GloBox A/B Homepage Analysis

By: Hugo Gomes.

Table of contents

Executive Summary

The data analysis conducted revealed several key insights regarding the impact of the new homepage (Treatment) on conversion rates and average spending.

Key Insights:

The introduction of the new homepage resulted in a higher conversion rate in the Treatment group compared to the Control group.
However, there was no significant difference in the average amount spent between the Control and Treatment groups.
Country-specific data indicated that the new homepage showed promising results in certain regions, suggesting the potential for selective implementation.

Recommendations:

Implement the new homepage selectively: Based on the analysis, it is recommended to implement the new homepage in specific countries or regions where it has shown promising results. This targeted approach allows for a more efficient allocation of resources and minimises risks associated with a broader implementation.

Continuously monitor and analyse: Regularly monitor the performance of the new homepage in the selected countries and assess its impact on conversion rates and user behaviour. Gather additional insights to inform future decision-making and optimise the user experience.

Explore other factors: Investigate additional factors that may contribute to improved conversions and revenue generation. Conduct further research, A/B testing, or analyse user behaviour to identify elements that can enhance the user experience and drive higher conversion rates.

Data Sources and Limitations:
The data used for the analysis primarily came from internal databases, which provided information on user behaviour, conversion rates, and average spending. However, it is important to acknowledge the limitations and assumptions made during the analysis, such as the reliance on historical data and the potential influence of external factors not captured in the dataset.

Please refer to the respective sections of the full report for more detailed information, including methodology, data analysis and additional insights.

Introduction

This report aims to analyse the impact of a new homepage on user conversions and average spending in an online platform. By conducting an A/B test, we compare the control group (Group A) with the treatment group (Group B) to evaluate the effectiveness of the novelty factor in driving conversions and influencing user behaviour. The analysis focuses on identifying any significant differences in conversion rates and average spending between the two groups. The insights derived from this analysis will provide valuable information for decision-making and potentially inform strategies to enhance user engagement and revenue generation.

Methodology

The analysis is based on an A/B test, a commonly used experimental design in which users are randomly assigned to either the control or treatment group. In this case, Group A represents the control group, where users experience the existing platform, while Group B represents the treatment group, where users are exposed to the platform with the new homepage. By comparing the outcomes between the two groups, we can assess the impact of the new homepage on conversions and average spending.

To measure conversions, we track the number of users who successfully complete a desired action, such as making a purchase. Additionally, we analyse the average spending per user as a key metric to understand the monetary impact of the novelty factor.

The A/B test is conducted over a 13-day period, during which we collect and analyse relevant data. The statistical significance of the results is assessed using hypothesis testing, considering a significance level of 5% ($p < 0.05$) as the threshold for determining the presence of a meaningful difference between the control and treatment groups.

Data Description

The data used for this analysis comprises user-level information, including demographic attributes, device type, and user activity logs. We obtain this data from the platform's internal databases, ensuring the privacy and confidentiality of user information.

To prepare the data for analysis, necessary cleaning and preprocessing steps are performed via SQL. This includes handling missing values, removing outliers, and aggregating the data at appropriate levels for analysis. The cleaned and transformed data is then used to calculate key metrics such as conversion rates and average spending per user in EXCEL.

It is important to note that the duration of the experiment is limited to 13 days, which may not capture the full range of dynamics in user behaviour. Additionally, as the experiment focuses on a specific novelty factor, the results may not be generalizable to other interventions or factors that can influence user conversions.

The following sections of this report will present the analysis and findings, recommendations based on the insights derived, and limitations to consider when interpreting the results.

Analysis and Findings

In the A/B test conducted, the Control group consisted of 24,343 users, while the Treatment group consisted of 24,600 users. The conversion rate observed in the Control group was 3.92%, whereas in the Treatment group, it was 4.63%. The overall conversion rate across both groups was calculated to be 4.28%.

Furthermore, the average amount spent by users in the Control group was $3.375, while in the Treatment group, it was $3.391.

To determine the statistical significance of the observed differences, a significance threshold of 5% was set. Using a two-sample z-test for comparing Conversion Rates, the resulting p-value was calculated to be 0.000111412. This p-value suggests strong evidence against the null hypothesis and indicates a significant difference in the conversion rates between the Control and Treatment groups.

The margin of error for the estimated difference in conversion rates was calculated to be 0.003583837. The lower bound of the confidence interval for the difference in conversion rates was found to be 0.003485985 (0.35%), while the upper bound was 0.01065366 (1.07%). These confidence interval bounds provide a measure of uncertainty around the estimated difference in conversion rates and indicate the plausible range of values for the true difference.

