

New Landing Page A/B Test Analysis

By: Asenogguan Eric

Date: 04/08/2023

Summary

We ran this A/B Test for two(2) weeks, and we had 48943 mobile users divided into two groups, control group with 24343 users and treatment group with 24600, the GloBox database used had three(3) tables which are users table with three columns(id,country,gender), groups table had four columns(uid,group,join_dt which is date user join the experiment,device IOS/Android), activity table had four columns(uid,dt purchase date, device, spent which is the purchase amount in USD).

Our key metrics were Conversion Rate and Average Total Spent, testing these metrics on both groups we saw statistically significant evidence that showed that there was no difference between both groups comparing Avg Amount Spent.

While comparing the Conversion Rate between both groups we found that there was a significant difference between the control and treatment group.

After carefully analyzing the novelty effect of the difference of both groups key metric and also conducting a power analysis we recommend running this experiment again on a larger scale.

Context

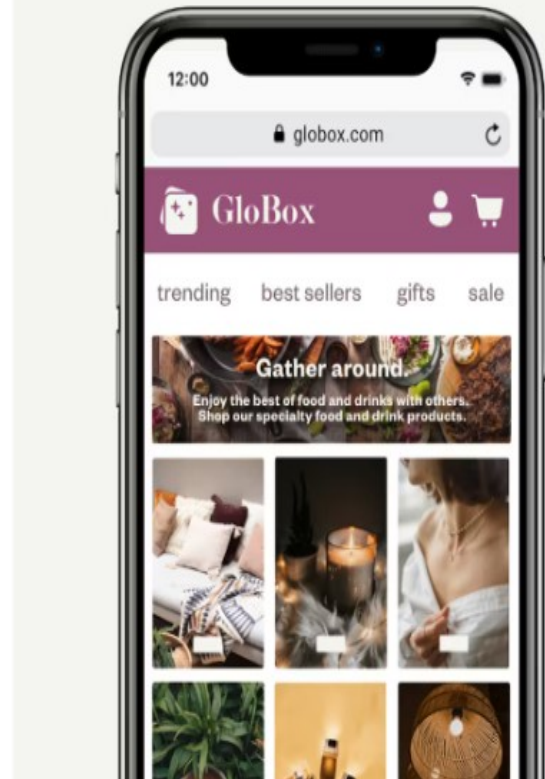
This experiment was conducted with the hope of seeing an increase in conversion rate and average total spent from the treatment group,

Those in the control(Group A) saw the existing landing page, while those in the treatment(Group B) saw the new landing page with food and drink banner

Group A: Control
existing landing page



Group B: Treatment
landing page with food & drink banner



Results

Conversion Rate

We wanted to be sure that there was a difference in conversion rate between the two groups, we ran a hypothesis test, we observed a statistically significant difference between the two groups at the 5% significance level [$P = 0.000101$] and the 95% confidence interval for the difference in conversion rate per users between the two groups is (0.0035, 0.0105), The control group had 3.9% conversion rate and the treatment group had 4.6% conversion rate, which we also visualized with Tableau.

A/B TESTING (Hypothesis Test for Conversion Rate)

Null Hypothesis	
Written	The conversion rate for Group A and Group B are equal
Notation	$H_0 : P_A = P_B$

Alternative Hypothesis	
Written	The conversion rate for Group A and Group B Are not equal
Notation	$H_1 : P_A \neq P_B$

Decision	Reject Null Hypothesis
Conclusion	With $[P < 0.05]$ we reject the null hypothesis that the conversion rates are equal between both groups, in favor of the alternative hypothesis that the conversion rates are not equal between both groups.

Formulas

P-Value = $2 * (1 - \text{NORM.S.DIST}(\text{ABS}(\text{Z-TEST})))$

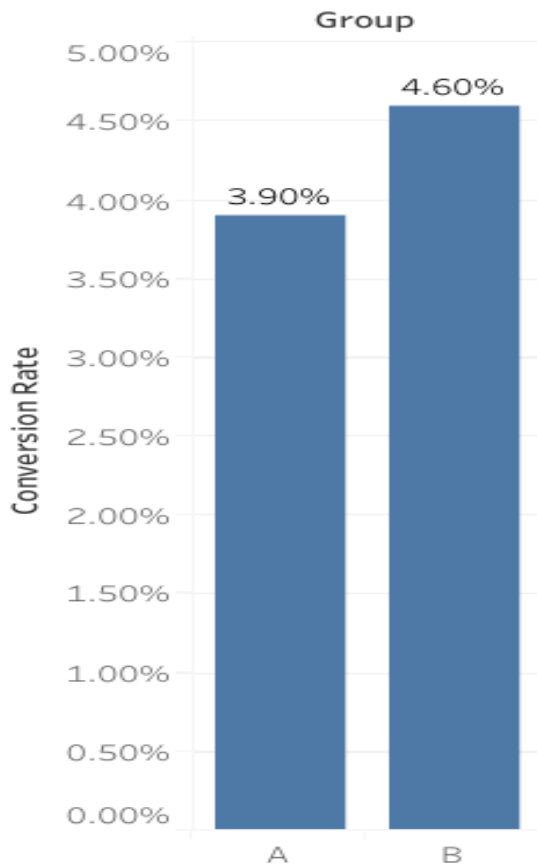
Z-TEST = Point Estimate (P.E) for conversion rate / Standard Error(S.E) for conversion rate

