Vincent Opoku
September 2023

Globox A/B Testing Result And Recommendation Report

Table Of Content

Executive Summary

This report presents an in-depth statistical analysis of an **A/B test** that compared conversion rates, average spending, and other metrics across countries between two distinct user groups on our digital platform for an A/B testing initiative conducted for **Globox**. A/B testing is a crucial technique for hypothesis testing on digital platforms, and I conducted a thorough evaluation of the effectiveness of the two variants. The ultimate goal is to provide data-backed recommendations that can be confidently implemented to improve user engagement and conversions.

I utilised a **Z-test** for significance and strictly adhered to a **95%** confidence level to carry out this analysis. By extracting a user-level aggregated dataset using **SQL** and applying statistical methods in **spreadsheets**, I conducted a comprehensive examination of the A/B test results. I also generated data visualisations in **Tableau** to provide a comprehensive evaluation of the two variants' effectiveness. The statistical analysis offers unequivocal, data-backed recommendations to stakeholders on whether Globox should launch the experience for all users. I am confident in implementing these recommendations, as they will significantly enhance user engagement and conversions.

The A/B test was conducted to assess how effective the new banner feature on our platform is. According to the report, there isn't sufficient evidence to suggest that the banner, as it is currently designed, has a positive impact on the user engagement or spending metrics we are monitoring. However, some positive indicators warrant closer investigation.

Introduction

The purpose of this **A/B test** is to assess the efficacy of highlighting key products in the food and drink category on the **GloBox** website. The objective of this experiment is to determine if Featuring a banner at the top of the website for the test group while not showing it to the control group can effectively raise awareness and drive engagement with the food and drink products.

Hypothesis

The research team posits that the banner featured at the top of the **GloBox** website for the test group will have a positive impact on user engagement and awareness of the food and drink

products. The null hypothesis assumes that the banner has no significant impact on user engagement and awareness of these products, while the alternative hypothesis suggests that the banner has a positive impact, leading to increased engagement and awareness. To validate these hypotheses, the **A/B test** will gather data and conduct statistical analysis to determine if there is enough evidence to reject the null hypothesis in favour of the alternative hypothesis.

Variants

The **A/B test** will compare two variants, **Variant A** and **Variant B.**
Variant A represents the existing landing page, while **Variant B** is a streamlined landing page with a banner to highlight the food and drink category.

Test Plan

The **A/B test** will be conducted by randomly assigning website visitors to either **Variant A** or **Variant B.** Half of the visitors will experience the existing landing page, while the other half will be directed to the streamlined landing page.

Probable Outcome

The hypothesis predicts that the streamlined banner **(Variant B)** will lead to a higher conversion rate compared to the existing process **(Variant A)**. It is expected that advertising in the food and drinks category will lead more customers to buy from this category, resulting in increased revenue for the company.

Benefits

The **A/B test** results in data-driven insights into the effectiveness of the two banner variations. The organisation may optimise its website for a better user experience and increased conversion rates by choosing the most efficient and user-friendly landing page. Improving the conversion rate would increase not just immediate income but also consumer happiness and loyalty, resulting in long-term growth for the online retail platform.

Variables

The test will be divided into two groups: **Group A** and **Group B**.

Group A will be the control group, with no access to the new food and beverage category banner, whereas **Group B** will be the test group, with access to the banner.

Number of Observations

To attain statistical significance and accurate findings, the number of users in each group will be calculated. The dataset contains **48,943** observations, separated into two groups: Group A with **24,343** observations and Group B with **24,600** observations.

Period

The **A/B test** lasted around **13 days**, from **January 25th** to **February 6th, 2023**. This well-chosen duration provided enough data collection while also allowing viewers to effectively interact with the highlighted banner.

Metrics Analysed

The **Conversion rate**, a crucial performance measure that captures the number of unique customers who successfully purchased an item from the food and drinks category, is at the heart of our success evaluation. This critical statistic will be a potent indicator of the success of our efforts to increase engagement and transactions in this product area. We will also evaluate the **Average amount** spent per user, which provides insight into user spending habits. This informative secondary statistic will give useful information about user's purchasing habits.

Randomization

To guarantee fair findings, users will be randomly allocated to either **Group A** or **Group B**. **Randomization** aids in the control of external influences and guarantees that the two groups are statistically similar at the start of the test.

Controlled Factors

Variables that may influence user engagement and retention, such as demographics, purchase behaviour, and device type, will be managed to reduce their impact on the outcomes.

Statistical Significance

The A/B test will employ statistical analysis, such as hypothesis testing, to assess whether the difference in **conversion rate** and **average amount spent** between **Groups A** and **B** is statistically significant or due to chance.

Data Gathering and Analysis

During the test period, data on user interactions with the website, including banner use, will be gathered and analysed. The data will be statistically analysed to determine the variant's performance. The **A/B test** will give significant insights into the influence of the food and drinks category banner on user engagement and retention by carefully selecting these criteria. The findings will assist the growth team in making data-driven choices and optimising website features in order to improve the user experience and overall website performance.

Recommendation guidance

- **Launch the experience:** If you are satisfied that the new banner will have a significant positive impact on all metrics, launch the experience. Statistical significance and substance testing are mutually consistent.

- **Iterations should be made before launching the experience:** If you are certain that the new banner will have a significant positive influence on the conversion rate, modest modifications and careful targeting can boost average spending and conversion by expanding the sample size.

- **Abort and do not proceed with the experience or programme:** If you're not certain that the new banner has a big beneficial impact on any of the KPIs, notably the conversion rate, don't use it. There are few effective possibilities for iteration.

