In this project, I analysed the A/B test results of a fictional company GloBox, and provided a recommendation to stakeholders about whether they should launch the experience. This project was done in conjunction with the Data Analytics program at Masterschool.

Company Profile

GloBox is an online marketplace that specialises in sourcing unique and high-quality products from around the world.

> We at GloBox believe that shopping should be an adventure, and we want to bring the world to your doorstep. From exotic spices and rare teas to handmade jewelry and textiles, we have a curated selection of products that you won't find anywhere else.

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, their food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue.

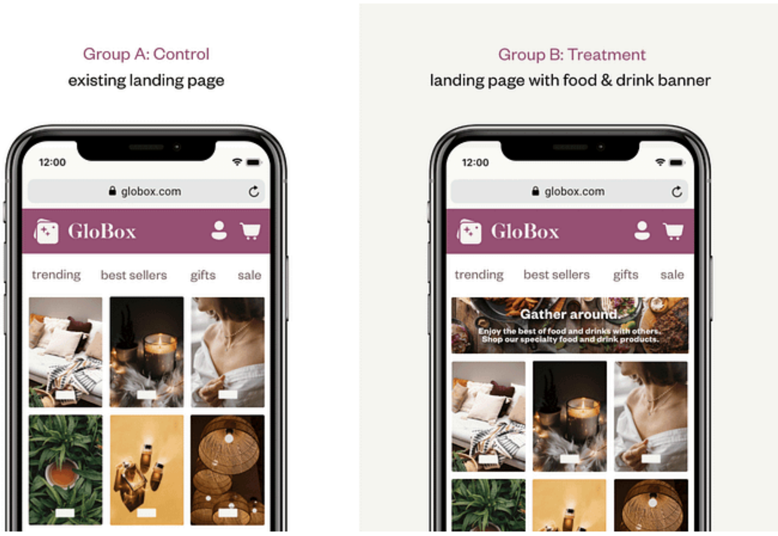
To this end, an A/B test that highlights key products in the food and drink category as a banner was run at the top of the mobile website.

What is an A/B test?

> An A/B test is an experimentation technique used by businesses to compare two versions of a webpage, advertisement, or product feature to determine which one performs better. By randomly assigning customers or users to either the A or B version, the business can determine which version is more effective at achieving one or many objectives.

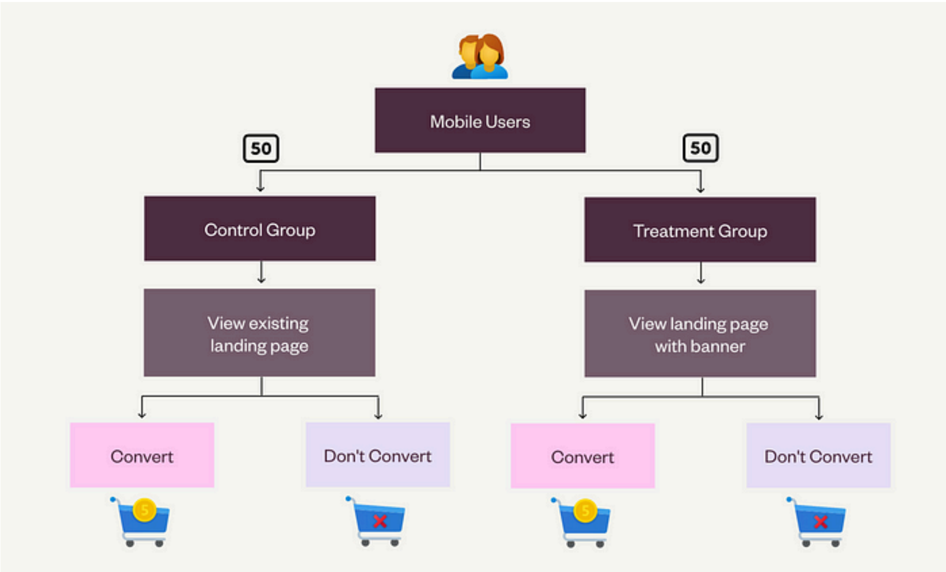
A/B Test Setup

The Growth team decided to run an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the test group sees it as shown below:



The setup of the A/B test is as follows:

1. The experiment is only being run on the mobile website for 13 days from 25th January 2023 to 6th February 2023.
2. A user visits the GloBox main page and is randomly assigned to either the control or test group. This is the join date for the user.
3. The page loads the banner if the user is assigned to the test group, and does not load the banner if the user is assigned to the control group.
4. The user subsequently may or may not purchase products from the website. It could be on the same day they join the experiment, or days later. If they do make one or more purchases, this is considered a “conversion”.



