GlowBox

Analysis of Mobile Banner A/B Test

Wandering Tern
Matt Oster, Owner
October, 24,2023

Summary

This report summarizes the analysis, by Wandering Tern, of an A/B test conducted on GlowBox's mobile landing page. The results show a statistically significant increase in the conversion rate for the group exposed to the banner. However, the impact on the average amount spent per user was not significant. Based on these findings, we recommend further testing, a larger sample size, and reiteration before implementing the banner for all mobile users.

Test Context

GlowBox is an online marketplace most known for their boutique fashion apparel and high end decor/textiles. In recent months there has been an increase in food and drink products offered to customers . To promote, the Growth Team at GlowBox wanted to see if highlighting these new product lines through a mobile landing page banner could increase profits and customer recognition of new diversified product lines.

The test data was collected from January 25, 2023 - February 6,2023 with 48,943 users randomly split into a Control group (A) and Treatment group(B). The dataset provided by Glowbox for analysis consists of only entries for the dates listed above, three worksheets with multiple columns containing ids, sales, group, etc.. The data set was in good condition, with only minor cleaning required like eliminating “null values” and addressing multiple purchases by discrete users.

Measures for Data Set			
	Full Population	Test Group A	Test Group B
Sample Size	48943	24343	24600
Mean Sales	3.3827	3.3745	3.3910
Conversions	2095	955	1139
Conversion ratio	0.04280	0.03923	0.04630

Standard Deviation	25.6749	25.9364	25.4146

Results

Beekeeper software was used to query the GlowBox database with PostgreSQL. Queries were made to have a usable dataset to do the analysis in Google Sheets, and visualizations in Tableau The focus was to organize the data focusing mostly on the two key metrics between the groups; the rate of conversion and the mean average spent per user between the two groups.

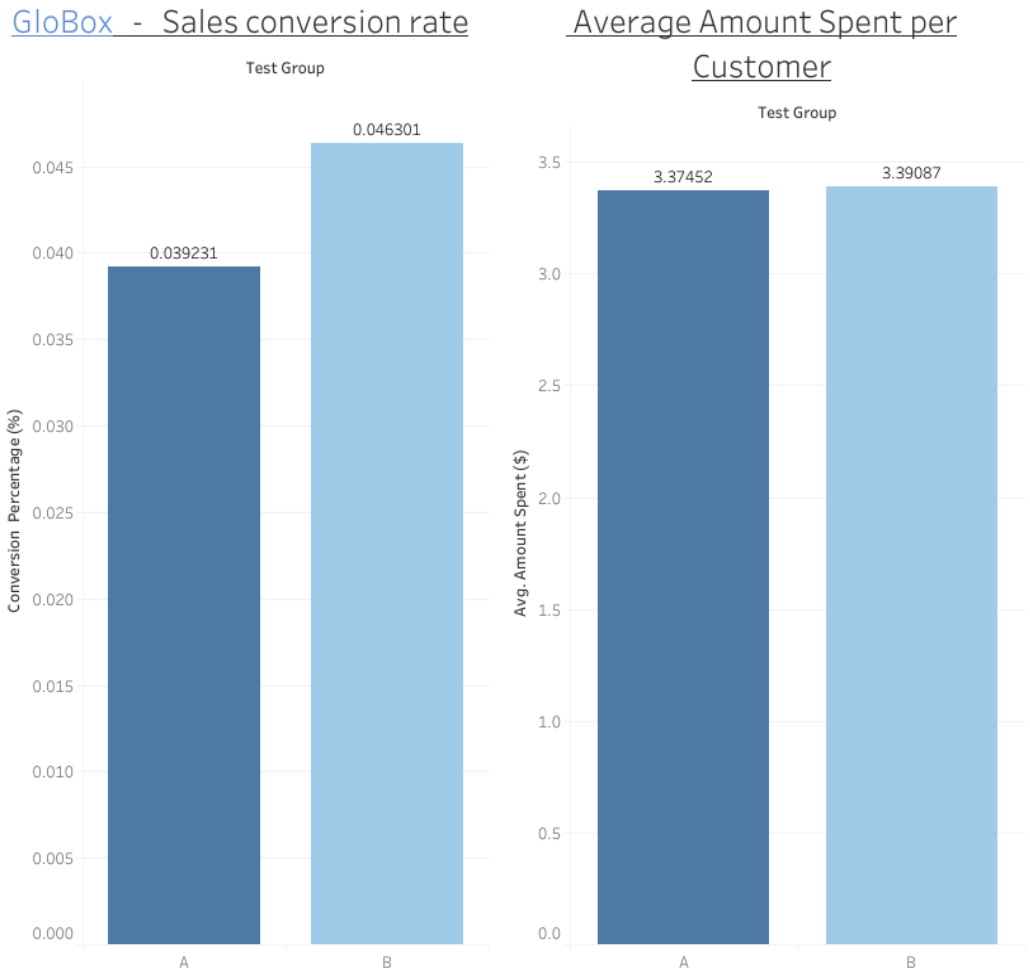
Hypothesis and confidence interval tests were done to compare the difference in conversion rate between the two groups. For the Hypothesis test we used a two-sample Z-test with pooled proportions. Below are the equations and results for this test.