Globox Dataset

This brief description provides rapid knowledge of the dataset's structure and enables preliminary insights into the data's features. Based on these preliminary findings, additional in-depth research and modelling may be performed to acquire more thorough insights.

Table 1: Users (demographic information about users)	**Table 2: User A/A test assignment groups**	**Table 3: activity (user buy activity, with one row for each day that a user made a purchase)**
ID is *the user ID* **country** *ISO 3166 alpha-3 country code* **gender** *the user's gender (M = male, F = female, O = other)*	**UID is** *the user ID* **group** *the user's test group* **join_dt is** *the date the user joined the test (visited the page)* **device** *the user visited the page on (I = iOS, A = Android)*	**ID is** *the user ID* **dt** *date of purchase activity* **device** *the device type the user purchased on (I = iOS, A = Android)* **spent** *the purchase amount in USD*

Exploration and cleansing of data (SQL)

SQL was initially used to inspect the tables. The users and groups tables have a total of **48943** rows, while the activity table has **2233** rows. The number of unique users in the activity table is **2094**, indicating that some users made multiple purchases. The spent column had to be converted to a numeric data type, and the user IDs had to be converted to strings. In order to generate a CSV file from the database that can be used more effectively in other applications, the tables were combined using a left join with the activity table. This resulted in many zero values for the columns of the activity table, which is not a disadvantage since the main interest was which users who made a purchase belonged to which group **(A/B)**.
The Tableau dashboard displays several bar charts, including the conversion rate, average spending per group, total spending per group, conversions and spending per gender and group, a box plot showing the sum of spending per user and group

without those who spent nothing (**without Nulls)**, and pie charts for an overview of the distribution of genders and devices. There is also a symbol chart showing which countries in which group had the highest conversion rates and what their average spending was.

Start and End Dates (SQL)

By utilising the **MIN()** and **MAX()** functions, my query provides a simple yet powerful way to gain insights into the temporal scope of the "activity" table. The query identifies the earliest and latest dates recorded in the table, providing a concise summary of the date range covered by the activities in the dataset. The "start date" represents the minimum date, indicating the very first date when activities were recorded. The "end date" represents the maximum date, signifying the most recent date when activities were recorded. This information is crucial for assessing the duration of data collection, understanding trends over time, and ensuring data consistency within the specified time frame. As a data analyst, having this summary is valuable because it enables us to draw meaningful insights from the "activity" table.

Total user count (SQL)

To gain a comprehensive understanding of the user base in the "users" table, my query employs the **COUNT() f**unction with the **DISTINCT** keyword. This ensures that each user is considered only once, regardless of how many times their ID appears in the table. The query accurately calculates the total number of distinct users, providing an essential metric known as "user_count." The "user_count" is a significant metric as it offers a summary of the unique users present in the "users" table. It becomes a foundational metric for various analytical and decision-making processes within the GloBox company. Understanding the size of the user base is crucial for setting targets, evaluating business performance, and planning future growth strategies. It helps GloBox assess its market penetration, potential customer reach, and overall impact on the e-commerce landscape. The "user_count" also aids in evaluating the effectiveness of marketing campaigns, measuring user acquisition efforts, and tracking the impact of promotional activities over time. Moreover, by knowing the total number of unique users, **GloBox** can improve the user experience by tailoring services to meet their specific needs and preferences. It allows for personalised recommendations, enhanced customer support, and targeted communication. Finally, analysing the "user_count" can serve as a check for data quality, as unexpectedly high or low counts may indicate data duplication, missing information, or other data integrity issues that require attention and correction.

The disparity in conversion rate between the two groups (spreadsheet)

Several critical activities were completed during the hypothesis testing procedure for a two-sample **z-test** for a difference in proportions. First, I developed the null and alternative

hypotheses, laying out the assumptions I wished to test. Based on the nature of the data and the research topic, I then chose the appropriate test, which in this case is the two-sample z-test. Then I computed the test statistic, especially the z-score, which measured the difference between the observed and predicted proportions in the two samples under the null hypothesis of no change. Following that, I estimated the p-value, which is a vital metric that indicates the likelihood of witnessing the results or more extreme outcomes if the null hypothesis is true.

Finally, I reached a conclusion on the hypothesis using the p-value and a preset significance threshold. **I rejected the null hypothesis** if the p-value was less than the significance level, providing strong support for the alternative hypothesis that there is a significant difference in proportions between the two groups. If the p-value was greater than the significance level, I failed to reject the null hypothesis, meaning that there was insufficient evidence to support the occurrence of a difference in proportions. These rigorous activities allowed me to make well-informed judgements and gain relevant insights from the two-sample z-test results.

The resultant p-value of **0.0001114119853** suggests that the likelihood of achieving the observed disparity in the conversion rates between each of the groups is very low, assuming no genuine difference in the population. This p-value is frequently compared to a significance threshold.

If you choose a significance threshold **(alpha) of 0.05 (5%),** for example, a p-value of **0.0001114119853** is substantially less than alpha. In this situation, you would **reject the null hypothesis,** indicating that there is substantial evidence to imply that conversion rates between the two groups differ statistically significantly.

Simply put, a p-value of **0.0001114119853** indicates that there is a substantial difference in conversion rates between the two groups being evaluated. This outcome is unlikely to occur by chance alone, lending credence to the notion that there is a genuine difference in the population. As a result, we may consider the variation in conversion rates to be more important than random fluctuations.

