New Banner for GolBox

A/B Test Analysis

By

Yuwani Wijerathne

Table of Contents

List of Tables

List of Figures

1. Project Background

1.1 Introduction

GloBox is well known among its customer base for its luxury home goods and designer clothes. However, the food and beverage section have significantly increased over the past few months, so our management wants to increase awareness of this product category in order to increase sales.

The Growth team decided to conduct an A/B test for this banner which at the top of the website that promotes important products in the food and beverage category. The treatment group (test group) has seen the banner while the control group did not.

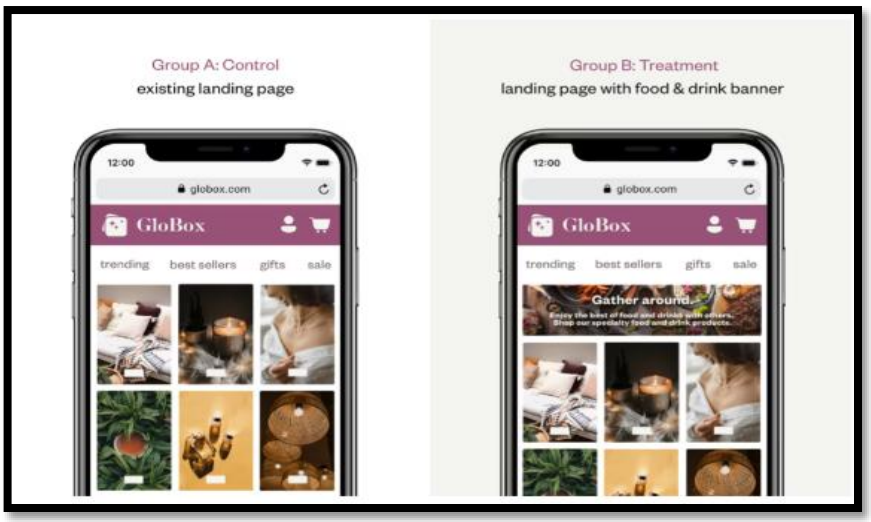

Figure 1 - Banner display

1.2 Experiment Implementation

The A/B test was set up as follows:

1. The experiment was only being run on the mobile website
2. When a user viewed the GloBox main page, he or she was randomly assigned to the control or treatment(test) groups. This was the user's joining date
3. If the user was allocated to the test group, the page has loaded the banner; otherwise, the page did not load the banner
4. The user subsequently may or may not have purchased products from the website. It could have been on the same day they joined the experiment, or days later. If they did make one or more purchases, this was considered a “conversion”

1.3 The task of the experiment

Analyze the A/B test findings and make a recommendation to the relevant stakeholders about whether the GloBox should launch the experience to all users.

1.4 The Dataset

Below is a description of each table and its columns,

Users – user demographic information
id **–** the user id
country **-** *ISO 3166 alpha-3 country code*
gender **-** *the user's gender (M = male, F = female, O = other, Non binary)*

Table 1 **-** *Users*

Groups – user A/B test group assignment
uid – the user ID
group – the user's test group
join_dt – the date the user joined the test (visited the page)
device – the device the user visited the page on(I – iOS, A – android)

Table 2 **-** *Groups*

Activity – user purchase activity, containing 1 row per day that a user made a purchase
uid – the user ID
dt – date of purchase activity
device – the device type the user purchased on (I – iOS, A – Android)
spent – the purchase amount in USD

Table 3 **-** *Activity*

The resulted data is provided in the SQL database hence the three tables were joined by using SQL (Appendix 01). For further Segment Analysis, the countries have grouped into each Continent.

CAN, USA and MEX have included into North America("NA")

FRA, GBR, DEU, ESP have included into Europe("EU")

TUR has included into Middle East ("Middle_East")

BRA has included into South America("SA")

AUS has included into Oceania("Oceania")

The NULL rows in country column has put into "Others"

The following columns were renamed as follows when joining the tables,

Id AS user_id, group AS user_group, join_dt AS user_join_date, device AS user_device, dt AS purchase_date, spent AS total_spent.

The dataset also has several null values as follows,

```
user_id        0
country      647
continent      0
```

gender 6882
user_group 0
user_join_date 0
user_device 295
purchase_date 0
total_spent 0

1.5 Context

Control group, also called as A

Treatment group, also called as B

The experiment lasted 13 days, from January 25th until February 6th, 2023. There were 48,943 users selected into two groups: 24,343 in the control group and 24,600 in the treatment group.

The control group (A) had 10,069 female users, 10,054 male users, and 808 non-binary users. On the other hands, the treatment group (B) consist of 10,061 female users, 10,235 male users and 861 non-binary users.

The experimenters have selected two measures for A/B testing,

<u>First measure</u>

1). What is the average amount spent per user for the control and treatment groups?

<u>Second measure</u>

2). What is the user conversion rate for the control and treatment groups?

1.6 First metric

Null Hypothesis (H0): There is no difference in the average amount spent per user between the two groups

Alternative Hypothesis (Ha): there is a significant difference in the average amount spent per user between the two groups

H0: $\mu A - \mu B = 0$

Ha: $\mu A - \mu B \neq 0$

For hypothesis testing, we used a two-sample t-test, which is a statistical test used to evaluate whether two sets of data have different means. It is specifically used to compare the means of two independent groups or samples to see if they are statistically substantially different.

1.6.1 Hypothesis testing results

Metric	Control Group A	Treatment Group B
Average amount spent per user ($)	3.375	3.391
95% confidence interval for the average amount spent per user	[3.049, 3.700]	[3.073, 3.708]