The calculated difference in conversion rates between the Control and Treatment groups was 0.007069823 (0.71%), indicating a higher conversion rate in the Treatment group compared to the Control group.

A two-sample t-test was performed to compare the average spent amounts between the Control and Treatment groups. The test was conducted with a confidence level of 95% to determine the statistical significance of the observed difference.
The standard deviation of the Control group was calculated to be 25.93585782, while the standard deviation of the Treatment group was found to be 25.41359305. Using these values, a pooled standard deviation of 0.232135815 was obtained.

The resulting p-value from the t-test was determined to be 0.943854898. This p-value indicates that there is no statistically significant difference in the average spent amounts between the Control and Treatment groups.

The standard error of the difference in average spent amounts was calculated to be 0.232135815.

Considering a confidence level of 95%, the critical value associated with this level was found to be 1.960061445. The margin of error, which represents the maximum likely difference between the sample means and the population means, was determined to be 0.455000461.

The lower bound of the confidence interval for the difference in average spent amounts was -0.438651983, indicating that the true difference may be negative but not statistically significant. The upper bound of the confidence interval was found to be 0.471348939, suggesting that the true difference may be positive but not statistically significant.

These findings suggest that there is no significant difference in the average spent amounts between the Control and Treatment groups. The results are subject to a confidence level of 95% and indicate that the true difference falls within the calculated confidence interval bounds.

Minimum Detectable Effect for the difference in means:

To determine the minimum detectable effect (MDE) for the **difference in means** between the Control and Treatment groups, an analysis was conducted specifically for the average values. Based on the obtained data, it was found that a sample size of at least 5240 users per group would be required, resulting in a total sample size of 10480.

By achieving this sample size, we can achieve a power of 95%, indicating a 95% chance of detecting a significant difference in the average values between the Control and Treatment groups if such a difference truly exists. A significance level of 5% was used as the threshold for detecting a true difference in means between the two groups.

It is important to note that the calculated pooled standard deviation of 0.232135815 played a vital role in determining the necessary sample size and ensuring an adequate power for the analysis.

Considering these factors, the chosen sample size and statistical parameters provide us with a robust framework to detect meaningful differences in the average values between the Control and Treatment groups in our case study.

Minimum Detectable Effect for the difference in conversions:

Determining an appropriate Minimum Detectable Effect (MDE) for the **difference in conversions** requires careful consideration of the desired level of sensitivity to detect changes in conversion rates. In this analysis, it is challenging to provide a conclusive recommendation for the MDE due to the lack of specific criteria defined by the company. However, I can outline the implications of setting different MDE values and their potential benefits in a technical manner:

Lower MDE: Setting a lower MDE allows for the identification of even small variations in conversion rates. This approach would be advantageous if there is an expectation that increased spending on the product would lead to consistent but potentially modest improvements in conversion rates. It enhances the sensitivity of the analysis to detect subtle changes, enabling a detailed evaluation of the impact of interventions or changes on conversions.

Higher MDE: Alternatively, a higher MDE indicates a focus on detecting larger changes in conversion rates. This approach assumes that only substantial differences in conversion rates will significantly impact revenue. Setting a higher MDE allows for a more conservative approach, targeting interventions or changes that are expected to yield significant improvements in conversion rates.