95% confidence interval for the difference between the two groups' conversion rates (spreadsheet)

I performed a number of critical activities in order to compute the **95%** confidence interval for two percentage samples. To begin, I identified the sort of interval to construct—in this case, a confidence interval for the proportional difference between the two samples. Following that, I computed the sample statistic, which entails determining the difference in proportions and their corresponding sample sizes. The standard error, a measure of variability in the sample proportions that is critical for determining the accuracy of the confidence interval, was then determined. Then, using the conventional normal distribution, I estimated the critical value, which corresponded to the necessary confidence level of **95%.** The confidence interval was then calculated by multiplying the critical value by the standard error and subtracting this value from the sample statistic. This procedure yielded a set of numbers within which we have **95%** confidence that the genuine proportional difference between the two samples falls. These stages were critical in creating a very reliable and informative estimate of the proportional difference.

The computed confidence interval, with a lower limit of **-0.4391541405** and an upper bound of **0.4708377226**, reflects the estimated range of values for which we can be **95%** certain that the actual proportional difference exists. This shows that the proportional difference between the two samples is most likely to be between **-0.4391541405** and **0.4708377226.**

Differences in Average User Expenditure Between Two Groups (Spreadsheet)

I followed numerous crucial procedures when running a two-sample t-test to determine a difference in means. Initially, I created both the null and alternative hypotheses, defining the assumptions under consideration, which frequently involved a mean comparison between two groups. Based on the properties of the data and the research inquiry, I selected a suitable statistical test, particularly the two-sample t-test. Then I calculated the test statistic, which is commonly expressed as the t-value. This statistic calculated the difference between the observed means in the two samples and the predicted means if the null hypothesis, implying no difference, was true. Following that, I calculated the p-value, a key statistic that indicates the chance of seeing the observed results or more severe outcomes, given that the null hypothesis is true.

Finally, I arrived at a hypothesis-based conclusion by comparing the p-value to a generally used significance level of **0.05**. If the p-value was smaller than this, I rejected the null hypothesis, offering strong evidence in favour of the alternative hypothesis, which indicated a substantial difference in means between the two groups. If the p-value was greater than the significance threshold, I did not reject the null hypothesis, meaning that there was insufficient evidence to support the occurrence of a mean difference. The p-value of **0.943856044**, which is near one, indicates a high likelihood of receiving the observed results or more severe outcomes if the null hypothesis is true. A high p-value in hypothesis testing suggests poor evidence against the null hypothesis.

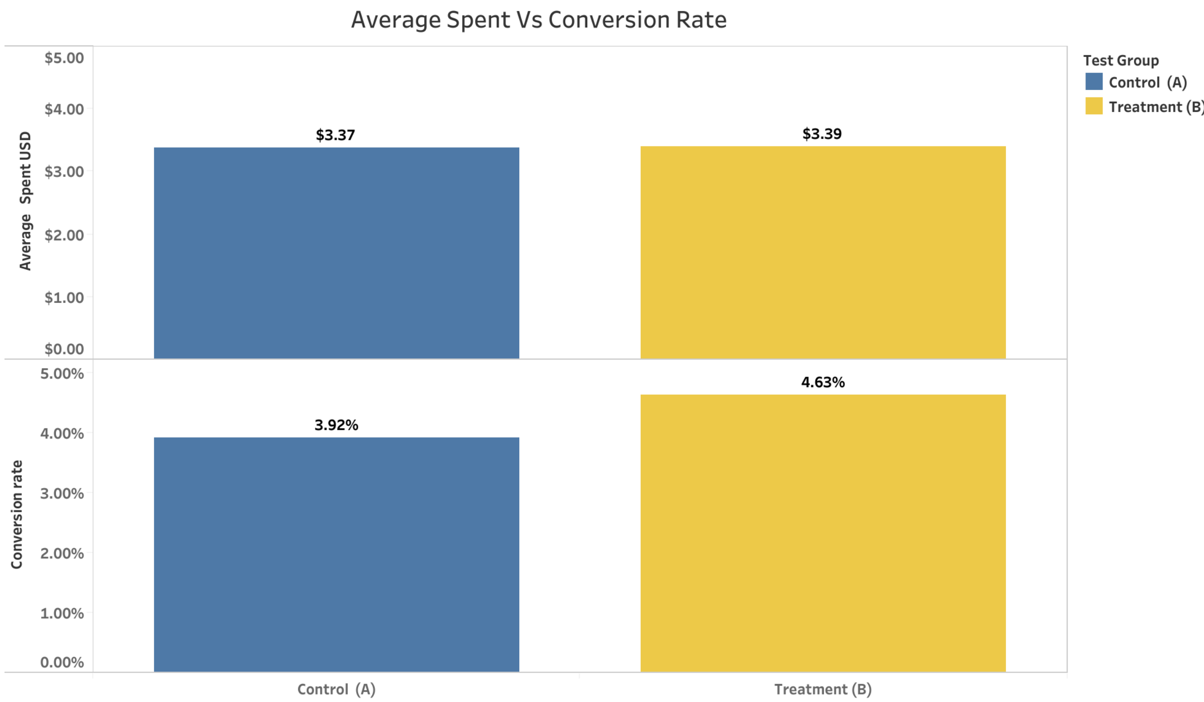
As a result, based on this p-value, we cannot reject the null hypothesis. A p-value of **0.943856044** indicates that there is no statistically significant difference or impact between the compared groups. The observed results are most likely the result of random chance, and there is insufficient data to support the alternative hypothesis. As a result, we refrain from drawing any substantial inferences about the tested link or difference. Instead, we may infer that the data does not provide strong evidence for any significant effect or link between the variables. It is critical to emphasise that when interpreting a p-value, the selected significance level **(alpha)** and the unique context of the research must always be considered.