Objectives

- ***Boost the visibility of the Food & Drink section***
- ***Increase revenue***

Key Metrics & User Attributes

The success of the test is determined by the following measures:

- *Conversion Rate / Conversion Lift (the increase in conversion rate comparing the two groups)*
- *Average Amount Spent*

48k users participated in the test split equally into treatment and control groups. The following user attributes were analysed:

- *Gender*
- *Device*
- *Country*

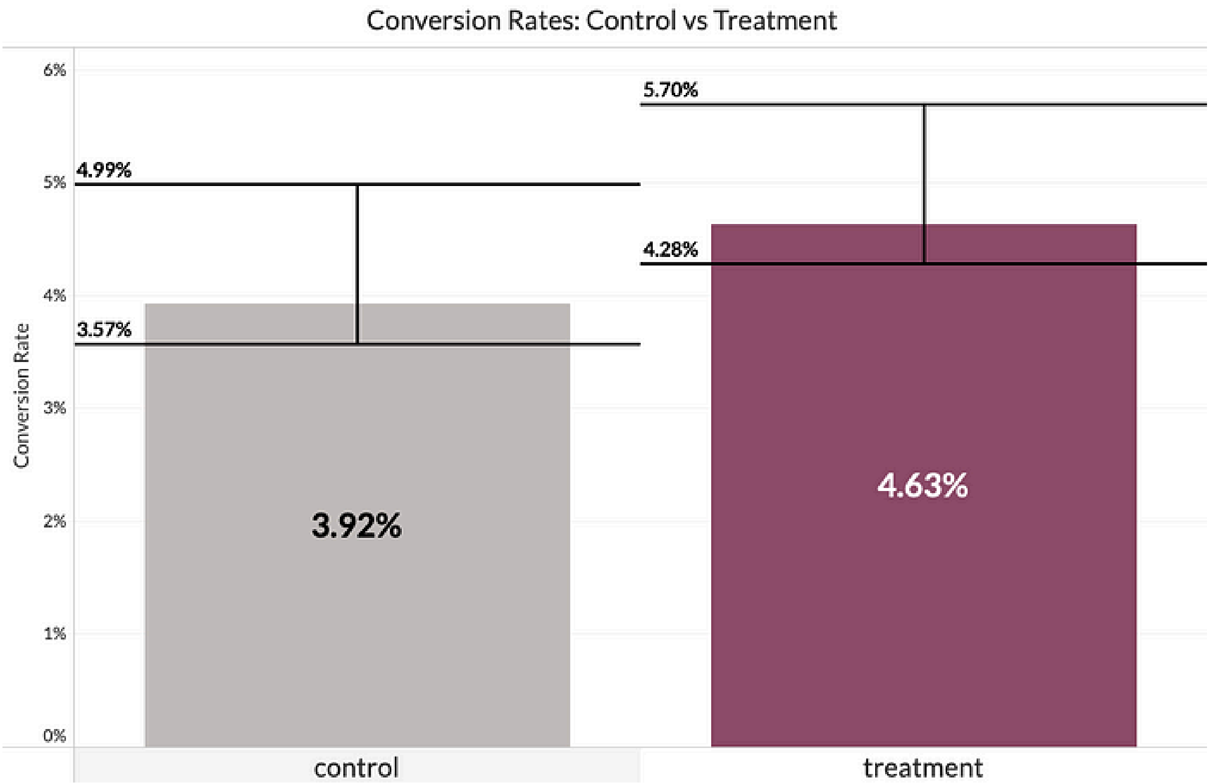
Key Insights

Key Metrics

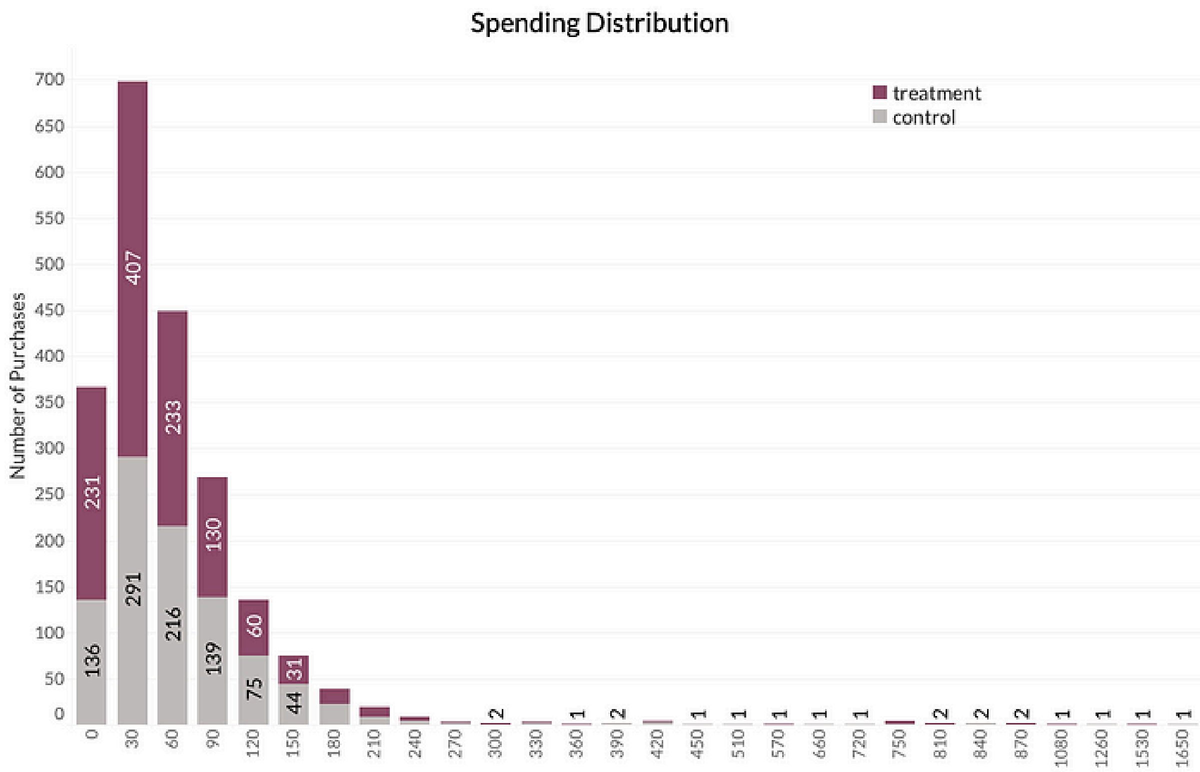
Hypothesis testing revealed a **statistically significant increase of 18% in the conversion rates of the treatment group** compared to the control group, but no such difference was found for the average amount spent per user.