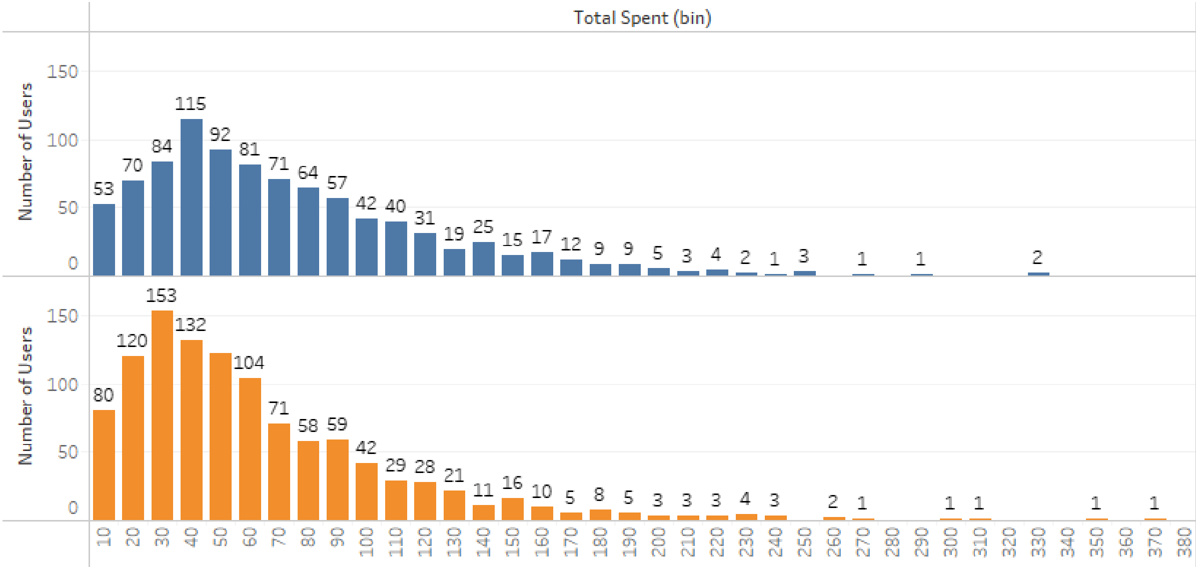
Given the absence of specific criteria defined by the company, it is challenging to determine the most appropriate MDE value. Therefore, further discussions and input from relevant stakeholders are recommended to establish the desired level of sensitivity and identify the appropriate MDE for conversions.

Please refer to the graphic below for a visualisation of the average spending per group. Blue represents the Control group and Orange the Treatment group.

In terms of the gross income, it is important to note that the Treatment group primarily consists of a larger number of users who make smaller individual purchases, resulting in a higher count of transactions. Conversely, the Control group exhibits a different pattern, with fewer users but higher average spending per user. To provide a visual representation of these insights, a bin bar chart has been created to illustrate the distribution of user purchases in both groups.

For a more comprehensive understanding of the average spending per user in each group, please refer to the visual representation provided below.

Visualisation Note: The bin bar chart uses blue for the Control group and orange for the Treatment group. Each bin represents a range of purchase values ($10 units), with the count of users displayed on the Y-axis and on the top of each bin. The exact purchase values are shown on the X-axis.

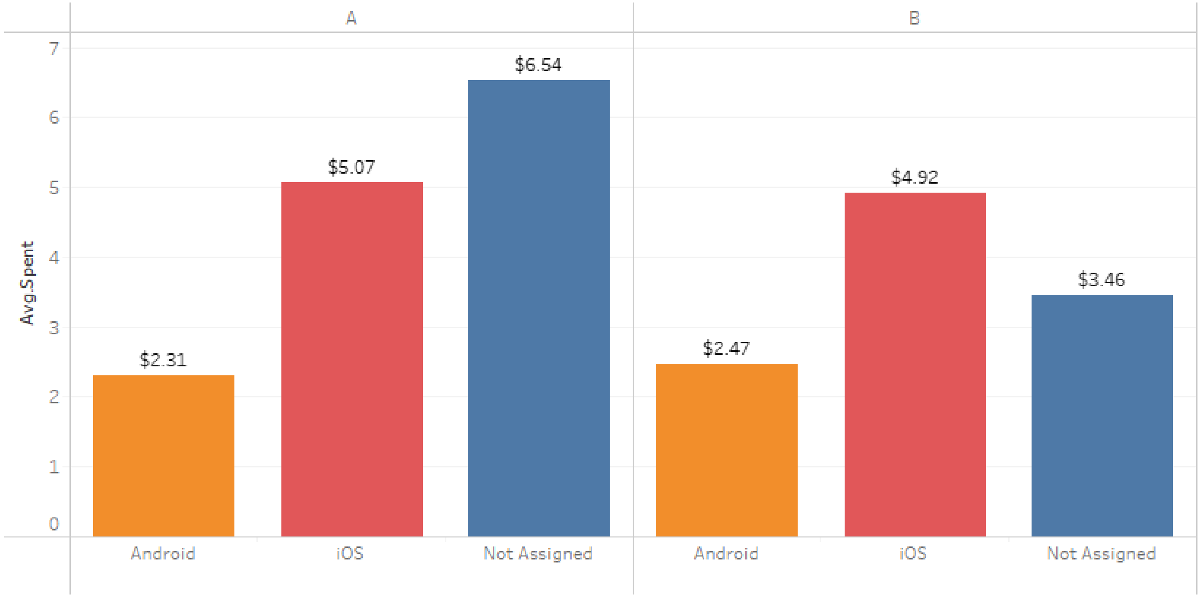


I conducted an analysis for the average spending per group on different devices, specifically Android, iOS and Not Assigned, were compared. The control Group and the Treatment group were examined to assess any differences in average spending patterns.