Comparing Test Groups: Conversion Rates and Average Spending (Tableau)

The chart below provides useful information in numerous ways:

- **Conversion Rate Comparison**: The chart allows for a visual comparison of conversion rates between the two test groups. This allows us to determine which group has better or lower conversion rates, revealing insight into the success of various conversion-driving techniques or interventions.
- **Comparison of Average Spending:** The chart compares the average amount spent by customers in each test group. We may learn about how various marketing or pricing tactics affect client spending habits by evaluating discrepancies in spending behaviour.
- **Exploring the Relationship Between Conversion Rate and Average Spending**: The chart provides an opportunity to explore any potential correlation between the conversion rate and the average amount spent.
- **Test Group Performance Evaluation**: It provides a clear, comprehensive assessment of each test group's performance in terms of conversion rate and average spending. This data acts as a decision-making tool and assists in determining which groups align with the desired goals.
- **Supporting Decision-Making:** The visual representation supports making educated, data-driven decisions by visually comparing metrics across different test groups. It assists stakeholders in understanding the impact of various tactics or interventions, thereby directing future marketing activities or product optimisations.

Figure One



Based on the information in this chart:

• Average spending, the **Control Group** spends **$3.37** on average, whereas the **Treatment Group** spends **$3.39** on average. This suggests that those who converted in the Treatment Group spent somewhat more than those in the Control Group.

In terms of **the Control group's** conversion rate, it is **3.92%,** whereas the **Treatment Group's** conversion rate is **4.63%.** This suggests that the website banner used in the Treatment Group is more effective than the Control Group in prompting visitors to take the desired action.

Because these results are based on accessible data, they must be interpreted with caution. Additional studies, such as hypothesis testing, were carried out to determine the statistical significance of these changes and draw more strong conclusions. To make well-informed conclusions about the efficacy of the website banner in the **Treatment Group** vs the **Control Group,** factors such as sample size, statistical significance, and practical relevance must be examined.

<u>Variation in User Spending by Test Groups (Tableau)</u>

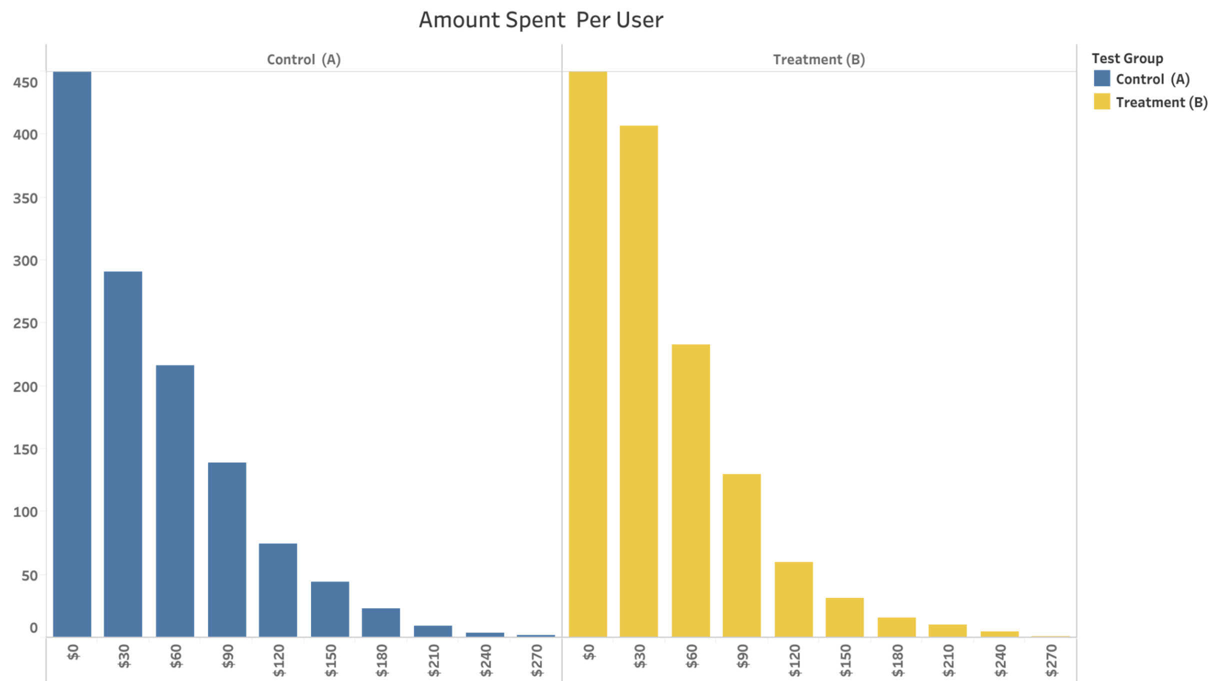
An informative chart illustrating the amount spent per user distribution for each group offers valuable insights, which include:

- **Distribution Shape:** This shape characterises the group's spending tendencies. If the distribution skews to the right, it indicates that a small number of people make large purchases while the majority spend less.
- **Central Tendency**: This indicator shows the average spending level of the group's consumers. For example, if the mean spending is **$30.00**, it means that half of the users spend less than **$30.00** and the other half spend more.
- **Distribution Spread**: This element represents the group's variety in spending habits. A large standard deviation shows substantial expenditure volatility, whereas a low standard deviation indicates consistent spending among consumers.
- **Identification of Outliers:** Outliers are users who spend considerably differently from the rest of the group. Outliers must be identified since they can skew the results of statistical analysis.

