GloBox A/B Test Analysis

Impact of New Banner on Mobile Website

Made by: Robert Tiger

Aug', 2023
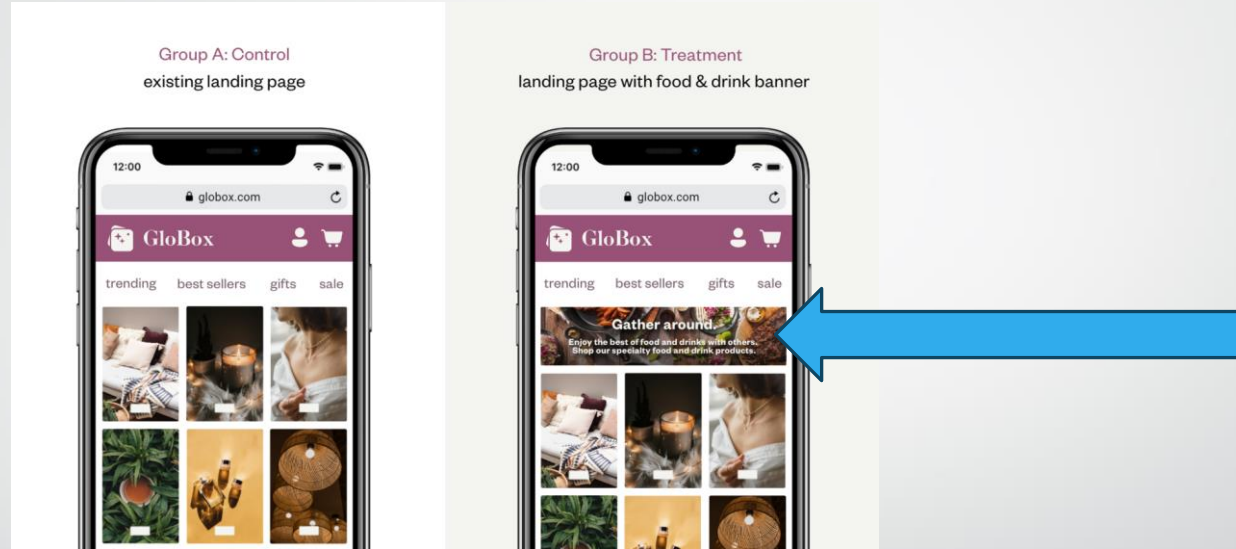
Dates of the test:
January 25th – February 6th , 2023

Control Group (Group A)
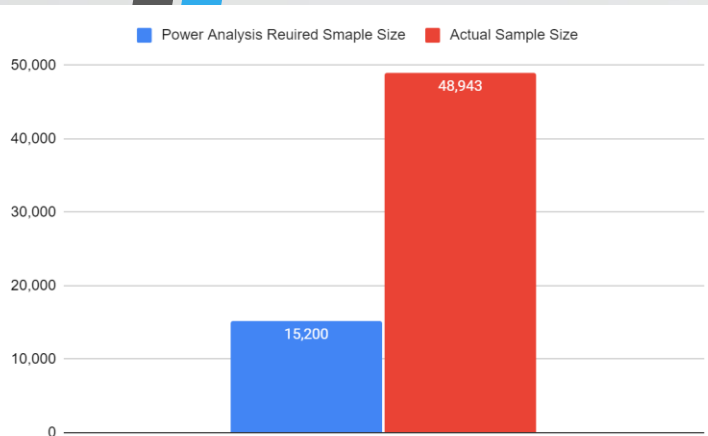Were presented with current site, without the new banner

Treatment Group (Group B)
Were presented with the new banner

The users were assigned randomly into their groups

A/B Test Sample Sizes: Ensuring Reliable Results

Big Sample Size = Reliable Results

Power analysis - **7,600 participants** in each group.

Test – More than **20,000 participants** in each group, totaling more than **40,000 participants**.

What it means ?

We can trust our test results even more because we had big enough groups to be really sure in what we found.

