Mandeep Toor

Date: 25,09,2023

Summary: For this assignment, the results of an A/B test done by Glo Box, an e-commerce firm focusing on unusual and high-quality products. The company intends to enhance revenue by promoting its food and beverage segment, which has recently seen strong growth. The A/B test entails showing a banner emphasizing food and drink products to the test group while hiding it from the control group.

assess the experiment's impact and make recommendations to relevant stakeholders. The Growth Product & Engineering Team, led by Leila Al-Farsi and Alejandro Gonzalez, and Mei Kim, the Head of Marketing, are among the decision-makers. The analysis will be critical in establishing whether Glo Box is viable.

Context:

- ✓ Data Exploration and Preparation:
- ✓ Can a user show up more than once in the activity table? Yes or no, and why?

You need to inspect the dataset to determine if there are duplicates in the activity table. This can be checked by examining unique identifiers such as user IDs.

- ✓ What type of join should we use to join the users table to the activity table?

The type of join (e.g., INNER JOIN, LEFT JOIN) depends on your analysis goals. If you want to include all users and their activities, use LEFT JOIN.

- ✓ What SQL function can we use to fill in NULL values?

You can use the COALESCE or ISNULL function in SQL to fill in NULL values with a specified default value.

- ✓ What are the start and end dates of the experiment?

Identify the minimum and maximum date values in your dataset to determine the experiment's start and end dates.

- ✓ How many total users were in the experiment?

Count the unique user IDs in the dataset.

- ✓ How many users were in the control and treatment groups?

Categorize users into control and treatment groups based on relevant columns, then count each group's size.

- ✓ **Conversion Rate Analysis:**
- ✓ What was the conversion rate of all users?

Calculate the overall conversion rate by dividing the number of users who converted by the total number of users.

- ✓ What is the user conversion rate for the control and treatment groups?

Calculate conversion rates separately for the control and treatment groups.

- ✓ **Average Amount Spent Analysis:**
- ✓ What is the average amount spent per user for the control and treatment groups, including users who did not convert?

Calculate the average amount spent per user in both control and treatment groups, considering all users.

- ✓ Why does it matter to include users who did not convert when calculating the average amount spent per user?

Including non-converters in the calculation provides a more accurate representation of user behavior and spending patterns, which can help in making informed decisions.

- ✓ **Hypothesis Testing:**
- ✓ Conduct a hypothesis test to see whether there is a difference in the conversion rate between the two groups.
    - Perform a hypothesis test (e.g., chi-squared test) to compare conversion rates between control and treatment groups. Calculate the p-value and draw a conclusion based on the significance level.
- ✓ What is the 95% confidence interval for the difference in the conversion rate between the treatment and control (treatment-control)?

Calculate the confidence interval for the difference in conversion rates.

Conduct a hypothesis test to see whether there is a difference in the average amount spent per user between the two groups. Perform a hypothesis test (e.g., t-

test) to compare average spending between control and treatment groups. Calculate the p-value and draw a conclusion based on the significance level.

- ✓ What is the 95% confidence interval for the difference in the average amount spent per user between the treatment and the control (treatment-control)?

Calculate the confidence interval for the difference in average spending.

- ✓ **Tableau Visualization:**
- ✓ Use Tableau to create visualizations that compare conversion rates and the average amount spent between test groups. You can also visualize the distribution of spending per user for each group and explore relationships with user attributes like device, gender, and country.
- ✓ To provide more specific SQL queries and analysis steps, please share the structure of your database tables and some sample data.

Dataset Overview:

In this comprehensive data analysis, explore user interactions with a product under test. examine user behavior, perform hypothesis tests on conversion rates and spending, and utilize Tableau for visualizations. Key steps include identifying data duplicates, selecting appropriate joins, handling NULL values, and calculating conversion rates and average spending. Inclusivity of all user types is essential for accurate insights. Hypothesis tests determine group differences and confidence intervals provide insights. Tableau is used to visualize metrics by user attributes.