95% CONFIDENCE INTERVAL:

LOWER BOUND = Point Estimate for conversion rate - Margin of Error

UPPER BOUND = Point Estimate for conversion rate + Margin of Error

Conversation Rate



Average Total Spent

We also tested the difference for the Average Total Spent between both groups by running a hypothesis test, we observed that there was no statistically significant difference between the two groups at 5% significance level [$p = 0.93$]. The 95% confidence interval for the difference in Average Total Spent per users between the two groups is $(-0.43, 0.47)$, the control group had 3.35 as the avg.total spent and the treatment group had 3.37 as the avg.total spent, we also visualized this metric with tableau.

Null Hypothesis	
Written	There is no difference in the average amount spent by users between group A and Group B.
Notation	$H_0 : X_A = X_B$

Alternative Hypothesis	
Written	There is a difference in the average amount spent by users between group A and Group B
Notation	$H_1: X_A \neq X_B$

Decision	Fail To Reject Null Hypothesis
Conclusion	With $[P > 0.05]$ we fail to reject the null hypothesis that there is no difference in the average amount between both groups, in favor of the alternative hypothesis that there is a difference in average amount spent by users between both groups.

Formulas

P-Value = (T.DIST.2T(T-TEST - DF)

DF= Degree Of Freedom = $(N_A - 1) + (N_B - 1)$

N = sample size

A = Group A

B = Group B

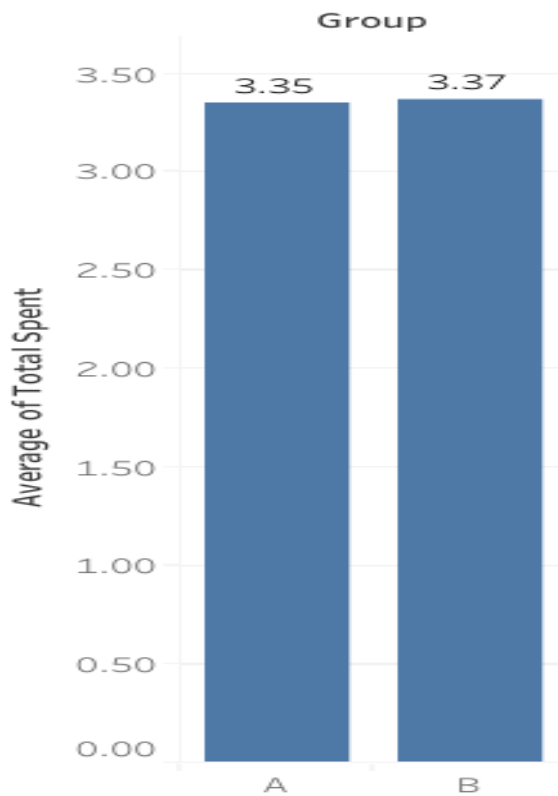
T-TEST = Point Estimate for avg.total spent/ Standard Error for avg.total spent

95% CONFIDENCE INTERVAL

LOWER BOUND = Point Estimate for avg.total spent - Margin of Error

UPPER BOUND = Point Estimate for avg.total spent + Margin of Error

Avg Total spent



Novelty Effect

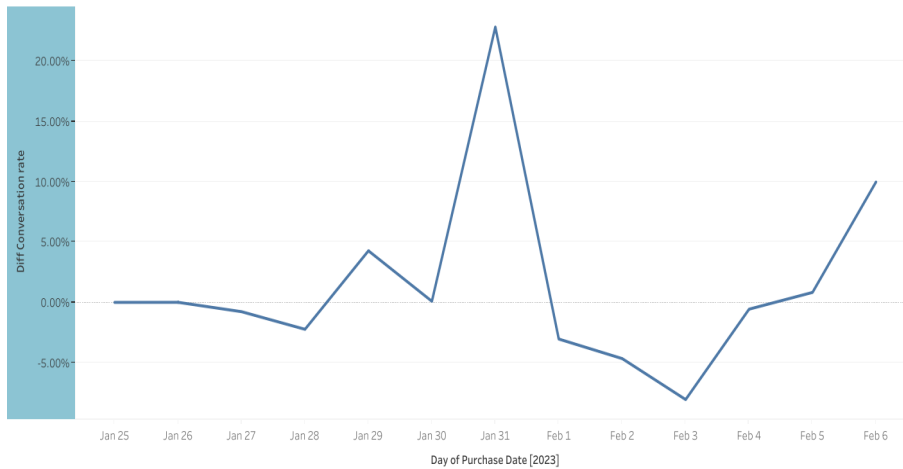
Novelty effect is a temporary change in users behavior that occurs when users encounter something new, users may present with interest, excitement or curiosity when presented with a new experience, as people are naturally drawn to new and unusual things.