Table 4 - Hypothesis testing results

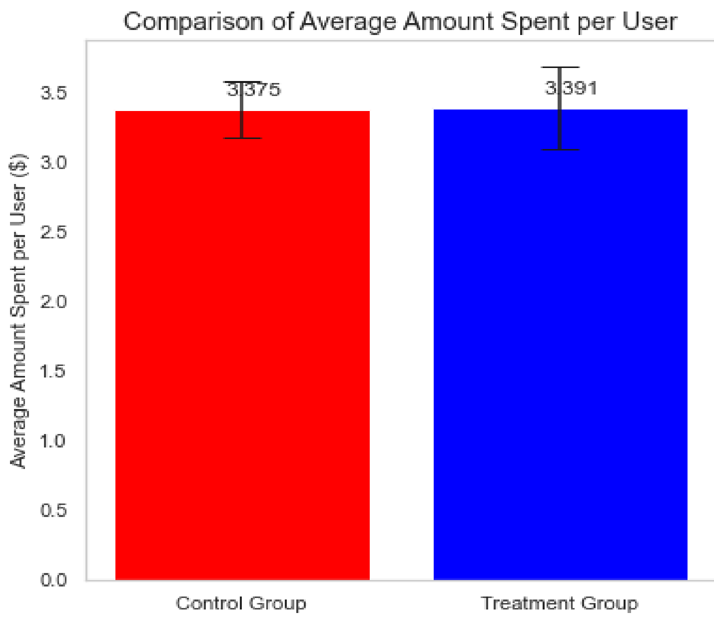

Figure 2 - Comparison of average amount spent per user

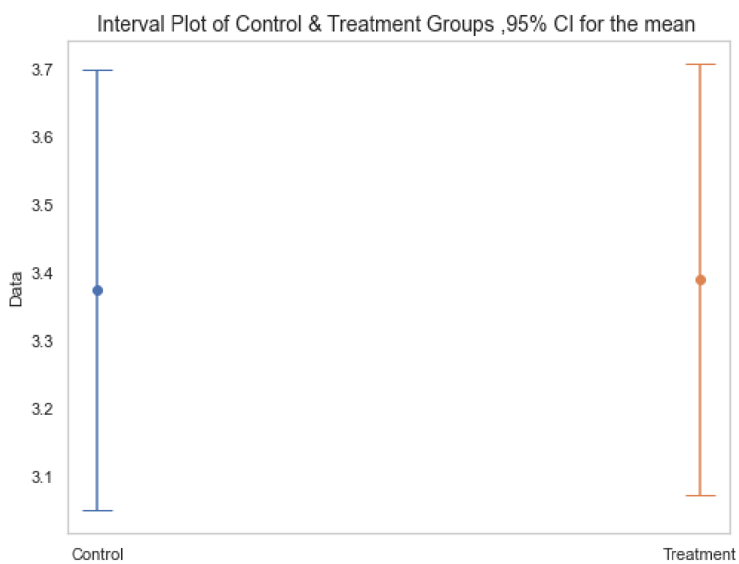

Figure 3 - Interval Plot of control & treatment groups, 95% CI for the mean

1.6.2 P values and 95% confidence interval results for treatment and the control groups

As mentioned, we were using a two-sided t-test for a difference in means. Assuming unequal variance, and we used the unpooled standard error. As result, when performing t test for both control and treatment data, we have obtained p value of 0.9438 which is greater than 0.05 significance level. Similarly, the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control is (-0.439, 0.471).

1.6.3 Conclusion

Based on the results of the hypothesis test, with a significance level of alpha=0.05 and the obtained p-value of 0.9438 we do not have sufficient evidence to reject the null hypothesis that there is no difference in the mean amount spent per user between the Control and Treatment. Therefore, we cannot conclude that there is a statistically significant difference or relationship between the variables being tested. Furthermore, the 95% confidence interval for the difference in the average amount spent per user between the two groups was calculated to be (-0.439, 0.471), which suggests that the true difference in means is likely to fall within this range with 95% confidence.

Since the confidence interval includes zero, there is no statistically significant difference between the two groups at the 5% level of significance. Therefore, we cannot make a definitive conclusion that the treatment has a significant effect on the amount spent per user compared to the control group.

1.6.4 Recommendations for First metric

After analyzing the results of the experiment, it appears that the treatment did not lead to a significant increase in average amount spent per user. it is important to note that failing to reject the null hypothesis does not necessarily mean that there is no difference between the groups. It could be that the sample size was too small or that the treatment did not have a significant effect. Therefore, it is important to interpret the results with caution and consider other factors that may have affected the outcome of the study.

Based on these findings only, we are not in position to take a decision. Therefore, we decided to conduct second metric and further segmentation analysis for data.

1.7 Second metric

Null hypothesis (H0): The conversion rate in the treatment group is equal to the conversion rate in the control group.

Alternative hypothesis (Ha): The conversion rate in the treatment group is different from the conversion rate in the control group.

H0: pA -pB = 0

Ha: pA - pB $\neq 0$

We used a two-sample two side z interval for a difference in proportions, assuming that the data is normally distributed and the variances in the two groups are equal. Moreover, supposing equal proportions, we used pooled standard error.

1.7.1 Hypothesis testing results

Metric	Control Group A	Treatment Group B
conversion rate (%)	3.92	4.63
the 95% confidence interval for the conversion rate of users	(0.0368,0.0417)	(0.0437,0.0489)

