

GloBox A/B Test Analysis

by

Christopher H. Kroll

A/B Test Setup

We ran an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the test group sees it as shown below:

Group A: Control
existing landing page

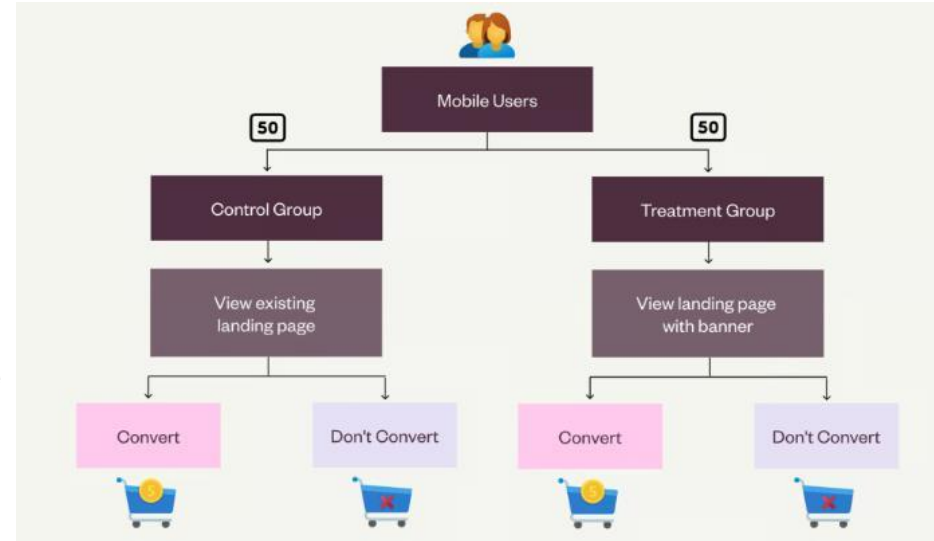


Group B: Treatment
landing page with food & drink banner



A/B test specifics:

- Experiment was conducted only on the mobile website.
- Visitors were randomly assigned to either a control or test group (join date).
- Users in the test group saw a banner, while those in the control group did not.
- A purchase on the website after joining, whether on the same day or later, was considered a "conversion".



Overview of the Experiment

The **goal** for us was to determine:

Should we launch the banner or not?

Characteristics

1. Two Groups ($N = 48,943$)
 - A: Control Group ($n_A = 24,343$)
 - B: Treatment Group ($n_B = 24,600$)
2. Time period (13 days)
 - Begin: 25.01.2023
 - End: 06.02.2023

Key Metrics

1. Conversion Rate (CR)
2. Average Amount Spent (AAS)

Success Criteria

1. At least one metric
 - increased
 - statistically significant
2. No observed negative impact

Execute Summary

The key findings of the A/B test are, in order of importance:

- **Conversion Rate increased largely and significantly**

Suggesting a large positive impact on conversions and potential customers.

- **Average Amount Spent did not change, neither largely nor significantly**

The observed effect was positive, yet very small and statistically insignificant.

There was no negative impact observed.

Stable AAS amidst higher conversion rates suggests a shift from luxury purchases to more frequent, lower-priced food items.

- **Stable Long-Term Conversion Without Novelty Effect**

Initial spikes attributed to sample variability, not novelty.

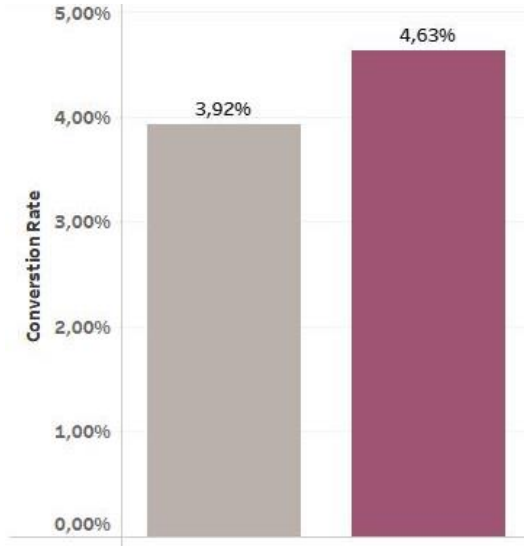
Consistent long-term rates indicate sustained benefits, discounting novelty effect.

