GloBox A/B Test Analysis

By: Levon Sebuhyan

Summary

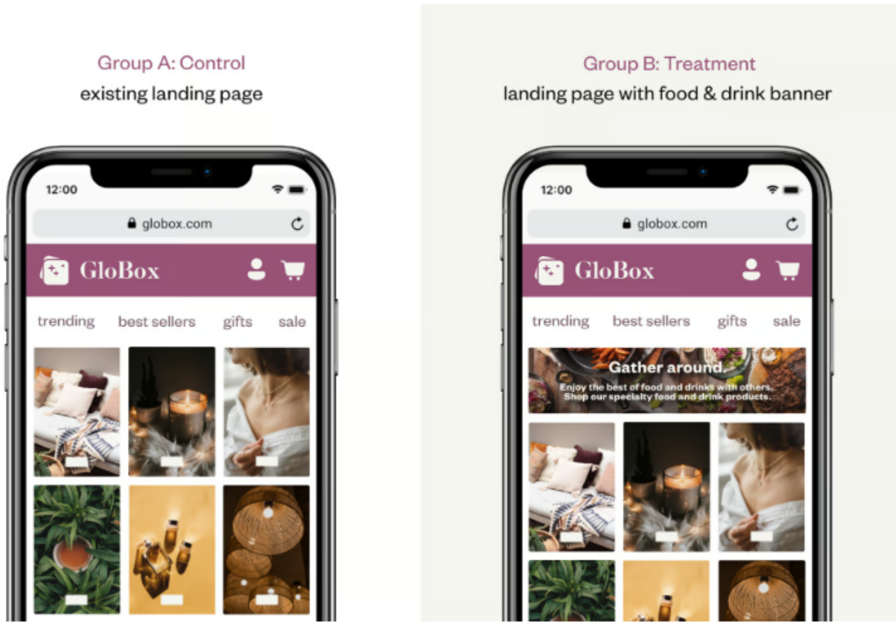
I recommend that we launch the new homepage because even though we did not observe evidence of a change in revenue, the significant change in the conversion rate proves that still is a good start. Maybe with additional future changes, we can make them generate more revenue later.

Context

Motivation

The Growth Product & Engineering Team is looking to compare two versions of a landing page for GloBox's website. The reason for this is that among Globox's customer base, they are primarily known for boutique fashion items and high-end decor. However, their food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue.

We ran an A/B test with a new design for the homepage to see if it would increase revenue. We can see the difference between the two designs below. The control group saw the old design, the treatment group saw the new design.



Test Groups

The test groups are broken into two categories; Those who didn't see the banner (control, 24,343 users) and those who did see the banner (treatment, 24,600 users).

Test Parameters

The experiment was run for 13 days from January 25 to February 6th. We also took into consideration whether the users were using Android, iOS, or some other device and which region the users were from: Australia, Europe, South America, or North America.

Success Metrics

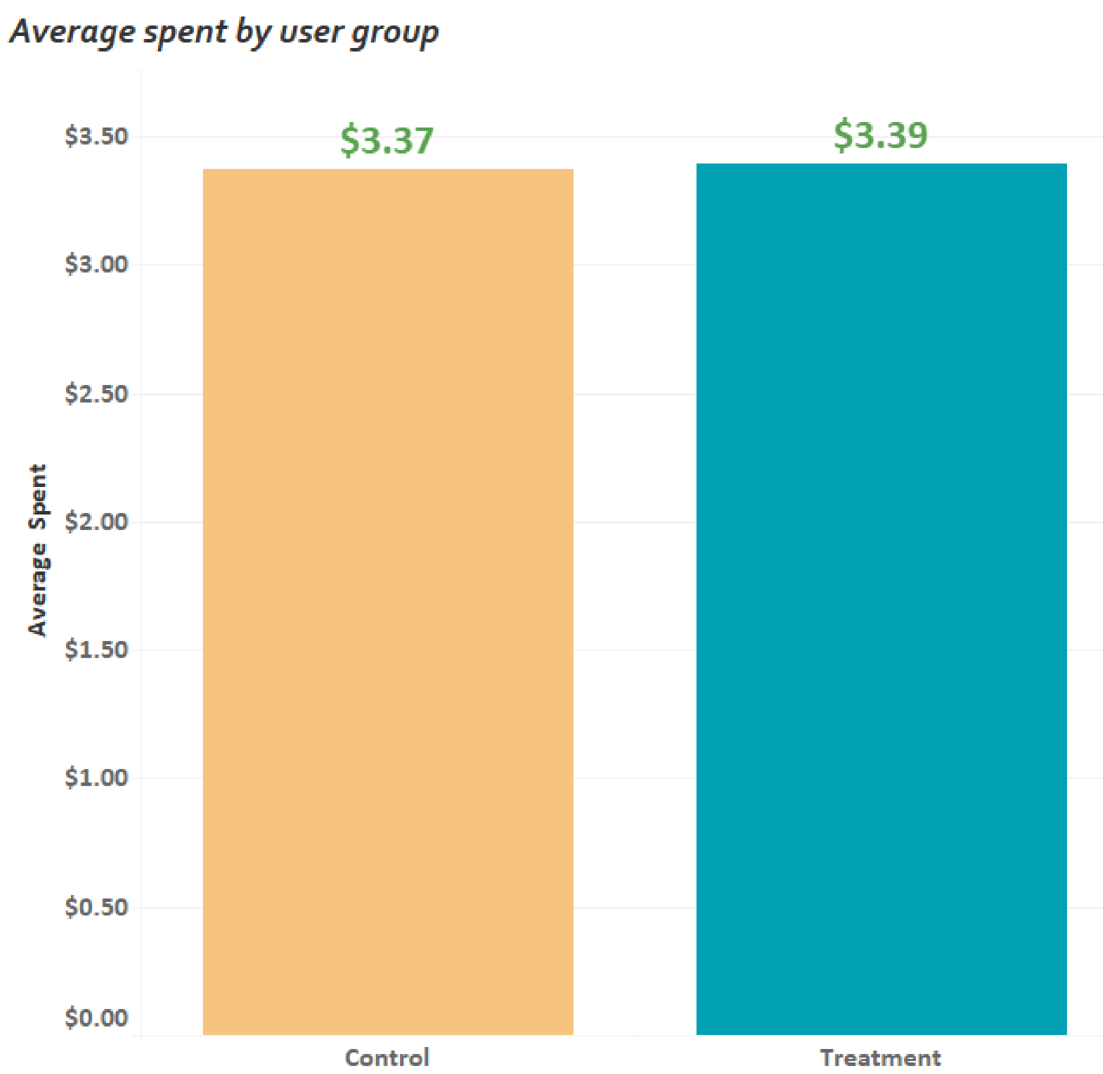
The success of our experiment is determined by two metrics, conversion rate and average amount spent per user. We need to see a significant rise in both or one of the metrics, but more importantly, the average amount spent per user, to justify launching the banner.

