GlowBox Data

Domanic Rincones
07/08/2022

Summary

Our A/B testing experiment yielded interesting results across the control and treatment groups. Notably, the treatment group experienced an 18% statistically significant increase in conversion rates. However, the average amount spent per user was not significantly different between the two groups.

These outcomes strongly suggest that the new features introduced to the treatment group are effective in driving conversion rates. Therefore, we recommend deploying these changes to a larger audience. Banners and similar features, not typically associated with high costs, present an affordable way to enhance key metrics. Any significant increase in conversion rates, as observed in our experiment, is a good reason to implement these features.

However, we also acknowledge the lack of significant difference in the average amount spent per user. Given this, we recommend conducting further experiments or analyses to identify strategies that could potentially increase user spending. This might involve adjusting the current new features or experimenting with completely new ones. The primary aim is to ensure that our efforts not only increase user conversion rates but also encourage higher spending.

In essence, we can view the results of our A/B testing experiment as a successful step towards improving our platform. It demonstrates that the changes made are effective in driving one of our key metrics (conversion rate). However, further investigation and adjustments are needed to drive improvement in average user spending.

Background and Dataset Overview

We launched an A/B testing initiative in our pursuit to boost user engagement and streamline the user experience on our platform. This experiment scrutinized the influence of new features or changes to the user interface on vital metrics like conversion rates and average user spending.

The dataset for the experiment consisted of a wide range of user behavior details from both control and treatment groups. This data was comprehensive, including information such as user IDs, demographic aspects, device types, test group assignments, conversion statuses, and user spending. This dataset was instrumental in highlighting statistically significant behavioral differences between the two groups and guiding the platform enhancement decisions.

Motivation

We primarily aimed to enhance the user conversion rates and average spending as these are the key metrics influencing our revenue stream. Any improvement in these areas indicates increased user engagement and satisfaction, boosting our business performance.

Test Groups

The experiment involved a control group that interacted with the existing platform interface and a treatment group that was introduced to new features. The control group constituted 24,343 users while the treatment group had 24,600 users.

Test Parameters

We introduced new interface elements and workflows to provide a more intuitive user experience in this experiment. We hypothesized that these changes would drive user conversion rates and average spending upwards.

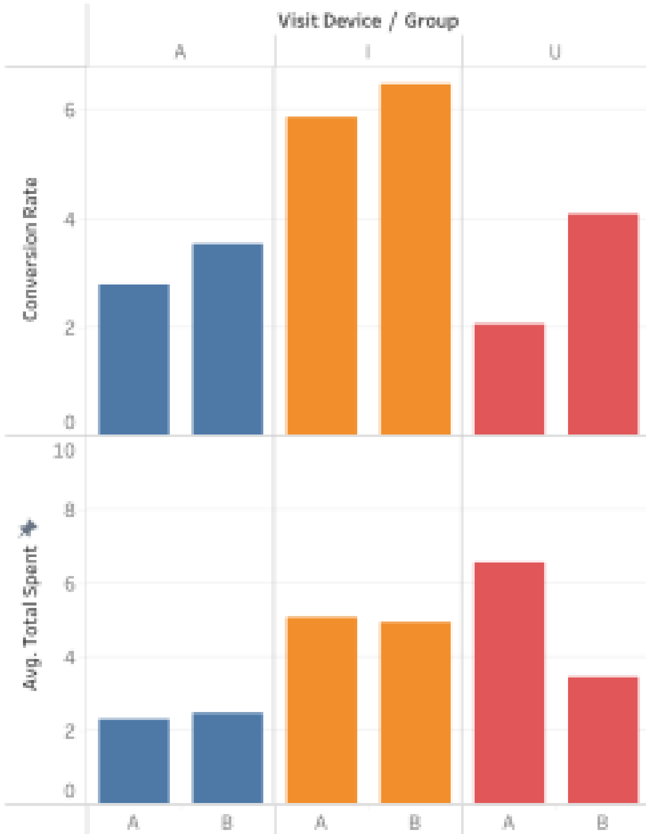
Success Metrics

The conversion rate and the average user spend were the focal metrics in this experiment. The control group had a conversion rate of 3.92%, while the treatment group showed a conversion rate of 4.63%. Additionally, the average spending for the control group stood at $3.375, marginally lower than the treatment group's $3.391.