Data Collected

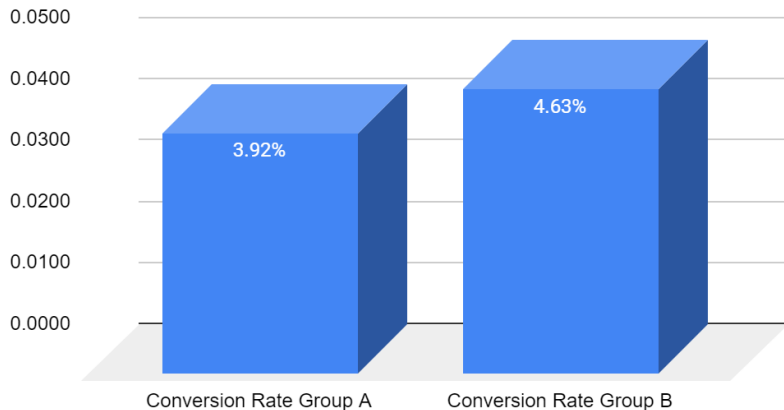**Date** **Amount Spent** **Device Type** **Gender** **Country**
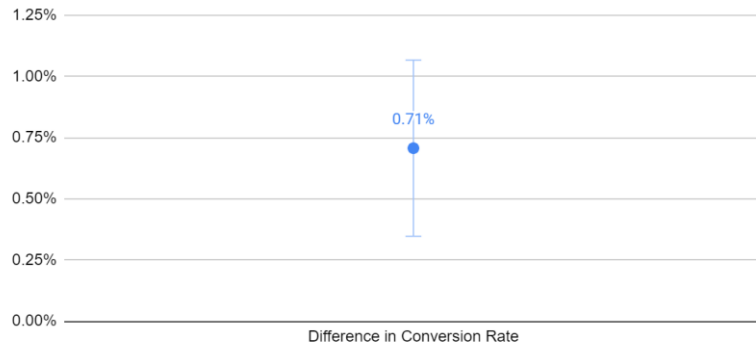
Difference in Conversion Rate

Conversion Rate (%)

18.11% percent increase

Confidence Interval (95%)

0.71%

True difference falls between **0.3%** to **1.06%** increase at a **95% certainty**.

Our tests made us clear that **there is** a significant difference between the groups.

This is not luck!

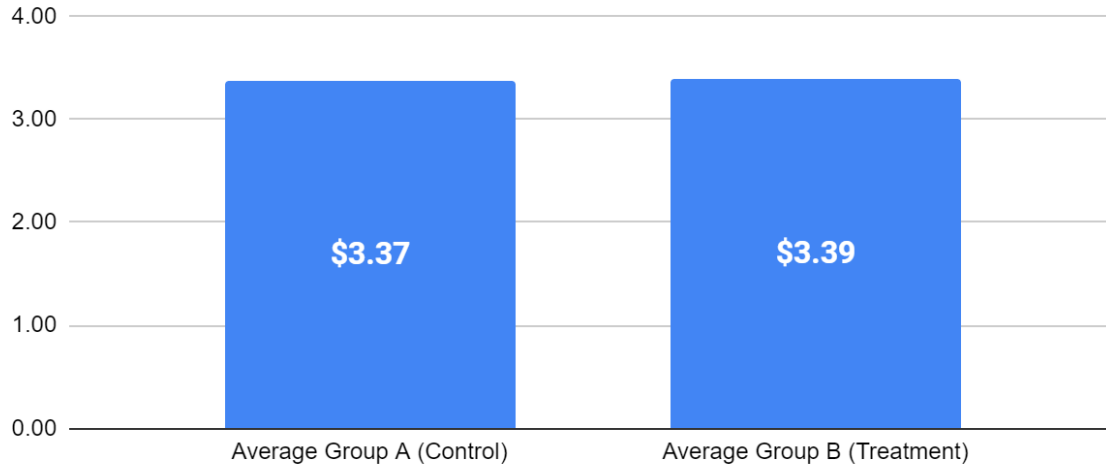# Average Spent per User

Average Amount Spent ($)

p-value of **0.9439:** higher than **0.05**.

No Significant Difference: No strong evidence to say the changes have made a real impact on the average spending per user in the two groups.

! **our analysis doesn't support the idea that the treatment has significantly increased the average spending per user between the two groups.**

Confidence Interval Analysis

Average Amount Spent and 95% Confidence Interval

Control   Treatment

$3.37 $3.39

Average Spent

What It Means? This interval represents where we are 95% <u>certain</u> the true difference lies.

95% confidence interval difference in average spending. Ranges from <u>**-0.44**</u> to <u>**0.47**</u>.

