**Summary**

*Batuhan Şengezer*

*11 May 2023*

*Analysis of GloBox A/B Test*

**The aim of this data analysis project was to evaluate the impact of applying a banner to the GloBox company's web page by analyzing the results of an A/B test. The dataset generated by the test was processed using Excel and analyzed using SQL. The results were then visualized using Tableau to generate insights for the company and explore the potential outcomes of implementing the treatment. The findings revealed that while the use of the banner did not significantly affect the average amount spent by users, it did have a significant impact on the conversion rate. Therefore, it is recommended that the banner be utilized since it is beneficial for the company. This study provides evidence that the use of the banner can potentially increase the company's conversion rate and improve its marketing strategy.**

**Starting Point**

**The experiment was conducted between January 25th, 2023, and February 6th, 2023, and involved randomly assigning main page visitors to either the control or the test group. Users in the test group were exposed to the banner, while those in the control group were not. The primary outcome measure was conversion, defined as users who made one or more purchases. All data were collected exclusively from the mobile website of GloBox. The study's design allowed for a direct comparison of user behavior between the two groups, making it possible to isolate the impact of the banner on user behavior. The findings of this analysis contribute to a better understanding of the effectiveness of banners, with potential implications for the development of marketing strategies by GloBox.**

**Dataset**

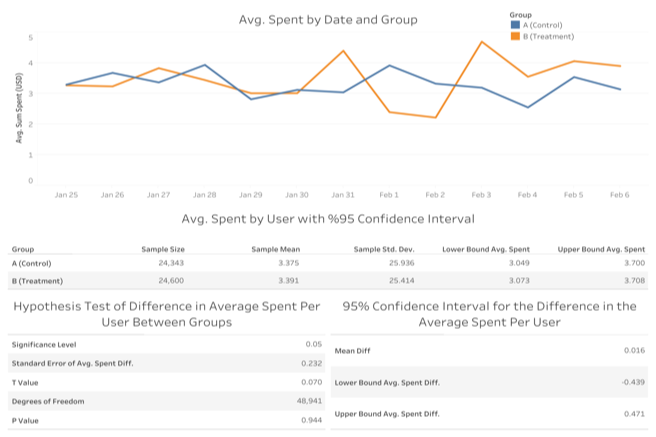
**The dataset used for this study consists of three primary tables: users, groups, and activity. The users table comprises the id, country, and gender columns, providing essential demographic information about each user. The groups table includes the uid (user id), group, join date (the date on which the user visited the web page), and device columns, which are useful for identifying the group to which each user was assigned, as well as their device type and date of joining. Finally, the activity table includes the uid, date (the date of each purchasing activity), device, and spent columns. This table provides valuable data on each user's purchase activity, including the date of each purchase, the device used for the transaction, and the amount spent. These three tables together comprise a comprehensive dataset for analyzing user behavior during the experiment period and determining the impact of the banner on user activity.**

**Data Processing**

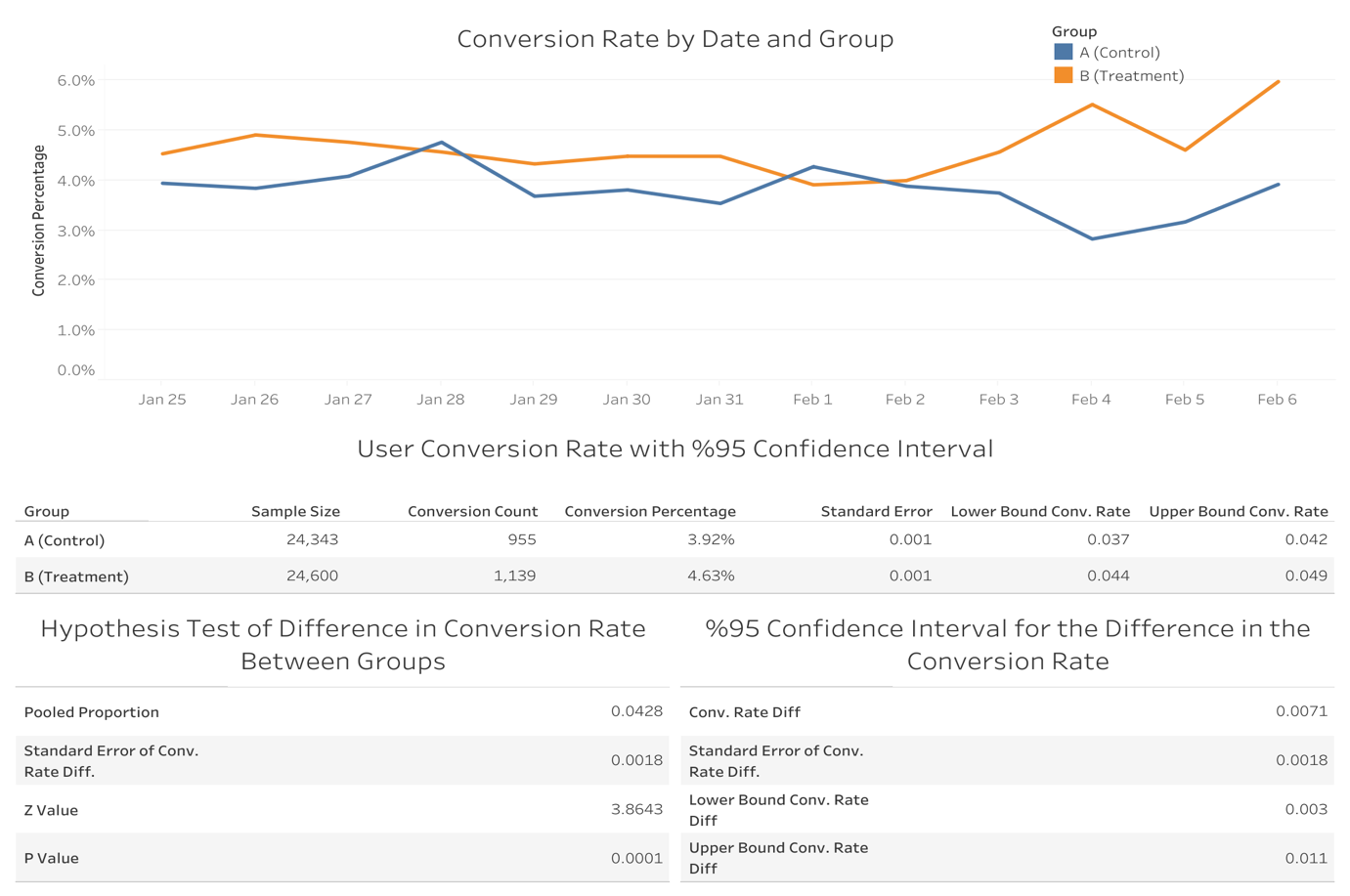
**Prior to data analysis, several data cleaning steps were taken to ensure data quality and accuracy. Firstly, the dataset was checked for duplicate rows, and none were found. Additionally, null and blank device types as well as blank country and gender cells were changed to “unknown” to ensure consistency in the dataset. The dataset was then checked to identify any negative values or outliers in the amount spent column that may negatively impact the accuracy of the analysis. However, no negative values were found, and it was decided to include outliers in the analysis, as they were deemed to not significantly affect the overall results. To prepare the data for analysis, data types were corrected, and formats were standardized. Specifically, two views were created by joining all tables in the dataset to have a comprehensive table for analysis, one including each individual row including multiple purchases from the same user, and another with the summing amount spent by each user. The data cleaning steps taken in this analysis are important to ensure that the analysis is conducted on high-quality data, minimizing the risk of biased results.**