However the novelty effect tends to diminish over time as users become more familiar with the new product.

In these cases differences in metrics arise from long-term impact.

Novelty Effect for Conversion Rate

Difference in Conservation Rate Novelty Effect



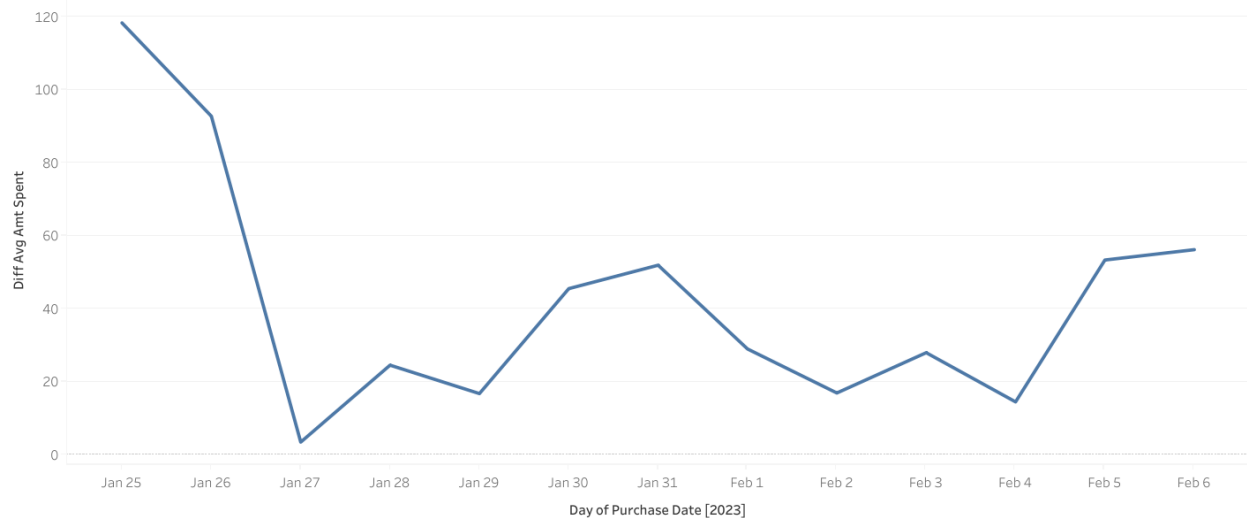
From the above chart we visualized the difference in conversion rate for both groups and I will say that in the first two days conversion rate was the same for both, meaning that users in the groups curiously explored both landing pages.

Also on January 31st when the difference in conversion rate was at its highest because users in group B converted most

We can also say that from February 1st to February 3rd users in group B had a higher conversion rate.

Novelty Effect for Average Total Amount Spent

Difference in avg amt novelty effect



From this chart above (novelty effect for difference in Avg, Total Amt spend) users from both groups were really curious to explore the landing pages reason for the high spending and also high difference between both groups, with users from group B having the highest avg, total spent.

We can also notice a sharp decrease in the difference of the avg, total spent on the third here we can say there was a decrease in the novelty effect

Power Analysis

Power analysis also known as sample size calculation or statistical power analysis is a statistical method used in research to determine the minimum sample size required to draw meaningful conclusions

Power Analysis For Conversation Rate

The screenshot shows the Statsig A/B Test Sample Size Calculator interface. The browser tabs include 'Advanced Tasks', 'A/B Test Sample Size Calculator', 'Google Translate', and 'Copy of Globex Hypothesis Test'. The URL is 'statsig.com/calculator'. The page features a navigation bar with 'Platform', 'Developers', 'Resources', and 'Pricing', along with 'Sign In' and 'Book a Live Demo' buttons.

