Nicholas Jones

07/07/2023

Globox A/B Test: Food and Drinks Banner

Abstract: After a thorough analysis of both the control group (the group without the banner) and the treatment group (the group with the banner), the data revealed the following: The difference of the conversion rates are statistically significant, while the difference of means is not; the treatment group spent more money on average; I-phone users outspent Android users; Women outspent and had higher conversion rates than men; some international markets spent more money than the US on average; the Novelty Effect is more present in the difference of means than in the difference of conversion rates. Therefore, my recommendation is that Globox should go back and conduct another experiment in order to sufficiently gain insights into the effectiveness of the banner on purchase activity because the experiment showed inconclusive, mixed results.

Introduction/Context: I had to acquire the dataset that contained the data that was most relevant to the objective of the experiment, which I obtained via the SQL code in Appendix #1 . After I queried this data and exported it in an Excel file (which can be accessed in the Appendix #2), I transferred this data set into Tableau for further analysis. The main objective of this analysis was to see if the new banner would increase product sales (conversion rates) in the treatment group (Group B) as opposed to the control group (Group A). At this stage of the experiment, I sequentially analyzed the data in the following way: comparing the conversion rates and average amount spent between the two groups via hypothesis tests and confidence intervals, the amount spent per user in each group, how the conversion rates and the average amount spent correlated with certain variables (types of mobile devices used, gender, country of origin), the novelty effect in the data, and the distributions of the confidence intervals.

Test Design: This was an A/B test in order to test the efficacy of the food and drinks banner on mobile devices. The control group consisted of 24,343 randomly assigned individuals that did not see the banner, but the treatment group consisted of 24,600 randomly assigned people that did see the banner.

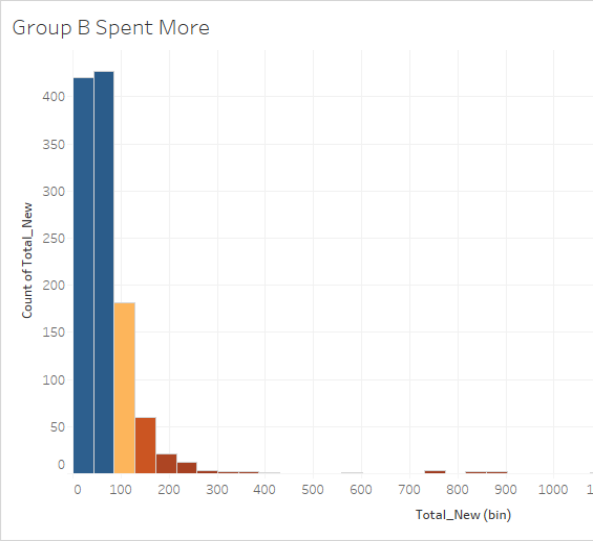
Results of the Analyses:

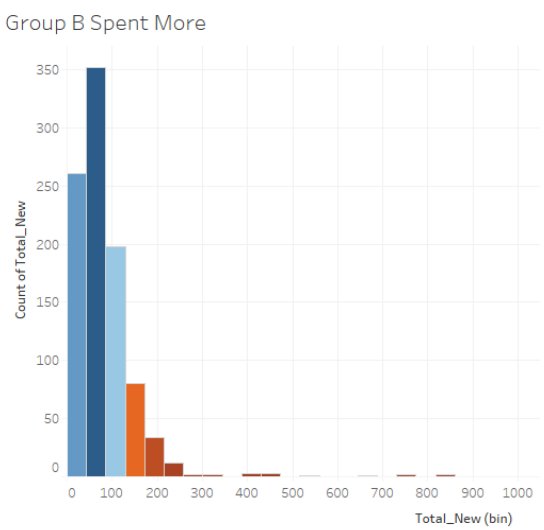
Context of Hypothesis Tests and Confidence Intervals: In order to compare the conversion rates and average amounts of money spent between the two groups, I performed two hypothesis tests, formulated two confidence intervals, and analyzed the two metrics in the provided visualizations. For the hypothesis test, comparing the difference of the conversion rates of Group A and Group B, the following hypotheses were tested against each other: The Null Hypothesis which is there is no difference of the conversion rates of the two groups was tested against the Alternative Hypothesis of there is a difference between the conversion rates of the two groups. Furthermore, these are the other assumptions that were made: There was a 5% significance level, a normal distribution was utilized, the test was two-tailed, Group A's conversion rate was 3.92%, Group B's conversion rate was 4.63%, and the standard error was pooled.