Results

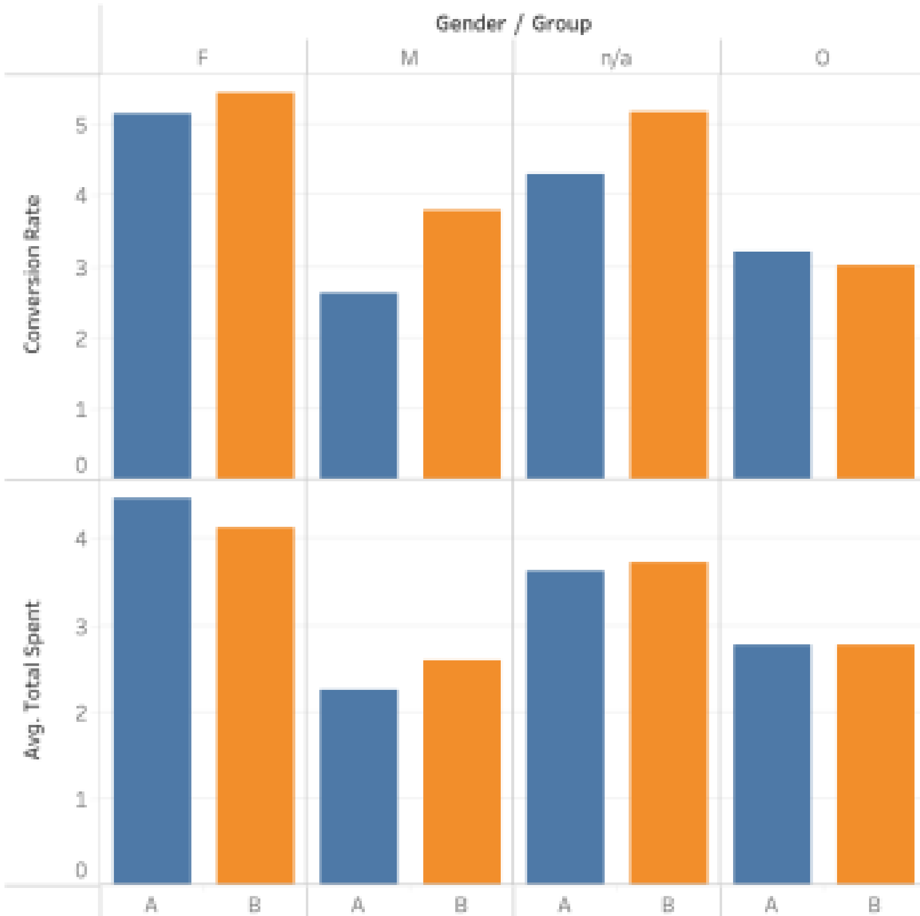
Overall, the treatment group demonstrated a significant 18% relative increase in the conversion rate, albeit without a substantial difference in the average user spend.

Results by Device



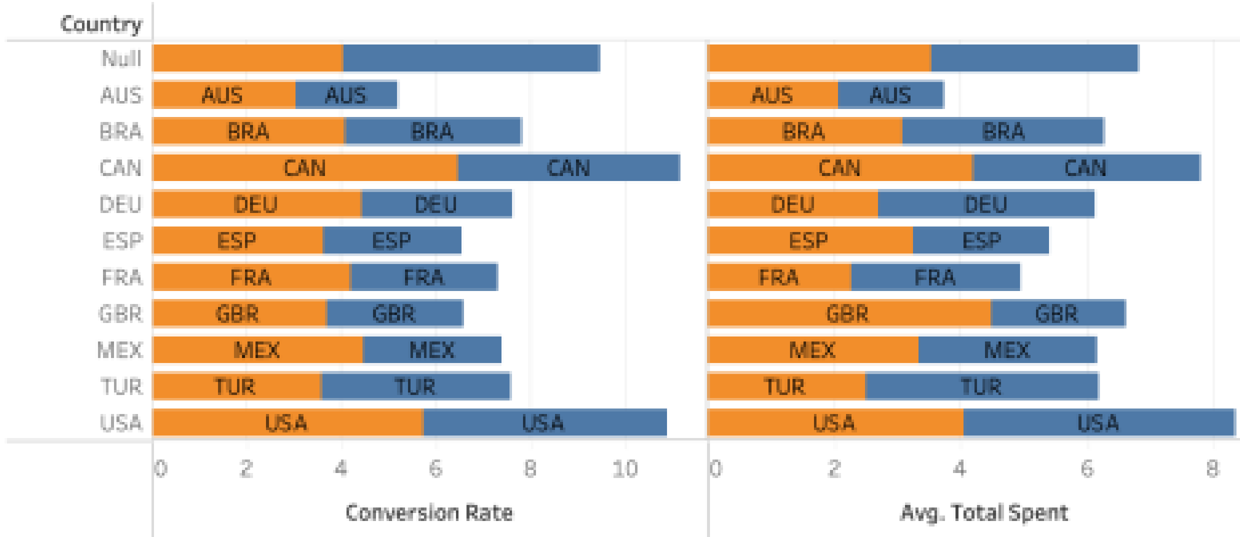
Our A/B test has shown distinct behavioral differences across device types. Android users in the treatment group spent more on average compared to those in the control group. This suggests that the new banner was effective in increasing average spending among Android users. However, iOS users in the control group spent more on average, with a lower average spend observed in the treatment group. This may indicate that the new banner did not work as well with our iOS user base. These results highlight the importance of considering user device type when implementing new features. Tailoring our approach to each platform may improve overall effectiveness.