<u>**No Clear Difference**</u>**:** Interval includes **negative** and **positive values**. Meaning the change because of the treatment can be positive as we would like, but it can be **also negative** at a 95% confidence level.
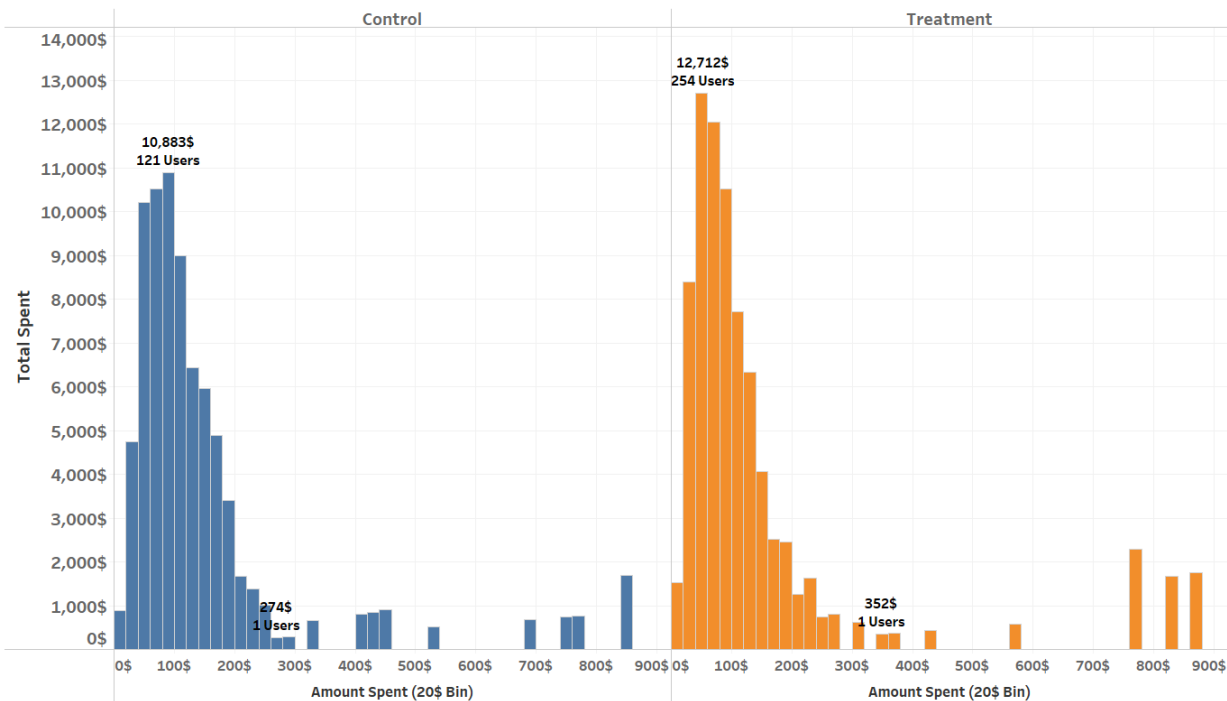
<u>**Conclusion**</u>**:** Considering this interval, caution is advised before proceeding with the new banner launch.

<u>**Not Conclusive:**</u> This analysis doesn't support launching the new banner with confidence in a significant change in user spending.

Amount Spent Distribution

Spending Distribution

4 outliers that spent over $1,000 removed.
Majority of spending falls below this 1,000$.

Both groups exhibit similar spending patterns.
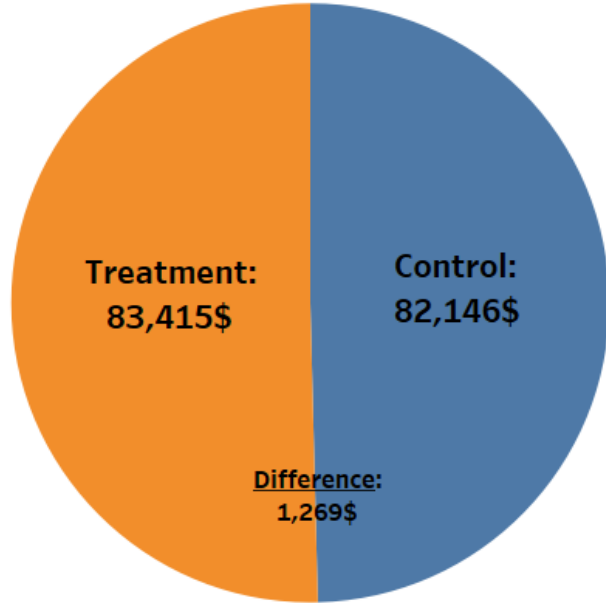Slight edge in spending for Group B.