For the Control group, the average spending per device was as follows: $2.31 for Android, $5.07 for iOS, and $6.54 for Not Assigned. In the Treatment group, the average spending per device was $2.47 for Android, $4.92 for iOS, and $3.46 for Not Assigned.

Although the average spending appears somewhat similar between the two groups, it is noteworthy that the Treatment group achieved higher total spending due to a greater number of converted users. The total spending for the Control group was $82,145.90, whereas the Treatment group recorded a higher total of $83,415.33.

Please refer to the graphic below for a visualisation of the average spending per device in a group. Group A represents the Control Group, while Group B represents the Treatment group.

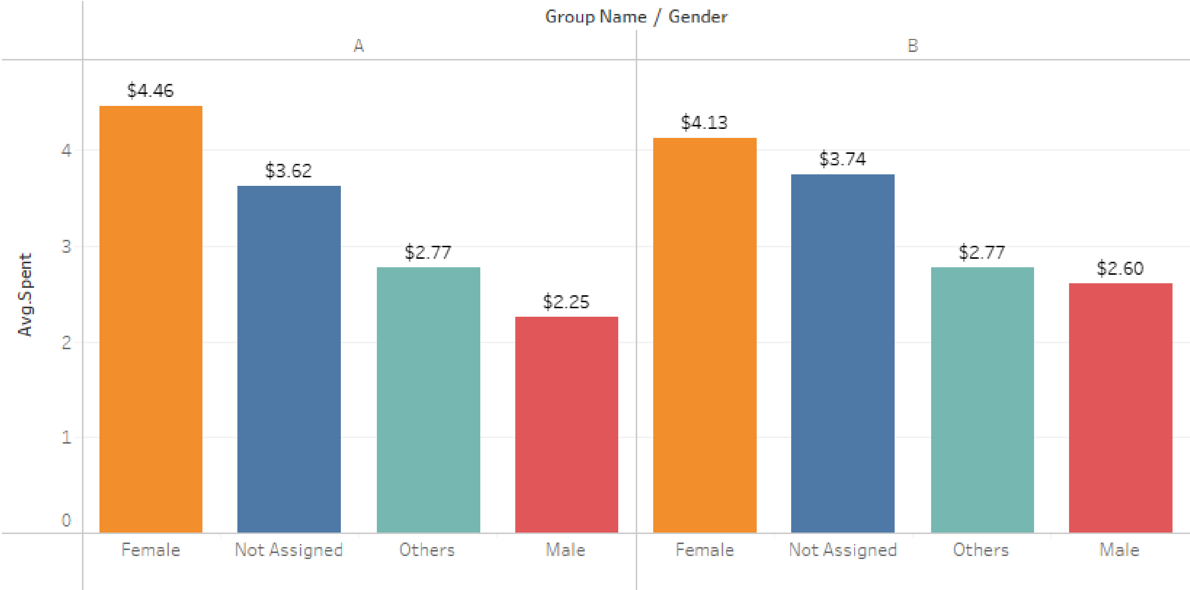


Furthermore analysis of the average spending per group based on different genders was made, specifically Female, Male, Others and Not Assigned, were compared. The control Group and the Treatment group were examined to assess any differences in average spending patterns.

For the Control group, the average spending per gender was as follows: $4.46 for Female, $2.25 for Male, $2.77 for Others, and $3.62 for Not Assigned. In the Treatment group, the average spending per gender was $4.13 for Female, $2.60 for Male, $2.77 for Others, and $3.74 for Not Assigned.

It is noteworthy to mention that the Treatment group presents better results for Male genders, where the converted number was 388 VS 264 in the Control group, resulting in a significantly higher average spend. Meanwhile it is the opposite for the Female genders, where it has a higher converted number 547 VS 518, yet it still displays a lower average spend.

Please refer to the graphic below for a visualisation of the average spending per gender in a group. Group A represents the Control Group, while Group B represents the Treatment group. F represents Female, M Male, O Others.



I also analysed the average spending per group based on different countries, specifically the United States of America, Canada, Great Britain, Brazil, Mexico, Turkey, Germany, Spain, France, Australia and Not Assigned, were compared. The control Group and the Treatment group were examined to assess any differences in average spending patterns.

