GLOBOX A/B TEST REPORT
Prepared by: Alaofin. O. David
Date: 28th October 2023

1.0 Summary

In an effort to increase revenue by raising awareness of various product categories, Globox launched an A/B test to evaluate two versions of the company's mobile website. The test results revealed a difference in conversion rates between the control and treatment groups, although no significant difference was found in the average amount spent per user. Indications of a possible Novelty effect were observed, and the sample size was deemed inadequate based on power statistics calculations. A longer test duration with a larger sample size is recommended for more robust results.

2.0 Context

2.1 Introduction

Globox aims to increase awareness of its diverse product categories, particularly the food and drink category, which has shown significant growth. To achieve this, the company considered launching a new version of its website featuring a prominent banner for the food and drink category. An A/B test was commissioned by the Growth Team to evaluate the impact of this change.

The A/B test involved two groups:

Control Group (A-Group): Did not see the banner.

Treatment Group (B-Group): Saw the banner.

The test was conducted exclusively on the mobile website, with users randomly allocated to each group.

2.2 Sample Size

Total Population: 48,943 users

 A-Group: 24,343 users

 B-Group: 24,600 users

2.3 Timing

Data Collection Period: 25th January 2023 to 6th February 2023

2.4 Methods

The following user records were obtained:

- Unique ID

- Device type

- Gender

- Country

- Join date

- Amount spent

- Group assignment

Basic Assumptions:

- Users who spent any amount were considered converted.

- Null values in the "spent" column were treated as zero.

- Null values in "gender" were treated as "Not Specified."

Data Cleaning:

- Utilized the COALESCE function to replace null values.

  - Nulls in "spent" were replaced with zero.

  - Nulls in "device", "gender", and "country" were replaced with "Not Specified."

Analysis:

Data extraction was performed using SQL (see Appendix). Analysis and visualization were conducted using Tableau, focusing on high-level results and segmented by device, gender, and country.

3.0 Results

Conversion Rates:

- The treatment group (B-Group) showed a slightly higher conversion rate compared to the control group (A-Group).

Average Amount Spent:

- No significant difference between the two groups, with both averaging approximately $3.40 per user.

Gender-Based Insights:

- Female users exhibited higher engagement and spending compared to male users.

- Users with unspecified gender also performed better in terms of conversion rates.

Device Usage:

- iPhone users showed higher performance and engagement compared to other mobile device users.

- Users with unspecified device types spent more in the control group.

Geographical Insights:

- Highest conversion rate observed in Canada.

- Highest average spending observed in the USA.

4.0 Further Statistical Analysis

Conversion Rate Hypothesis Test:

Null Hypothesis (H0): No difference in conversion rates between control and treatment groups.

Alternative Hypothesis (H1): Difference in conversion rates between control and treatment groups.

P-value: 0.001 (significant at the 0.05 level)

95% Confidence Interval: (0.0035, 0.0107)

Conclusion: Reject H0, indicating a significant difference in conversion rates.

Average Amount Spent Hypothesis Test:

Null Hypothesis (H0): No difference in average amount spent per user between control and treatment groups.

Alternative Hypothesis (H1): Difference in average amount spent per user between control and treatment groups.

P-value:1.048 (not significant at the 0.05 level)

95% Confidence Interval: (-0.457, 0.430)

Conclusion: Fail to reject H0, indicating no significant difference in average spending.

Novelty Effect:

- Initial spike in engagement observed at the start of the experiment, normalizing over time with a subsequent rise from February 3rd, 2023. Suggests a possible novelty effect.

Power Statistics:

- Suggested a total sample size of 90,100 users for a robust result, indicating the current sample size is inadequate.

5.0 Recommendations

- Extend the experiment duration to at least one month.

- Increase the sample size to achieve more reliable results.

- Further analyze potential novelty effects by monitoring user behavior over an extended period.