Analysis shows this difference isn't significant.

Given the lack of statistical significance and the similarities in spending behavior, it is advisable to delay the launch of the new banner.

Total Revenue Between Groups

Treatment: 83,415$
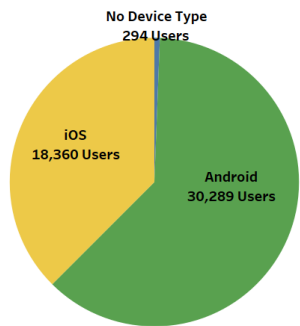
Control: 82,146$

Difference: 1,269$

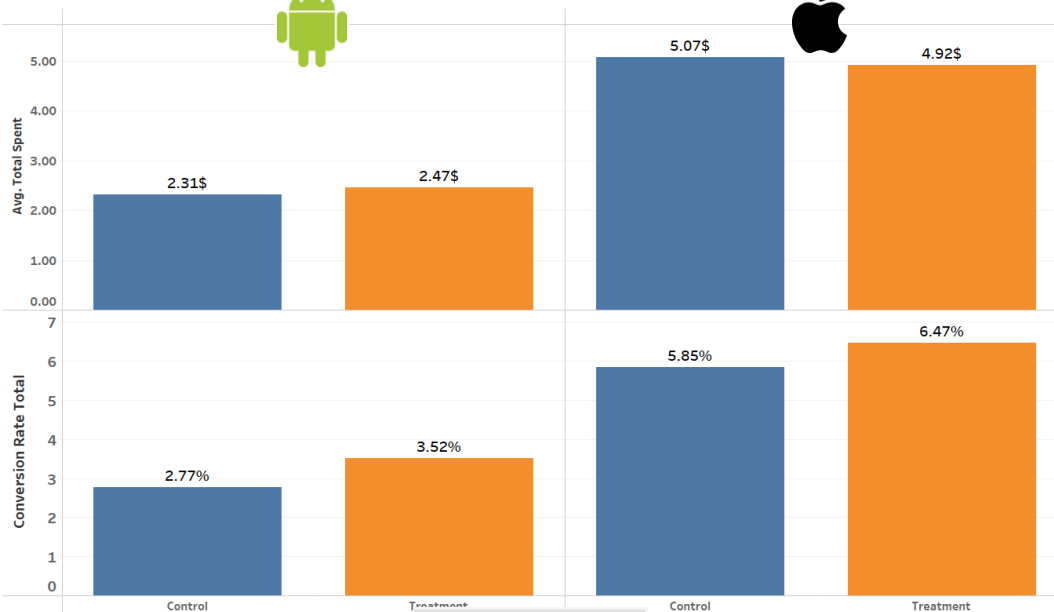**Insufficient Evidence:** Not enough evidence to conclude that the treatment applied to Group B had a significant effect on revenue when compared to Group A.

Random Variation: Results suggests that any difference in revenue may be due to random variation rather than the treatment.

Based on these findings, we should exercise caution in attributing the revenue difference to the treatment.

Platform Comparison - Android vs. iOS

Average Spending per User:
Both groups on Android devices showed **similar results** in terms of the average amount spent per user.

Reduced Spending in Treatment Group:
iOS - **$4.92** average spending.
→ Less than the control group - **$5.07**.

Banner may have had **a negative impact on user spending behavior among iOS users.**

Consider re-evaluating the banner's impact on iOS users and potentially refining the treatment strategy for this platform.

Correlation Between Avg. Spent & Con. Rate

Highest Spending Country: **United States**
Out of **14,772** users:
4.17$ average spent per user
5.44% conversion rate.