The chart below shows the average conversion rate for the groups, along with a range above and below that average. This range is called an “error bar,” and it helps us understand the variability or uncertainty in our measurements.

- Control Group: The average conversion rate is 3.92%, with a range going from about 3.57% to approximately 4.99%.
- Treatment Group: The average conversion rate is 4.63%, with a range from around 4.28% to about 5.70%.



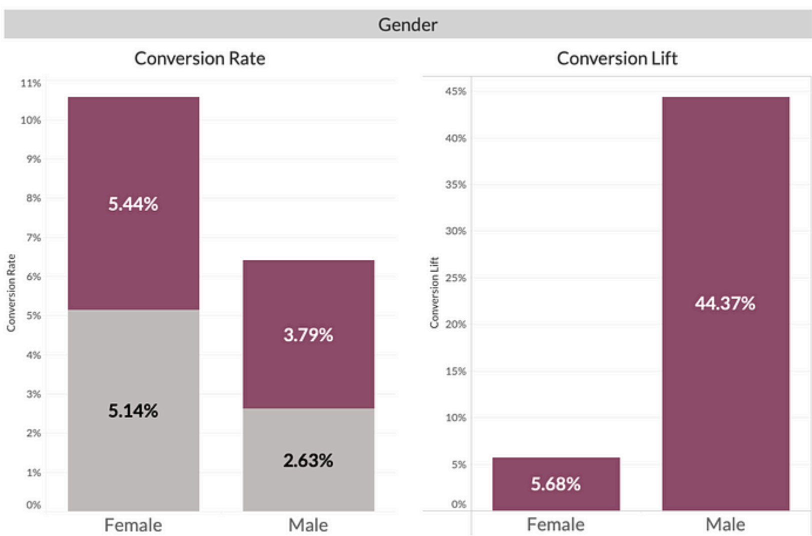
Most purchases were between 30–60 USD, and the majority of purchases were below 150 USD for both groups. While less common, there are a few purchases that reach up to 1650 USD. We need to understand these high-value transactions as well, even though they occur less frequently.



