A/B Test Analysis on New Homepage

BY: Sandhya Bamaniya

Data Analyst

Summary:

I recommend that we should launch the new homepage banner to all users because we observed strong evidence that there is higher conversion rate in the treatment group. This will help in increasing the revenue. It will have a significant impact in the increase of revenue per user in upcoming times.

Context:

This report represents the results of A/B test conducted for GloBox. An A/B test was performed with an updated landing page layout to see whether it would enhance revenue. The difference between the two designs is shown here.
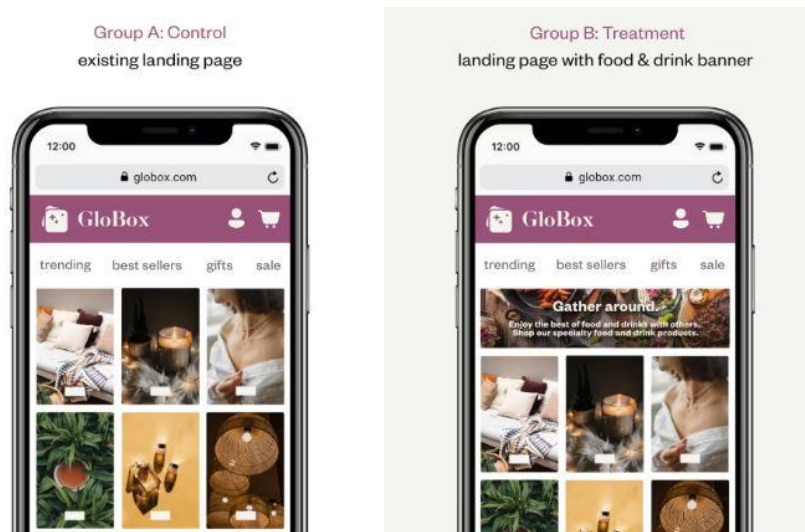
The existing layout has been presented in the control group, whereas the updated layout with the food and drinks banner has been introduced in the treatment group.

A randomised controlled trial design was used for the A/B test. Random assignment ensured that the treatment and control groups were comparable. Data on the selected metrics were collected during a specified duration for subsequent analysis.

The experiment continued for 13 days in Q1 2023. There were 24343 control users and 24600 treatment users, having a total number of 48943 users.

Metrics:

This A/B test is mainly focused on two primary metrics to evaluate treatment's effectiveness.

a. Revenue per user: This metric measures the average revenue generated per user.
b. Conversion Rate: This metric calculates the percentage of users who converted or completed a desired action, such as making a purchase.
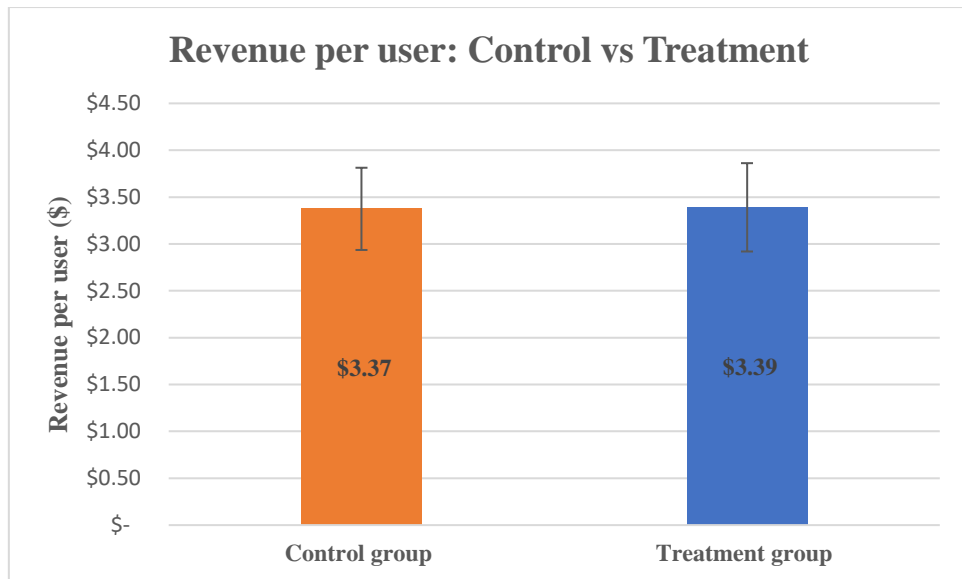
Results:

a. Revenue per user:
Statistical Analysis: A hypothesis test was conducted to see if there was a difference in revenue per user between the two groups. At the 5% significance level, there was statistically insignificant difference between the two groups [**p=0.944**]. The 95% confidence interval for the difference in revenue per user between two groups is (-0.439, 0.471).
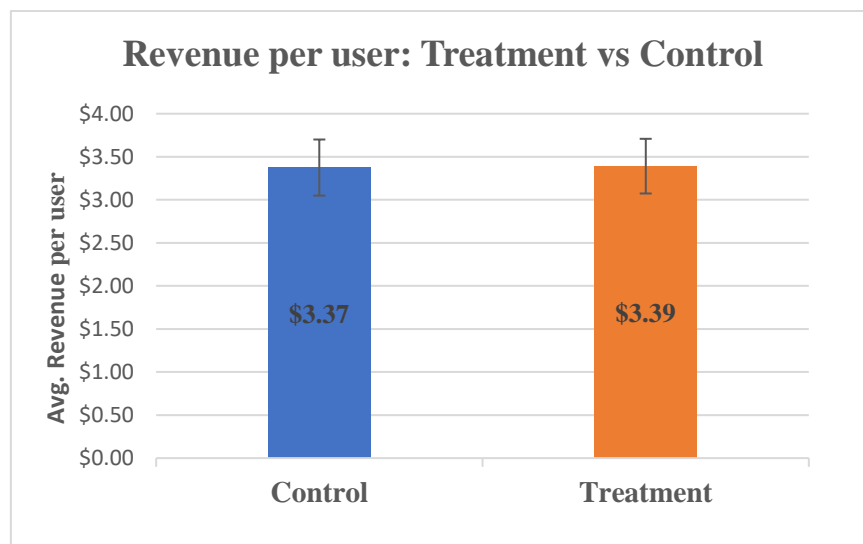
b. Conversion rate:
Statistical Analysis: A hypothesis test was conducted to see if there was a difference in conversion rate between the two groups. At the 5% significance level, there was statistically significant difference between the two groups [**p=0.0001**]. The 95% confidence interval for the difference in revenue per user between two groups is (0.0035, -0.0107).

Revenue per user: Control vs Treatment

Recommendation:

Based on the results above, we should launch the new homepage banner as the conversion rate shows statistical significant difference in the treatment group. The new banner will raise awareness of the new product category in GloBox which will help to increase the revenue. However, we also noted that there is statistically insignificant difference in the revenue per user between control and treatment group. On the contrary if it was an enhancement in the existing category such as boutique product then it would have been a different scenario and since we are dealing with a new product category awareness to the users is a must.