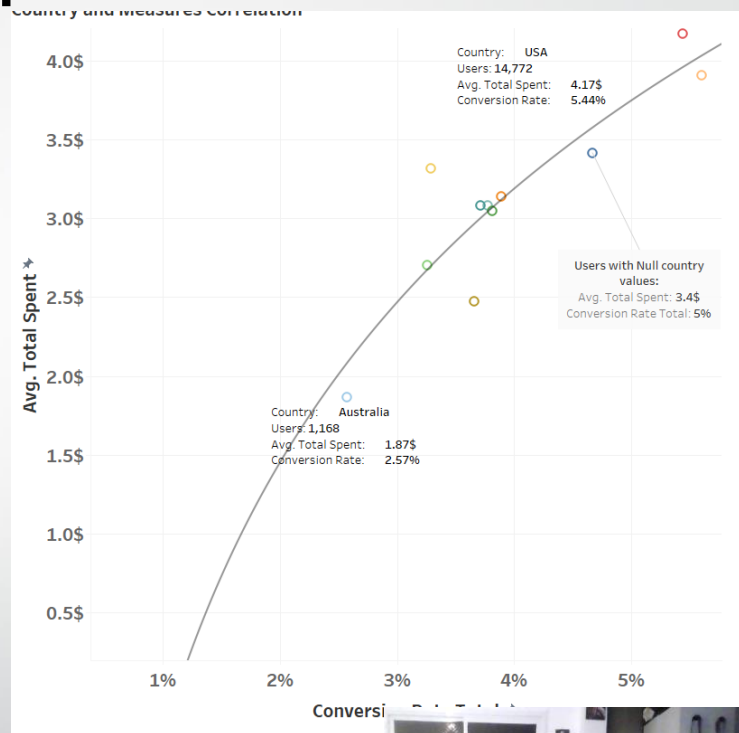Lowest Spending Country: **Australia**

2.5% average spent per user

1.86$ conversion rate.

❗ 643 users were not assigned country.
3.41$ average spent
4.67% conversion rate.
This can cause a slight difference in the data.