Calculations	Equation
Sample proportion	p1_hat=Conversions/Sample Size
Standard Error	SE=SQRT(((.03923*(1-.03923)/24343)+((.04630*(1-.03630))
Test Stat	T=(p1_hat-p2_hat)/ (SQRT(p_hat*(1-p_hat))*((1/n1)+(1/n2)))
p-value	p_val=T.DIST.2T(3.86351, 24342)

Calculations	Notations	Value
Sample Size (control)	n1	24343
Sample Size (Treatment)	n2	24600
Sample Proportion(Control)	p1_hat	0.03923
Sample proportion (treatment)	p2_hat	0.04630
Pooled Proportions	p_hat	0.04280
Standard Error	SE	0.00124
Test Statistic	T	-3.86351
p-value	p-val	0.00011

The null hypothesis was that the conversion rate between the control and treatment groups were equal, while the alternative hypothesis was that the treatment group

conversion rate was greater than the control group. Alpha was set to $\alpha = .05$. With a resulting p-value of .00011, we reject the null hypothesis and accept the alternative hypothesis where H1>H0. This signals that the difference between the two conversion rates were statistically significant.



For the confidence interval for the difference in conversion rates test we used a two-sample Z-interval test with unpooled proportions Again, the Alpha=.05 . Below are the results and formulas used in this test.

Calculation	Notation	Equation
Sample statistic	stat	.=p1_hat-p2_hat
Standard error	SE	SQRT(((p1_hat*(1-p1_hat))/N1)+(p2_hat*(((1-p2_hat))/N2)))
Z-score	z	z=(p-hatA-p-hatB)/SE

Critical value	cv	.=Z(1-(a/2)
Margin of error	ME	ME=SE*Z
Lower bound		.=stat-ME
Upper bound		.=stat+ME

Calculation	Notation	Value
Sample size (control)	n1	24343
Sample size (treatment)	n2	24600
Sample proportion(control	p1_hat	0.03923
Sample proportion treatment)	p2_hat	0.04630
Sample statistic	stat	0.00707
Standard error	SE	0.00183
Z-score	z	-3.86652
Critical value	cv	-1.9816
Margin of error	ME	-0.0036
Lower bound		0.00345
Upper bound		0.01069

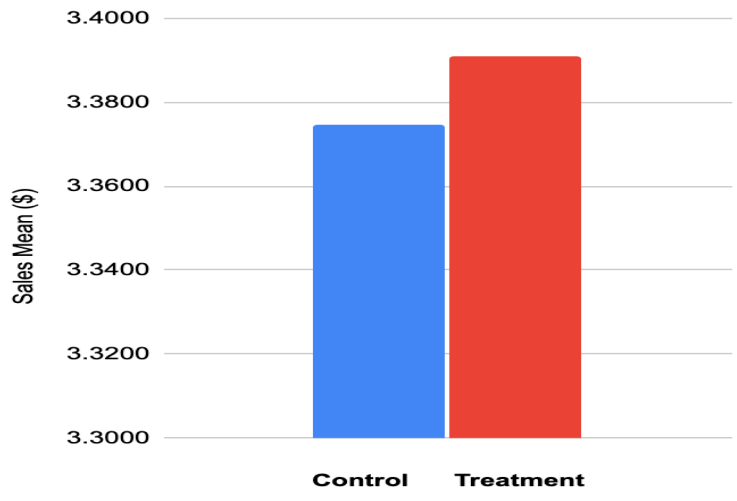
We aimed to determine if there was a significant difference in conversion proportions between the control and treatment groups. Analysis at the 95% confidence interval revealed a range from 0.00345 to 0.01069, which did not encompass zero, indicating a significant distinction in conversion proportions between groups.

For the Hypothesis Test for the difference in the average amount spent per user in each group. The Alpha was set to 0.05. Below are the results and equations for this test.

Hypothesis	
NULL	H0: mu1_bar = mu2_bar
Alternate	H1: mu1_bar <mu2_bar

With an alpha threshold of 0,05 we are testing the null hypothesis where sample mean1 is equal to sample mean2 to decide to reject or fail to reject H0. If we reject the null hypothesis we will adopt the alternative hypothesis where the sample mean2 is greater than sample mean 1.	