**Analysis and Results**

**To determine the effectiveness of the banner, the average amount spent per user for the control and treatment groups was statistically analyzed. The results revealed that the average amount spent per user for the control group was 3.375, while that of the treatment group was 3.391. To establish the 95% confidence interval for the average amount spent per user in both groups, a t-test was conducted. The resulting confidence interval for the control group was (3.049, 3.700), while that of the treatment group was (3.073, 3.708). A subsequent t-test was performed to test the hypothesis and determine whether there was a significant difference in the mean amount spent per user between the control and treatment groups. The resulting p-value of 0.944 indicated a failure to reject the null hypothesis, leading to the conclusion that there is insufficient evidence to suggest a significant difference in the mean amount spent per user between the two groups. Finally, another t-test was conducted to establish the 95% confidence interval for the difference in the average amount spent per user, which was found to be (-0.439, 0.471). The complete analysis parameters and results of the GloBox A/B test for the average amount spent per user are presented in the following figure.**



**As another way to assess the effectiveness of the banner, the conversion rates for users in both the control and treatment groups were analyzed. The user conversion rate for the control group was found to be 3.923%, while that of the treatment group was 4.630%. To determine the 95% confidence interval for the conversion rate of users in the control and treatment groups, a z-test was conducted, yielding results of (0.037, 0.042) for the control group and (0.044, 0.049) for the treatment group. Additionally, a z-test was performed to test the hypothesis of a difference in conversion rate between the two groups, resulting in a p-value of 0.0001. Thus, it was determined that there is a statistically significant difference in conversion rates between the control and treatment groups. Finally, a z-test was conducted to find the 95% confidence interval for the difference in conversion rate between groups, resulting in a confidence interval of (0.003, 0.011). Overall, these results suggest that the banner may have a significant impact on user conversion rates. The following figure shows the complete analysis parameters and results of the GloBox A/B test for the user conversion rate.**



**Recommendations**

**The results of the statistical analysis indicate that the banner had no significant effect on the average amount spent by users, while it did have a statistically significant effect on the conversion rate. Therefore, it can be concluded that applying the banner on the GloBox mobile webpage can be considered a successful strategy for increasing the conversion rate and potentially boosting business performance.**

**To further improve the accuracy of the analysis and gain deeper insights, several suggestions for future analysis have been proposed. Firstly, extending the test period would provide more data and a better understanding of the long-term impact of the banner on user behavior. Secondly, analyzing the results by country, gender, and device type could reveal potential differences in the effectiveness of the banner among different user segments. Additionally, extracting additional data such as age could provide a more comprehensive understanding of the user demographics and help identify potential areas for improvement. Lastly, running tests on alternative banner designs could help optimize the effectiveness of the banner and improve business outcomes. Overall, these suggested analyses could provide valuable insights and aid in making informed decisions regarding the implementation of banners on the website.**

**Appendix**

**Link to complete project repository:**

<https://github.com/BatuSengezer/GloBox_A-B_Data_Analysis>

Link to tableau story:

<https://public.tableau.com/views/GloBox_A_B_Data_Analysis/Story1?:language=en-US&:display_count=n&:origin=viz_share_link>