Country and Measures Correlation

Country: USA
Users: 14,772
Avg. Total Spent: 4.17$
Conversion Rate: 5.44%

Users with Null country values:
Avg. Total Spent: 3.4$
Conversion Rate Total: 5%

Country: Australia
Users: 1,168
Avg. Total Spent: 1.87$
Conversion Rate: 2.57%

Avg. Total Spent

Conversion Rate Total

UHHH...
I DON'T KNOW
makeameme.org

Gender Comparison
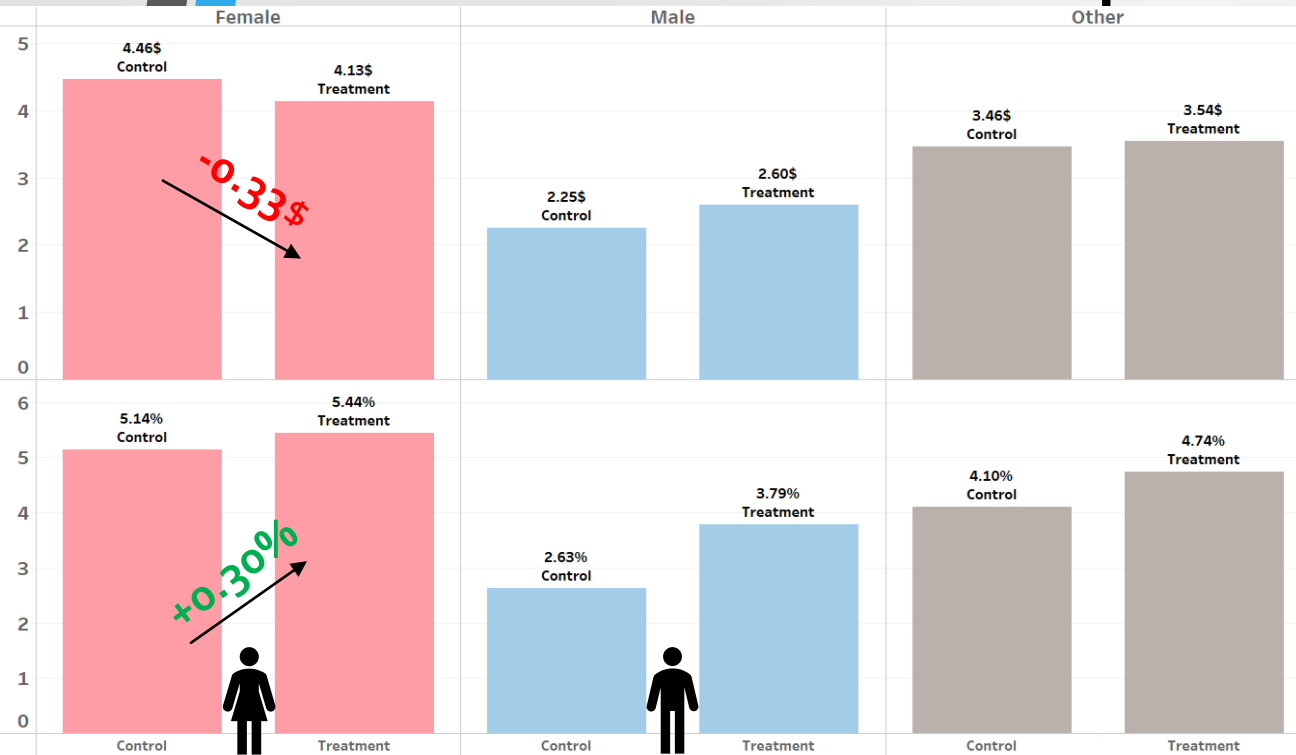
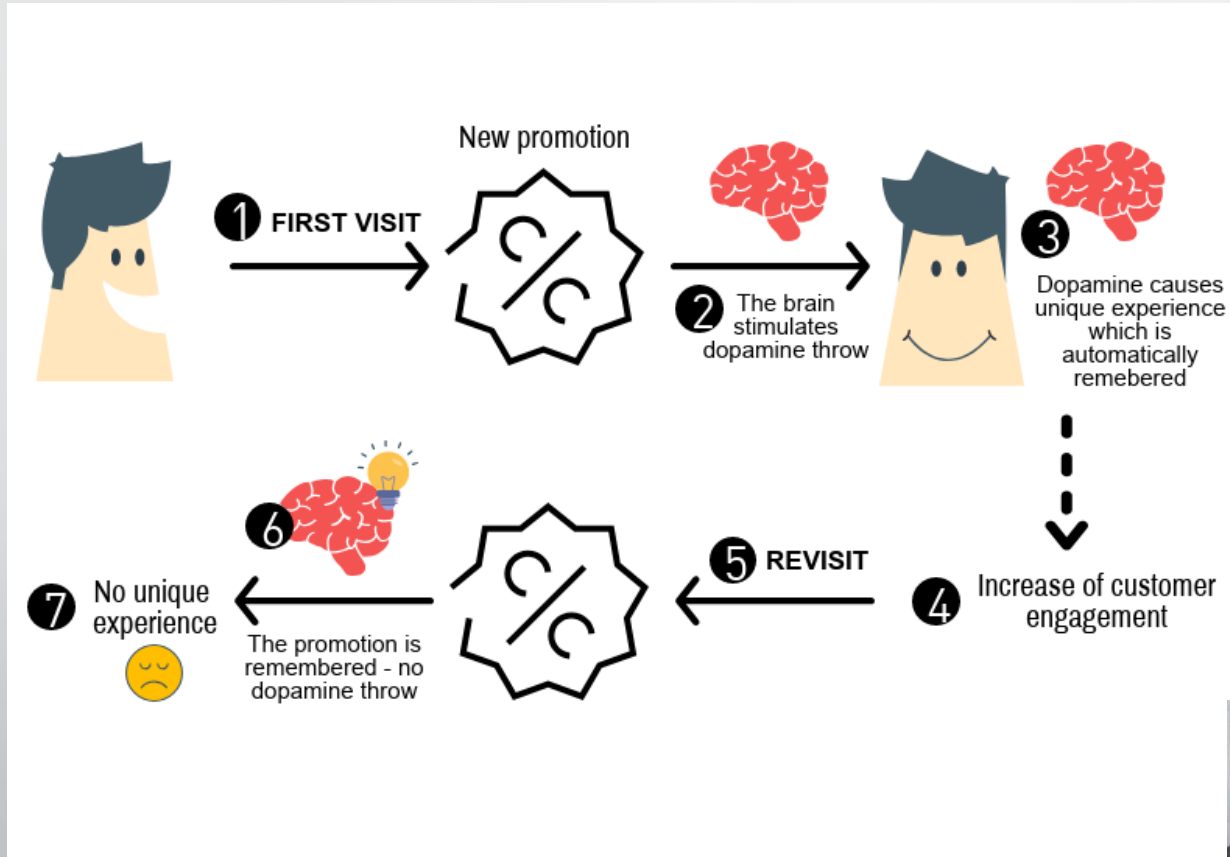Females tend to spend more than males in both the control and treatment groups.

Females showed a **decrease** in **average spending** but an increase in conversions.