Results by Gender



In our A/B test, the treatment positively impacted both male and female conversion rates, with males at 3.79% (from 2.62%) and females at 5.43% (from 5.14%). The average spending also increased for males ($2.60 from $2.24), but slightly decreased for females ($4.13 from $4.46). This suggests the treatment was more beneficial for male users, improving both their conversion rates and spending.

Regional Analysis of User Conversion and Spending



Our A/B testing results varied across different regions:

Canada : This treatment group showed a higher conversion rate compared to the control group, with an average spending of $4.19 in the control group and $3.60 in the treatment group.

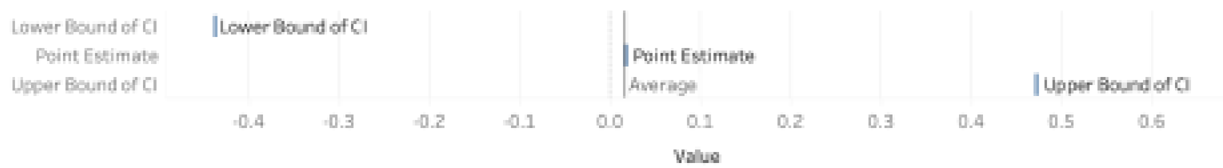
Turkey: The control group surpassed the treatment group in conversion rates and average spending, indicating less effectiveness of the banner ad.

USA: the new banner boosted the conversion rate, indicating its effectiveness in engaging users. However, it didn't noticeably affect the average spend per user. This suggests that the banner is good at prompting initial purchases, but doesn't necessarily lead to increased spending.