Calculation	Notation	Value	Equation
Sample size (control)	n1	24343	
Sample size (treatment)	n2	24600	
Sample mean (control	μ1_bar	3.3745	
Sample mean (treatment)	μ2_bar	3.3910	
Sample STD (control)	s1	25.9364	
Sample STD (treatment)	s2	25.4146	
Standard error	SE	0.2321	= √(s1^2/n1 + s2^2/n2)
Test statistic	T	0.0711	.=(μ_bar-μ2_bar)/SE
Degrees of freedom	df	24342	.=n1-1
p-value	p-val	0.9433	.=T.TEST.2T(T, df)



Difference in Average Amount Spent

In our two-sample t-test with unpooled means, we investigated whether there is a statistically significant difference between the two groups. The test yielded a p-value of 0.9439, which is significantly higher than the alpha level of 0.05. As a result, we fail to reject the null hypothesis, suggesting that there is insufficient evidence to conclude that the means of the two groups are different. Therefore, we can reasonably assume that the two samples are not significantly distinct in their means.

For the confidence interval for the difference in average amount spent per user test we used a two-sample T-interval test with unpooled proportions. Below are the formulas and results used in this test.

Calculation	Notation	Value	Equation
Sample size (control)	n1	24343	
Sample Size (treatment)	n2	24600	
Sample Mean(control	x1_bar	3.3745	
Sample Mean (treatment)	x2_bar	3.3910	
Sample std dev (control)	s1	25.9364	
Sample std dev (treatment)	s2	25.4146	
Sample statistic	stat	0.0165	.=x1_bar-x2_bar
Standard Error	SE	0.2321	.= √(s1^2/n1 + s2^2/n2)
Degrees of Freedom	df	24342	n1-1
Critical Value	T	1.9601	.=T.INV(probability, df)
Margin of Error	ME	0.4550	.=T*SE
lower bound		-0.4385	.=stat-ME
upper bound		0.4715	.=stat+ME

We conducted a two-sample t-test to assess potential differences between two groups. The resulting confidence interval ranged from -0.4386 to 0.4715. This interval reflects the likely range of the true difference in means between the two groups. As it includes both positive and negative values around zero, it suggests that we cannot assert a statistically significant difference between the groups at the chosen 95% confidence level. Our analysis indicates that, with the available data and confidence level, the means of the two groups are consistent with being equal.

Additional comparisons

First we did a power analysis for the tests we were trying to perform to determine an appropriate sample size to have confidence in making a conclusion. For conversion tests we used Statsig sample size calculator. Using a Minimum Detectable Effect (MDE) of 6% and a base conversion rate of 4%, we should use a sample size of 164,900 users in this experiment to maintain confidence in our conclusions. We ran these tests with a sample size of 48,943 users.

For means testing we used Statulator difference between two means calculator. Setting the difference of means at 5 and the expected standard deviation of 25. A sample size of 786 is large enough to maintain confidence in our conclusions around difference of means between the two groups.

Data collection happened in a relatively short time of two weeks. Still we looked to see if there was any “novelty effect” of an early spike in sales which dies off. No such effect was apparent.

Gender, home country, and type of mobile device were all compared with both conversion and average sales in both groups. Distribution of average sales was also looked at. Links to visualizations will be on the last page. No big insights were found for the questions at hand, but if you launch or partially launch the banner some demographics definitely outperform others.

Recommendations

We do not recommend launching the banner at this time if revenue growth is the main motivator. We believe there is possible potential in the banner, but additional iteration and data collection for use in further analysis is the route we suggest.

From what we have gleaned is conversions (more sales) increased in the treatment group, although we can not have confidence in our statistics due to the small sample size. We can

say with confidence the average sales per user can be considered the same with or without the banner

Moving forward we suggest a longer test with more users or activity be performed. We also suggest iterating the banner with more analysis on demographics to be able to take advantage of your already diverse audience.

Links

Visualizations -

https://public.tableau.com/shared/ZNJSP5YXF?:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/GlowBoxDashes/StoryGlowbox?:language=en-US&:display_count=n&:origin=viz_share_link

Data set

https://docs.google.com/spreadsheets/d/1G5OMh_tpoUjqIPT0_OOZWW7ktU9fYFXzTOUve8mnjco/edit?usp=sharing

Sheets
https://docs.google.com/spreadsheets/d/1w8ucypyJMfDbYZ72mx4EEWk7_HAS0p3zPjrv5OUqMXY/edit?usp=sharing