Results

To determine whether there was a difference in revenue per user and a difference in conversion rates, we ran a hypothesis test.

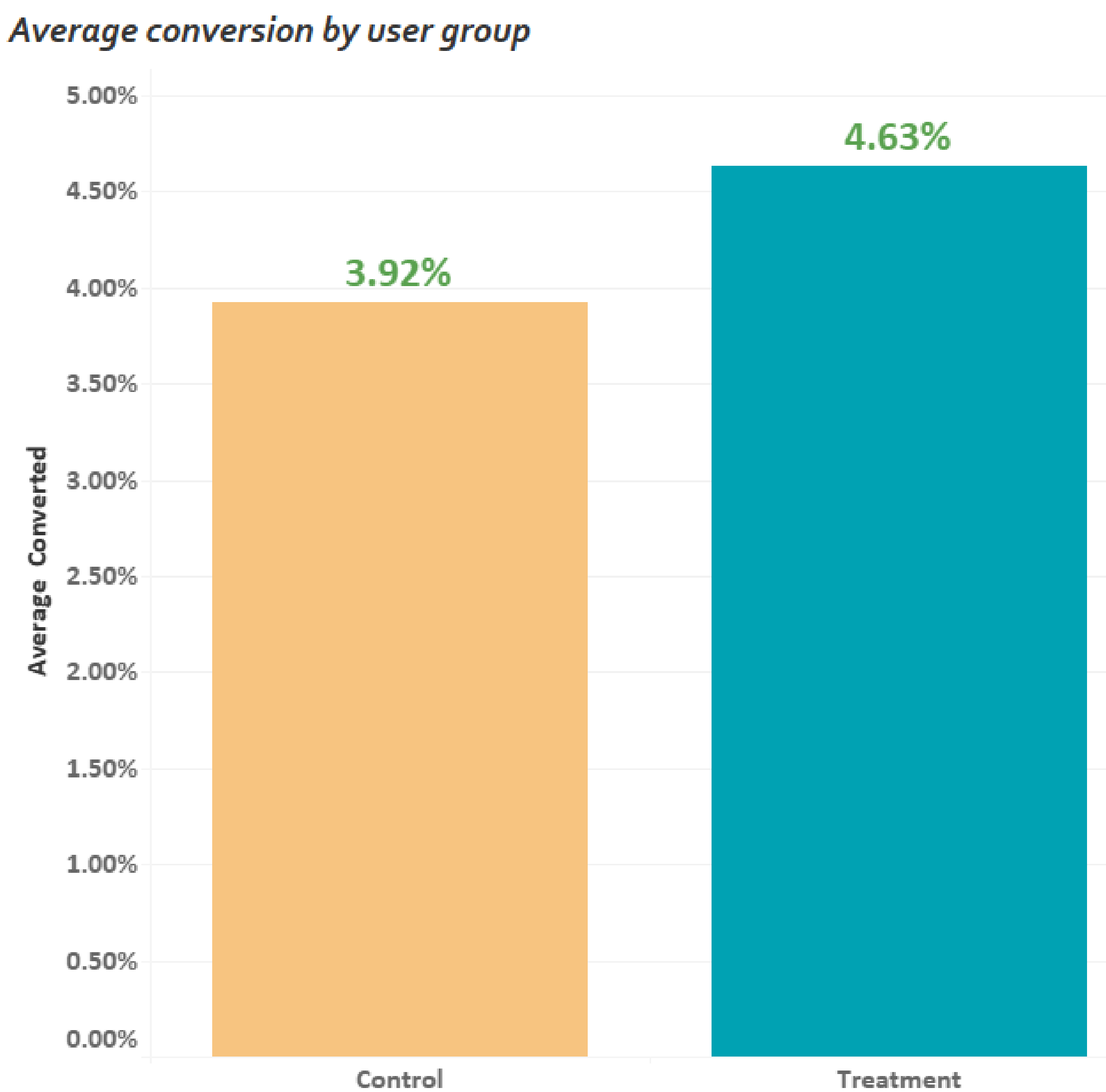
Hypothesis Tests for the Difference Between Average Spent

We used a two-tailed student's t-test for means to see if we could reject our H0 (µA = µB). With p-value = 0.94, we failed to reject H0. In addition to our t-test, the below chart indicates that the average spent difference between the control and treatment groups is insignificant with the **treatment** group's average spent being 1% higher than the **control** group. These numbers show us that our changes are not working to the average spent end.



