobile website. A user visits the GloBox main page and is randomly assigned to either the control or test group.

Keywords: A/B group; total_spent; users_iud; Converted; gender; device_group

Context

To evaluate the impact of highlighting key products in the food and drink category on the Globox website an A/B test was conducted on the Globox mobile website over a time span of 12 days, from January 26 to February 6. The purpose of the test was to evaluate the impact of a new banner promoting food and drink offerings on user behavior.

The new food and drink banner was proposed by the team and introduced with the expectation that it would enhance user engagement and increase purchases by highlighting the variety of offerings available. While the Users visiting the site during the test period were randomly assigned to the control(A) or treatment group(B). The control group saw the existing landing page and the treatment group saw a streamlined landing page with a banner to highlight the food and drink category. The assignment of users was randomly to ensure that the two groups were comparable in terms of user characteristics as described below.

Control group A: The existing landing page.
Treatment group B: A streamlined landing page with a banner to highlight the food and drink category.

Must be assured that the hypothesis predicts that the treatment group (B) with the streamlined banner will lead to a higher conversion rate compared to the existing process on control group(A). It is expected that the advertisement of the food and drinks category will lead more customers to buy from this category resulting in increased revenue for the company.

Metrics and Analytics

The primary metrics take on for the test were the conversion rate and the average amount spent per user which the dataset used for the analysis includes data on approximately total sample size of 48,943, users who visited the site during the test period. The conversion rate was defined as the proportion of users who made a purchase during their visit or subsequent visits to the site. The average amount spent was calculated as the total amount spent divided by the number of users. Both metrics were calculated for each group separately. These metrics are important as an increase in one or both would be an overall increase in revenue for GloBox.

The data collected was the group they were assigned to, whether they made a purchase or not, the amount they spent, and other user characteristics such as gender, device, and country of residence.

Querying on PostgreSQL

After Querying the dataset which contains Uid – users ID, Gender, Ab_group, Total spent, converted and Country, which are significant piece of information as it offers a summary of the unique users present in the 'users' table with direct relation with a Gender table and with the Ab_group. It becomes a foundational metric for various analytical and decision-making processes within the company.

A/B Test on Excel

After querying the database using Beekeeper Studio, I obtained the values of the Control group which had a conversion rate of 3.923% while the Treatment group had a conversion rate of 4.630%. The average spent in the Control group was $3.37 and $3.39 in the Treatment group.

1. **Difference in conversion rate between the two groups:**

test group	notation	proportion
A	24343	0.0392
B	24600	0.0463
calculation	Notation	value
Sample size (control)	n1	24343
Sample size (treatment)	n2	24600
Sample proportion (control)	p1^	0.0392
Sample proportion (treatment)	p2^	0.0463
pooled proportion	p^	0.0428
Test statistic	T	3.8815
p-value	pval	0.0001

Throughout this hypothesis testing process for a two-sample z test for a difference in proportions since the p-value =0.0001< 0.05, we reject the null hypothesis that conversion rate is the same between the two groups in favor of the alternative that there exists a difference between the groups.

2. **Is the 95% confidence interval for the difference in the conversion rate between the treatment and control (treatment-control)?**

Two-samples z-interval with unpoled proportion

Confidence interval: (0.349%, 1.065%)

calculation	Notation	value
sample size (control)	n1	24343
sample size (treatment)	n2	24600
sample proportion (control)	p1^	0.039
sample proportion (treatment)	p2^	0.046
sample statistic/ point estimate	p^	0.007
standard error	Z	0.002
critical value	moe	1.959963986
margin of error	0.004	
lower bound	0.0034	
upper bound	0.0106	

The confidence interval result lower (0.0035) and upper bond (0.0107) represents the estimated range of values within which we can be 95% confident that the true difference in proportions lies. Also, it suggests that the difference in proportions between the two samples falls between 0.0035 and 0.0107. This range provides us with a measure of the precision and uncertainty associated with our estimate. We can conclude that the proportion in the first group is likely higher than that in the second group. In summary, the confidence interval (0.0035, 0.0107) provides valuable information about the estimated difference in proportions and reinforces the confidence in the observed effect.