Results: The results were statistically significant. So, the Null Hypothesis should be rejected in favor of the Alternative Hypothesis (p-value = .000111). As further confirmation of the hypothesis test, the resulting 95% confidence interval contained the difference of the conversion rates but did not contain zero, which implies the invalidity of the Null Hypothesis: (-.01065365985, -.00345985309).

Context of Hypothesis Tests and Confidence Intervals (continued): For the hypothesis test comparing the difference in the average amounts of money spent between group A and group B, the following hypotheses were tested against each other: The Null Hypothesis of there is no difference in the average amount of money spent between the two groups was tested against the Alternative Hypothesis of there is a difference between the average amount of money spent between the two groups. Furthermore, these are the other assumptions that were made: There was a 5% significance level, a Student's T-distribution was utilized, the test was two-tailed, Group A's average spending is 3.37, Group B's average spending is 3.39, Group A's sample standard deviation is 25.93, Group B's sample standard deviation is 25.41, and the variance was unequal.

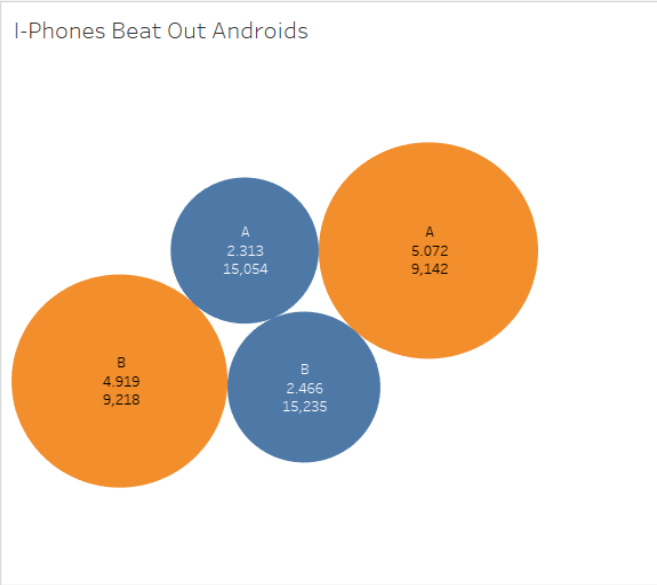
Results: The results were statistically significant. So, the Null Hypothesis should be rejected in favor of the Alternative Hypothesis (p-value=.94438). As further confirmation of the failed hypothesis test, the resulting 95% confidence interval did contain the difference of the means and zero, which implies the Null Hypothesis might be true: (-.076, .043). Furthermore, I created a bar chart with filters that can give all of the aforementioned averages and conversions rates (Appendix #3). This visualization shows the average total amount spent per user, and you calculate the conversion rates for both groups with the help of the filter. Two important facts that this visualization bears out is that Group A spent a little less on average than Group B, and Group B had a higher conversion rate.





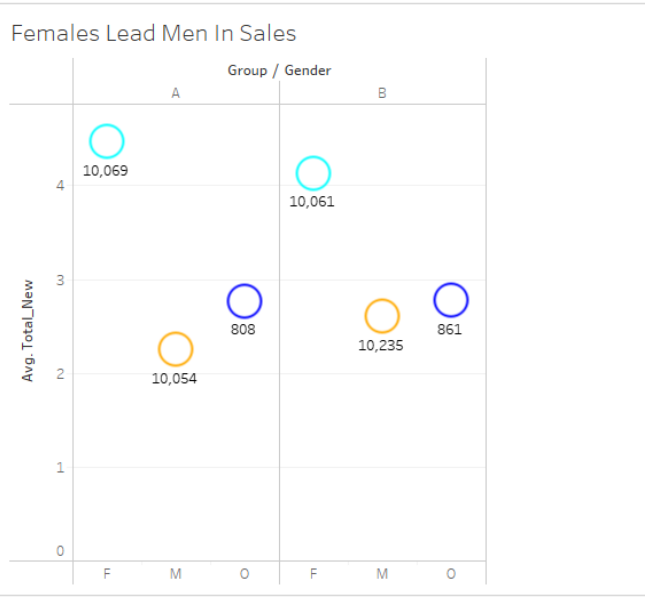
Context of The Amount Spent Per User In Each Group: The next crucial part of my analysis was to consider which group the amount spent per user in each group. In order to visualize this, I decided to create two histograms of the total amount spent in the Group A and Group B (Appendix #4). The histograms show the distributions of the sums of the total amounts spent per group.