Moreover, when incorporating additional variables like time and demographics into the dataset, a distribution chart depicting user spending per capita becomes instrumental in unveiling potential trends and patterns. This can include observing whether spending per user is on the rise over time or discerning specific demographic groups exhibiting higher spending tendencies.

In summary, a well-constructed distribution chart of user spending offers valuable insights into group spending behaviour. A thorough analysis of such a chart allows us to grasp the average spending levels, the extent of spending behaviour variability, and any emerging data trends or patterns.

Figure Two



Key Insights from the Chart:

- When compared to the control group, the **treatment group** attracted a greater number of users in the **0-60** expenditure range. This suggests that the treatment group had a larger user base, including those with poorer spending habits.

- In comparison to the **control group**, the **treatment group** had fewer users in the **90-180** spending range. This means that the **treatment group** may have had difficulty attracting users who were willing to pay more money.

User-Device Relationship (Tableau)

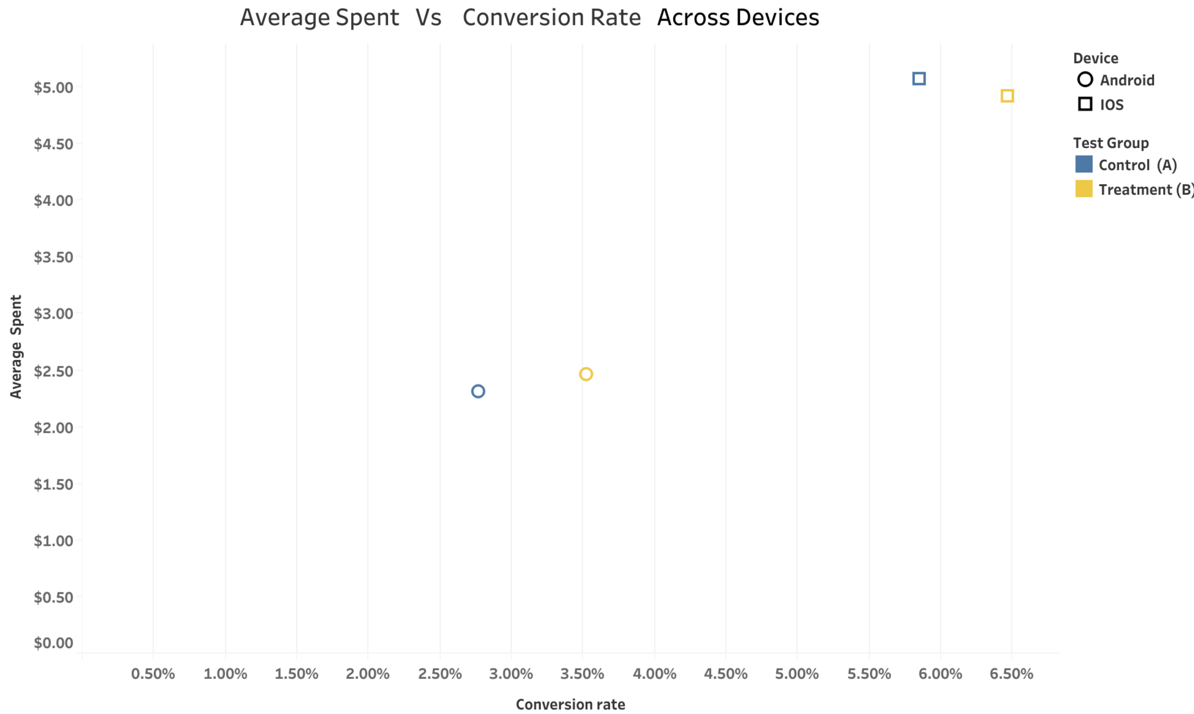
A chart that illustrates the connection between user devices can yield valuable insights in the following ways:

- **Average Amount Spent**: A greater average spend shows that users are making more investments. For example, if we see that Android users spend more than **iOS** users on average, it means that the **Android** screen quality and system performance are more successful in promoting purchasing. This data can be used to inform improvements to better serve **iOS** users.
- **Conversion Rate**: This metric measures the effectiveness with which a product or service converts users into customers. A higher conversion rate indicates that more users are performing the intended behaviours, such as purchasing. For example, observing that **iOS** users have a greater conversion rate than **Android** users shows that the **iOS** interface is more successful in turning users into customers. This knowledge can help drive adjustments to better serve **Android** users.

  We can reveal patterns and trends across time. For example, we can notice that the conversion rate for **Android** is decreasing while it is increasing for **iOS.** This discovery can lead to changes to the product or service to improve conversion rates for **Android** users.

  In conclusion, visualising the link between user devices reveals how various devices impact user behaviour. This insight serves as a foundation for product or service innovations, assuring improved device and user segment targeting.