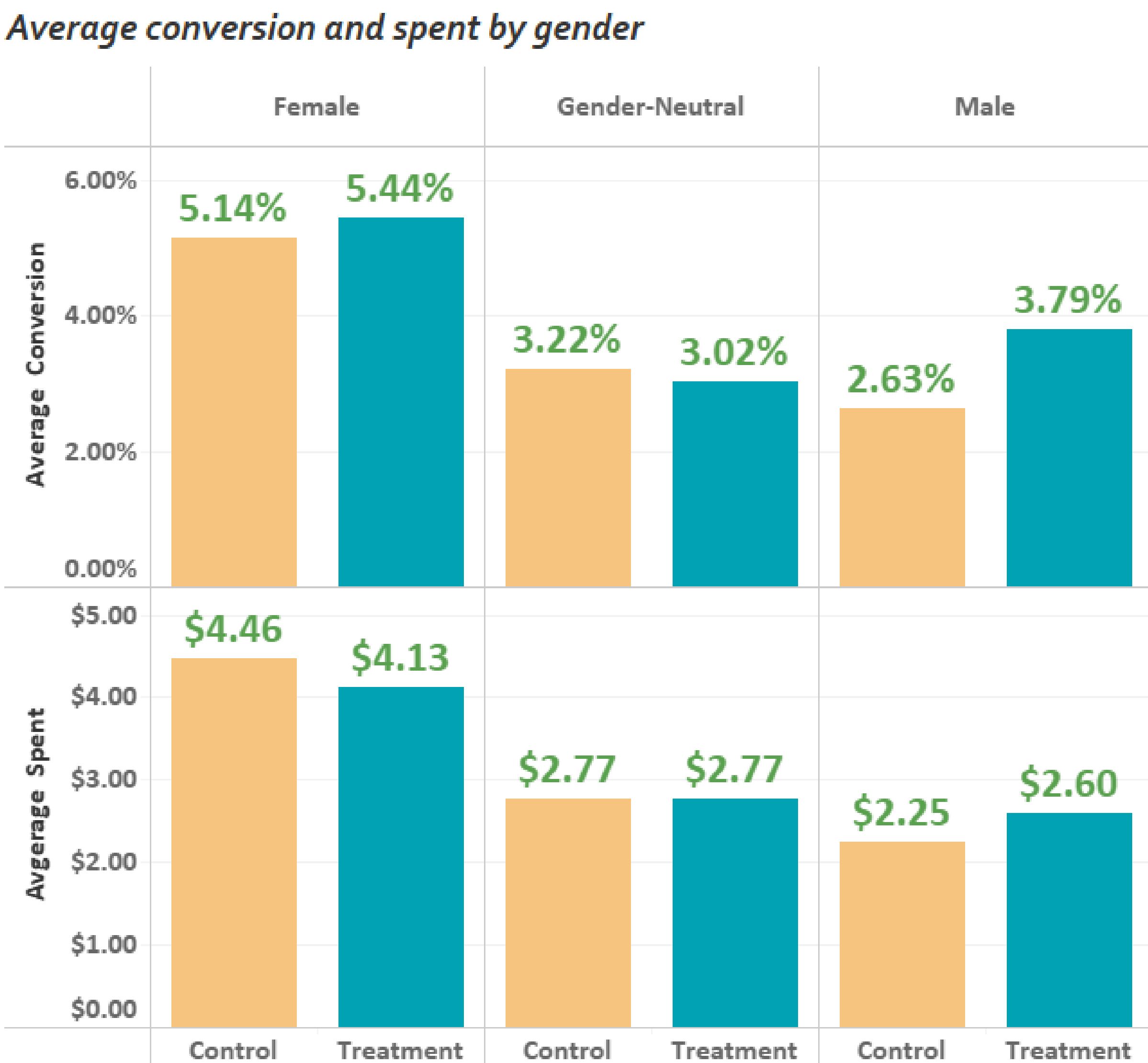
Hypothesis Tests for the Difference Between Conversion Rates

We used a two-tailed z-test for proportions to see if we could reject our H0 (µA = µB). With p-value = 0.0001, we reject H0. In addition to our z-test, the below chart indicates that the average conversion difference between the control and treatment groups is significant with **treatment** group conversion being **18%** higher than the **control** group. These numbers also show us that the changes we have made are working to the user conversion end.