- **Power Analysis suggest that the experiment should be run for a longer time period**

Larger Sample numbers needed to be certain about the effects on both metrics.

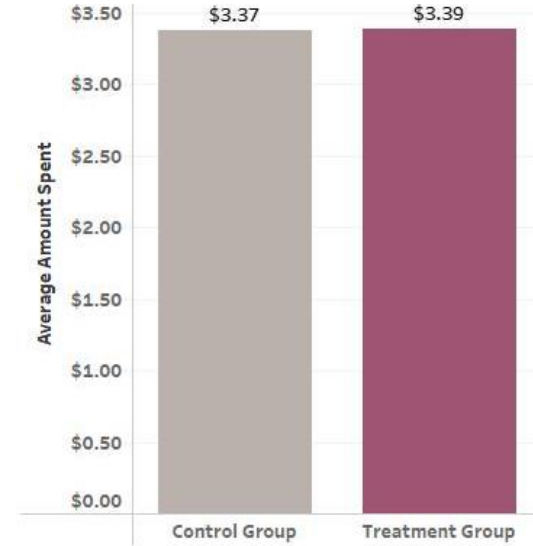
Impact on Key Metrics

Effects on Conversion Rate and Average Amount Spent

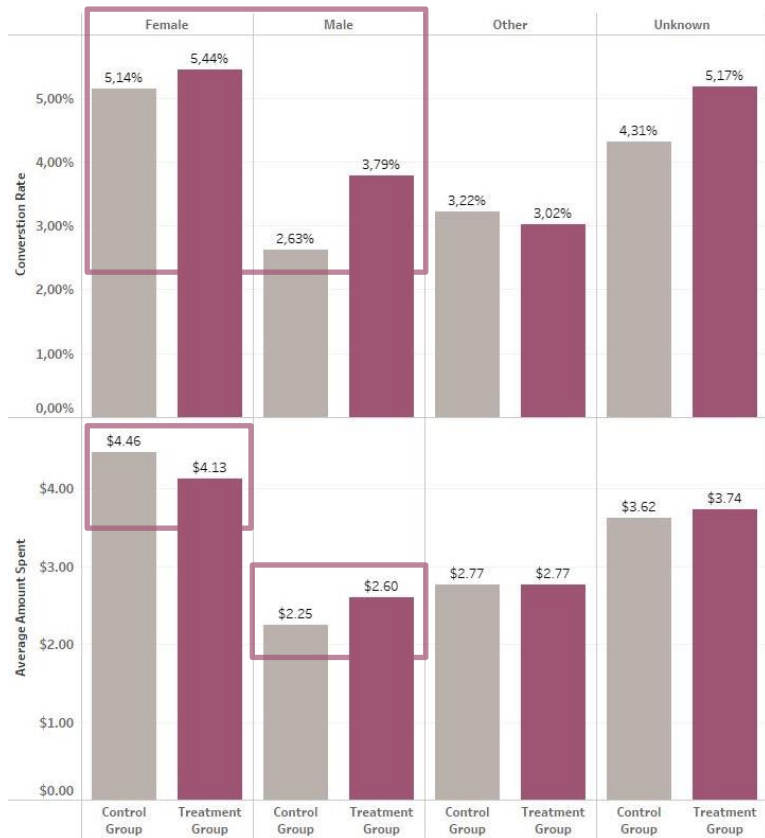


Takeaways

- Large, significant positive effect on Conversion Rate ($\Delta = 0.71$)
- Very small, insignificant, positive effect on Average Amount Spent ($\Delta = 0.02$)
- Shifting Buyer Behavior: Stable AAS amidst higher conversion rates suggests a shift from luxury purchases to more frequent, lower-priced food items, indicating potential product cannibalization.