For the Control group, the average spending per country was as follows:

- United States of America(USA): $4.30.
- Canada(CAN): $3.60.
- Great Britain(GBR): $2.11.
- Brazil(BRA): $3.21.
- Mexico(MEX): $2.81.
- Turkey(TUR): $3.69.
- Germany(DEU): $3.40.
- Spain(ESP): $2.18.
- France(FRA): $2.68.
- Australia(AUS): $1.67.
- Not Assigned(N/A): $3.26.

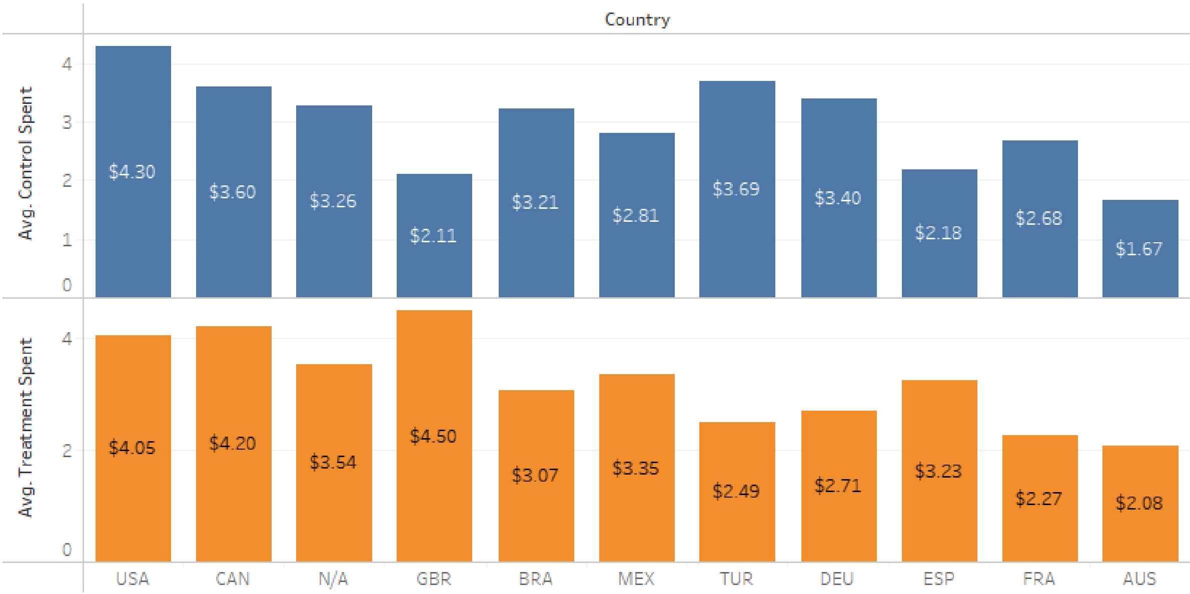
For the Treatment group the average spending per country was as follows:

- United States of America(USA): $4.05.
- Canada(CAN): $4.20.
- Great Britain(GBR): $4.50.
- Brazil(BRA): $3.07.
- Mexico(MEX): $3.35.
- Turkey(TUR): $2.49.
- Germany(DEU): $2.71.
- Spain(ESP): $3.23.

- France(FRA): $2.27.
- Australia(AUS): $2.08.
- Not Assigned(N/A): $3.54.

It is noteworthy to mention that the average spending in the Treatment group was higher than that of the Control group in five countries, including, Canada, Great Britain, Mexico, Spain and Australia.
On the other hand, the Control group had higher average spending in the other five countries such as the United States of America, Brazil, Turkey, Germany and France.

Please refer to the graphic below for a visualisation of the average spending per country in a group. Blue represents the Control Group, while Orange represents the Treatment group.



Novelty effect:

The analysis of the daily purchase data during the novelty effect revealed significant insights, including the presence of the Simpson's Paradox and its relation to country-specific average spending.

Note: The Simpson's Paradox is a phenomenon where the overall trend of a data set contradicts the trends observed within subgroups of that data set which we will follow in order to explain the novelty effect.