Additional Findings

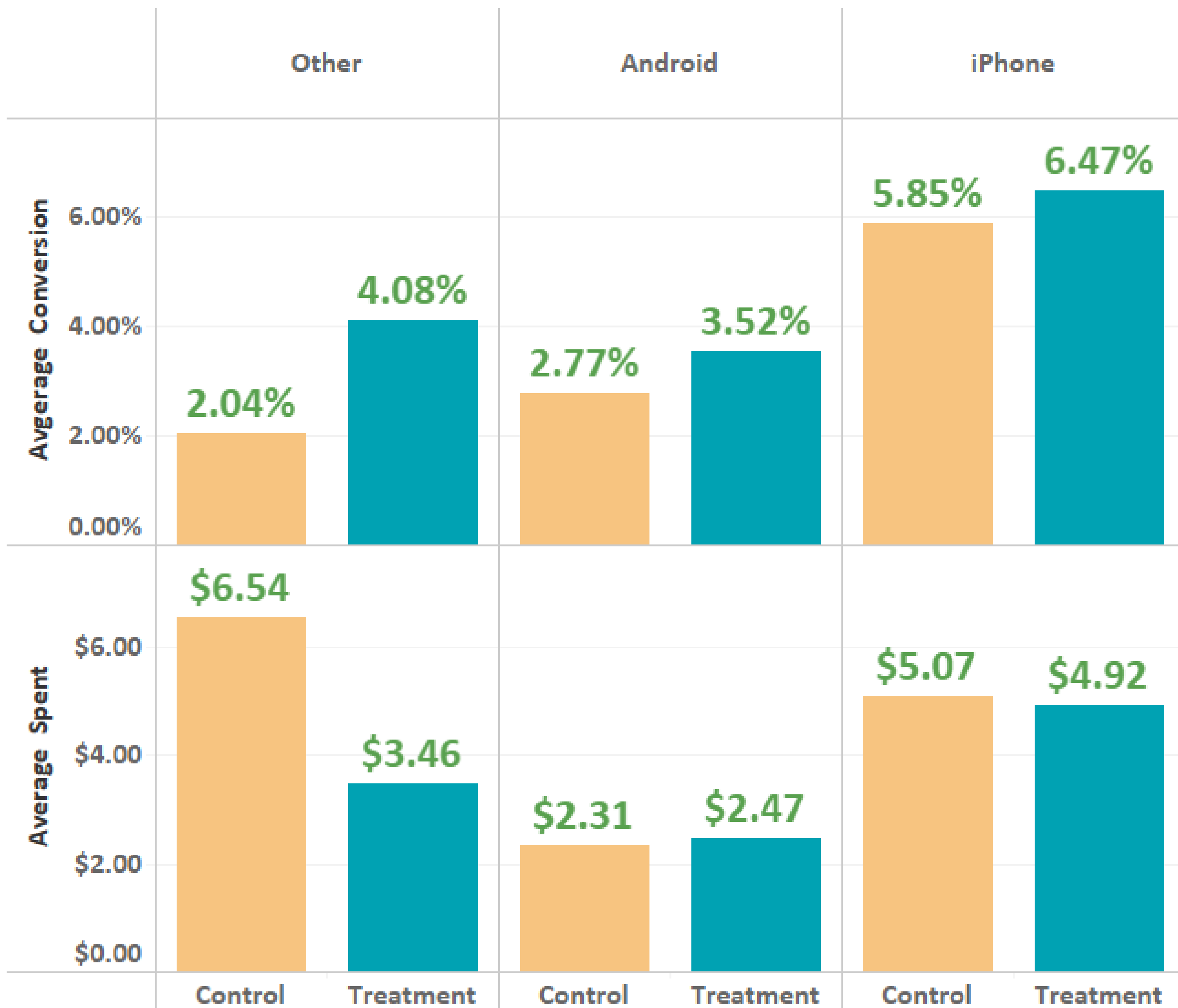
The below chart indicates that male users have a remarkable 44% increase in conversion rate. And a significant increase on average spent by 15%. On the other hand, the impact on female users was less. While females are gaining a 6% conversion rate boost, on the other hand, they gave up on average spending by 8%.



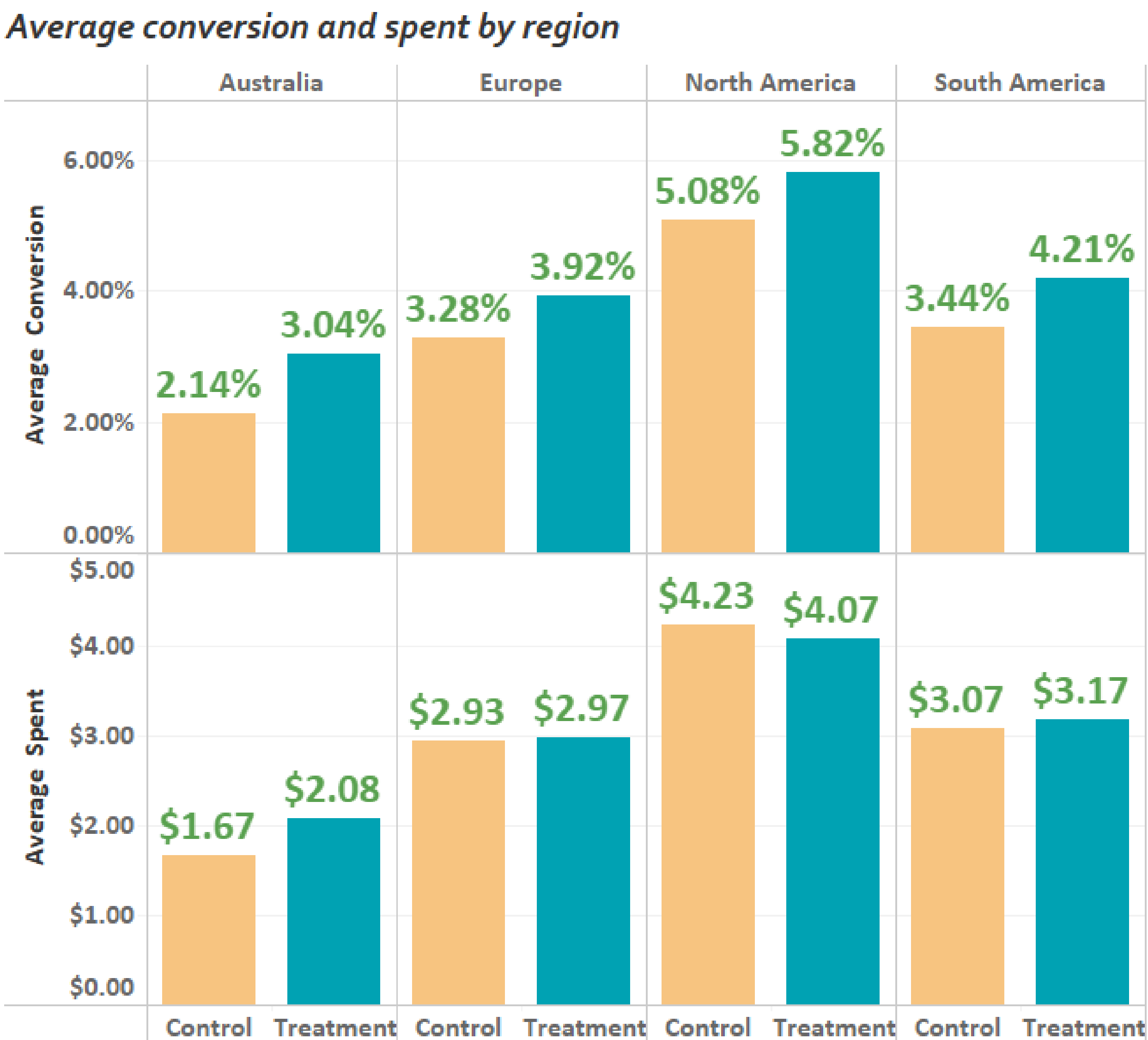
When it comes to findings on devices, the banner had a positive impact on iPhone users, while the treatment group gained an approximately 11% boost in conversion rate, unfortunately, It had a slight decrease in the average amount spent among iPhone users, a decline of approximately 3%.