Gender Segmentation



Takeaways

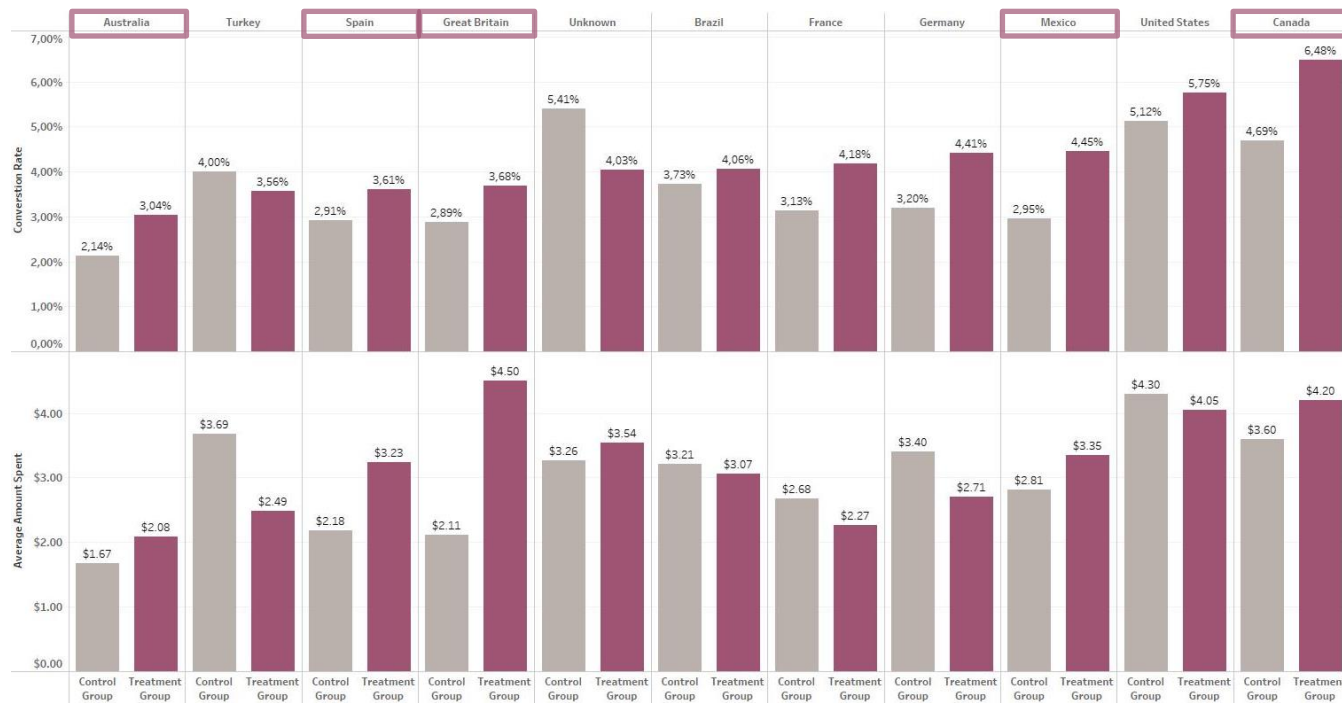
- Positive effect on Conversion Rates of Females and Males
- Postive effect on Average Amount Spent of Males
- Negative effect on Average Amount Spent of Females

Device Segmentation



Only large negative effect on Average Amount Spent of Unknown Devices

Country Segmentation



Takeaways

- Positive effect on both metrics in Australia, Spain, Great Britain, Mexico and Canada
- Positive effect on CR and negative effect on AAS in Turkey, Unknown countries, Brazil, France, Germany, and United States