Results:

User Engagement: **48943**

The total users were 48943, 24343 and 24600 were for control and treatment respectively. Perhaps, the total average spent is 3.38, whereas total sum is 165561.

test_group	COUNT of user_id	AVERAGE of total_spent	SUM of total_spent	STDEV of total_spent	SUM of total_spent
Control	24343.00	3.37	82145.90306	25.94	82145.90306
Treatment	24600.00	3.39	83415.32687	25.41	83415.32687
Grand Total	**48943**	**3.38273563**	**165561.2299**	**25.67494579**	**165561.2299**

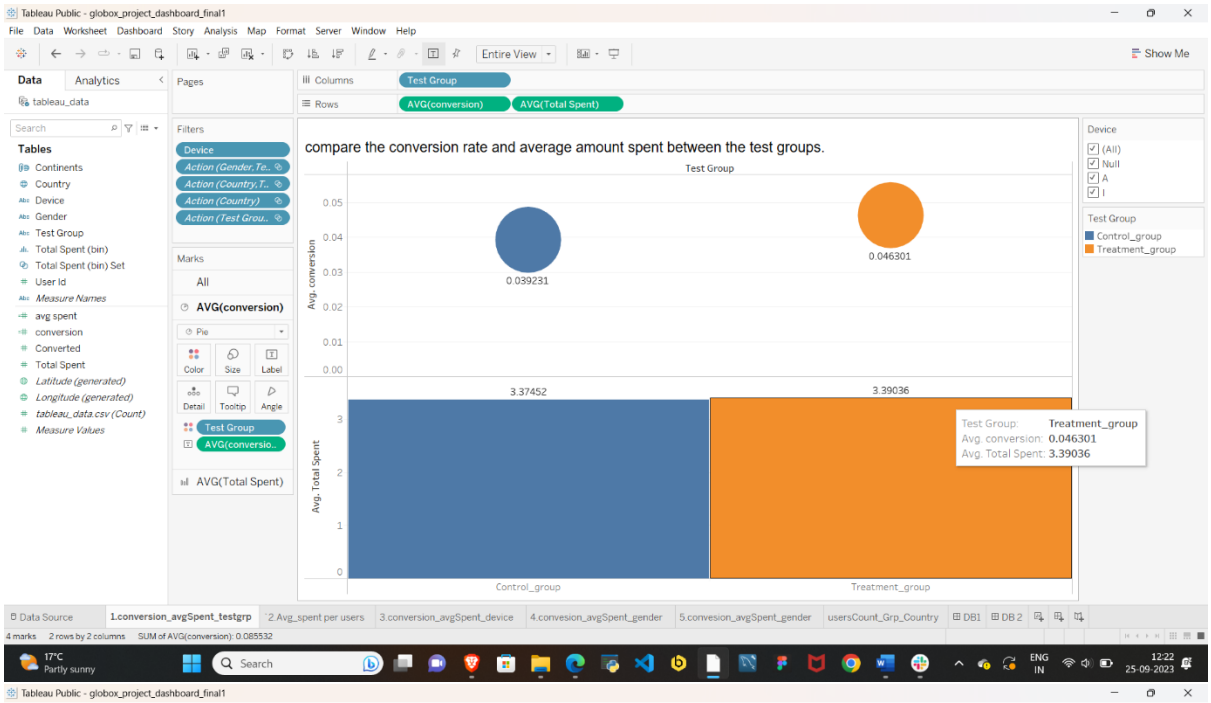


Tableau Public - globox_project_dashboard_final1
compare the conversion rate and average amount spent between the test groups.
Test Group
Avg. conversion
0.039231
0.046301
Avg. Total Spent
3.37452
3.39036
Control_group
Treatment_group
Test Group: Treatment_group
Avg. conversion: 0.046301
Avg. Total Spent: 3.39036