Appendix:

```sql
--1. How many users in the control group were in Canada?
SELECT COUNT(*)
FROM groups AS g
JOIN users AS u
ON g.uid = u.id
WHERE u.country = 'CAN' AND g.group = 'A';

--2. What was the conversion rate of all users?
WITH A AS (
    SELECT COUNT(DISTINCT uid) :: NUMERIC AS c1
FROM activity),
U AS (
    SELECT COUNT(DISTINCT id) :: NUMERIC AS c2
```

```sql
    FROM users
)
SELECT ROUND((c1/c2)*100, 2)
FROM A, U;

--3. As of February 1st, 2023, how many users were in the A/B test?
SELECT COUNT(*)
FROM groups AS g
JOIN users AS u
ON g.uid = u.id
WHERE join_dt <= '2023-02-01';

/*1. What is the average amount spent per user for the control and treatment
groups?
This question is required.*/
SELECT g.group, SUM(a.spent) / COUNT(DISTINCT u.id) AS avg_amt
FROM users AS u
LEFT JOIN activity AS a
ON a.uid = u.id
JOIN groups AS g
ON g.uid = u.id
GROUP BY g.group;

/*2. What is the 95% confidence interval for the average amount spent per user
in the control?
This question is required.*/
SELECT (SUM(spent)/COUNT(DISTINCT u.id)) - (1.960061 *
(STDDEV(COALESCE(spent,0))/SQRT(COUNT(DISTINCT u.id)))) AS lower_bound,
(SUM(spent)/COUNT(DISTINCT u.id)) + (1.960061 *
(STDDEV(COALESCE(spent,0))/SQRT(COUNT(DISTINCT u.id)))) AS upper_bound
FROM users AS u
JOIN groups AS g
ON u.id = g.uid
LEFT JOIN activity AS a
ON a.uid = g.uid
WHERE g.group = 'A';

/*3. What is the 95% confidence interval for the average amount spent per user
in the treatment?
This question is required.*/
SELECT (SUM(spent)/COUNT(DISTINCT u.id)) - (1.960061 *
(STDDEV(COALESCE(spent,0))/SQRT(COUNT(DISTINCT u.id)))) AS lower_bound,
(SUM(spent)/COUNT(DISTINCT u.id)) + (1.960061 *
(STDDEV(COALESCE(spent,0))/SQRT(COUNT(DISTINCT u.id)))) AS upper_bound
FROM users AS u
JOIN groups AS g
ON u.id = g.uid
LEFT JOIN activity AS a
```

```sql
ON a.uid = g.uid
WHERE g.group = 'B';

/*4. Conduct a hypothesis test to see whether there is a difference in the
average amount spent per user between the two groups. What are the resulting
p-value and conclusion?
This question is required.*/



--6. What is the user conversion rate for the control and treatment groups?
WITH conversion AS (
    SELECT g.group, COUNT(DISTINCT a.uid)/ COUNT(DISTINCT u.id) :: NUMERIC AS
conversion_rate
FROM users AS u
JOIN groups AS g
ON g.uid = u.id
LEFT JOIN activity AS a
ON a.uid = g.uid
GROUP BY g.group)
SELECT ROUND(conversion_rate*100, 2)
FROM conversion;

--7. What is the 95% confidence interval for the conversion rate of users in
the control?
SELECT COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC AS
conversion_rate,
    (1.96 * sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) /COUNT(DISTINCT
u.id))) AS margin_of_error,
    COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC - (1.96 *
sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) / COUNT(DISTINCT
u.id))) AS lower_bound,
   COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC + (1.96 *
sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) / COUNT(DISTINCT
u.id))) AS upper_bound
FROM users AS u
JOIN groups AS g
ON g.uid = u.id
LEFT JOIN activity AS a
ON a.uid = g.uid
WHERE g.group = 'A';

--8. What is the 95% confidence interval for the conversion rate of users in
the treatment?
SELECT COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC AS
conversion_rate,
```

```sql
    (1.96 * sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) /COUNT(DISTINCT
u.id))) AS margin_of_error,
    COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC - (1.96 *
sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) / COUNT(DISTINCT
u.id))) AS lower_bound,
    COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC + (1.96 *
sqrt((COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC * (1 -
COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id) :: NUMERIC)) / COUNT(DISTINCT
u.id))) AS upper_bound
FROM users AS u
JOIN groups AS g
ON g.uid = u.id
LEFT JOIN activity AS a
ON a.uid = g.uid
WHERE g.group = 'B';
```

The excel file attached below has calculations and visualisations.

Mastery Project
1.xlsx