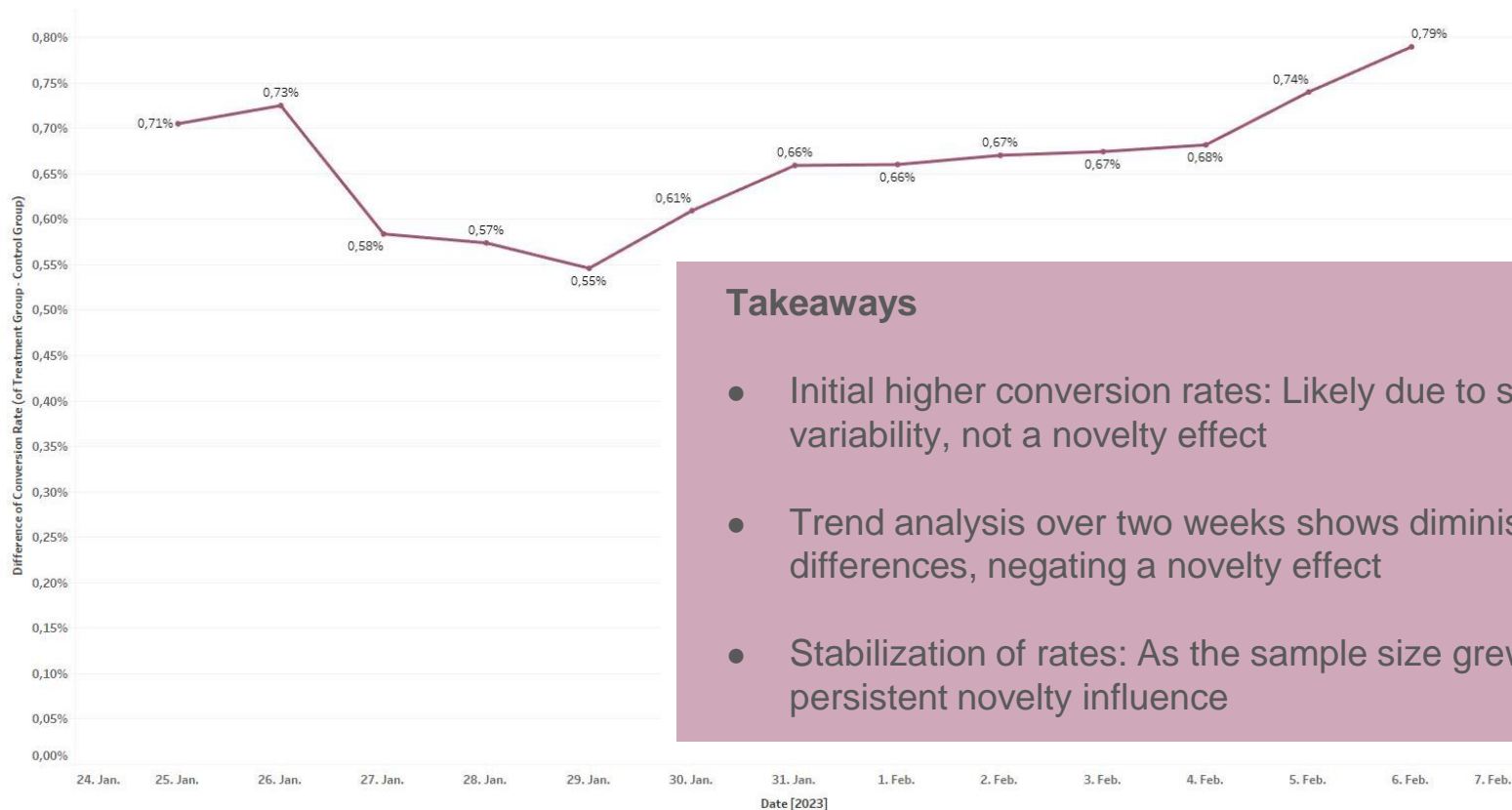
Distribution of Amount Spent



Takeaways

- Strongly right-skewed distribution of AAS indicating that most users spent small amounts
- Small contraction of the distribution of ASS of the Treatment Group