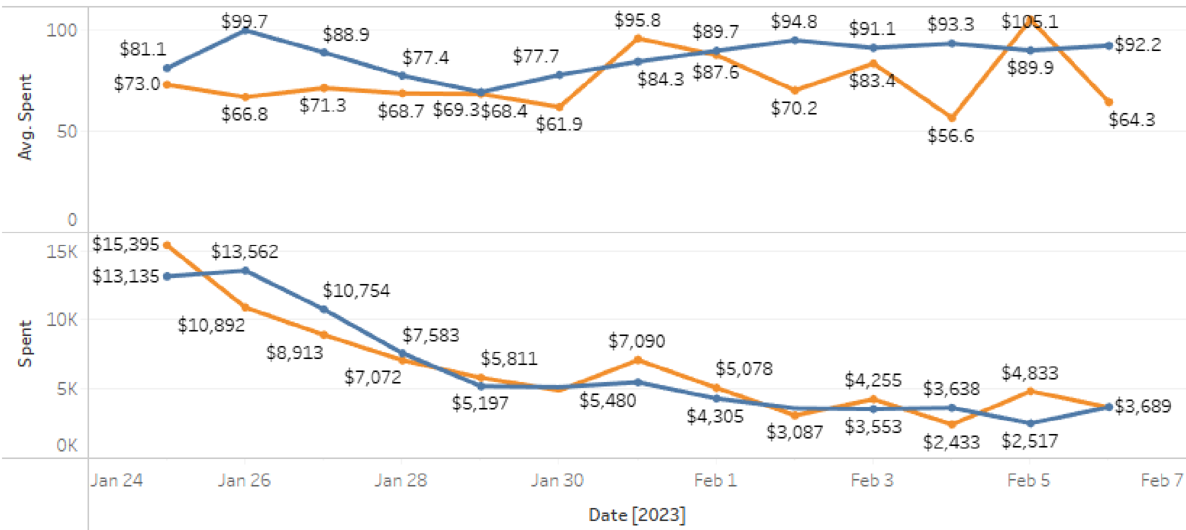
Upon examining the average spending data for the Control group and the Treatment group, a peculiar phenomenon known as Simpson's Paradox emerged.

Control group, on its own, exhibited a stable average spent over the 13-day period, with an average of $86.94 per day.
On the contrary, the Treatment group did not show a stable average spent over the same period, constantly fluctuating between $106.1 and $56.6 with an average of $74.85 per day.
However, when the total spent for the entire period was considered, the Treatment group had a higher total spend of $83,417 compared to the Control group of $82,147.
This paradoxical observation indicates that the overall trend differs from the individual group trends, warranting further investigation, which we then did such using the Country information we have and its average spending.

Country-Specific Average Spending:

The average spending per country in the Control and Treatment groups provided insights into how the novelty effect varied across different countries.
Notably, the average spending in the Treatment group was higher than that of the Control group in five countries, including, Canada, Great Britain, Mexico, Spain and Australia.
On the other hand, the Control group had higher average spending in the opposite five countries such as the United States of America, Brazil, Turkey, Germany and France.
These country-specific variations in average spending suggest that the impact of the novelty factor may be influenced by cultural or regional factors.
The presence of Simpson's Paradox highlights the complexity of analysing the novelty effect on purchasing behaviour. Although the Control group showed consistent average spending, the Treatment group exhibited fluctuating average spending over the given period. As a result, further investigation is required to understand the overall trend and the influence of the novelty factor.

Please refer to the graphic below for a visualisation of the average spending and total spent per day in each group. Blue represents the Control Group, while Orange represents the Treatment group.



Recommendations

Implement the new homepage selectively: Given that the new homepage (group B) did not show a significant increase in average spending compared to the Control group (group A), it may be advisable to avoid a full-scale implementation. Instead, consider implementing the new homepage selectively in countries or regions where it has shown promising results. This targeted approach will allow for a more efficient allocation of resources and minimise potential risks.

Monitor and analyse country-specific data: Utilise the country-specific data to identify regions or markets where the new homepage had a positive impact on conversion rates or average spending. Continuously monitor and analyse the performance of the new homepage in these countries, gathering additional insights to inform future decision making.

Evaluate other factors: While the new homepage did not result in a significant increase in average spending overall, consider exploring other factors that may contribute to improved conversions and revenue generation. This could involve conducting further research, A/B testing, or analysing user behaviour to identify additional elements that can enhance the user experience and drive more conversions.

Continual optimization: Use the findings from the A/B test as a starting point for continual optimization of the user experience and conversion strategies. Experiment with different homepage layouts, features, or promotional activities to further enhance user engagement and drive higher conversion rates.