On Android devices, the treatment Group indicates an uplift of approximately 27% in conversion rate among Android users. It also had a positive impact on the average amount spent among Android users, with the treatment Group showing an increase of 7% in the average amount spent.

Average conversion and spent by device

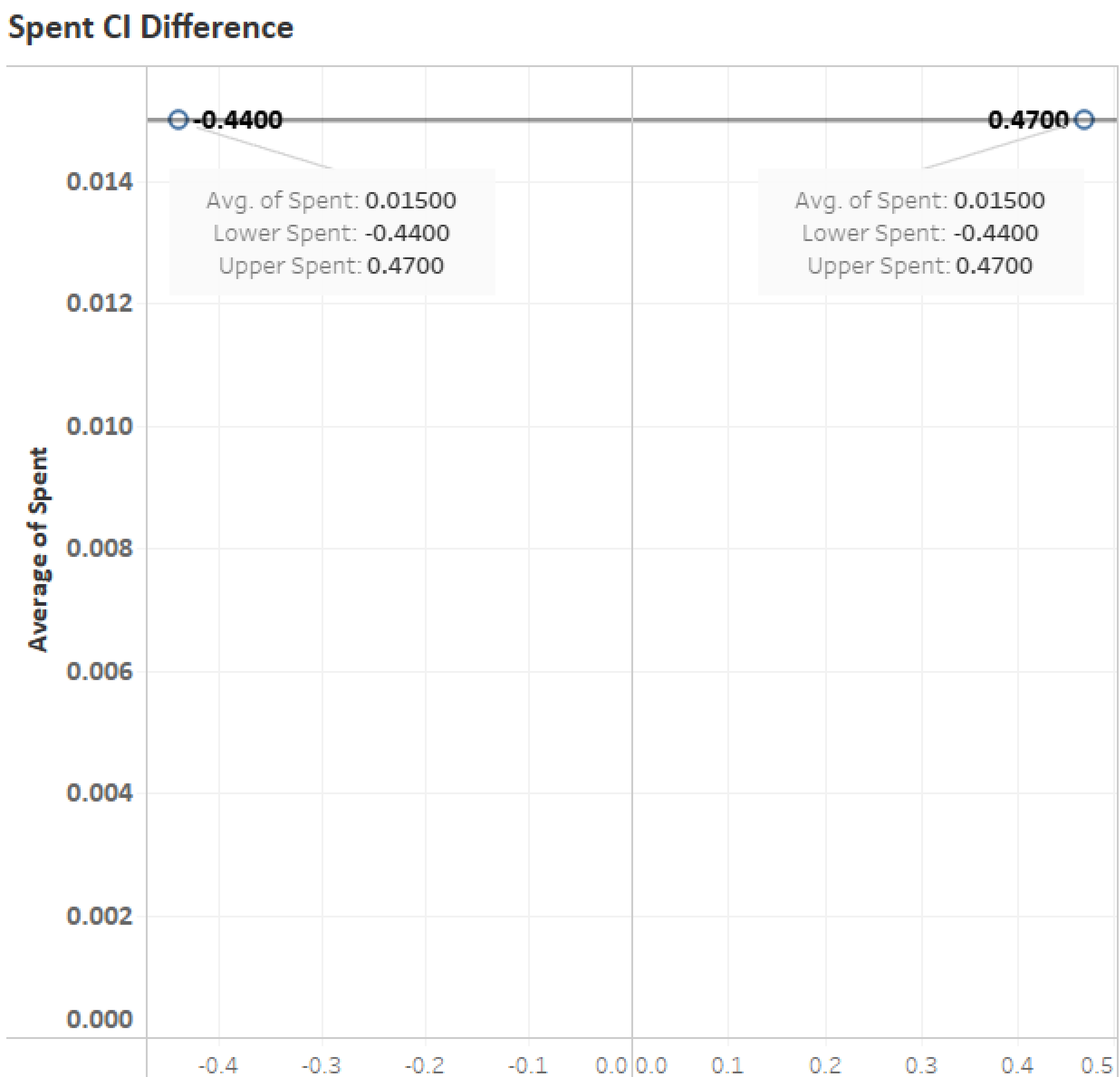


Last but not least when we scoop over regions we see Australia is the clear winner with a 42% increase in conversion rate and a 24% gain in average spent. This was followed by South America, where the treatment group showed a significant 22% improvement in conversion rate, followed by Europe, where the treatment group also experienced a positive impact with a 20% gain in conversion rate.