Results: The important takeaways of this analysis are that both distributions are right-skewed, and Group B has a higher propensity to spend money in the price range of $0 to $86 than Group A.



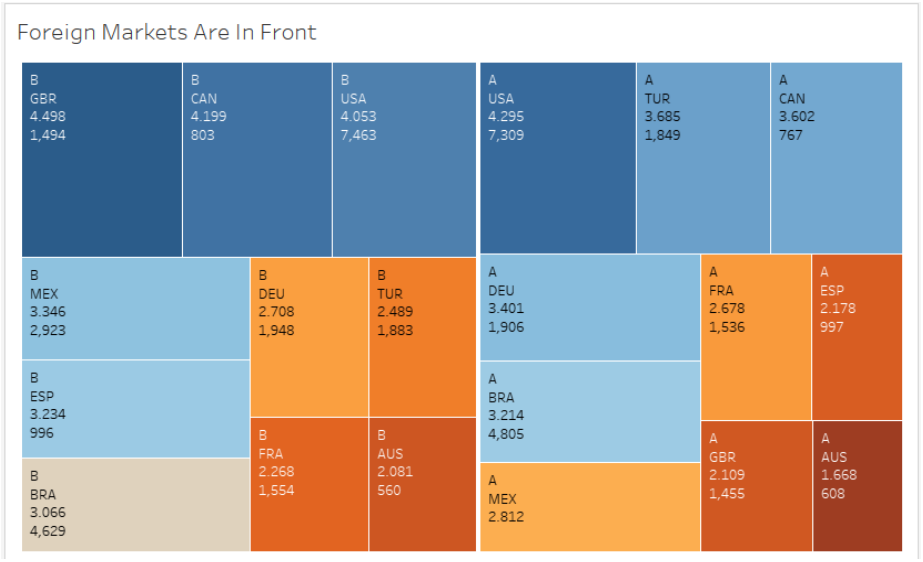
Context of The Mobile Device: Next, I decided to see if there is a relationship between the average amount of money spent per user, the conversion rates, and the mobile devices that were used to make these purchases by making a bubble chart (Appendix #5).

Results: Interestingly, in both groups, users with I-Phones spent more money and outnumbered the users with Android phones. Furthermore, the I-Phone users in the Group A spent more on average than I-phone users in the Group B ($5.07 compared to $4.91), even though the size of the Group B's I-Phone users was bigger than the Group A's I-Phone users (9218 users compared to 9142 users).