'Other' gender, control and treatment groups have similar spending patterns.
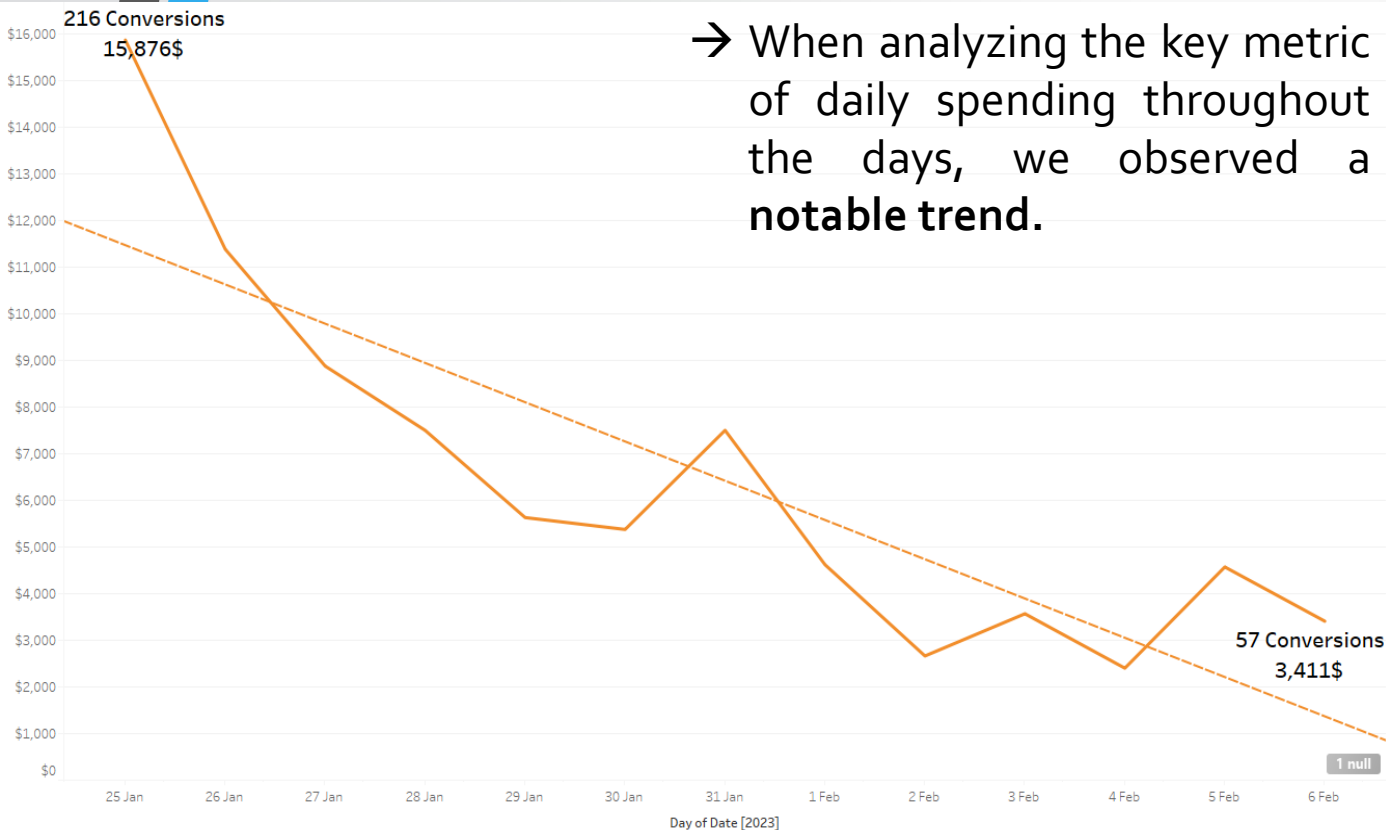
The A/B test results indicate variations in spending behavior among different gender groups. **Further analysis is required.**