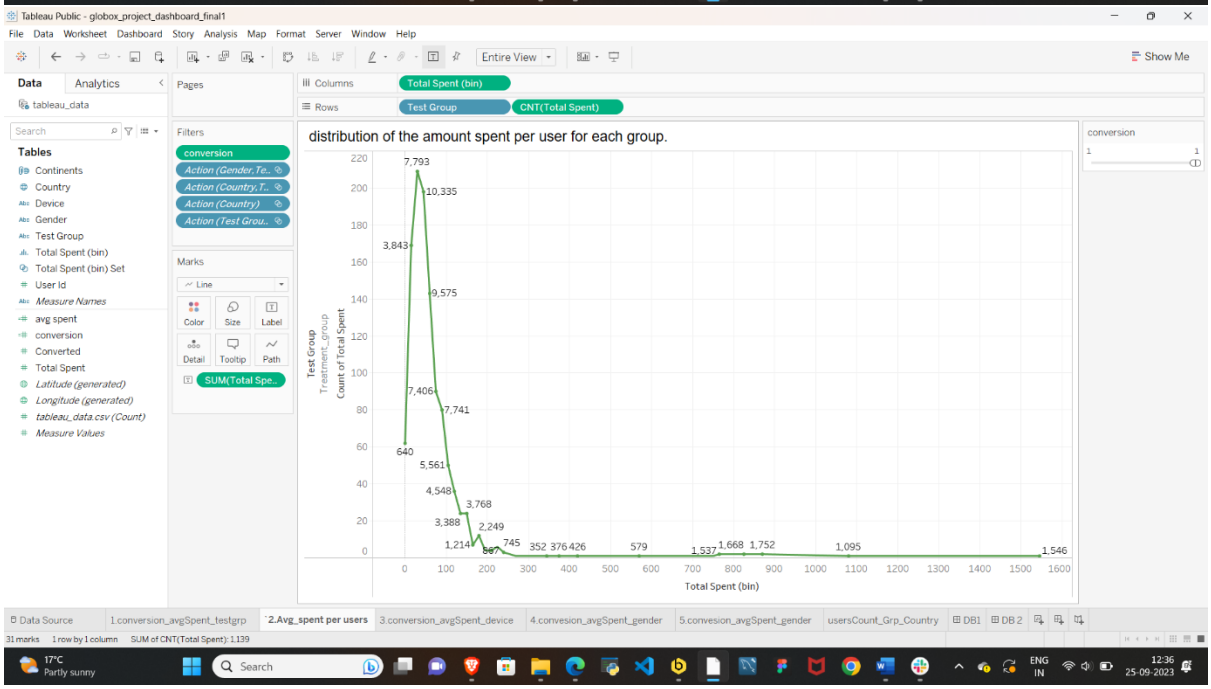


Tableau Public - globox_project_dashboard_final1
distribution of the amount spent per user for each group.
Test Group
Treatment_group
Count of Total Spent
Total Spent (bin)
7,793
10,335
3,843
9,575
7,406
7,741
640
5,561
4,548
3,768
3,388
2,249
1,214
745
352
376
426
579
1,537
1,668
1,752
1,095
1,546

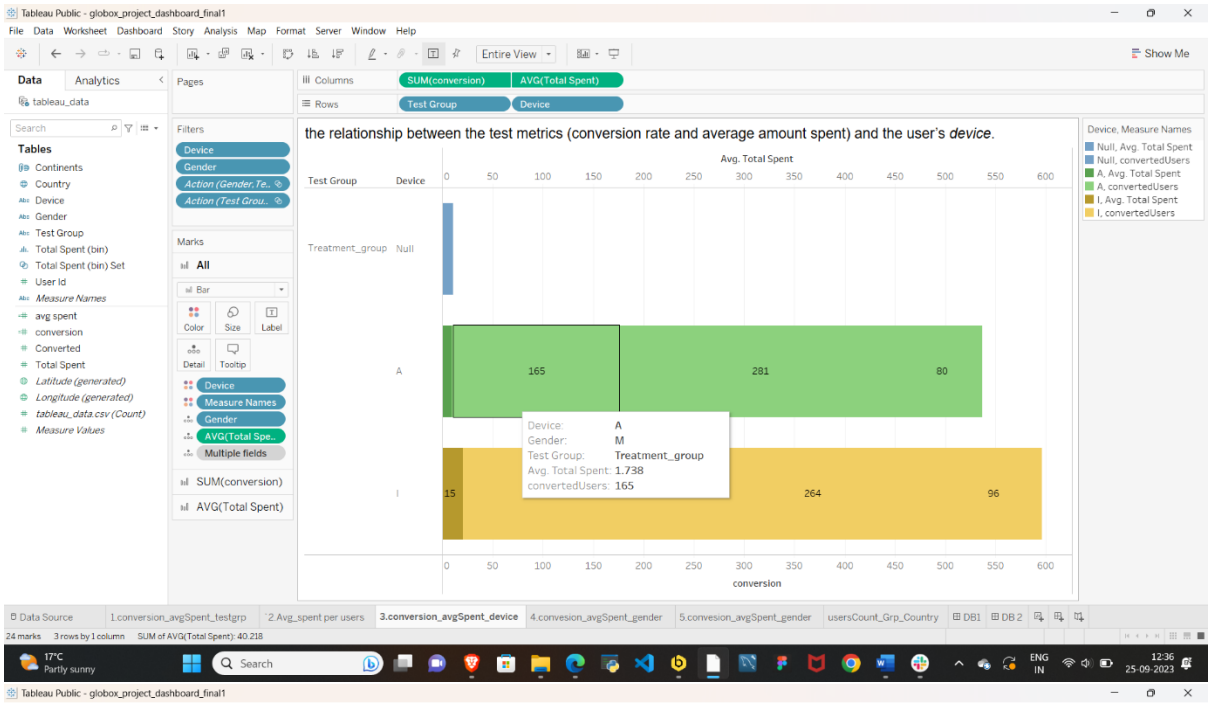


Tableau Public - globox_project_dashboard_final1
the relationship between the test metrics (conversion rate and average amount spent) and the user's device.
Avg. Total Spent
conversion
Device, Measure Names
Null, Avg. Total Spent
Null, convertedUsers
A, Avg. Total Spent
A, convertedUsers
I, Avg. Total Spent
I, convertedUsers
Device: A
Gender: M
Test Group: Treatment_group
Avg. Total Spent: 1.738
convertedUsers: 165