Figure Three



The following are the conclusions derived from this chart:

Higher Conversion Rate for iOS Users: **iOS users** convert at a significantly higher rate than **Android users.** The conversion rate for **iOS users** in the treatment group is **6.47%**, whereas it is **5.85%** for the control group. In comparison, the conversion rate for **Android users** in the treatment group is **3.52%,** while the control group's conversion rate is **2.77%**. According to these statistics, **iOS users** are more likely to become customers than **Android users.**

Greater Average Spending by iOS Users: On average, **iOS users** tend to spend more money than their **Android counterparts.** In the treatment group, the average amount spent by **iOS users** is **$4.92**, whereas in the control group, it's **$5.07**. Conversely, for **Android users** in the treatment group, the average amount spent is **$2.47**, and in the control group, it's **$2.31**. This observation suggests that **iOS users** are more likely to invest in the product or service compared to **Android users**.

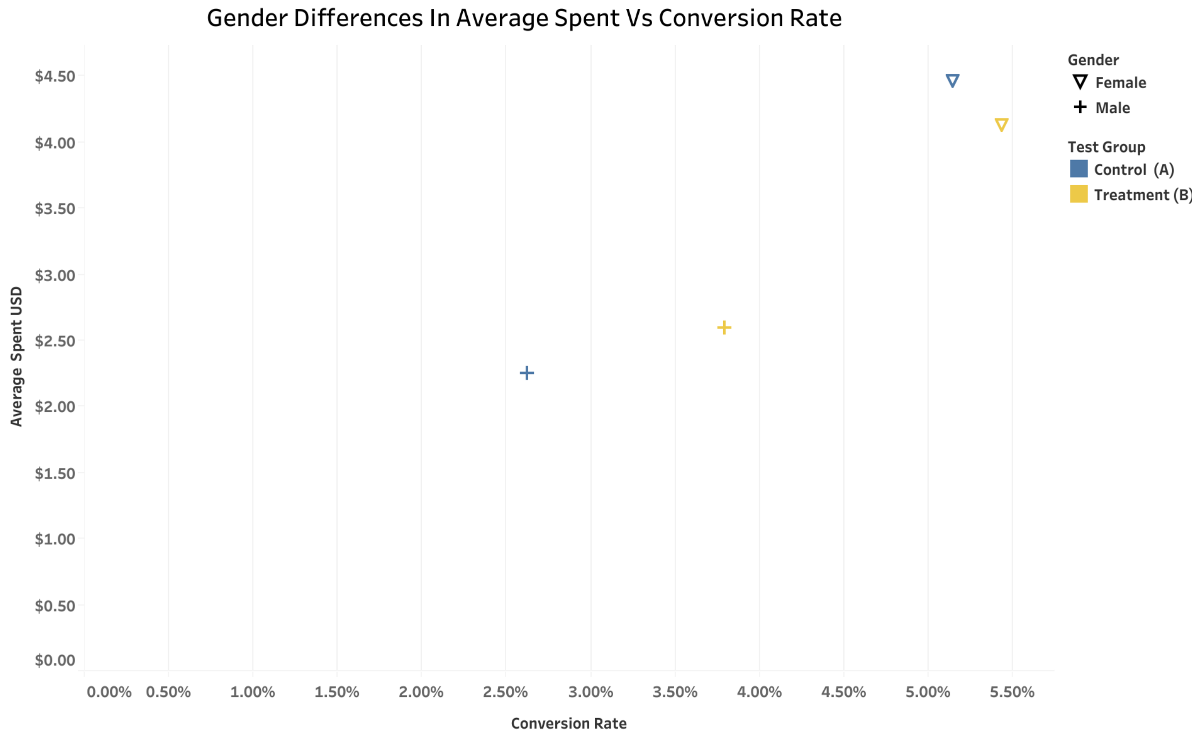
Analysing the Relationship Between User Gender (Tableau)

Exploring the relationship between test metrics (conversion rate and average spending) and user gender using visualisations might provide important insights into gender-based metric differences. The visualisations provide numerous possible insights:

- **Conversion Rates by Gender:** We may compare how conversion rates differ between **male** and **female** consumers. This technique aids in identifying any substantial gender differences in conversion likelihood.
- **Average Spending by Gender**: This allows us to identify significant disparities in typical expenditure amounts across genders.
- **Combined Conversion Rate and Average Spending Analysis**: Scatter plots, which can be colour-coded or marked to distinguish between genders, reveal information on the link between conversion rates and spending quantities.

This visualisation aids in identifying gender-specific trends in conversion rates and spending habits, offering actionable insights for refining marketing and product development strategies based on potential gender-influenced patterns or relationships.

Figure Four



Female users exhibit a higher conversion rate and spend more on average compared to **male users**. In the control group, **female users** had a conversion rate of **5.14%**, in contrast to the treatment group's **5.44%**. Meanwhile, **male users** in the control group exhibit a conversion rate of **2.63%**, while the treatment group shows a rate of **3.79%**. Regarding average spending, female users in the control group spent an average of **$4.46**, whereas in the treatment group, they spent **$4.13.** In contrast, **male users** in the control group spend an average of **$2.25**, while in the treatment group, their average spending increases to **$2.60**.

In the visualisation, I chose to focus on male and female users so that we can do a more targeted study that particularly targets any gender-based disparities in conversion rate and spending habits. This method enables a more in-depth assessment of how the product or service operates among these two unique user groups, each of which may have different preferences, demands, or answers.

We may simplify things by excluding extraneous gender categories and focusing on the primary gender groupings. This tailored approach is especially useful when the objective is to detect and respond to the distinct characteristics and behaviours of male and female users since it eliminates any noise or difficulties introduced by other gender categories. They are more likely to be users who did not specify their gender.