Gender

There was a much higher representation of male and female users in the test than other genders. Female users had the highest overall conversion rate of 5% in both groups. Female users had the highest overall average spend of 4 USD which decreased in the treatment group.

However, **male users in the treatment group converted 44% more than the control group and also spent 15% more.**

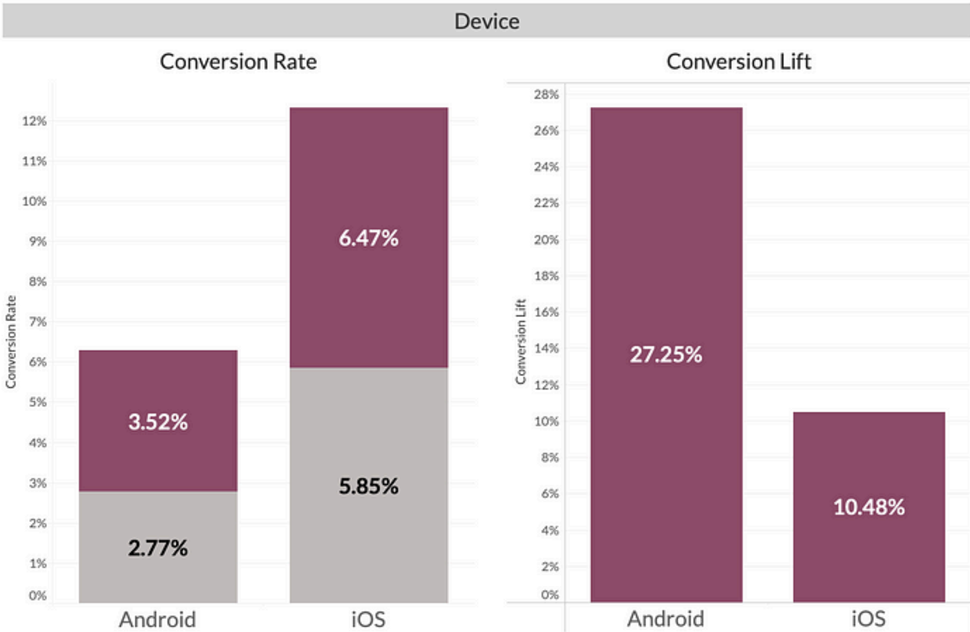


Device

There were more Android users than iOS. iOS users had the highest overall conversion rate of 6%. They also had the highest overall average spend of 5 USD, but this was lesser in the treatment group.

Similar to men, **android users who saw the banner converted 25% more than those who did not and also spent 7% more.**

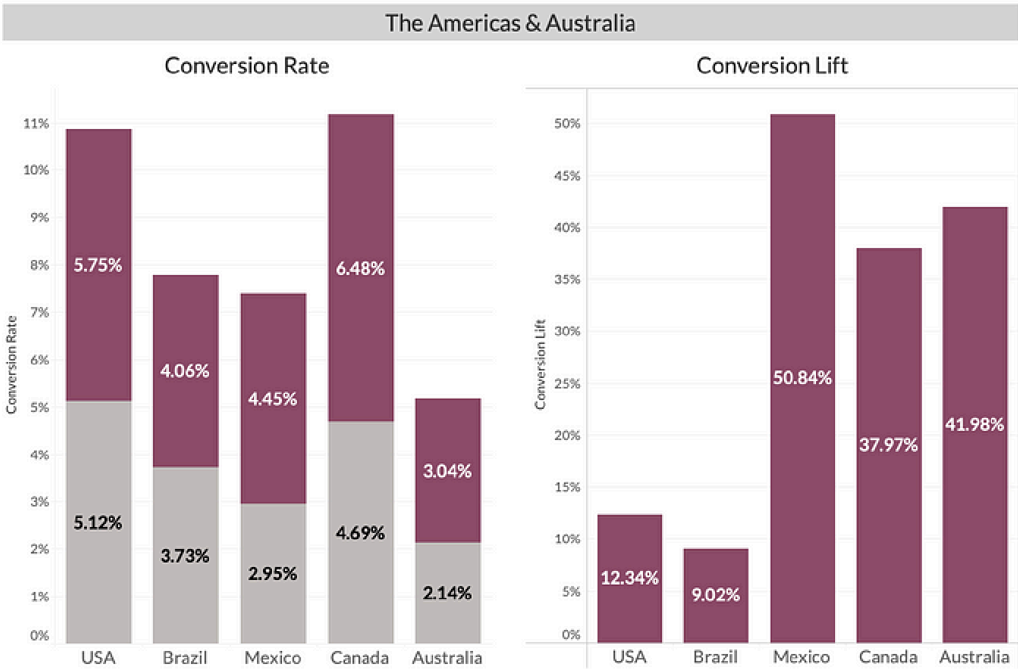
It is interesting that despite having a bigger sample size than the iOS segment, android users have seen a higher conversion rate and average spending. As only the iOS banner is available in this study, it would be a good idea to compare the two versions and see if there are differences in accessibility and user interface of the banners.