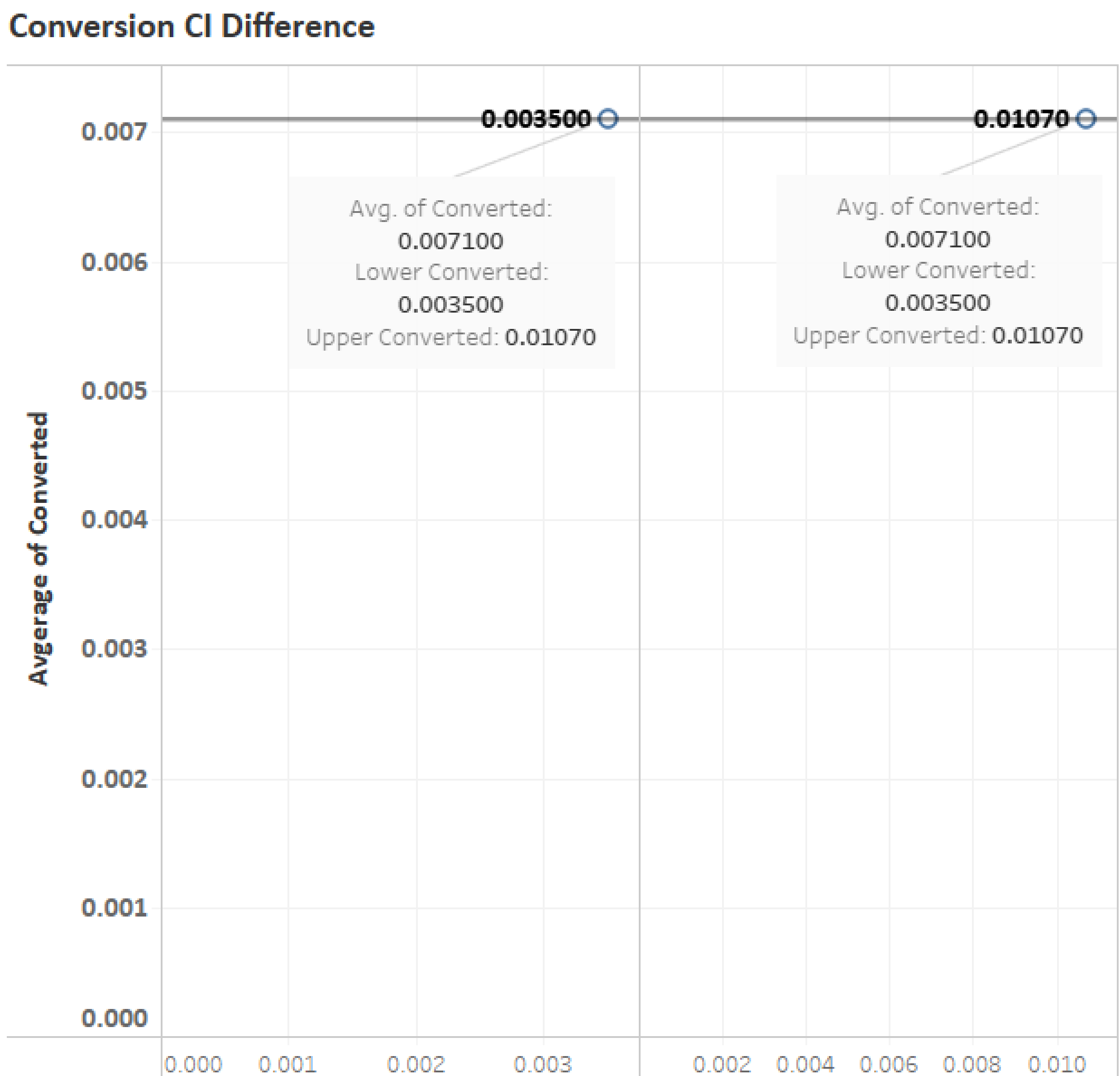


Confidence Intervals for the Differences Between the Two Groups

Based on our confidence interval tests, the average amount spent between the two groups ranges between **-$0.44** and **$0.47**. Also, we estimated that the difference in conversion rate between the two groups ranges between **0.35%** and **1.07%**.