Introduction: Novelty Effect

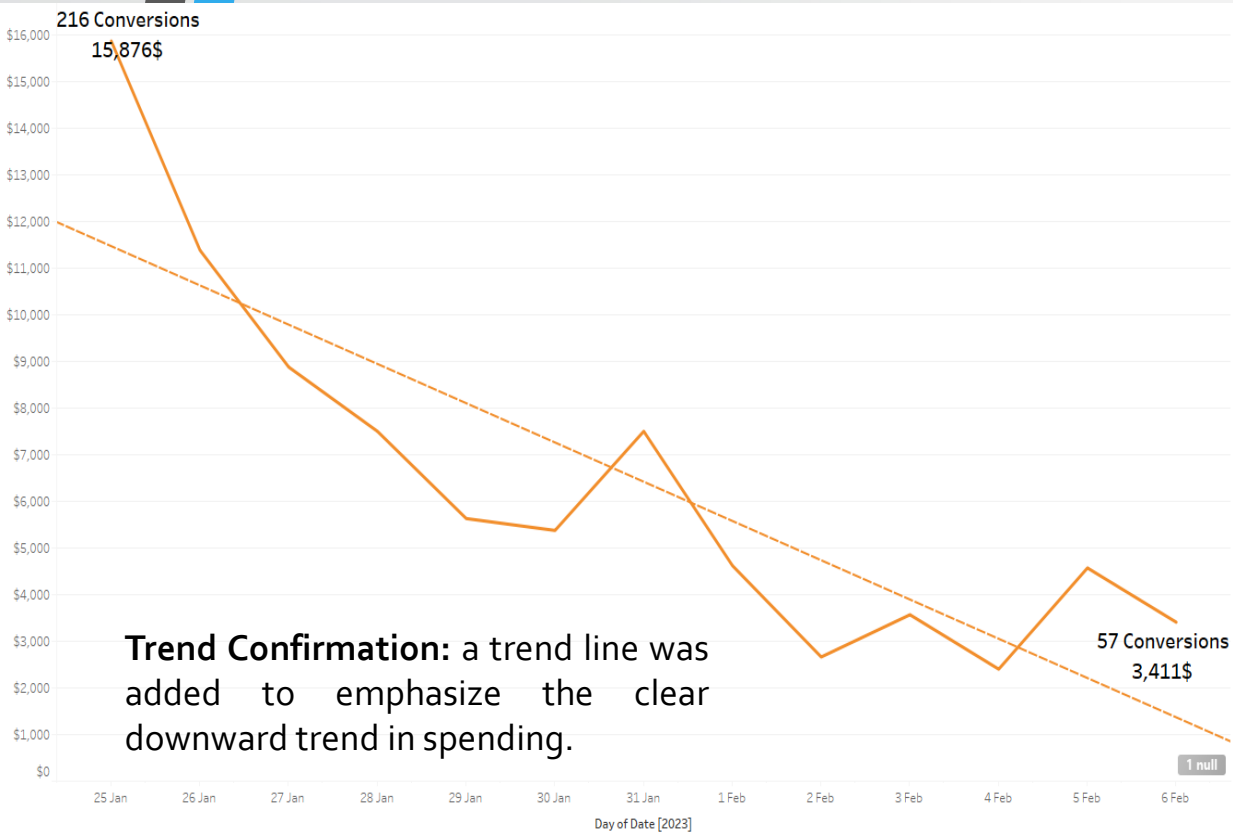# Novelty Effect - User Spending Trends

→ When analyzing the key metric of daily spending throughout the days, we observed a **notable trend.**

Group B showed a **consistent decline in daily spending & conversions** over the course of the experiment.

Clear Decline Pattern

216 Conversions
15,876$

Trend Confirmation: a trend line was added to emphasize the clear downward trend in spending.

57 Conversions
3,411$

1 null

Day of Date [2023]

The **best-performing** days occurred early in the experiment:
Treatment group reached **15,876$** in sales with **216** conversions.

However, the **worst days**, roughly a month later, saw a **significant drop**, with **$2,403** for the day and only **57** conversions.

Recommendations and Future Guidance

Hold Your Horses

Based on the comprehensive analysis of the A/B test results, it is recommended that we **refrain from launching the new banner on the mobile website**.

Key Findings:

1. Increased conversion rates **did not** significantly impact revenue or average spending.
2. **No statistically significant difference in average spending** between the groups.
3. Users may have initially engaged with the new feature but didn't sustain higher spending levels, possibly due to a novelty effect.
4. The A/B test had a larger sample size than needed, ensuring reliable results.

Conclusion

Final Verdict:

<mark>it is advisable to hold off on the widespread implementation of the new feature and consider alternative approaches to achieve meaningful revenue growth.</mark>

REJECTED

Alternative Strategies:

Instead of implementing the new feature universally, the company may want to explore alternative revenue growth strategies that are more likely to yield a substantial return on investment.

Thank you for your time