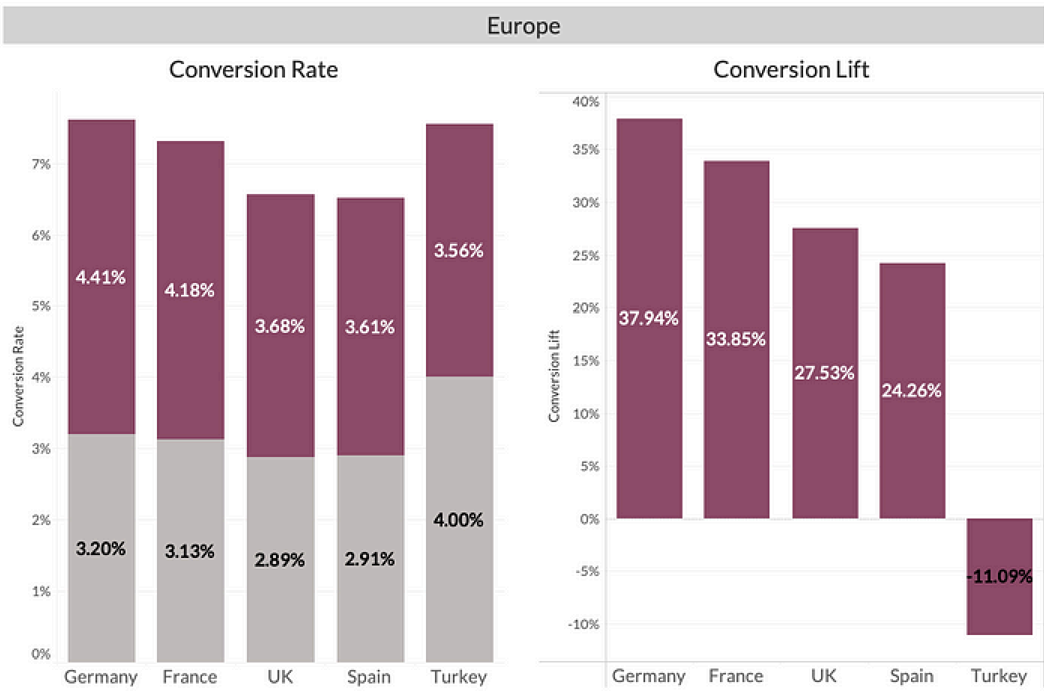
Country

Ten countries took part in the test across America, Australia and Europe.

Canada had the highest overall conversion rate of 6% but has the third smallest sample. In contrast, **Mexican users in the treatment group converted 51% more than the control group.** Mexico also has the third-largest number of users.