Consider customer feedback: Seek feedback from customers to gain insights into their preferences, pain points, and suggestions for improving the purchasing experience. Incorporate this feedback into future iterations of the homepage or other strategies to align with customer needs and expectations.

Regularly assess and review results: Continually evaluate the performance and impact of the new homepage or any subsequent changes. Regularly review the metrics related to conversion rates, average spending, and overall revenue to assess the effectiveness of implemented strategies and make data-driven decisions for future optimizations.

By following these recommendations, you can make informed decisions based on the data and analysis conducted, improving the chances of success in driving conversions and maximising revenue generation.

Conclusion

Based on the analysis conducted, there was no significant increase in the average amount spent in the Treatment group (group B) compared to the Control group (group A), it suggests that the new homepage may not have a significant impact on overall spending.

Given this insight, an alternative approach could be to consider implementing the new homepage (group B) selectively in specific countries where it has shown success. By leveraging the country-specific data, you can identify regions or markets where the new homepage had a positive influence on conversion rates or average spending.

This targeted approach allows for a more tailored implementation, focusing resources and efforts on regions where the new homepage is likely to yield better results. By applying the new homepage in specific countries, you can gather more data and insights to further evaluate its effectiveness and make informed decisions on a larger scale in the future.

This strategy provides an opportunity to maximise the benefits of the new homepage in regions where it has shown promise, while minimising potential risks associated with a broader implementation without clear positive outcomes.

It is important to continuously monitor and assess the impact of the new homepage in these selected countries, adjusting the strategy as necessary based on ongoing analysis and user feedback.

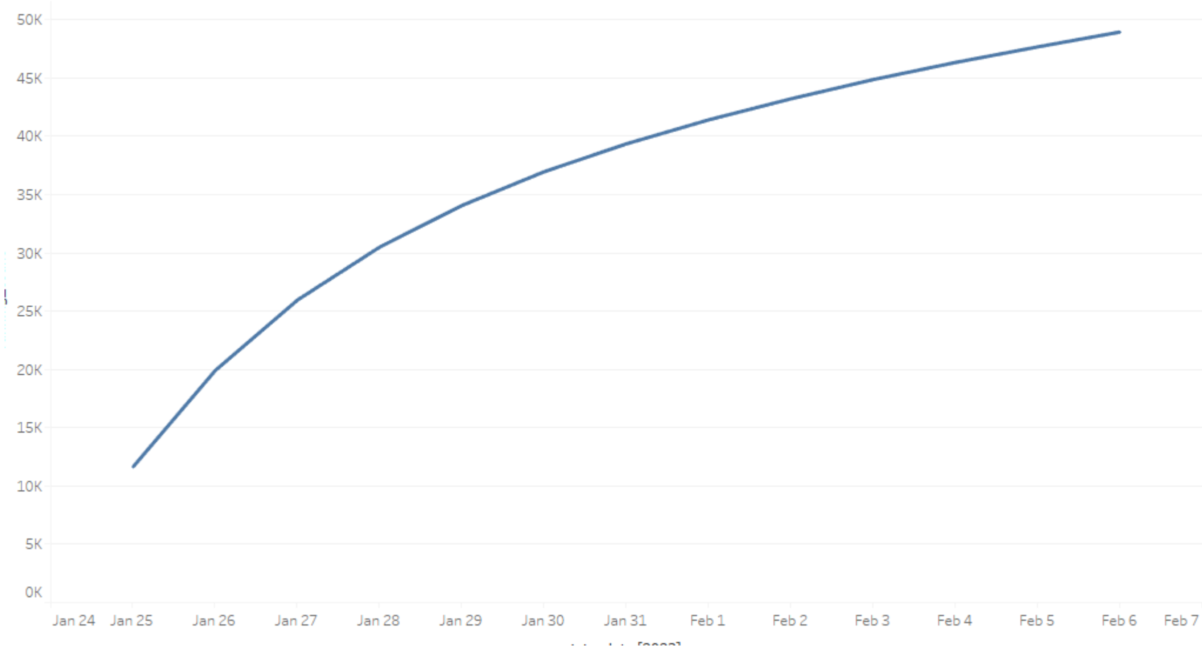
Limitations and Assumptions

Assumptions of Conversion Rate Changes: The MDE analysis assumes that changes in conversion rates are of interest and can be reliably measured. It is essential to recognize that this analysis is based on the assumption that conversion rates reflect the desired outcome and provide meaningful insights for decision-making. Other factors that may influence user behaviour or conversion rates, such as external market conditions or competitor activities, should be considered but were not explicitly addressed in this analysis.