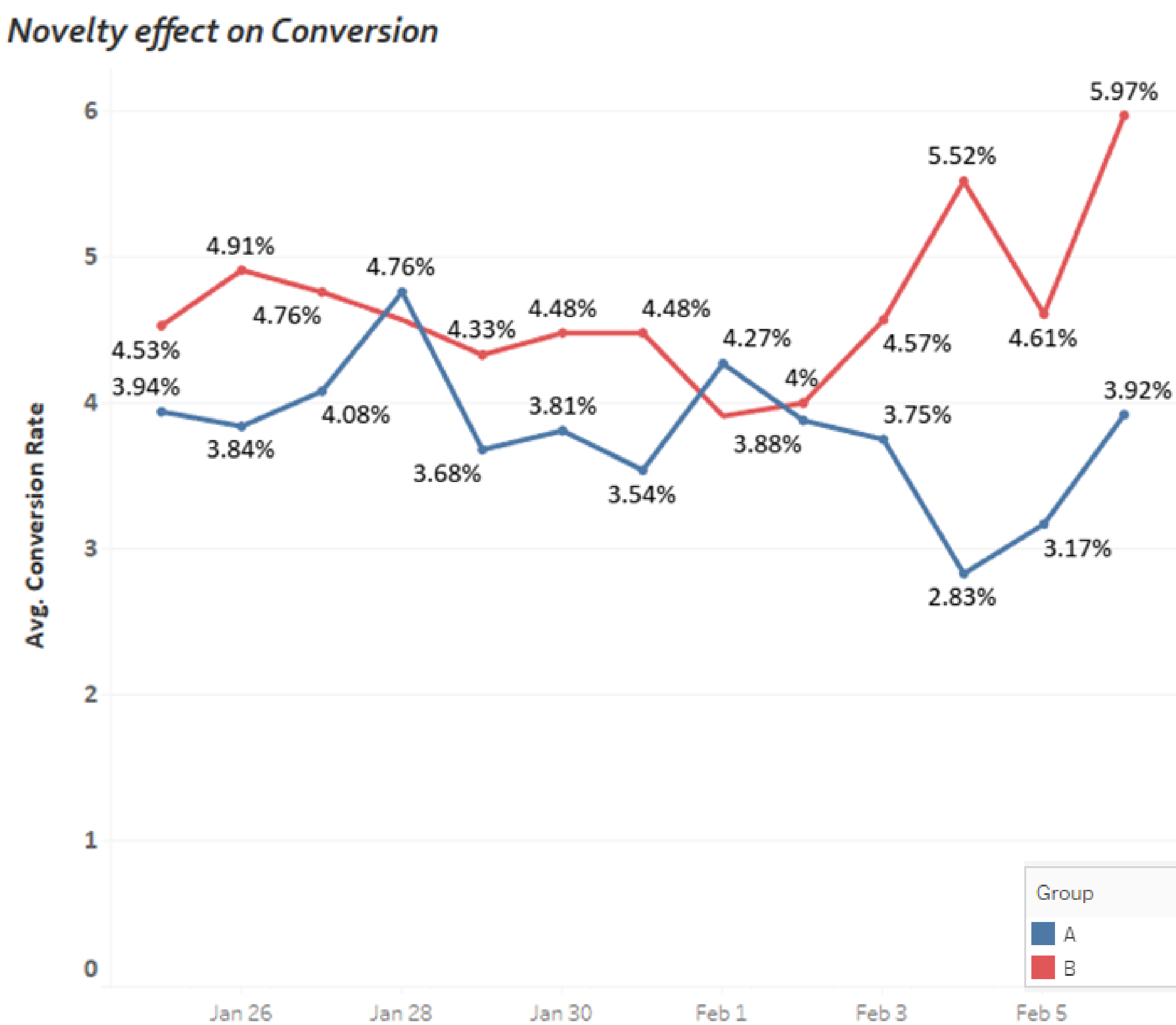
The above graph shows the confidence interval difference for the average spent by users.



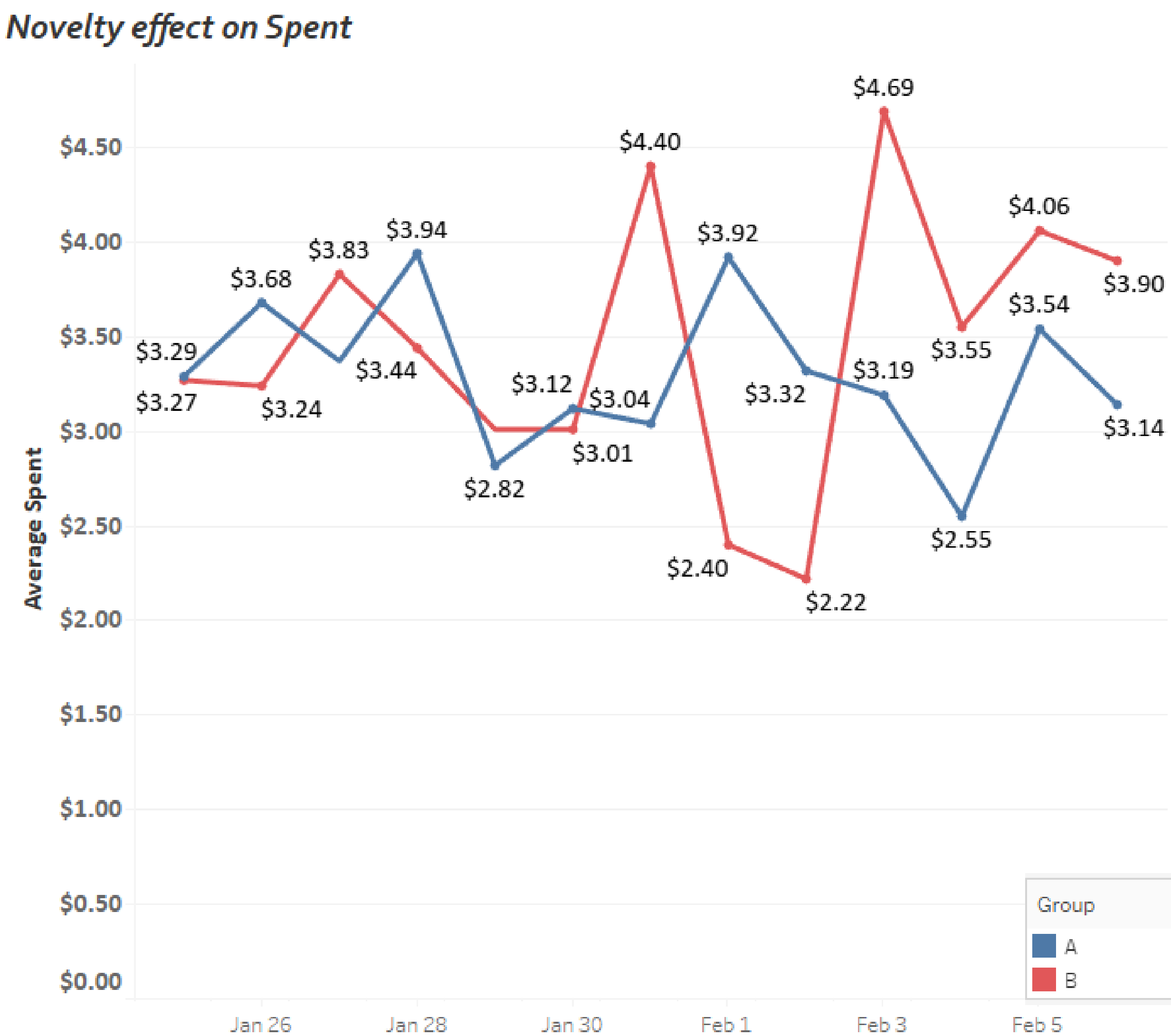
The above graph shows the confidence interval difference for the conversion rate by the users.