3. **hypothesis test to see whether there is a difference in the average amount spent per user between the two groups.**

H0: µ1 = µ2 (The average amount spent per user in the control and treatment groups is equal)

HA: µ1 ≠ µ2 (The average amount spent per user in the control and treatment groups is not equal)

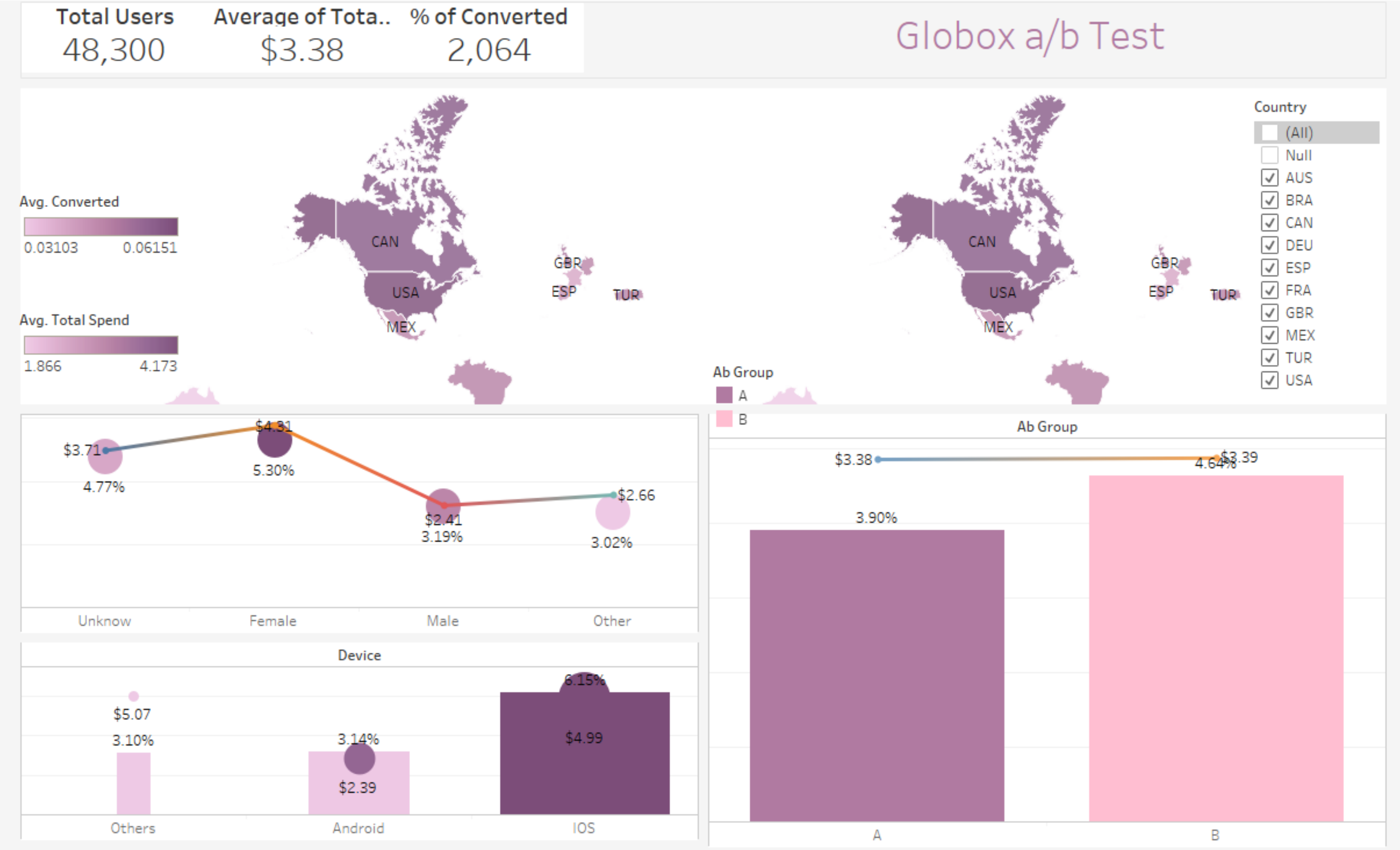
calculation	Notation	value
Sample Size (Control)	n1	24343
Sample Size (Treatment)	n2	24600
Avarage Spent (Control)	p1^	3.37
Avarage Spent (Treatment)	p2^	3.39
Standard Deviation (Control)	sdC	25.94
Standard Deviation (Treatment)	sdT	25.41
Standard Error	sE	0.2321405588
Test Statistic	T	0.07042491002
H0 Mean	H0 X	0
Degree Of Freedom	dF	24342
P-Value	pVal	0.9438560437
	α = 0.05.	