Confidence Intervals

Results Breakdown
Conversion Rate

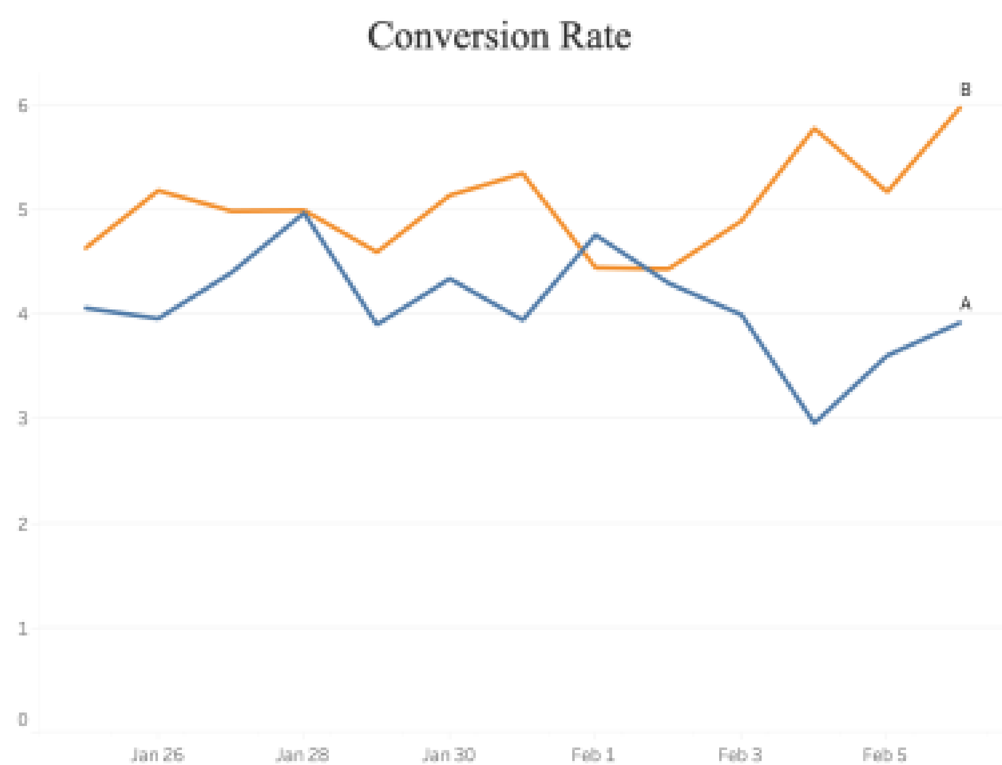
Average Amount Spent



In terms of conversion rate, our point estimate represents the observed difference in conversion rates between our control and treatment groups. The associated confidence interval is a range which, based on our sample data, we are 95% certain contains the true difference in conversion rates if we were to carry out the experiment with the entire population.

Similarly, when discussing average spending, the point estimate is the difference in average spending between the treatment and control groups in our sample. We've also constructed a confidence interval for this measure, which is the range that, with 95% certainty, we think would encompass the true difference in average spending if the experiment was conducted on the entire population.

Novelty Effect



In our observations, we found that the conversion rate for the treatment group remained consistently higher than the control group throughout the experiment, except for a slight dip at one point. This pattern suggests that the new features did enhance the user experience and contributed to an increased conversion rate. Notably, there were no significant spikes in the conversion rate for the treatment group, which one might expect if a novelty effect were in play.

The novelty effect typically manifests as an initial spike in engagement or conversion rates following the introduction of a new feature, followed by a decrease as users lose interest over time. However, our results showed a consistent and sustained improvement in conversion rates in the treatment group, suggesting that the observed improvement was likely due to the actual utility and added value of the new features rather than their novelty.

Power Analysis

Our A/B testing experiment included a total of 48,943 participants, which is a substantial number. However, after conducting a power analysis using a control conversion rate of 3.92% and a minimum detectable effect (MDE) of 10%, we found that ideally, we would require a sample size of approximately 60.6k to be more confident in detecting the expected changes. Despite this, we managed to reject the null hypothesis in our experiment, which indicates that we found statistically significant differences between the control and treatment groups.

Though our actual sample size is smaller than the calculated ideal, our statistical analysis showed a significant increase in conversion rates for the treatment group. This indicates that our experiment, despite a smaller sample size, was sensitive enough to detect the impact of the new feature.

Considering the practical significance of our findings, the increase in conversion rates translates to an improvement in user engagement, which can potentially boost our revenue. Additionally, since the costs associated with the implementation of the new feature (banner) are expected to be low, the benefit-cost ratio favors the change.

Therefore, while the power analysis provides us with valuable insights about the sensitivity and potential limitations of our experiment, it does not change our recommendation. We still believe the launch of the new banner is a favorable decision. This decision is based on our observed significant increase in conversion rates, practical significance, and alignment with the company's goals to enhance user experience and engagement.

Recommendation

Given the encouraging conversion rate improvement, we suggest deploying the new features. Banners aren't associated with high costs, so a substantial metric increase justifies its launch. But since there's no significant difference in average spend, further feature adjustment could be explored to encourage user spend.

In conclusion, the experiment's success in improving the conversion rates proves the effectiveness of the new features. However, future iterations should focus on enhancing the average user spend.

Appendix

Data Extraction Query

```
SELECT
  u.id AS user_id,
  u.country,
  u.gender,
  g.device AS visit_device,
  a.device AS purchase_device,
  g.group,
  CASE WHEN SUM(a.spent) > 0 THEN 1 ELSE 0 END AS converted,
  COALESCE(SUM(a.spent), 0) AS total_spent
FROM
  users u
LEFT JOIN
  groups g ON u.id = g.uid
LEFT JOIN
  activity a ON u.id = a.uid
GROUP BY
  u.id,
  u.country,
  u.gender,
  g.device,
  a.device,
  G.group;
```

Novelty Effect Query

```
SELECT
  groups.join_dt,
  groups.group,
  COUNT(*) AS total_users,
  SUM(CASE WHEN activity.spent > 0 THEN 1 ELSE 0 END) AS
  total_conversions, SUM(activity.spent) AS total_amount_spent
```

```
FROM
  groups
LEFT JOIN activity ON activity.uid = groups.uid
GROUP BY
  groups.join_dt,
  groups.group
ORDER BY
  groups.join_dt ASC;
```

Power Analysis Calculator

https://www.statsig.com/calculator

Excel Spreadsheet

https://docs.google.com/spreadsheets/d/1ZTPVX8vRkVtU1eTZSNDiudKOsJu-LIqp7K-gzR6U4FQ/edit#gid=248000973