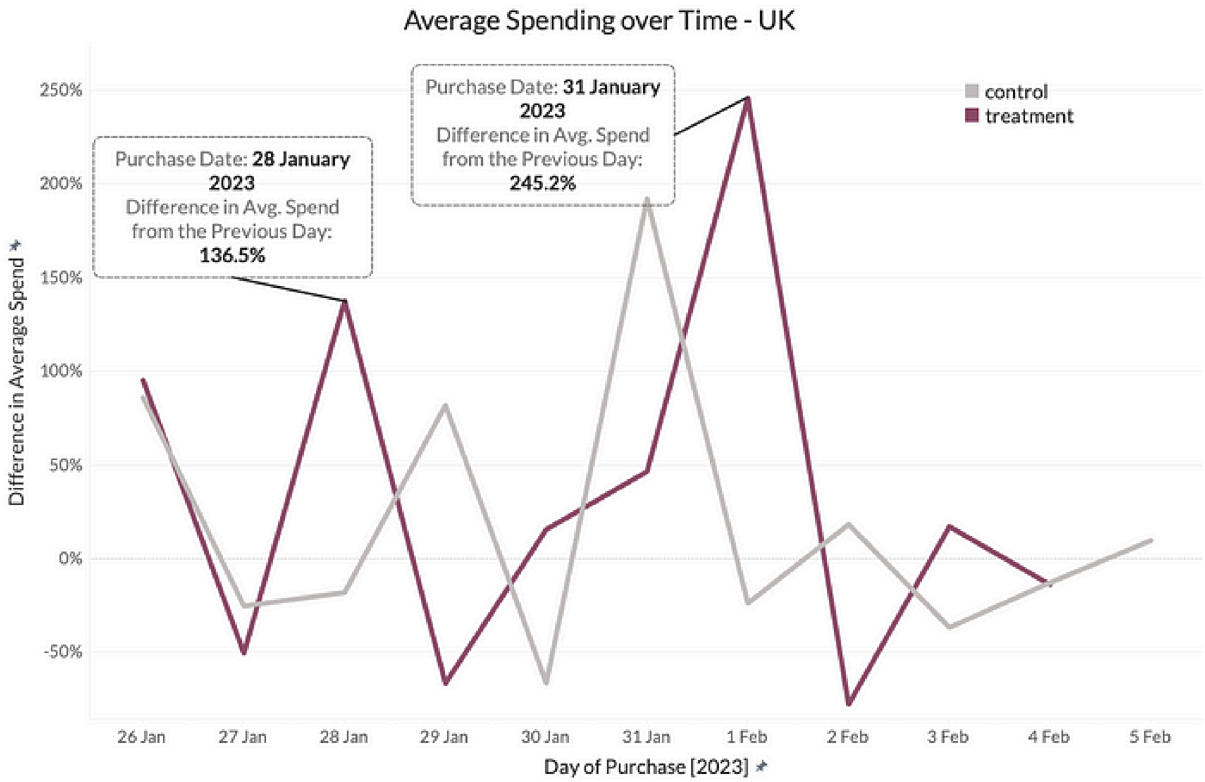
Turkey is the only country with both a decreased conversion rate and average spending in the treatment group. Germany, France and Brazil also saw a decrease in average spending.



Since Turkey is the only country that has a decreased conversion in the treatment group, it should be investigated. Maybe there are some technical problems or cultural context issues with the image shown.

Interestingly, the UK saw a 113% increase in spending among those who saw the banner. According to UK spending trends, this increase is due to two spikes in spending on the following days.

1. 28th January — Average spending increased by 136.5% compared to the previous day.
2. 1st February — Average spending increased by 245.2% compared to the previous day.



This increase in spending coincides with the timing of payday, which may be the reason for these spikes. This trend of increased spending between 28th January and 2nd February is also seen in the spending patterns of most countries.