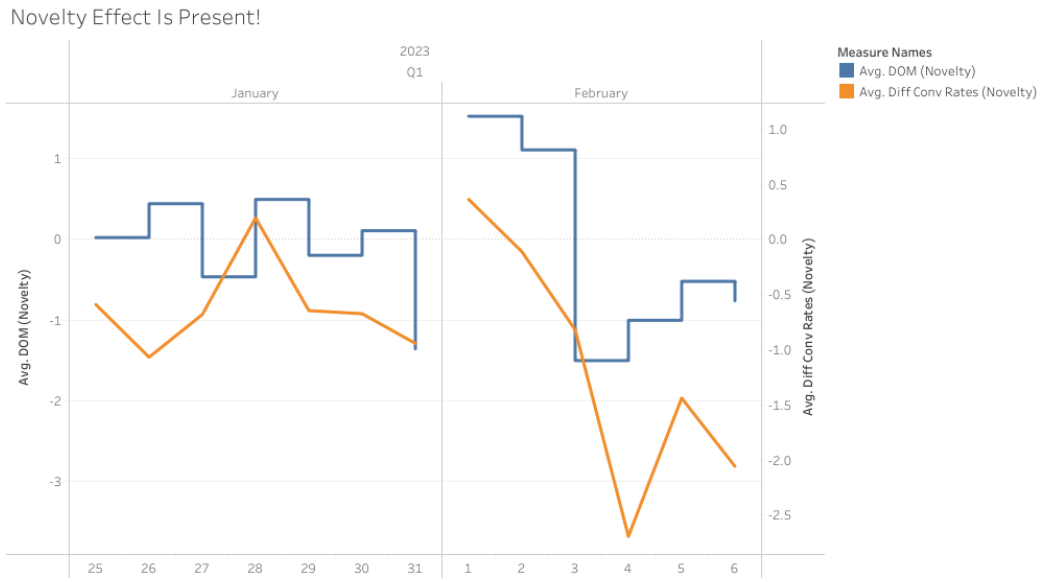
Context of Gender: Another important factor that I wanted to explore was how the conversion rates and the average amounts of money spent were related to the genders of the buyers. So, I decided to create a circular views plot to show several variables at the same time. This information can be found in the provided link (Appendix #6).

Results: Some important trends that this plot brought out was that even though the number of men and women in this experiment were almost equal (20,289 compared to 20,130, respectively), women spent over 50% more money than men on average. Also, women had a higher conversion rate than men (5.29% compared to 3.21%).



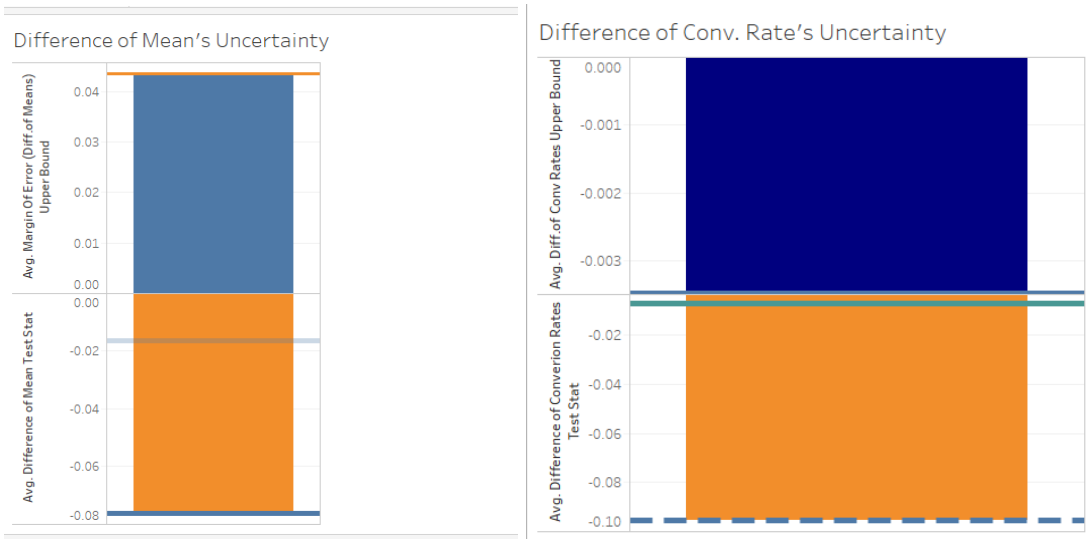
Context of Country of Origin: The last demographic variable that I considered was the country of origin: I wanted to see if certain countries had higher conversion rates and average amounts of money spent than others. To illuminate any hidden relationships, I decided to create a tree map. The visualization can be viewed here (Appendix #7).

Results: The United States had the highest conversion rates in both groups: Group A had a conversion rate of 5.1%, and Group B had a conversion rate of 5.7%. Conversely, in Group B, international markets had higher average amounts of money spent on goods as compared to the United States: In Group A, The USA spent $4.29 on average, Turkey spent $3.68 on average; in Group B, The GBR spent $4.49 on average, and the United States spent $4.05 on average.



