**A/B Test Results for Food & Drink Initiative**

Tyler Barrow

Growth Product & Engineering Team, GloBox

**Summary:**

The intention of this report is to bring clarity to the A/B test we conducted between January 25th and February 6th on the new banner to increase the user funnel to the food and drink section on our website. My findings in utilizing this dataset were gathered using multiple statistical methods, including hypothesis testing via significance level, visualizing confidence intervals to show variance in conversion rates and average spending per user, and practical statistics to show the surface level differences via conversion rate differences. My findings should show themselves to be conclusive to increase overall revenue for the company, which was the main goal of the initial testing. Based on the research and testing conducted, I’ve found that the basis of the test, A vs B, was not convincing enough statistically, so I looked further and found that females on our website are the highest converting customers we have, by a significant margin.

**Context:**

My initial approach to this project was to be able to provide concise results to the a/b test to give a clear plan going forward on how to utilize this information that was gathered. The initial hypothesis that I explored when viewing this dataset, was to test if there was any difference between A and B, with the primary goal of B providing more revenue overall. The data collected for this test was gathered separately from the typical data we keep in our dataset due to needing this data to be clear of data not included in this test. This was housed in our SQL database, and I utilized beekeeper studio for the cleaning and preparation of the data for review. Much of the research conducted in this test was completed via excel calculations and transferred to tableau to visualize for better visibility of the specific metrics captured. The primary motivation of this a/b test to be conducted was on the basis that we wanted to test the functionality of adding a banner on our main product website to drive more traffic to the food and drink category of our website. The primary parameters focused on in this a/b test are the means of both groups average spending per user, and conversion rate per group, as the variance between these 2 variables are the most imperative to determine the success or failure of the test.

**Determining the Validity of Data Utilized:**

My goal in this section is to validate the data collected to further prove the validity of recommendations seeing as the data has no significant faults like, novelty effect, skewed user distribution, or low sample size utilizing a power analysis. Based on the power analysis completed, even though typically this would be completed prior to testing, states that for us to reach a lower standard deviation of 15 for mean of average spent per user, we would need to increase sample size to nearly 8 million, however, this seems to be unrealistic to achieve based on current customer base. To have a test where the minimum variance in conversion rates is at least 5%, we'd have to have a sample size of at least 242,500.

In the below chart, you will see an evaluation of the novelty effect, or the effect of seeing the addition of the banner on the website over time could affect the overall interest in website due to customer liking the prior iteration of the website. This effect does not seem to be the case in this test, as the average spending per user over the course of the test varied, but ended higher than it started for group B.

Table 2.1: Novelty Effect visualization from January 25th to February 6th:

A graph of blue and orange lines

Description automatically generated

As you can see from the upcoming table, the distribution of male and female was nearly equal statistically showing minimal variance. Another layer to confirm the validity of the data for use in testing to produce accurate results.

Table 2.2: Number of Users based on Gender:

A graph of a bar

Description automatically generated

**A/B Testing Results:**

The below two charts are regarding the confidence intervals, or a visual representation of the variance between the conversion rates and average spending of groups A and B. These are important to show the differences in both groups, or lack thereof, as you can see with the difference in average spending between groups A and B, there is no significant difference. With the conversion rates, there is a visible difference and increase in conversion for group B, however, this difference is teetering on the edge of becoming significant statistically. Due to the minimal variance of both values, I concluded that there is not enough variance to be significant enough for us to be complacent with launching the change on these terms. I don’t recommend that we ignore a .71% increase in conversion based on practical statistical principles this is still increases in revenue. Though, as you read further, I have reason to believe we have more opportunity to improve these numbers.

Table 1.1: 95% Confidence Interval visual representation of Conversion Rate:

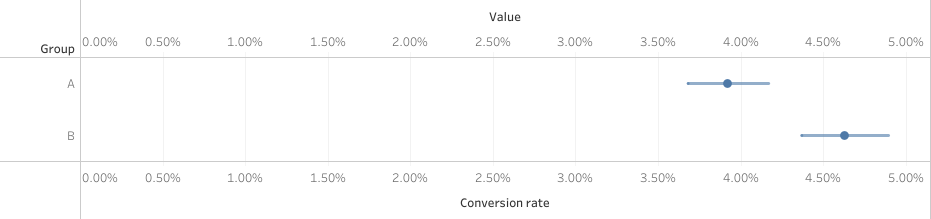
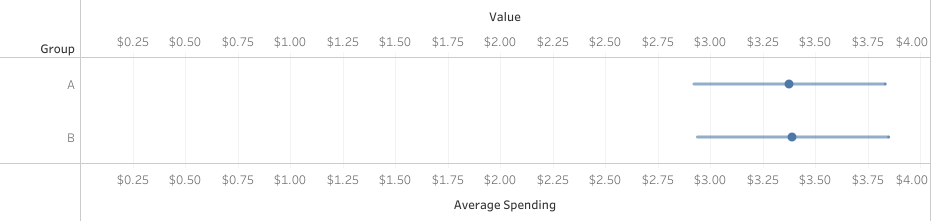