The calculator interface includes the following fields and results:

- Baseline Conversion Rate (%):** 3.9
- Minimum Detectable Effect (%):** 10
- Advanced Settings:**
 - Hypothesis:** Two-sided Test (Selected). One-sided Test (Recommended) is also visible.
 - A/B Split Ratio:** 0.5
 - Significance (α):** 0.05
 - Statistical Power ($1 - \beta$):** 0.8
- Results:**
 - TEST SIZE: 38.7k
 - CONTROL SIZE: 38.7k
 - TOTAL SAMPLE SIZE: 77.4k

Conducting a power analysis to see what should have been our minimum sample size Using our preexisting or expected conversion rate for our control group as the Baseline conversion rate and minimum detected effect 10%, Our minimum sample size should be 77400 users against 48943 users used in our experiment.

Power Analysis For Mean

Statulator Sample Size Calculator interface showing input values and results.

Input Values

Select one of the two options to specify input values. Hover over the [?](#) sign to obtain help.

☐ Expected Means

☒ Expected Difference between Means

Difference between Two Means: 0.336

Expected Standard Deviation: 25.38

Click the Options button to change the default options for Power, Significance, Alternate Hypothesis and Group Sizes. Use the Adjust button to adjust sample sizes for t-distribution (option applied by default), and clustering.

[Calculate](#) [Options](#) [Adjust](#) [Reset](#)

Results and Live Interpretation

Assuming a pooled standard deviation of 25.38 units, the study would require a sample size of:

89566

for each group (i.e. a total sample size of 179132, assuming equal group sizes), to achieve a power of 80% and a level of significance of 5% (two sided), for detecting a true difference in means between the test and the reference group of 0.336 units.

In other words, if you select a random sample of 89566 from each population, and determine that the difference in the two means is 0.336 units, and the pooled standard deviation is 25.38 units, you would have 80% power to declare that the two groups have significantly different means, i.e. a two sided p-value of less than 0.05.

Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Comparing Two Independent Means. Accessed 3 August 2023 at <http://statulator.com/SampleSize/ss2M.html>

Note: Statulator used the input values of a power of 80%, a two sided level of significance of 5% and equal group sizes for sample size calculation and adjusted the sample size for t-distribution. You may change the options by clicking [here](#) or the 'Options' button and the adjustments by clicking [here](#) or the 'Adjust' button.

© Statulator 2023 All Rights Reserved

[About](#) [Blog](#) [Contact us](#) [Privacy](#) [Terms](#)

Having entered the difference in our mean value and the pooled value for our standard deviation in our sample size calculator, we got 89566 users for each group making a total of 179132 users for our experiment

Recommendation

Due to the information we received from our power analysis, we recommend that this experiment should be run again and on a much larger scale

Appendix

SQL Code For GloBox(used to join and extract the needed columns)

```
create view globox_table_1 AS
select u.id,u.country,u.gender,g.device,g.group,a.spent
from users u
inner join `groups` g
on u.id = g.uid
left join activity a
on g.uid = a.uid;
```

```
create view globox_ease_table as
with cte_table AS(
```

```

select id,case
      when country = '' then 'NA'
      else country
end as country,
case
when gender = '' then 'NA'
else gender
end as gender,
case
      when device = '' then 'NA'
      else device
      end as device,`group`,coalesce(spent,0)as spent
from globox_table_1 gt)
select id,country,gender,device,`group`,sum(spent) as total_spent,case
      when (sum(spent)>0) then 1
      else 0 end as converted
from cte_table
group by id,country,gender,device,`group`;

```

SQL Code GloBox Table For Novelty Effect

```

create view date AS
select u.id, g.group,g.join_dt AS join_date,a.spent, a.dt AS purchase_date
from users u
inner join `groups` g
on u.id = g.uid
left join activity a
on g.uid = a.uid;
with test as
(select id, `group`, join_date ,coalesce(spent,0)as spent,
coalesce(purchase_date,0)as purchase_date
from date)
select id,`group`,join_date,sum(spent) as total_spent,case
      when (sum(spent)>0) then 1
      else 0 end as converted, purchase_date
from test
group by id,`group`,join_date,purchase_date
order by join_date

```

Tableau Link (Basic Metric Visualization)

<https://public.tableau.com/authoring/ABTestTableauVisualisation/Story1#1>

Tableau Link For Novelty Effect

<https://public.tableau.com/authoring/GloboxNoveltyEffect/Sheet2#1>

Tableau Link For Confidence Interval

<https://public.tableau.com/authoring/95C0nfidenceInterval/95ConfidenceInterval-AvgAmtSpent#1>

Spreadsheet Link for key metric and confidence interval

<https://docs.google.com/spreadsheets/d/169-vxE4mJ5-l4Qn6xLRIF7ppOYSegJMv2VSLs1o7k5o/edit#gid=994343830>

Spreadsheet links for Novelty Effect

<https://docs.google.com/spreadsheets/d/1IUj8QvJWNITHUAIZbz-FryA-mucrrK69NY6g-c4XzDo/edit#gid=125610113>