Context of the Novelty Effect (Advanced Task): At this point, I wanted to consider another aspect of the data, the novelty effect. After generating this new data set with the code in Appendix #7, I had to create several new variables in order to correctly do the analysis: "DOM (Novelty)" measures the daily Difference of means, and "Diff Conv Rates" measures the daily Difference of conversion rates. Afterwards, I created dual-axis line chart that contains two line graphs that display the changes of these two key metrics over time (Appendix #9).

Results: The novelty effect is present in both of the metrics, but it appears to be more present in the difference of the means than the conversion rates.



Context of Distributions of Confidence Intervals (Advanced Task): After analyzing the novelty effect, I wanted to visualize the confidence intervals of both statistics. In order to create the confidence intervals, I had to create some new variables: "Difference of Means" and "Diff of Conv Rates" calculate the respective sample statistics; for both statistics, there are also upper-bound and lower-bounds calculations for the confidence intervals. Also, I included a bar chart that visualizes the means of both groups (Appendix #10 and #11).

Results: This shows that the difference of means interval contains zero, which inhibits me from asserting the inaccuracy of the Null Hypothesis.

Recommendation: I feel confidently that Globox should redo the testing until both of the metrics show statistical significance because the data has not clearly shown a picture of the usefulness or inferiority of the banner. However, based on the data, Globox should retry the test with a focus on international markets, women, and I-Phone users. Lastly, Globox should redo the experiment because the difference of the conversion rates is statistically significant, but this statistic is in a nonzero, positive range. In order for Globox to see that its banner is working, we need both of the key statistics to be statistically significant within negative ranges that exclude zero because this will indicate that the treatment group spends more money and has a higher conversion rate.

Appendix

1. SQL Code
    a. SELECT G.uid, U.country, U.gender, G.device, G.group, SUM(A.spent) AS Total
    b. FROM groups as G
    c. LEFT JOIN users as U
    d. ON G.uid=U.ID
    e. LEFT JOIN activity as A
    f. ON G.uid=A.uid
    g. GROUP BY G.uid, U.country, U.gender, G. device, G. group
    h. ORDER BY Total
2. Excel file with the data and calculations [https://docs.google.com/spreadsheets/d/14WTbO7aEIxIcAuXFdP4B_tLU8jlxHn9mtMVEtGkOUvQ/edit?usp=sharing]
3. "Group Stats" Bar Chart [https://public.tableau.com/views/GroupStats/GroupStatistics?:language=en-US&:display_count=n&:origin=viz_share_link]
4. "Metric histograms" [https://public.tableau.com/views/MetricsHistogram/GroupBSpentMore?:language=en-US&:display_count=n&:origin=viz_share_link]
5. "Phones" Bubble chart [https://public.tableau.com/views/Phones_16898938068350/I-PhonesBeatOutAndroids?:language=en-US&:display_count=n&:origin=viz_share_link]
6. "Gender" Circular Views Chart [https://public.tableau.com/views/Gender_16898938281590/FemalesLeadMenInSales?:language=en-US&:display_count=n&:origin=viz_share_link]
7. Foreign Market Tree Map [https://public.tableau.com/views/ForeignMarketTreeMap/ForeignMarketsAreInFront?:language=en-US&:display_count=n&:origin=viz_share_link]
8. SQL Code:
    a. SELECT G.join_dt, G.uid, G.group, SUM(A.spent) AS Total
    b. FROM groups as G
    c. LEFT JOIN users as U
    d. ON G.uid=U.ID
    e. LEFT JOIN activity as A
    f. ON G.uid=A.uid
    g. GROUP BY G.join_dt, G.uid, G. group
    h. ORDER BY G.join_dt, Total
9. "Novelty" Line Charts [https://public.tableau.com/views/Novelty_16898892193230/NoveltyEffectIsPresent?:language=en-US&:display_count=n&:origin=viz_share_link]
10. "CI Part 1" Line Chart [https://public.tableau.com/views/CIPart1/DifferenceofMeansUncertainty?:language=en-US&:display_count=n&:origin=viz_share_link]
11. "CI Part 2" Line Chart [https://public.tableau.com/views/CIPart2/DifferenceofConversionRatesUncertainty?:language=en-US&:display_count=n&:origin=viz_share_link]