Table 5 - Hypothesis testing results

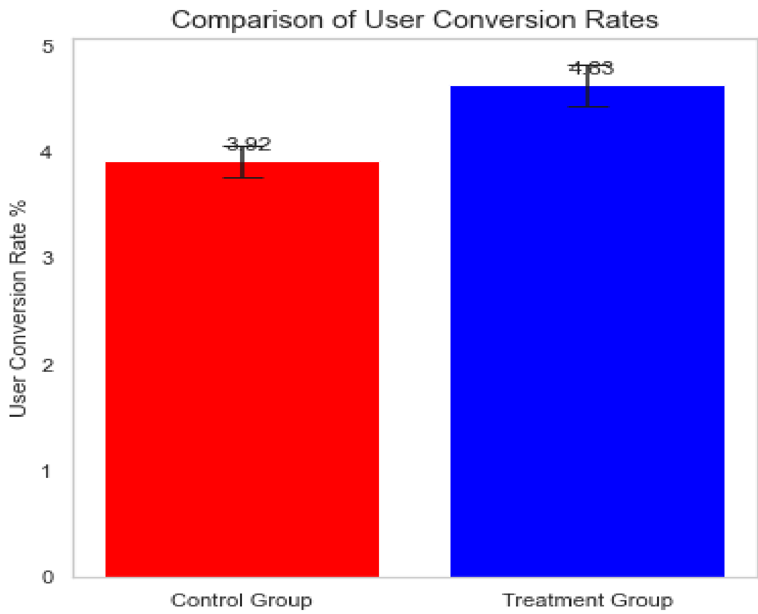

Figure 4 - Comparison of average amount spent per user

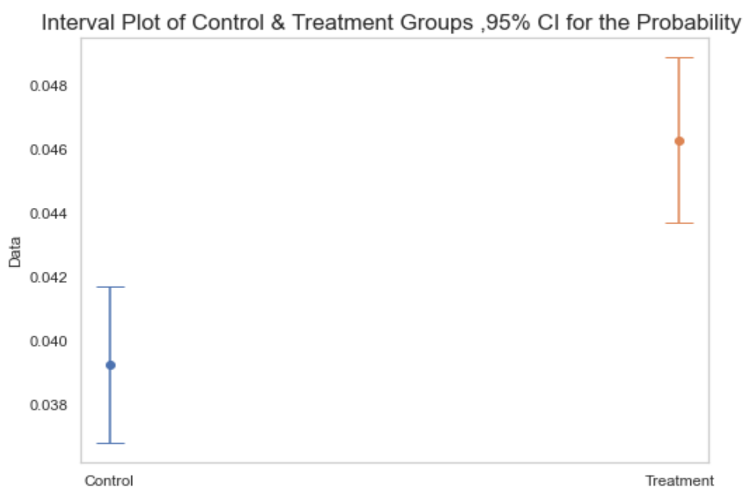

Figure 5 - Interval Plot of Control & Treatment Groups, 95% CI for the Probability

1.7.2 P values and 95% confidence interval results for treatment and the control groups

After conducting the hypothesis test to see whether there is a difference in the conversion rate between the two groups, we have obtained the p values of 0.00011 which is lower than 0.05 significance level. Similarly, 95% confidence interval for the difference in conversion rate between the treatment and control is (0.0035, 0.0107).

The 95% confidence interval (0.0035, 0.0107) suggests that the true difference in the conversion rate between the treatment and control groups is likely to fall within this range with 95% confidence. Specifically, the interval indicates that the treatment group's conversion rate is likely to be between 0.35% and 1.07% higher than the control group's conversion rate.

Since the confidence interval does not include zero, we can conclude that the difference in conversion rate between the treatment and control groups is statistically significant at the 5% level of significance. This means that there is strong evidence to suggest that the treatment has a significant effect on the conversion rate compared to the control group.

1.7.3 Conclusion

Based on the results of the hypothesis test and confidence interval analysis, we have evidence to reject the null hypothesis and conclude that there is a statistically significant difference in the conversion rate between the treatment and control groups. The obtained p-value of 0.00011 is lower than the significance level of 0.05, indicating that the probability of observing such a difference in conversion rate between the two groups by chance alone is very low.

Additionally, the 95% confidence interval for the difference in conversion rate between the treatment and control groups is (0.0035, 0.0107), which does not include zero. This suggests that we can be 95% confident that the true difference in conversion rate between the two groups lies within this range. Therefore, we can conclude that the treatment has a significant impact on the conversion rate, and it is likely to result in a higher conversion rate compared to the control group.

1.7.4 Recommendations for Second metric

Based on the significant difference observed in the conversion rate between the treatment and control groups, it can be recommended to implement the treatment to improve the overall conversion rate. However, it is also important to carefully monitor and evaluate the impact of the treatment on other relevant metrics such as customer retention, lifetime value, and profitability to ensure that the treatment is not negatively affecting these metrics.

Furthermore, it is recommended to conduct additional testing and experimentation to further optimize the treatment such as segment analysis and identify potential variations that could yield even better results. It is also recommended to investigate the factors that are driving the difference in conversion rate between the treatment and control groups to gain insights into customer behavior and preferences, and to use this knowledge to further improve the effectiveness of the treatment.

2. Part II – Segment Analysis

we conducted a comprehensive segment analysis, which involved categorizing users based on various factors such as their demographic information, purchasing behavior, and unique characteristics. Our primary objective was to evaluate the conversion rates of these different segments and use this information to make informed decisions regarding A/B testing in second metric. By leveraging this data-driven approach, we can optimize our strategies and tailor our offerings to specific customer groups, ultimately driving better results and maximizing our overall performance.

2.1 Analysis of distribution of gender data and assumptions for sample and population

The distribution of the 'gender' column data have analyzed to understand the gender split in our entire population. Based on the data, it appears that the proportion of males and females is approximately equal. This observation holds true even when a random sample of 3000 or more observations is taken, as the gender split remains at 50/50.

It is reasonable to assume that similar observations can be made for other attributes as well, which enables us to make inferences about the larger population based on the sample data.

Gender	Total Counts	Percentages
Female	20,130	41.13%
Male	20,289	41.45%

Table 6 - Gender Distribution

Note that there are 6,882 nulls and 1,669 of non-binary values in gender column.

2.2 Conversion Rates by Gender

	gender	user_group	Total_users	Coverted_users_g	Conversion_rates_gender_wise %
0	F	A	10069	518	5.14
1	F	B	10061	547	5.44
2	M	A	10054	264	2.63
3	M	B	10235	388	3.79
4	O	A	808	26	3.22
5	O	B	861	26	3.02

Table 7- Conversions rates by gender

Summary
The control group (A) has 10,069 female users, 518 converted with a conversion rate of 5.14%
The treatment group (B) has 10,061 female users, 547 converted with a conversion rate of 5.44%

The control group (A) has 10,054 male users, 264 converted with a conversion rate of 2.63%
The treatment group (B) has 10,235 male users, 388 converted with a conversion rate of 3.79%

The control group (A) has 808 non-binary users, 26 converted with a conversion rate of 3.22%
The treatment group (B) has 861 non-binary users, 26 converted with a conversion rate of 3.02%