Novelty Effect

Conversion rate by all users: The below charts show the conversion rates for all users in both groups over time. As you can see conversion rates do not show any significant differences between the Control and Treatment groups. Even though there are minor changes, still there is no clear indication of a novelty effect affecting the conversion rates.



Average amount spent by all users: The below charts show the average amount spent by all users rates for all users in both groups over time. As you can see the average amount spent does not show any significant differences in time. While there are minor changes, the differences between the control and treatment groups are not large enough to suggest a novelty effect.



Power Analysis

Through power analysis, it was revealed that we did not reach the minimum sample size of **182,164** to detect a 5% change in the average amount spent per user. Also, Power analysis recommends we have a sample size of **60,600** users to detect the effect of a 10% conversion rate. Our observed conversion rate is even larger (18%) which makes the reliability of the results even more questionable.

Recommendation

DO NOT LAUNCH!

We have a good start with the conversion rate being 18% higher in our test group. However, the change in average spent was not significant enough (p-value being 0.94).

Even though there is sufficiently strong evidence for an increase in conversion rate (p-value being 0.0001) in order to recommend launch we rely on a significant change in both metrics. Having almost no change in the average amount spent per user indicates changes we are trying to make are not helping on that metric. So in conclusion, we recommend that you do not launch the banner.

However, during our tests, we encountered interesting evidence for us to consider suggesting to tailor targeted marketing in one or more combinations of the following fields;

- Male users with a substantial 44% increase in conversion rate
- Australia with a 42% increase in conversion rate and 24% in average spent
- Europe with a 20% increase in conversion rate
- South America with a 22% increase in conversion rate
- Android users with a 27% increase in conversion rate and 7% in average spent

Appendix

Link to spreadsheet

https://docs.google.com/spreadsheets/d/1vhY562IyIbTScWvTkiUQiBuT1ohamO5E80_Kk7Qp8Rc/edit#gid=645741699

Link to Tableau

https://public.tableau.com/app/profile/levon.sebuhyan/vizzes

SQL Queries;

The short query for Tableau:

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

Detailed query to see everything separately:

```
SELECT
  g.uid AS user_id,
  CASE
    WHEN u.country IS NULL THEN 'N/A'
    ELSE u.country
  END AS user_country,
  CASE
    WHEN g.group = 'A' THEN 'A'
    ELSE '-'
    END AS user_group_a,
  CASE
    WHEN g.group = 'B' THEN 'B'
    ELSE '-'
    END AS user_group_b,
  CASE
    WHEN u.gender IS NULL THEN '-'
    ELSE u.gender
  END AS user_gender,
  CASE
    WHEN a.device IS NULL THEN '-'
    ELSE a.device
  END AS user_device,
  CASE
    WHEN g.group = 'A' and a.spent > 0 THEN 1
    ELSE 0
  END AS user_converted_a,
  CASE
    WHEN g.group = 'B' and a.spent > 0 THEN 1
    ELSE 0
  END AS user_converted_b,
  COALESCE(SUM(spent) FILTER (WHERE g.group = 'A'), 0) AS user_spent_a,
  COALESCE(SUM(spent) FILTER (WHERE g.group = 'B'), 0) AS user_spent_b,
  COUNT(DISTINCT CASE WHEN g.group = 'A' THEN a.uid END) AS user_spent_count_a,
  COUNT(DISTINCT CASE WHEN g.group = 'B' THEN a.uid END) AS user_spent_count_b,
  COUNT(DISTINCT CASE WHEN g.group = 'A' THEN g.uid END) AS total_user_a,
  COUNT(DISTINCT CASE WHEN g.group = 'B' THEN g.uid END) AS total_user_b
 FROM
  users u
LEFT JOIN
  groups g ON u.id = g.uid
LEFT JOIN
  activity a ON u.id = a.uid
GROUP BY 1, 2, 3, 4, 5, 6, 7, 8;
```

Query for the Novelty effect:

```
SELECT
    n.join_date,
    n.group,
    ROUND(CAST(SUM(n.paid_users) / MAX(n.total_users) * 100 AS
    DECIMAL(10,2)), 2) AS conversion_rate,
    ROUND(CAST(SUM(n.total_spent)/MAX(n.total_users) AS DECIMAL(10,2)),2) AS
    average_spent
FROM(SELECT
        g.join_dt AS join_date,
        g.group,
        COUNT(DISTINCT g.uid) AS total_users,
        COUNT(DISTINCT a.uid) AS paid_users,
        SUM(a.spent) AS total_spent
    FROM
        groups AS g
    LEFT JOIN activity AS a ON g.uid = a.uid
    GROUP BY
          g.group,
          g.join_dt
    ORDER BY 1) AS n
GROUP BY 1, 2;
```