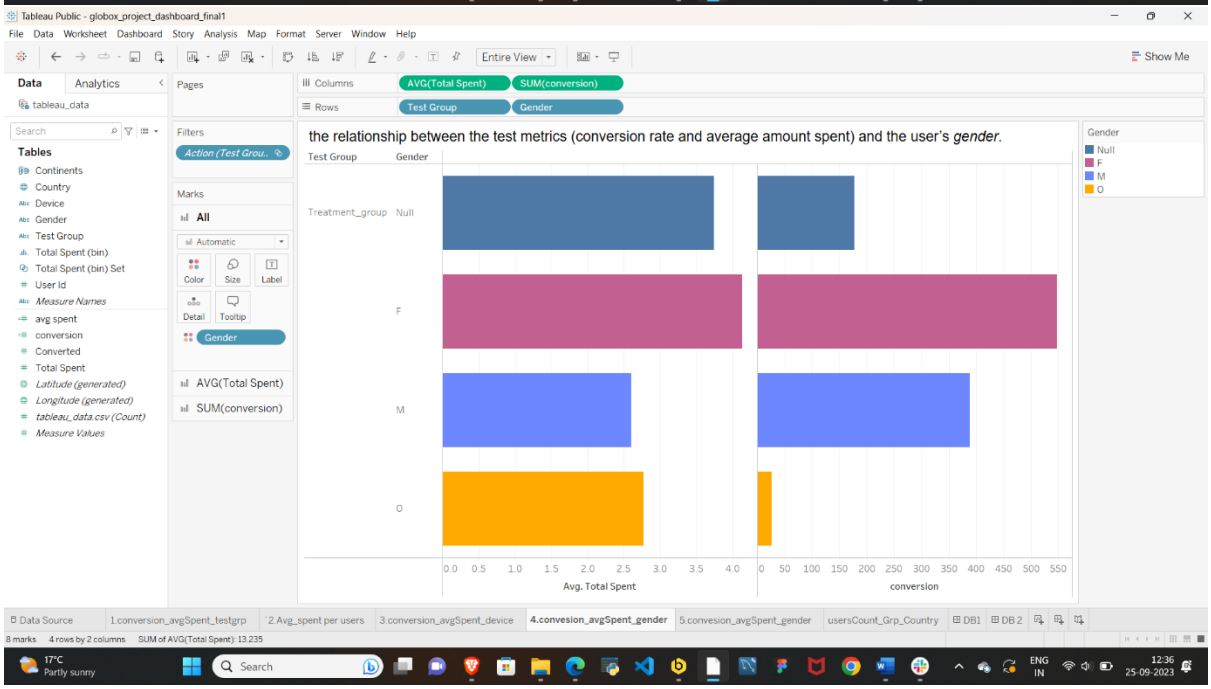
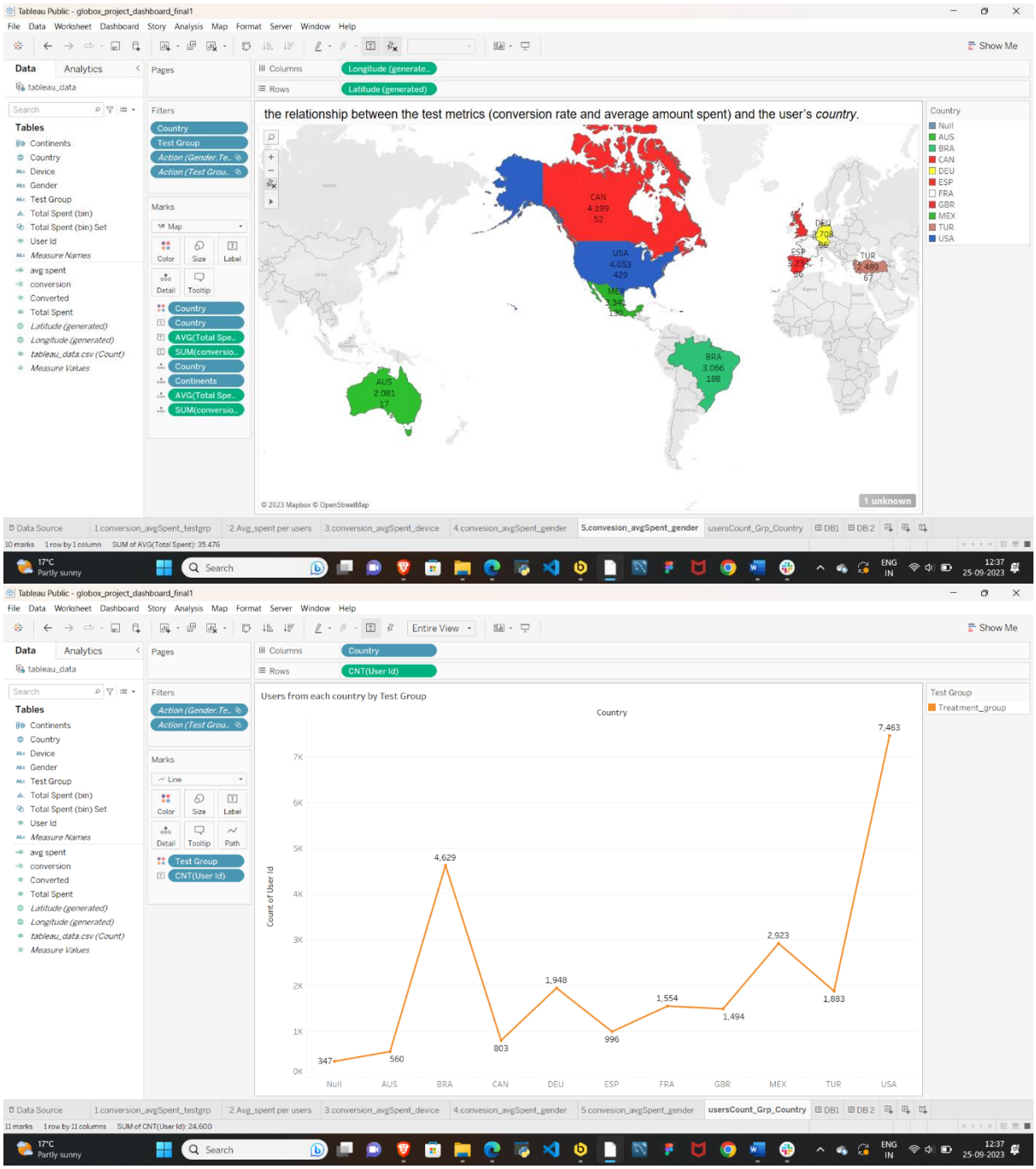