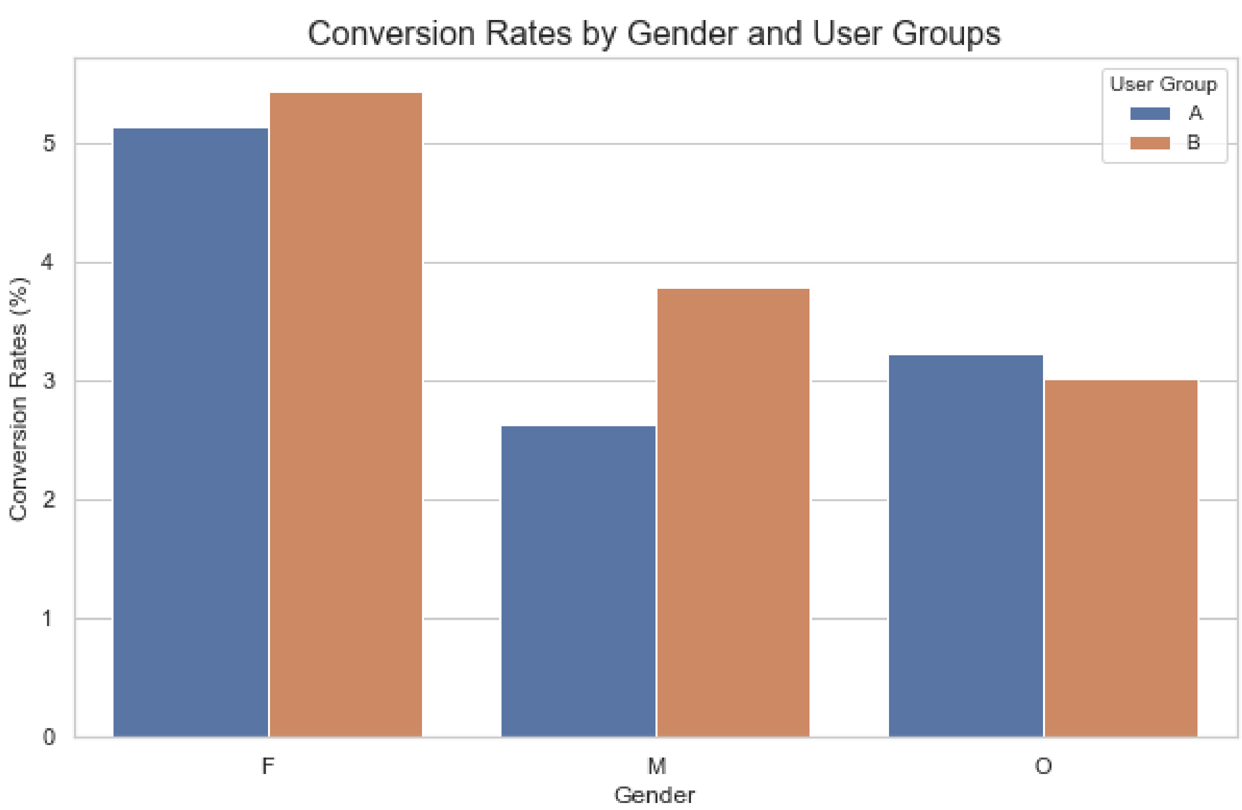

Figure 6 - Conversion rates by gender

From this summary of conversion rates for different groups, we can conclude the following,
The treatment group (B) has a higher conversion rates than the control group (A) for all gender groups (among non-binary users, there is no significant difference in conversion rates between the treatment and control groups). Additionally, it was observed that female users had higher conversion rates compared to users of opposite genders. As a result, it may be worthwhile to consider focusing marketing campaigns on opposite genders to potentially increase conversion rates.

2.3 Identify what kind of device that users were utilized to see the website/banner and Device wise conversion rates

	user_device	user_group	Total_devices	Coverted_users_d	Conversion_rates_device_wise %
0	A	A	15054	417	2.77
1	A	B	15235	537	3.52
2	I	A	9142	535	5.85
3	I	B	9218	596	6.47

Table 8 - Device wise conversion rates

Summary

The control group (A) has 15,054 Android users, 417 converted with a conversion rate of 2.77%
The treatment group (B) has 15,235 Android users, 537 converted with a conversion rate of 3.52%

The control group (A) has 9,142 iOS users, 535 converted with a conversion rate of 5.85%
The treatment group (B) has 9,218 iOS users, 596 converted with a conversion rate of 6.47%

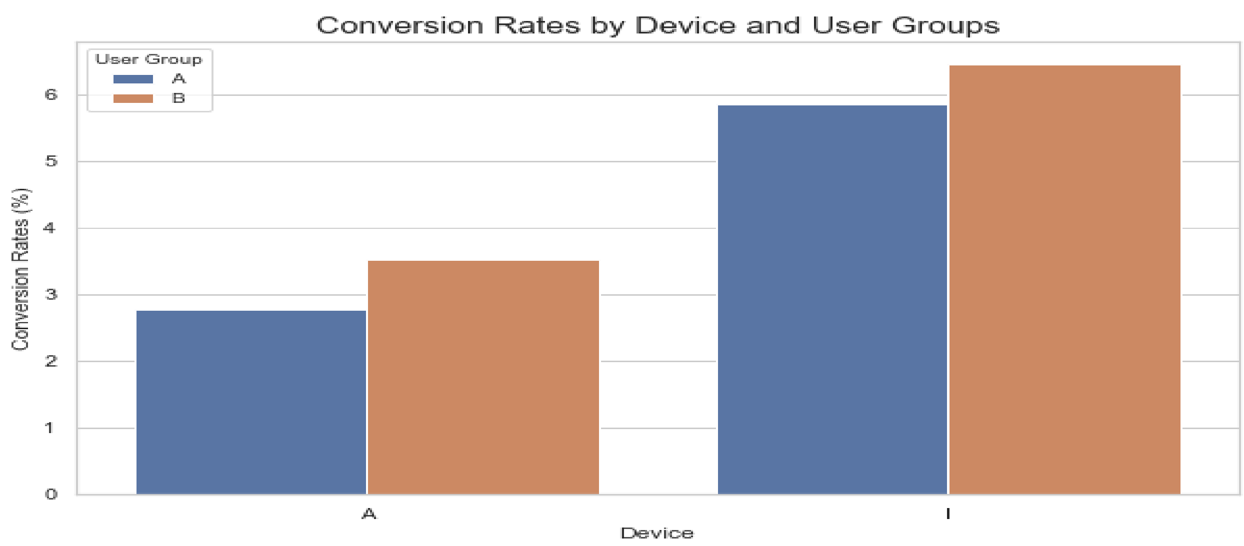

Figure 7- Conversion rates by Device

The conversion rates for both Android and iOS users are higher in the treatment group (B) than the control group (A). Moreover, most people were Android phone users.