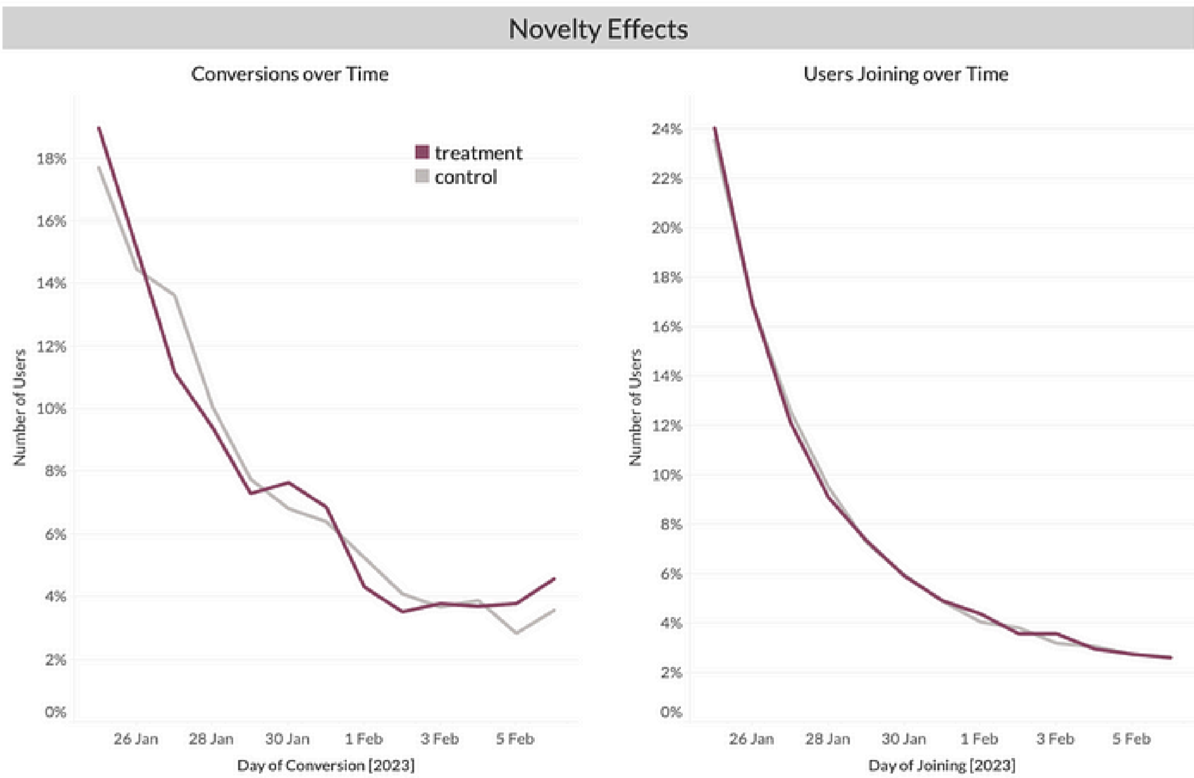
Novelty Effects

What are novelty effects?

> Novelty effects refer to initial, short-term trends that occur when users exhibit when they encounter something new and unfamiliar. These effects are typically characterised by increased attention, engagement, and interest in the novel element, but they may not necessarily reflect the long-term or sustained impact on user behaviour or preferences.

For example, when a new website banner is introduced, there may be an increase in click-through rates, longer time spent on the website, or higher levels of interaction with the banner, especially in the first few days. However, this heightened engagement could be temporary. Over time, users may become used to the banner, and the initial excitement can fade. These initial short-term trends are called novelty effects.

In this study, almost 50% of the conversions occurred in the first three days irrespective of whether a user saw the banner or not. However, around 50% of the users joined the test on the first three days as well. **A novelty effect is unlikely** because the number of conversions seems to be proportional to the number of users joining the test per day.



Power Analysis

What is power analysis?

Power analysis is a crucial process that ensures the reliability of our results. It is to test whether we have a sufficiently large sample size to accurately determine whether the differences we observe between two groups are meaningful or merely due to chance.

1. For the conversion rate results to have sufficient power, the test would require a sample size of 77k users split evenly into both groups. The result was calculated using this calculator.
2. For the average spending results to have sufficient power, the test would require a sample size of 51,720,108 users split evenly into both groups. The result was calculated using this calculator.

Based on these calculations, we see that there were not enough users for the results to have a meaningful impact.

Summary

1. When comparing the key metrics of the control and treatment groups:

- *There was a **significant increase in the conversion rates of users who saw the banner.***
- *There was **no significant difference in revenue**.*

2. Users from **Mexico, men and Android users converted the most**.

3. **No novelty effects** were detected.

Recommendation

The banner is easy to launch and maintain. However, it takes up high-value real estate on the main page. It is essential to have a high degree of confidence regarding the impact of the banner before launching it.

As only one of the objectives was met, **I recommend that we iterate the test** with the following considerations:

*1. A **sample size of at least 77K users** split equally for sufficient power.*

*2. A longer **duration of six weeks** to ensure that conversion and revenue increases are not due to the timing of payday.*

*3. Include **the category and time of purchase in data collection**.*

*4. **Equal number of Android and iOS users**.*

*5. **Equalised number of users joining every day** to check for novelty effects.*