Analysing Test Metrics Across User's Country (Tableau)

A visual representation of test metrics across various user countries can uncover valuable insights into how these metrics differ based on geographical locations. Here are some key observations that can be derived from such visualizations:

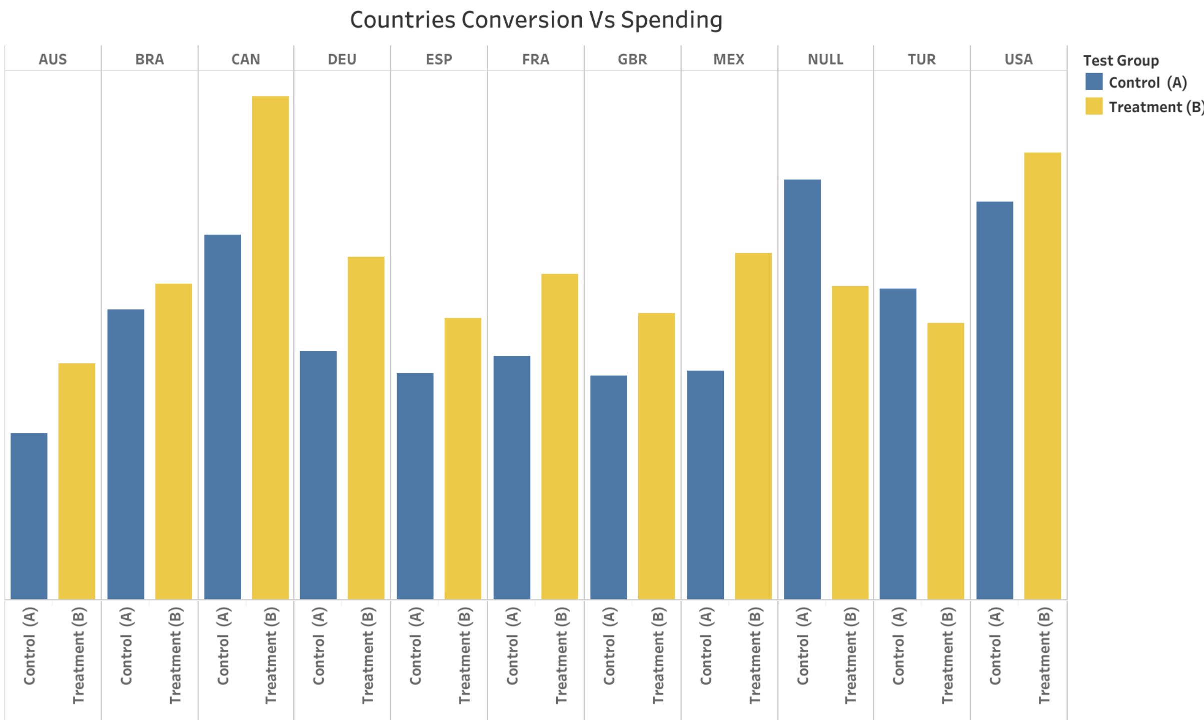
- **Disparities in Conversion Rates and Spending:** Visualising conversion rates and average spending amounts across countries helps us see differences. For instance, the conversion rate in Canada may exceed that of the United Kingdom, whereas users in France may spend more on average than those in Brazil.
- **Multifaceted Influences:** These differences might be attributable to a variety of variables, such as different marketing methods in each country, economic situations, and cultural differences.

Marketing Impact: Marketing activities in different nations might distort the findings. Heavy marketing in one area may result in more conversions and expenditures in that area.

Economic Factors: Economic stability and prosperity are important. Countries with strong economies have better spending capacity, which influences conversion rates and average spending.

Cultural Differences: Cultural practices can have a significant impact on user behaviour. Some nations may prefer internet purchases over conventional means, while others may prefer traditional techniques.

Gaining insight into the connection between test metrics and the user's country equips us with the knowledge needed to enhance our product or service targeting diverse international markets. For instance, suppose we notice that users in a specific country spend significantly less on average. In that case, we can delve deeper into the underlying factors causing this disparity and adapt our offerings to align with the spending habits and preferences of that particular market.