2.4 How many unique users saw the banner each day and purchases per day / Daily user conversion rates?

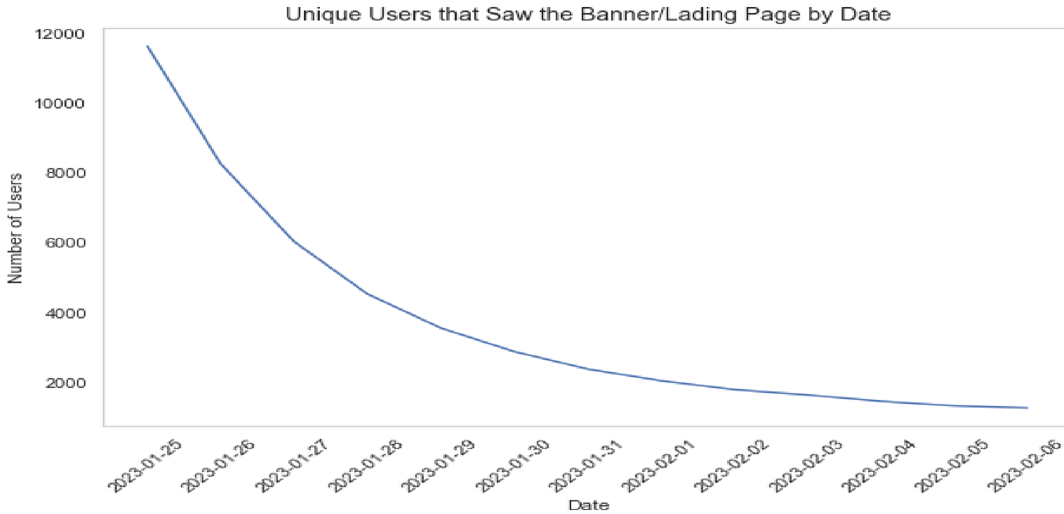

Figure 8 - unique users saw the banner/landing page by date

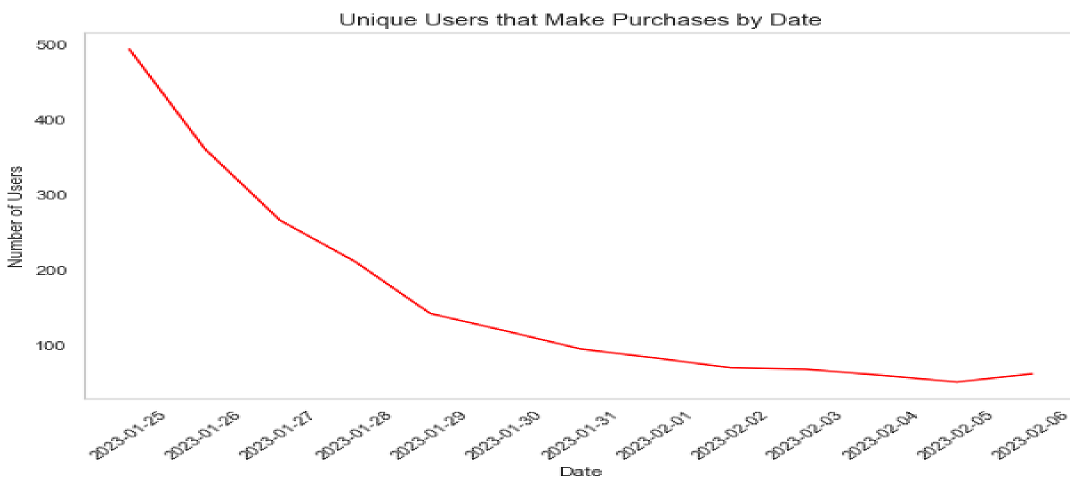

Figure 9 - Unique users that make purchases by date

Based on the figures, it appears that the number of users as well as the number of users who made purchases declined over time.

Conversion Rate over time

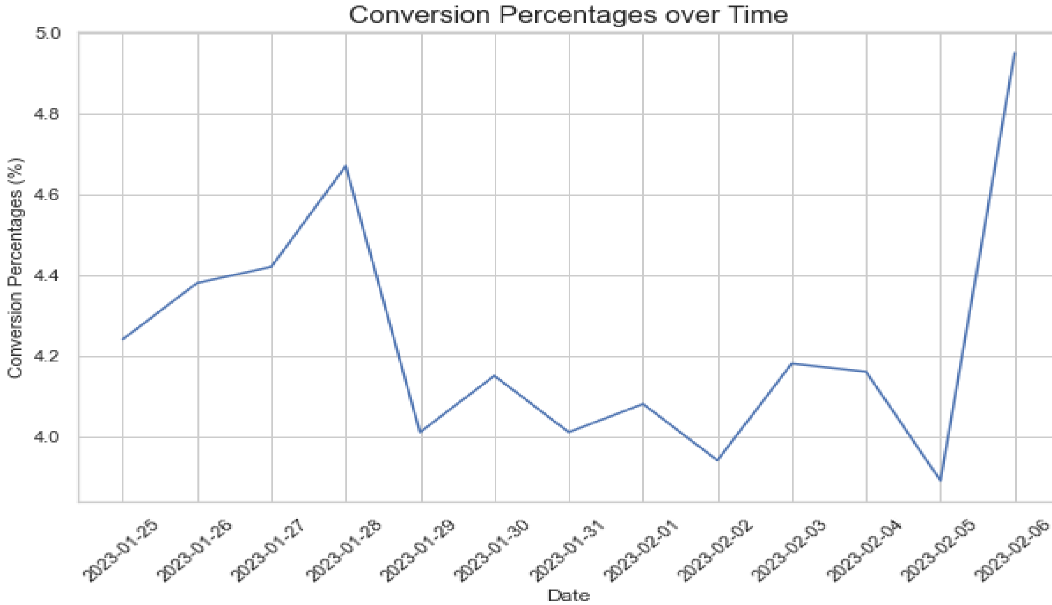

Figure 10 - Conversion Rate over time

It is apparent that there were higher conversion rates on certain days such as January 28th and February 2nd.

Conclusion

Based on the given information, it can be concluded that there has been a decline in the number of users as well as the number of users who made purchases over time. However, it is important to note that there may be various factors that contribute to this decline, such as changes in market trends or competitor activity.

Additionally, the higher conversion rates on certain days such as January 28th and February 2nd suggest that there may be certain factors that were driving user engagement and purchase behavior. It was possible that promotional campaigns, product launches, or other events may have contributed to these higher conversion rates.