Table 1.2: 95% Confidence Interval visual representation of Average Spending:



**Gender’s Effect on A/B Testing Results:**

Based on my further research into this topic of A/B testing for inclusion of the banner on the website, I have found another route to focus my efforts to increase revenues. The below charts represent the confidence intervals of conversion rate and average spending, similarly to the charts seen above. You will see there is quite a bit more variance to be seen with these charts that I find more telling of the future of the improvements we will be able to make to the website. My study focused on the business we receive from our female customer base, as you will see, they are by far the highest converting group within our sample. If you will also notice, the difference in conversions were not affected in near the same way as we see from the above graphs, showing a .71% increase in conversion overall, but females conversion numbers increased very minimally, by .3%. I find this as a fault of our advertising, as we should be focusing more so on increasing the business we receive from our female clientele. There is also another case to be discussed utilizing practical statistics where the average spending per female user went down by $0.33, another case to where we should be more so focusing on our female clientele and having our advertising be more so aimed at this client base. Based on this data, you can see that the difference in pooled conversions between female and male is 2.08%, following the difference in pooled average spent per user being $1.865. The other group has a much smaller sample size and much more variance as you can see, which is why I do not think it is statistically correct to compare. Similarly, to other, statistically we cannot compare to N/A conversions either, as this is a pooled average that could be inclusive of males, females, and other genders.

Table 3.1: 95% Confidence Interval visual representation of Conversion Rate based on Gender:

A graph with blue and white lines

Description automatically generated

Table 3.2: 95% Confidence Interval visual representation of Average Spending based on Gender:

A graph with blue dots and numbers

Description automatically generated

**Recommendation:**

Based on the research results gathered above, the understanding of the data is that there is a slight increase in conversion rate from A to B, but this is not significant enough to be complacent with current results. That is why I recommend, based on our customer base being made up of mostly females who convert to purchasing our products, that this advertising is hindering the amount individuals within the female demographic are spending on our website. We need to revamp our advertising to better fit our female demographic, as this should increase our overall revenue when we better take care of our ideal client. Alongside this, the conversion rate did increase, which is why I propose that the banner is a good idea, but not yet statistically significant enough, which is why I purpose that we need marketing team to revamp the design to better resonate with our female audience to increase our overall revenue more effectively. Based on this, I recommend that the next steps be, we get the marketing team to make a new advertisement based on the above and relaunch the project for another A/B test. Based on my suggestions, the cost of this change should be extremely minimal as we’d have our internal marketing team to create the new advertisement, and we already have the technology in place to accurately track the performance of the next generation of a/b testing.

**Appendix.01: SQL Queries in Report:**

1. Start and end dates of the experiment
   1. SELECT dt, MIN(dt) as start\_date, MAX(dt) as end\_date
   2. FROM activity
   3. GROUP BY dt
   4. ORDER BY dt;
      1. 1/25/2023 start date
      2. 2/6/2023 end date
2. SQL function used to replace null values with n/a
   1. UPDATE users
   2. SET gender = 'n/a'
   3. WHERE gender IS NULL
3. Formatting the data to pull from the database to a CSV file
   1. SELECT
      1. DISTINCT(u.id), u.country, u.gender, g.device, g.group,
      2. CASE WHEN a.spent > 0 THEN 'yes' ELSE 'no' END AS converted,
      3. SUM(a.spent) as total\_spent, g.join\_dt
   2. FROM
      1. users as u
      2. LEFT JOIN activity as a ON u.id = a.uid
      3. LEFT JOIN groups as g ON u.id = g.uid
   3. GROUP BY
      1. u.id, u.country, u.gender, g.device, g.group, a.spent, g.join\_dt
   4. ORDER BY u.id
4. (see appendix cont for unrelated SQL queries)

**Appendix.02: Links & Workbooks:**

Tableau: <https://public.tableau.com/views/GloBoxProject_16887010913620/GenderDashboard?:language=en-US&:display_count=n&:origin=viz_share_link>

Excel: included in github

CSV (tableau import): included in github

Presentation: included in github

Video Presentation:

SQL dataset: postgres://Test:bQNxVzJL4g6u@ep-noisy-flower-846766-pooler.us-east-2.aws.neon.tech/Globox