Duration of the Experiment: One limitation to consider is that the experiment was conducted over a 13-day period. This relatively short timeframe may not capture the full range of dynamics and variations in user behaviour. In particular, it is important to note that the joint curve, representing the stabilisation of user behaviour, was still in the early stages during this period. As a result, the observed effects and patterns may not fully reflect the long-term trends or the complete impact of the novelty factor.

Please refer to the graphic below for a visualisation of the joint curve.

Lack of Specific Criteria: The limitations of this analysis are compounded by the absence



of specific criteria provided by the company for determining the appropriate MDE for conversions. The lack of explicit guidance regarding the desired level of sensitivity to detect changes in conversion rates makes it challenging to establish a definitive MDE value and may introduce uncertainties in the interpretation of the results.

It is important to consider these limitations and assumptions while interpreting the results of the MDE analysis for conversions. They should be taken into account when making informed decisions and developing strategies based on the analysis outcomes.

References

- https://public.tableau.com/app/profile/hugo.gomes/viz/ABTestProject/ConversionAVGSpent
- https://www.statsig.com/calculator
- https://statulator.com/SampleSize/ss2M.html#

Appendix

SQL query to obtain the data:

```
SELECT
  COALESCE(u.id, 0) AS user_id,
  COALESCE(u.country, '') AS country,
  COALESCE(u.gender, '') AS gender,
  COALESCE(g.device, '') as device,
  COALESCE(g.group, '') AS group_name,
  COALESCE(a.total_spent, 0) AS total_spent,
  COALESCE(g.join_dt::text, 'N/A') AS join_date,
  COALESCE(a.purchase_date::text, 'N/A') AS purchase_date
FROM users u
LEFT JOIN (
  SELECT uid, SUM(spent) AS total_spent, MAX(dt) AS purchase_date
  FROM activity
  GROUP BY uid
) a ON a.uid = u.id
FULL OUTER JOIN groups g ON g.uid = u.id
ORDER BY 1 ASC;
```

Excel used formulas:

```
=COUNTIFS(G:G, "A",A:A, "<>")
=COUNTIFS(G:G, "A",H:H, "<>0")
=COUNTIFS(G:G, "A",A:A, "<>",F:F, "A")
=AVERAGEIFS(H:H,G:G, "A",H:H, "<>")
=IF(AND($G4="A", NOT(ISBLANK($H4))), $H4, "")
```

This was used for the Sample Size Conversions.

Difference Conversion Rates Two headed sample test:

Z-Statistic

$$Z = (\hat{p}_1 - \hat{p}_2) / \sqrt{(\hat{p} * (1 - \hat{p})) * (1/n_1 + 1/n_2)}$$

P-Value

```
2 * (1 - NORM.S.DIST(ABS(Z), TRUE))
```

Standard Error Control

$SE_1 = \sqrt{(\hat{p}_1 * (1 - \hat{p}_1) / n_1)}$.

Standard Error Treatment

$SE_2 = \sqrt{(\hat{p}_2 * (1 - \hat{p}_2) / n_2)}$.

Difference Standard Error

$SE\_diff = \sqrt{(SE_1^2 + SE_2^2)}$

Margin of Error

$ME = Z * SE\_diff$.

$Z = 1.96$

LowerBound

$CI\_lower = (\hat{p}_1 - \hat{p}_2) - ME$.

UpperBound

$CI\_upper = (\hat{p}_1 - \hat{p}_2) + ME$.

Difference Average Amount Two headed sample test:

To calculate the standard deviation I used the formula STDEVP to ignore empty values from the columns Control_Spent and Treatment_Spent.

STDEVP(Control) and STDEVP(Treatment)

Pooled Standard Deviation

$SQRT(((s_1^2) / n_1) + ((s_2^2) / n_2))$

T-Statistic

$(\bar{x}_1 - \bar{x}_2)$ / pooled_stdev

Degrees of Freedom

df = min(n1 - 1, n2 -1)

P-Value

2 * (1 - T.DIST(ABS(t), df, TRUE))

Difference Average Standard Error

SQRT((s_1^2 / n_1) + (s_2^2 / n_2))

Critical Value

T.INV(1 - α/2, df)

MarginOfError

t_critical * SE

Lower Bound

CI_lower = $(\hat{p}_1 - \hat{p}_2)$ - ME.

Upper Bound

CI_upper = $(\hat{p}_1 - \hat{p}_2)$ + ME.