To gain a better understanding of the factors that are driving user behavior, it may be helpful to conduct further analysis and gather additional data. This could include conducting surveys or interviews with users, analyzing competitor activity, or tracking changes in market trends. By gathering more information, it may be possible to identify areas where improvements can be made to increase user engagement and drive sales.

2.5 Group wise (control & treatment) conversion rates by each dates

Following table shows the conversion rates for control(A) & treatment(B) users in day wise

	Date	Total_spent_users_A	Total_spent_users_B	Total_users_by_day	A users conversion rate by day	B users conversion rate by day
0	2023-01-25	226	268	11646	1.94	2.30
1	2023-01-26	158	204	8270	1.91	2.47
2	2023-01-27	125	142	6043	2.07	2.35
3	2023-01-28	110	102	4543	2.42	2.25
4	2023-01-29	65	78	3567	1.82	2.19
5	2023-01-30	55	65	2894	1.90	2.25
6	2023-01-31	42	54	2392	1.76	2.26
7	2023-02-01	42	42	2057	2.04	2.04
8	2023-02-02	36	35	1803	2.00	1.94
9	2023-02-03	29	40	1650	1.76	2.42
10	2023-02-04	21	40	1468	1.43	2.72
11	2023-02-05	21	31	1336	1.57	2.32
12	2023-02-06	25	38	1274	1.96	2.98

Table 9 - Group wise conversion rates by each day

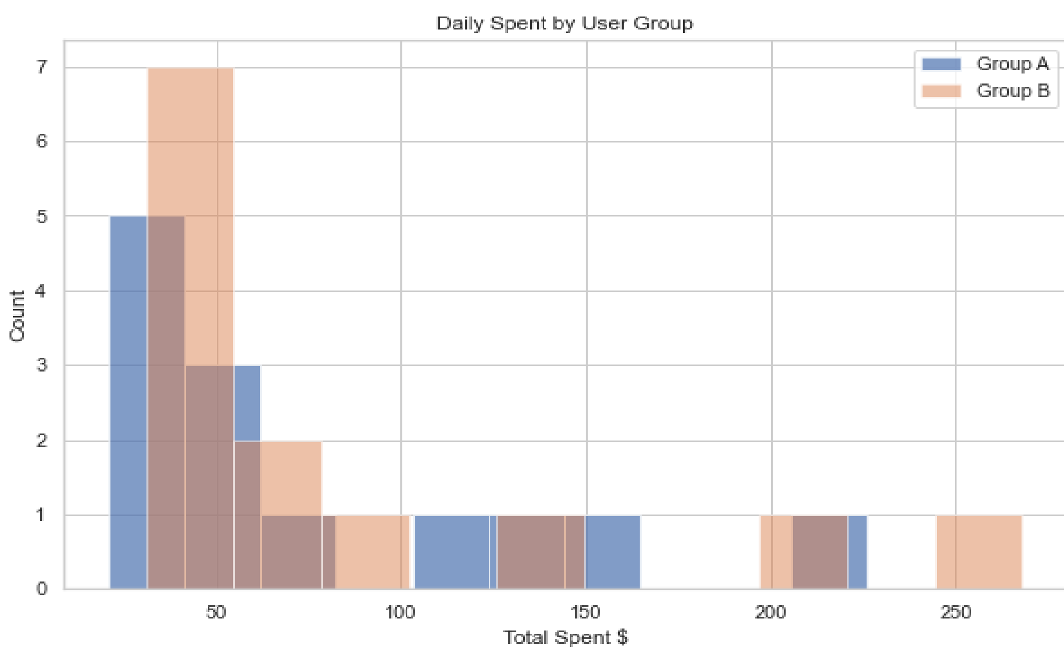

Figure 11 - Histogram of spent users for control & treatment

Based on the histogram presented, it appears that the group receiving treatment has made a greater number of purchases compared to the control group. Furthermore, it seems that a relatively small number of users in both groups spent over $100.

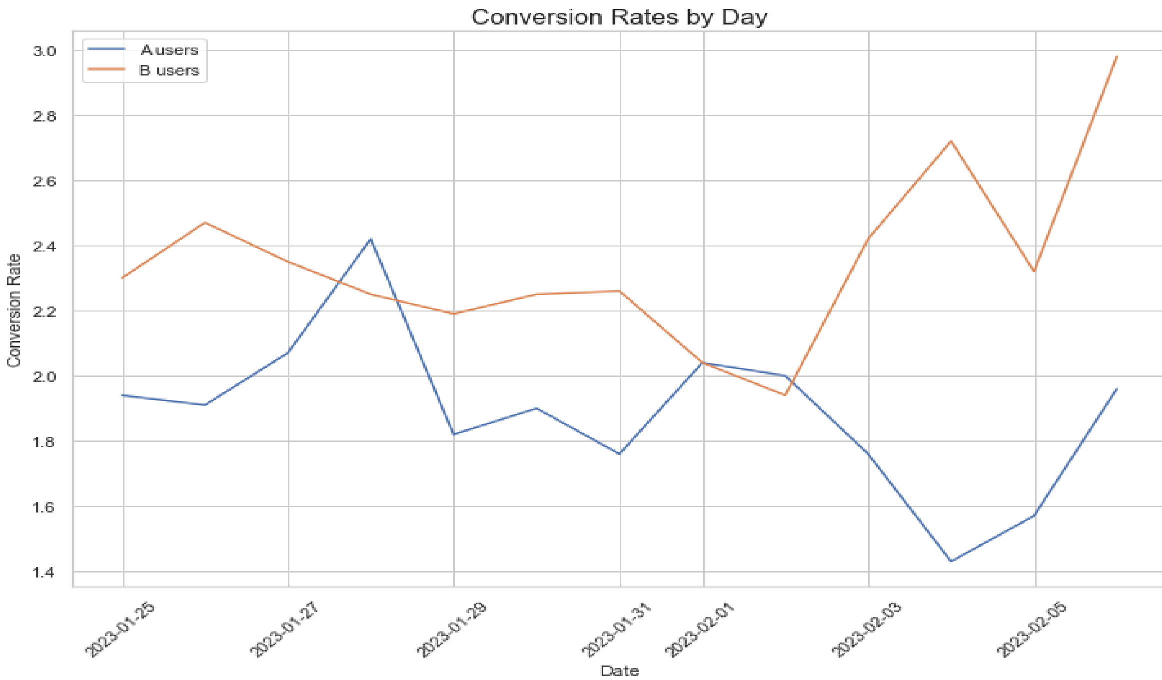

Figure 12 - Conversion rate by day for control & treatment groups

According to the information presented in the line chart, it appears that the daily conversion rates for the treatment group users are higher compared to the control group users.