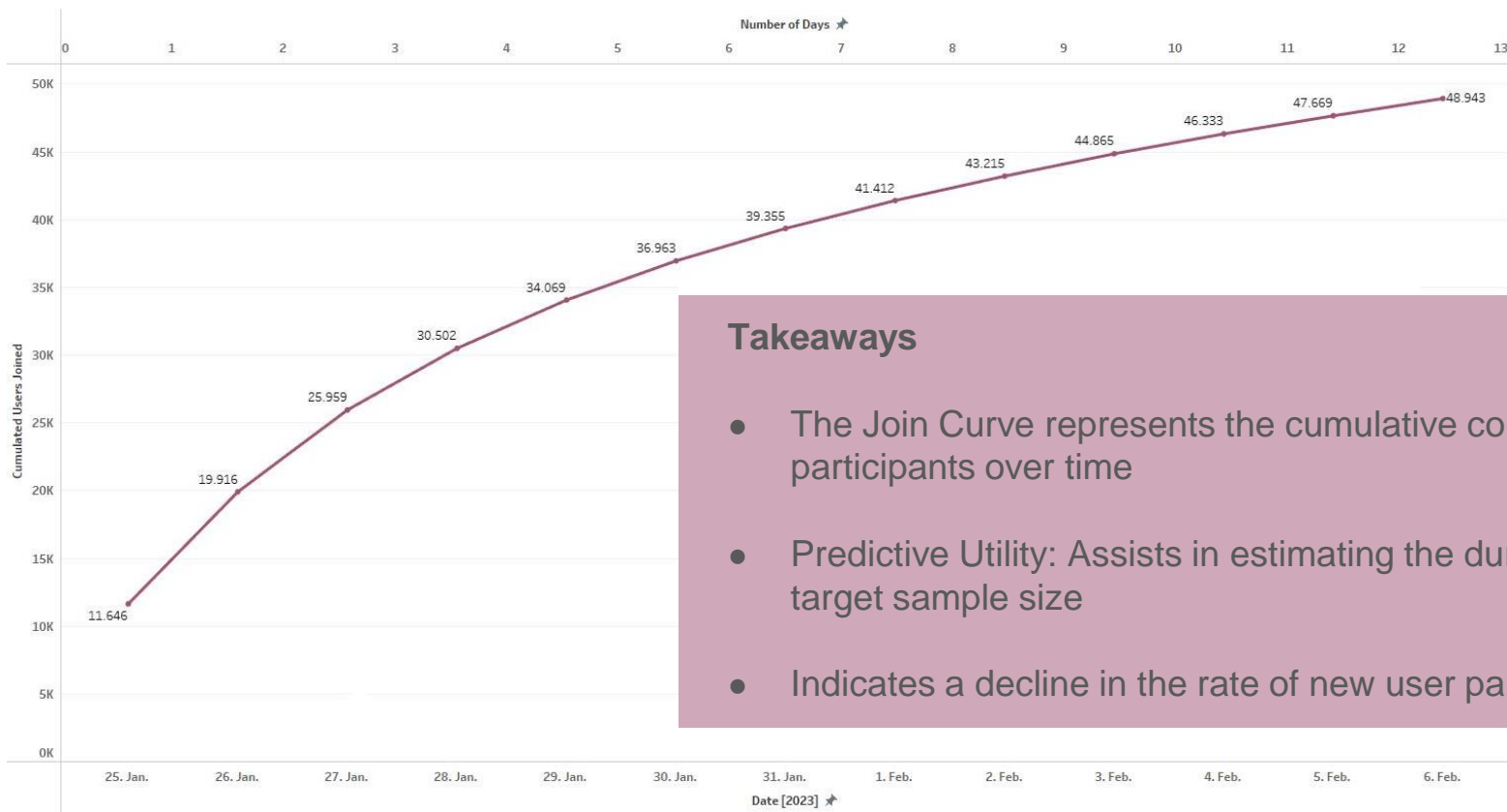
Stable Long-Term Conversion Without Novelty Effect



Takeaways

- Initial higher conversion rates: Likely due to small sample variability, not a novelty effect
- Trend analysis over two weeks shows diminishing differences, negating a novelty effect
- Stabilization of rates: As the sample size grew, suggesting no persistent novelty influence