Throughout this hypothesis testing process Since this P-value is greater than α we fail to reject the null hypothesis. This means that there is not a statistically significant difference in the average amount spent between the two groups.

4. **What is the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control)?**

calculation	Notation	value
Sample size (control)	n1	24343
Sample size (treatment)	n2	24600
Sample proportion (control)	p1^	0.039
Sample proportion (treatment)	p2^	0.046
standard deviation (control)		25.94
standard deviation (treatment)		25.41
standard error		0.2321405588
sample statistic	stat	0.007
degree of freedom	dF	24342
test statistic	T	0.0301541447
Critical value	t	2.241541405
margin of error		0.520
Lower bound	-0.513	
Upper bound	0.527	

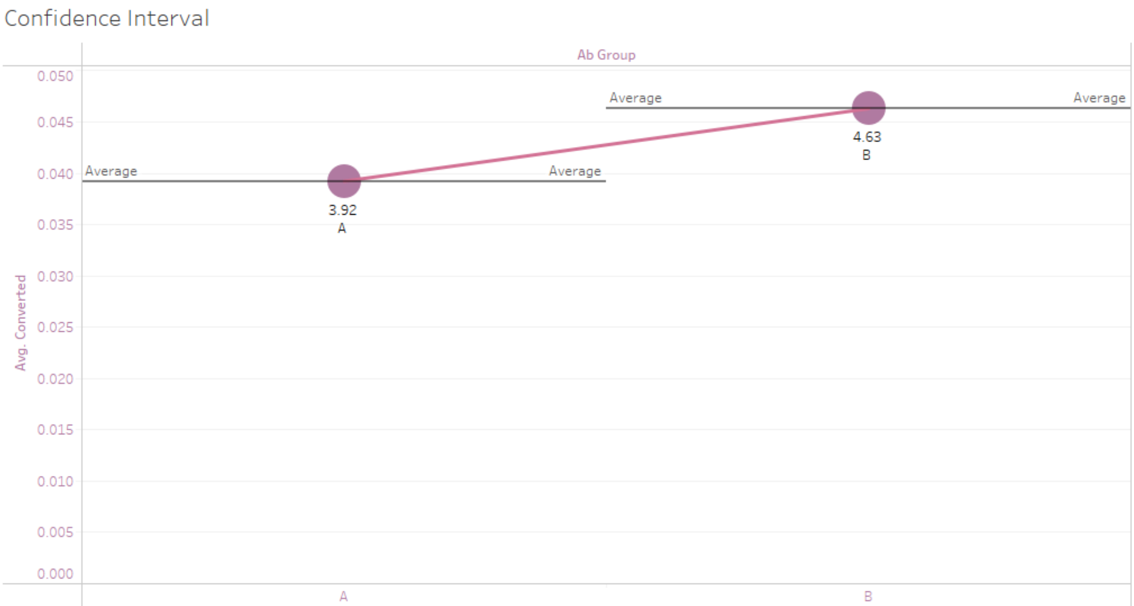
Considerations of Visualization Tableau



1. Conversion rate and average amount spent showed some differences by country, with Australia having the lowest and the United States and Canada having the highest metrics.
2. Among users, females have a higher conversion rate and average amount spent compared to male, unknown, or other users. Same happens to other tests metrics align with the performance of female users.
3. IOS users have a higher conversion rate and spend more money on average than Android or Unknown device users, and the test metrics align with IOS users' performance.
4. The treatment group outperformed the control group in both conversion rate and average amount spent, attracting users in higher spending brackets.
5. The treatment group had a larger number of users in lower spending brackets but had fewer users in higher spending brackets compared to the control group.

The website banner used in the Treatment Group was more effective in encouraging visitors to convert, resulting in a slightly higher average amount spent compared to the Control Group.

Confidence Interval



When interpreting confidence intervals, it's crucial to consider that the intervals are associated with a level of confidence (in this case, 95%). Higher confidence levels, such as 99%, will result in wider intervals, while lower confidence levels, such as 90%, will yield narrower intervals. The larger the sample size and the smaller the variability in the data, the narrower the confidence interval will be, providing more precise estimates of the population parameter.

Power analysis

A power analysis helps us understand the necessary sample size in order to achieve our desired minimum detectable effect and statistical power. The required sample size to achieve a specific statistical power and level of significance is 76,900 samples, which are more than our test sample. This calculated required sample size is necessary to attain the desired statistical power of 80% and a level of significance of 5% ($\alpha = 0.05$) for detecting a minimum detectable effect of 10% in the conversion rate.