Conclusion

Based on the information provided in both the histogram and line chart, it can be concluded that the treatment group has a higher conversion rate compared to the control group. The treatment group users made a greater number of purchases, as indicated by the histogram, and this is further strongly supported by the line chart (Figure 12), which shows a consistently higher conversion rate for the treatment group over time. However, it is important to note that both groups had relatively low numbers of users who spent over $100, suggesting that there may be room for improvement in terms of increasing overall revenue generated by the users.

2.6 Where the users are designated/located at & Conversions rates for each Continent?

Following table represents the Continent wise representation of conversion rates for each group, control & treatment.

	continent	user_group	Total User Counts	Coverted_users_continents	Conversion_rates_continenet_wise %
0	EU	A	5894	180	3.05
1	EU	B	5992	242	4.04
2	Middle_East	A	1849	74	4.00
3	Middle_East	B	1883	67	3.56
4	North America	A	10891	493	4.53
5	North America	B	11189	611	5.46
6	Oceania	A	608	13	2.14
7	Oceania	B	560	17	3.04
8	Others	A	296	16	5.41
9	Others	B	347	14	4.03
10	SA	A	4805	179	3.73
11	SA	B	4629	188	4.06

Table 10 - Continent wise conversion rates

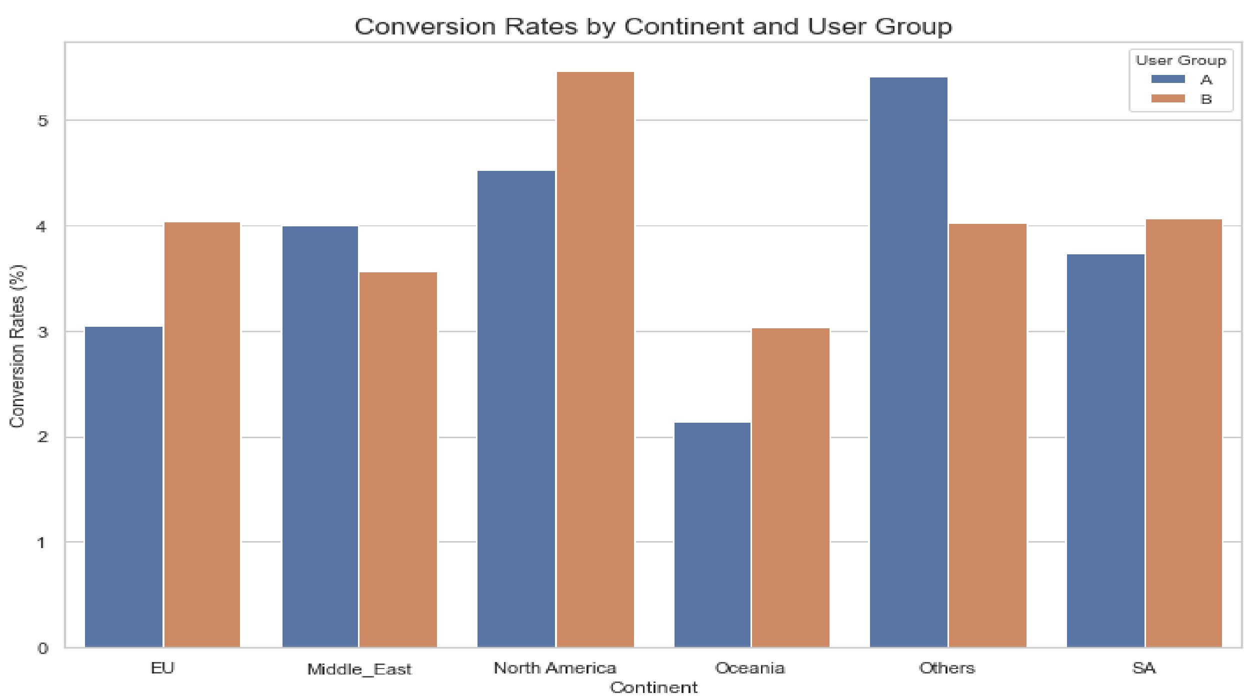

Figure 13 - Continent wise conversion rates

Comparing conversion rates between different continents can be a bit challenging, as there may be other factors at play that affect the rates. One way to make a fairer comparison would be to compare conversion rates within each group, rather than between groups.

For example, we could compare the conversion rates for each group within North America, and then do the same for Oceania. This would give a clearer picture of which group has higher conversion rates within each continent, and allow for a more meaningful comparison.

Alternatively, we could also compare conversion rates for each group across different countries within each continent. This would take into account any regional differences that may exist, and provide a more nuanced view of the conversion rates for each group.

Ultimately, it's important to consider multiple factors when comparing conversion rates, and to avoid making conclusions based on surface-level comparisons.

2.7 Country wise Analysis

country	std	max	sum	mean
DEU	31.12	1659.40	11756.87	3.04
GBR	34.67	1546.30	9788.24	3.31
USA	25.50	1266.80	61645.72	4.16
BRA	26.86	1094.80	29636.00	3.14
MEX	23.46	844.70	17694.54	3.08
TUR	20.91	745.80	11500.97	3.07
FRA	15.42	333.79	7637.84	2.47
ESP	17.35	267.80	5393.16	2.70
CAN	18.63	198.60	6134.11	3.90
AUS	13.48	164.58	2179.53	1.87

Table 11 - Country wise purchases (mean, sum, max & SD)