Beekeeper Studio (utilized for SQL queries): <https://www.beekeeperstudio.io/>

Mean power analysis calculator: [https://statulator.com/SampleSize/ss2M.html#](https://statulator.com/SampleSize/ss2M.html)

Conversion rate power analysis calculator = <https://www.statsig.com/calculator>

**Appendix.03: Exploratory SQL queries:**

1. Can a user show up more than once in the activity table? Yes or no, and why?
   1. Yes, because this represents each day that a customer made a purchase, and they could have made 2 purchases over 2 days
2. What type of join should we use to join the users table to the activity table?
   1. LEFT JOIN
3. What SQL function can we use to fill in NULL values?
   1. UPDATE users
   2. SET gender = 'n/a'
   3. WHERE gender IS NULL
4. What are the start and end dates of the experiment?
   1. 1/25/2023 start date
   2. 2/6/2023 end date
   3. SELECT dt, MIN(dt) as start\_date, MAX(dt) as end\_date
   4. FROM activity
   5. GROUP BY dt
   6. ORDER BY dt;
5. How many total users were in the experiment?
   1. 48943
   2. SELECT DISTINCT(COUNT(id))
   3. FROM users
6. How many users were in the control and treatment groups?
   1. A: 24343
   2. B: 24600
   3. SELECT groups.group, COUNT(groups.group)
   4. FROM groups
   5. GROUP BY groups.group
7. What was the conversion rate of all users?
   1. 4.28%
   2. SELECT CAST(COUNT(a.uid) AS float) / CAST(COUNT(u.id) AS float) as conversion\_rate
   3. FROM users as u
   4. LEFT JOIN activity as a ON u.id = a.uid
8. What is the user conversion rate for the control and treatment groups?
   1. A: 3.92%
   2. B: 4.63%
   3. SELECT g.group, CAST(COUNT(a.uid) AS float) / CAST(COUNT(u.id) AS float) as conversion\_rate
   4. FROM users as u
   5. LEFT JOIN activity as a ON u.id = a.uid
   6. LEFT JOIN groups as g ON u.id = g.uid
   7. GROUP BY g.group
9. What is the average amount spent per user for the control and treatment groups, including users who did not convert?
   1. A: $3.37
   2. B: $3.39
   3. SELECT g.group, AVG(a.spent)
   4. FROM users as u
   5. LEFT JOIN activity as a ON u.id = a.uid
   6. LEFT JOIN groups as g ON u.id = g.uid
   7. GROUP BY g.group
10. Why does it matter to include users who did not convert when calculating the average amount spent per user?
    1. Because this will give a true value to the average amount spent either way. IE, A, they spent more on average, but had lower user conversions.

**Appendix.04: Query data to pull into a CSV:**

1. Formatting the data to pull from the database to a CSV file
   1. SELECT
      1. DISTINCT(u.id), u.country, u.gender, g.device, g.group,
      2. CASE WHEN a.spent > 0 THEN 'yes' ELSE 'no' END AS converted,
      3. SUM(a.spent) as total\_spent, g.joint\_dt
   2. FROM
      1. users as u
      2. LEFT JOIN activity as a ON u.id = a.uid
      3. LEFT JOIN groups as g ON u.id = g.uid
   3. GROUP BY
      1. u.id, u.country, u.gender, g.device, g.group, a.spent, g.joint\_dt
   4. ORDER BY u.id

**Appendix.05: A/B test statistics using Spreadsheets:**

1. Conduct a hypothesis test to see whether there is a difference in the conversion rate between the two groups. What are the resulting p-value and conclusion?
   * Use the normal distribution and a 5% significance level. Use the pooled proportion for the standard error.
   * **Null (A) – There is no change in business, or A = B**
   * **Alternative (B) – There is a difference, or A <> B (2 sided)**
   * P value = -0.00015791, significant, reject null value
2. What is the 95% confidence interval for the difference in the conversion rate between the treatment and control (treatment-control)?
   * Use the normal distribution and unpooled proportions for the standard error.
   * Lower = 0.00348599
   * Upper = 0.01065366
3. Conduct a hypothesis test to see whether there is a difference in the average amount spent per user between the two groups. What are the resulting p-value and conclusion?
   * Use the t distribution and a 5% significance level. Assume unequal variance.
   * Null: no change, A = B
   * Alt: A <> B, there is a difference in the average amount spent
   * P value = .943854898, insignificant, fail to reject null value
4. What is the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control)?
   * Use the t distribution and assume unequal variance.
   * Lower = -0.47133468
   * Upper = 0.438637719