Tableau Public - globox_project_dashboard_final1
the relationship between the test metrics (conversion rate and average amount spent) and the user's gender.
Avg. Total Spent
conversion
Gender
Null
F
M
O



Recommendation:

Statistical Significance: Determine whether the observed differences in conversion rates and average spending between the control and treatment groups are statistically significant. This can be done through hypothesis testing, as mentioned earlier. If the results show statistical significance, it's a positive indicator.

Confidence Intervals: Assess the 95% confidence intervals for both conversion rates and average spending differences. If these intervals do not include zero, it suggests that the observed differences are unlikely due to random chance.

User Feedback: Consider any qualitative user feedback and insights gathered during the test. Feedback can provide valuable information about user preferences and satisfaction with the food and beverage segment.

Long-Term Impact: Evaluate whether the observed impact on revenue and user engagement is sustainable over time. Consider the potential for seasonality or external factors that may have influenced the results.

Iterate and Test: If the results are inconclusive or if there's room for improvement, consider iterating on the banner strategy and running the A/B test again. This allows you to refine the approach and gather more data for a more informed decision.

Collaborative Decision: Involve relevant stakeholders, including the Growth Product & Engineering Team and the Head of Marketing, in the decision-making process. Their expertise and insights are vital in determining the best course of action.

Risk Assessment: Assess the potential risks associated with launching the food and beverage segment more widely. Consider factors like inventory management, marketing costs, and customer retention.

Appendix:

Github: mandeeptoor10/globox: Integrating SQL, Python, and Tableau, this project showcases adept data analysis. Complex datasets are explored, analyzed with Python, and visualized through interactive Tableau dashboards, emphasizing informed decision-making. (github.com)

Spreadsheets Globox_database_q12_s_t - Google Sheets