- Baseline conversion rate (p1) = 3.92% = 0.0392

- Minimum detectable effect (d) = 5% of the baseline conversion rate = 0.05 * 0.0392 = 0.00196
- Desired significance level (α) = 0.05Desired statistical power ($1 - \beta$) = 0.8

The practical significance of the study's findings might be limited due to the smaller sample size. Although the analysis might have detected statistically significant results, the small sample size could raise concerns about the real-world impact and meaningfulness of the observed effect.

Recommendation

In this analysis we found that there was a statistically significant difference in conversion rate between the control and test group, suggesting that the new banner had a positive impact on user behavior which is great.

While there was no significant difference in the average amount spent per user between the two groups, the increase in the number of users making a purchase in the test group implies a potential for increased revenue.

Based on these findings, we would recommend the banner be launched for all users as there appears to be an overall positive effect on conversions and by extension, revenue. Even though the sample size is smaller than the ideal calculated size there was an immediate impact on the conversion rate that persisted during the test. If time and resources permit, it would be beneficial to conduct further testing with a larger sample size to confirm these findings.

Additionally, our analysis revealed a higher conversion rate among female users compared to male and other gender users. Understanding the factors behind this difference could provide valuable insights for improving targeting strategies and further increasing conversion rates. Future research could focus on exploring these gender differences in more detail.

Appendix I: SQL queries

--Can a user show up more than once in the activity table? Yes or no, and why?

-- Yes - they can make multiple purchases

```
select  uid,
  count(*)
from    activity
group by uid
order by count desc;
```

--**What type of join should we use to join the users table to the activity table?**

--LEFT JOIN

--**What SQL function can we use to fill in NULL values?**

--COALESCE()

--**What are the start and end dates of the experiment?**

-- Start: 2023-01-25. End: 2023-02-06

```
select min(join_dt) as min, max(join_dt) as max from groups
```

--**How many total users were in the experiment?**

--48943

```
select count(distinct uid) from groups;
```

--How many users were in the control and treatment groups?

```
select "group", count(*) from groups group by "group";
```

--**What was the conversion rate of all users?**

```
with cte_users_with_purchases as
(
 select
        u.id as uid,
        a.dt as dt,
        coalesce(a.spent, 0.0) as spent
        from
        users u
        left join
        activity a
        on
        u.id = a.uid
 ),
 cte_user_purchases_agg as
 (
  select
   uid,
   sum(spent) as total_spend,
   case when sum(spent) > 0 then 1 else 0 end as converted
  from cte_users_with_purchases
```

```
 group by uid
 )
 select round(avg(converted)*100,2) from cte_user_purchases_agg
```

--**What is the user conversion rate for the control and treatment groups?**

```
with cte_users_with_purchases as
(
 select
        u.id as uid,
        a.dt as dt,
        coalesce(a.spent, 0.0) as spent,
        g.group as ab_group
        from
        users u
        left join
        activity a
        on
        u.id = a.uid
        left join
        groups g
 on
        u.id = g.uid
 ),
```

```
cte_user_purchases_agg as
(
  select
   uid,
        ab_group,
   sum(spent) as total_spend,
   case when sum(spent) > 0 then 1 else 0 end as converted
 from cte_users_with_purchases
 group by uid, ab_group
)
select
        ab_group,
 round(avg(converted)*100,2)
from
        cte_user_purchases_agg
group by
        ab_group
```

--What is the average amount spent per user for the control and treatment groups, including users who did not convert?

```
with cte_users_with_purchases as
(
 select
        u.id as uid,
```

```
        a.dt as dt,
        coalesce(a.spent, 0.0) as spent,
        g.group as ab_group
        from
        users u
        left join
        activity a
        on
        u.id = a.uid
        left join
        groups g
 on
        u.id = g.uid
),
cte_user_purchases_agg as
(
  select
   uid,
        ab_group,
   sum(spent) as total_spend,
   case when sum(spent) > 0 then 1 else 0 end as converted
 from cte_users_with_purchases
 group by uid, ab_group
```

```
)
select
	ab_group,
  round(avg(converted)*100,2) as conversion_rate,
  round(avg(total_spend), 2) as avg_spend
from
	cte_user_purchases_agg
group by
	ab_group
```

--Why does it matter to include users who did not convert when calculating the average amount spent per user?

-- Because these users matter too for our business