The Join Curve

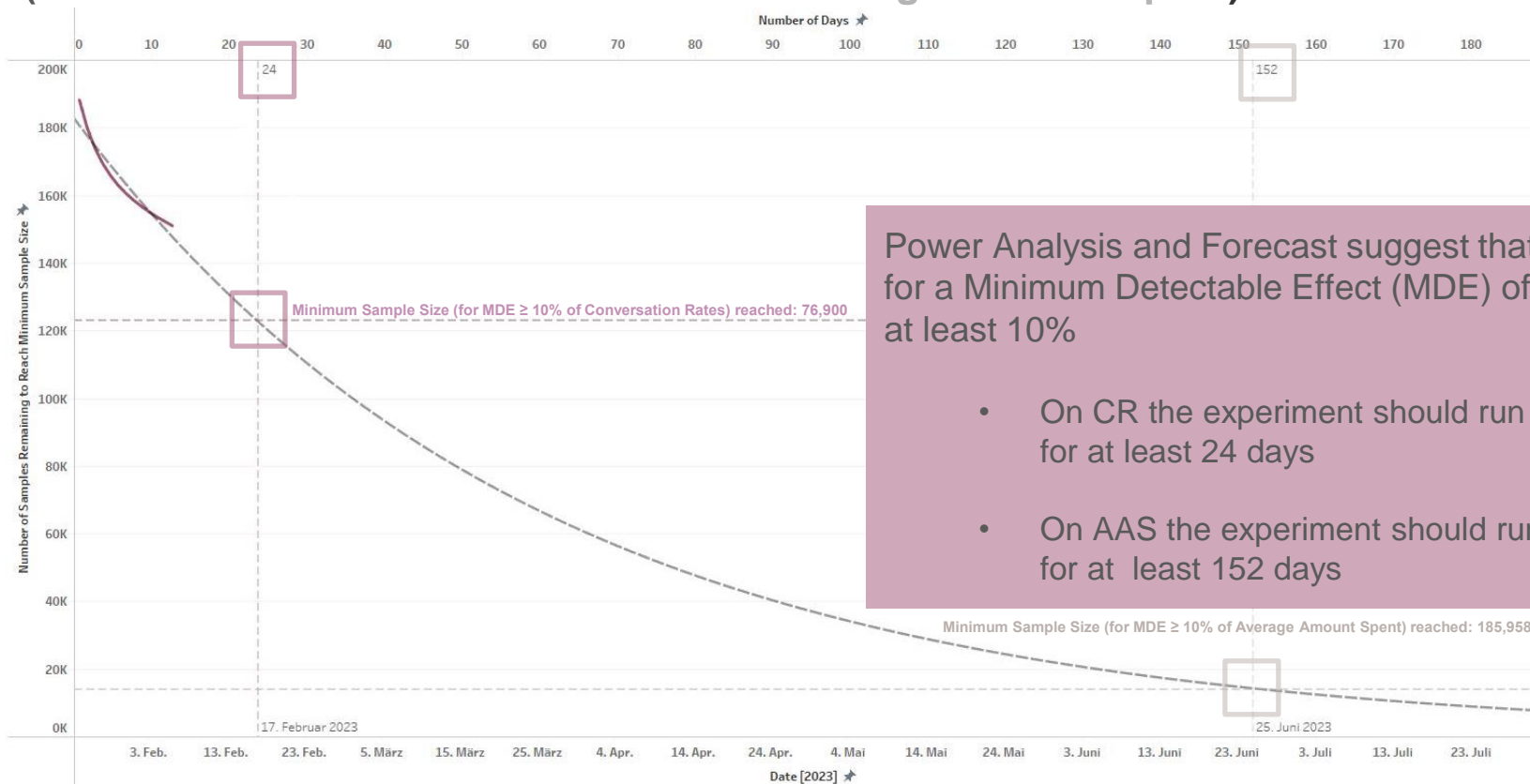


Takeaways

- The Join Curve represents the cumulative count of participants over time
- Predictive Utility: Assists in estimating the duration to achieve target sample size
- Indicates a decline in the rate of new user participation

Extrapolation of the Join Curve

Forecast on Experiment Duration necessary to reach Minimum Sample Sizes
(for MDE $\geq 10\%$ of **Conversion Rate** and **Average Amount Spent**)



Power Analysis and Forecast suggest that for a Minimum Detectable Effect (MDE) of at least 10%

- On CR the experiment should run for at least 24 days
- On AAS the experiment should run for at least 152 days

Next Steps

Recommendations

1. Launch the banner based on
 - Significant increase in the Conversion Rate.
 - No observed negative impact on the Average Amount Spent overall.
 - Business considerations suggesting value for both users and the business.
2. Continuously learn from user interactions with the banner in a real-world setting.
3. Reassess and adjust if future metrics indicate any unforeseen negative effects.

Analysis and Monitoring Suggestions

1. Analyze gender-, device- and country-specific effects on CR and AAS
2. Analyze contraction of the distribution of Average Amount Spent (AAS)
3. Continue monitoring the banner's impact on the Conversion Rate (CR) and Average Amount Spent (AAS).
4. Observe the banner's performance over a longer period (24 days for CR and 152 days for AAS) to determine the Minimum Detectable Effect (MDE) of at least 10%.
5. Pay attention to any unforeseen negative implications on metrics in the future, especially AAS.

Thank You!