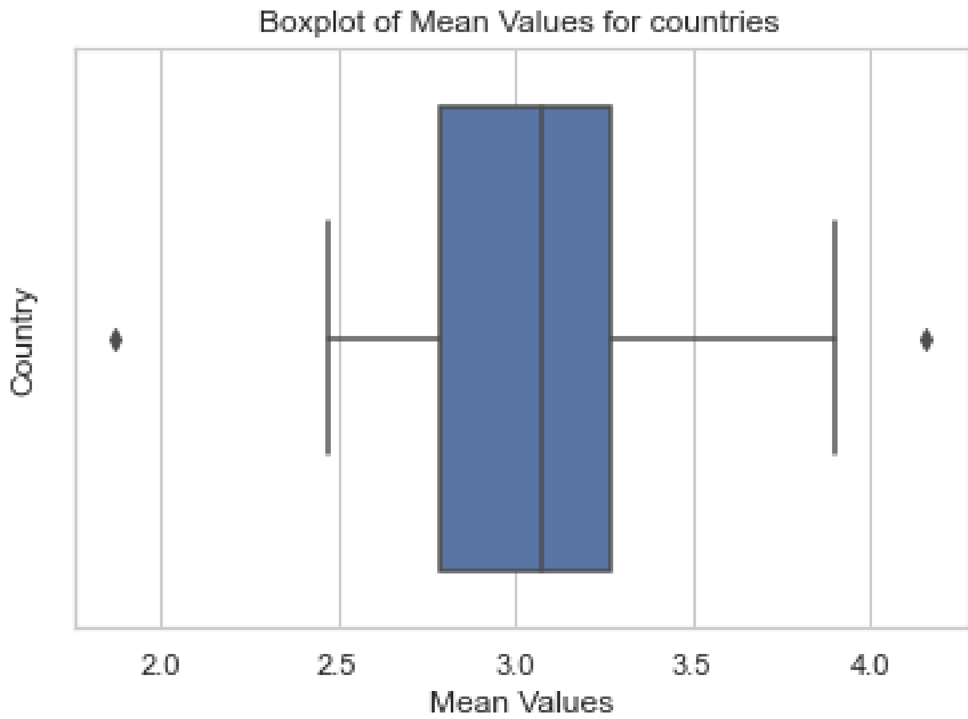

Figure 14 - Box plot for mean spent

According to the box plot there are two outliers which represents the USA mean purchase value of $ 4.16 and Australia mean purchase value of $ 1.87.

2.8 USA analysis

	gender	user_device	user_group	User Counts_USA	User spent Counts_USA	Conversion_rates_USA %
0	F	A	A	1845	91	4.93
1	F	A	B	1935	94	4.86
2	F	I	A	1167	93	7.97
3	F	I	B	1081	94	8.70
4	M	A	A	1825	53	2.90
5	M	A	B	1928	83	4.30
6	M	I	A	1161	67	5.77
7	M	I	B	1163	73	6.28
8	O	A	A	153	8	5.23
9	O	A	B	180	3	1.67
10	O	I	A	82	1	1.22
11	O	I	B	71	3	4.23

Table 12 - USA analysis

The treatment groups show the highest number of conversion rates, specifically among female users who own iPhones, with a conversion rate of 8.70%. On the other hand, the lowest conversion rates can be observed in the control group, particularly among non-binary users who own iPhones, with a rate of 1.22%.

2.9 AUS Analysis

	gender	user_device	user_group	User Counts_AUS	User spent Counts_AUS	Conversion_rates_AUS %
0	F	A	A	174	6.0	3.45
1	F	A	B	134	5.0	3.73
2	F	I	A	89	4.0	4.49
3	F	I	B	86	2.0	2.33
4	M	A	A	150	1.0	0.67
5	M	A	B	134	2.0	1.49
6	M	I	A	93	5.0	5.38
7	M	I	B	99	1.0	1.01
8	O	A	A	17	NaN	NaN
9	O	A	B	16	NaN	NaN
10	O	I	A	12	NaN	NaN
11	O	I	B	10	NaN	NaN

Table 13 - AUS analysis

The control group has the highest conversion rates among male users who own iPhones, with a rate of 5.38%. Conversely, the lowest conversion rates in the control group can be seen among male users who own Androids, with a rate of 0.67%. Additionally, there were no conversion rates for any non-binary users in Australia.

Conclusion

Overall, the results suggest that the treatment group's approach may be more effective in converting users into customers, especially among female iPhone users. It is also worth noting that there were no conversion rates for non-binary users in Australia. This conversion rates fluctuation could be due to various reasons such as a population volume in each country, small sample size, lack of interest in the product, or other factors that need to be investigated further.

Therefore, further analysis and investigation are needed to determine the significance of these findings and identify the factors that contributed to the observed differences in conversion rates.

3. Experimental design: power analysis

Based on the power analysis we conducted, we determined that a sample size of 48943 has an actual standardized effect size of 0.035. Assuming an effect size of 1%, we calculated that a new sample size of 156978 would be required to ensure a high probability of correctly rejecting the Null hypothesis that there is no difference between the control and treatment groups.

In conclusion, our power analysis suggests that a larger sample size is needed to detect a smaller effect size with a high level of confidence. The calculated sample size of 156978 can be used as a guide for future studies aiming to detect similar effect sizes between control and treatment groups.

4. Summary and Final Conclusion

Based on the strong evidence we have observed; I would highly recommend to launching the banner for all users. Our analysis indicates that there was a significant increase in user conversion rates. Therefore, it is reasonable to assume that the banner will have a positive impact on the overall conversion rate of the platform.

5. Appendix

Idle
PING 186ms
Run query

```sql
SELECT
  u.id AS user_id,
  u.country AS country,
  CASE
      WHEN u.country IN ('CAN','USA','MEX') THEN 'North America'
      WHEN u.country IN ('FRA', 'GBR','DEU','ESP') THEN 'EU'
      WHEN u.country IN ('TUR') THEN 'Middle_East'
      WHEN u.country IN ('BRA') THEN 'SA'
      WHEN u.country IN ('AUS') THEN 'Oceania'
  ELSE
      'Others'
  END as continent,
  u.gender AS gender,
  g.group AS user_group,
  g.join_dt AS user_join_date,
  g.device AS user_device,
  COALESCE(a.dt::VARCHAR(10),'0') AS purchase_date,
  COALESCE(SUM(a.spent),0) As total_spent
FROM users AS u
INNER JOIN groups AS g
    ON u.id = g.uid
LEFT JOIN activity AS a
    ON u.id = a.uid
GROUP BY user_id,country,continent,gender,g.group,user_group,user_join_date,
        user_device,purchase_date;
```

Appendix 1 - SQL code

For A/B test results and segment analysis is done by using Python. The PDF file is provided here, also the Python programming file – "GolBox_AB_Testing_Analysis.ipynb" file attached separately with the zip file.
GolBox_AB_Testing_Analysis.pdf