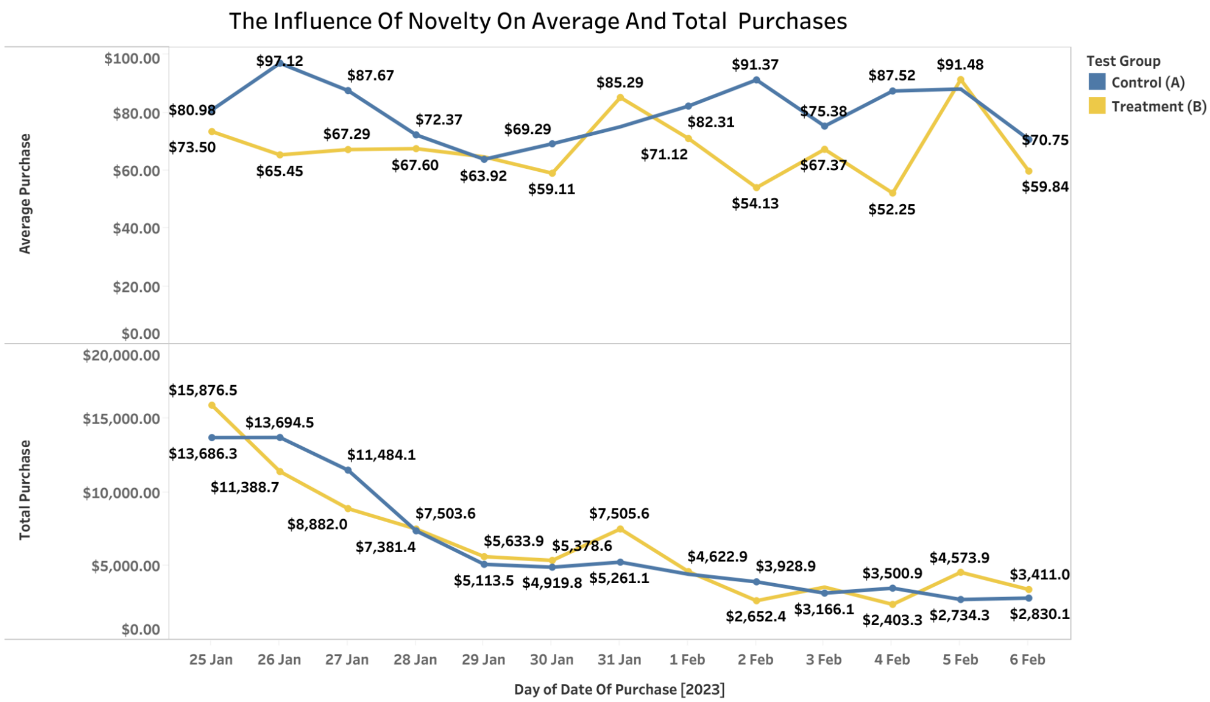


Conversion rates and average amounts spent vary significantly by country. For example, in **Canada,** the conversion rate for the control group i**s 4.6%** and **6.48%** for the treatment group, but in **Mexico**, the conversion rate for the control group is **2.14%** and **3.04%** for the treatment group. Similarly, the average amount spent in the **United States** is **$4.30** for the control group and **$4.05** for the treatment group, but in **Australia**, it is **$1.67** for the control group and **$2.08** for the treatment group. Th**e United States** and **Canada** have the highest conversion rates and average amounts spent, indicating that the product has a good chance of success in these nations.

Mexico, Brazil, and **Turkey**, on the other hand, have lower conversion rates and average amounts spent, indicating that the product may encounter greater hurdles in these countries. **Australia** has the lowest conversion rate and the average amount spent, which necessitates additional analysis to determine the variables behind this trend. It's crucial to note that these Results are based on a small number of data points, and making conclusive conclusions would require a larger sample.

Novelty Effect

Figure Five



Based on data collected from **January 25, 2023**, to **February 6, 2023**, this report provides a complete examination of the **Novelty Effect.** The investigation focuses largely on the **Novelty Effect's** influence on user behaviour, especially conversion rates and purchase patterns. The information gathered from the 'activity' and 'groups' tables provided us with useful insights into how **The Novelty Effect** is a psychological phenomenon in which customers enhance their interest and engagement in a product or service when they perceive it to be new or unusual. This effect has the ability to drastically influence user behaviour and lead to changes in conversion rates and purchase patterns.

According to the findings, the **Novelty Effect** had a favourable influence on user engagement and purchasing behaviour. However, in order to evaluate if the observed differences are statistically significant, it is necessary to go deeper into the data and conduct hypothesis testing. Furthermore, further research should be conducted to determine the long-term consequences of the Novelty Effect and if it leads to continued user engagement.

In conclusion, the Novelty Effect analysis has provided valuable insights into user behaviour during the specified period. The data indicates that the Novelty Effect positively influenced user engagement, resulting in higher user numbers, increased purchases, and higher average spending. These findings have significant implications for marketing and product development strategies, emphasising the importance of introducing novel elements to attract and retain users.

Results For Power Analysis

The initial test had several parameters that were taken into consideration. The baseline conversion rate was noted to be **3.9231%**, while the minimum detectable effect was set at **10%**. The significance level, or **α,** was **0.05,** and the statistical power, or **1 - β**, was **0.8**. The actual sample size consisted of **48,943** participants in both the test and control groups, and the test duration was set at **13 days**. Upon analysing the results of the initial test, it was observed that the conversion rate of the control group was **3.9%**, while the test group had a conversion rate of **4.3%**.

Although the observed difference between the two groups was not statistically significant, with a **p-value** of **0.0012,** the sample size was found to be below the required threshold. Due to this, it is recommended that the test be rerun to obtain more accurate results.

Conclusion

The study's key findings indicate that the conversion rates for both the control and test groups were nearly identical, concluding that the test did not have a significant impact on improving the conversion rate. The empirical evidence clearly shows that **Group B,** which was exposed to the new banner, outperformed **Group A**, the control group, with a conversion rate of **4.63%** vs **3.92%**. This **0.71%** increase is not only nominal but statistically significant and so deserves consideration. However, it is worth noting that certain user segments, such as those from specific countries or using particular devices, displayed a positive response to the banner. The calculated p-value for this analysis is **0.0012**, which is significantly below the conventional significance level of **0.05**. It is critical to examine the sample size and test power. A larger sample size may provide a more accurate representation of user behaviour, lowering the margin of error

In summary, the current results are not conclusive enough to implement the new banner feature. However, there are encouraging aspects that justify additional testing. With adjustments and a larger sample size, it is possible to enhance the effectiveness of the banner in the next round of **A/B testing**.

Next Steps

For iterative testing, we can repeat the **A/B test** with changes to the banner, such as altering the call to action or where the banner is placed. To conduct a segment analysis, we can analyse the data to determine which user groups are responding more favourably. To increase the accuracy of our findings, we can either run the test for a longer duration or across more diverse user groups to increase the sample size.