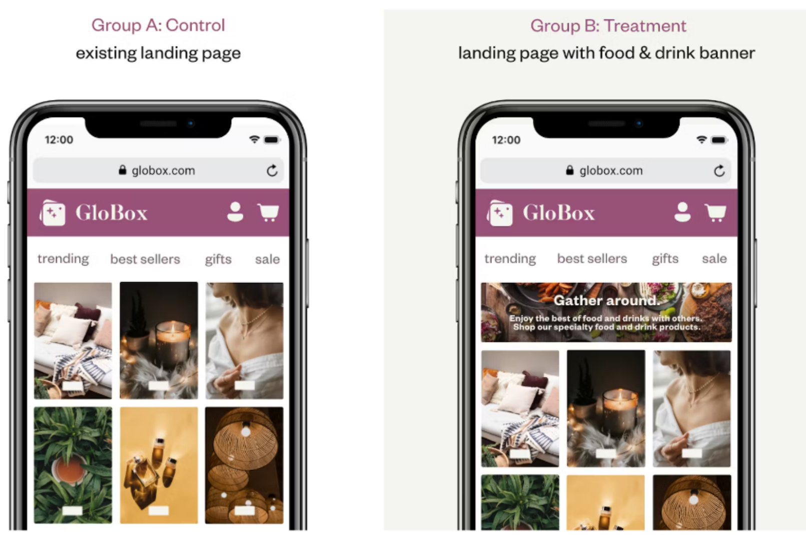
GloBox Banner A/B Test Report



Summary

Do not launch the new banner until further tests have been conducted.

We conducted an **A/B test** to see how our audience responded to the **new banner** leading them to the **Food & Beverages** product category.

In conclusion our results do suggest an **increase in user conversion** when interacting with the banner, however the **average spend** between the two groups **did not increase** enough to be statistically viable.

Context

GloBox offers an online marketplace which specialises in finding unique and high-quality products from around the world and allowing users to find it all in one place.

The Growth Team has seen the food and drink offerings have grown tremendously in the last few months. The company wanted to increase audience awareness of this category to increase the revenue in this area of the business.

A/B testing is a process of comparing two variables and comparing how the difference in the variable affects other metrics and determining which variable is more beneficial to our objective through statistical tests. We conducted an A/B test which displays the best performing products in the food and drink category as a banner at the top of the mobile site. The control group is not exposed to the updated website, and the test group sees it as on the first page of the report.

The A/B Test is set up so that it follows these 4 key points;

- The experiment is only ran on mobile websites
- When a user visits the website they are **randomly** assigned to either the control or test group. We classify this as the date that the users join.
- The website **only** loads the banner if the user is **assigned** to the **test group** and explicitly **does not** load the banner if the user is **assigned** to the **control group**.

- Afterwards the user may purchase or may not purchase products from the website. It could be either on the first day they joined the experiment or at a later date in which the experiment is still running. If they do make a purchase it will be considered a ‘conversion’.

The variable we were testing was how effective a new website header was at increasing conversion rate and the total average spend across the GloBox website.

We ran the A/B Test from 25/01/2023 till 06/02/2023, a 13-day period that includes two weekends, allowing insight into spending habits during both week and weekend. The timing coincides with the post-Christmas spending retraction typical in Western cultures, current rise in living costs post-pandemic, and economic impact of the ongoing conflict in Ukraine.

This could have led to **overall lower conversion and spending** during the experiment as users in **both control and treatment** groups are **affected by these financial pressures**.

For the A/B test, we conducted an experiment with a sample size of 48,943 users. The A/B test was aimed to compare two variations (A /B) of our mobile website to assess the impact on user behaviour. The sample was randomly split into two equal groups, with each group exposed to one of the variations of the website.

The large sample size of 48,943 users was chosen deliberately to achieve more precise estimates and increase the likelihood of detecting smaller effects, if they exist. By involving a substantial number of users, we can better understand potential trends within our sample group, which could provide insights into the behaviour of a larger population of users on our site.

With a larger sample size it allows us to capture a diverse range of user attributes, such as Country, Gender Identities, Device usage, Amount spent in each transaction, and purchase activity across multiple dates. This diversity in data ensures that our findings can be more robust and provide meaningful insights that may guide future improvements to the mobile website.

Hypotheses

- H_0: The implementation of the banner will have no effect on the conversion rate and average spend per user.
- H_1: The implementation of the banner will have an effect on the conversion rate and average spend per user.

The rationale behind this hypothesis is that a clear and prominent Food and Drinks banner can enhance the category's visibility and encourage engagement. By making it easier for users to find and click on the desired action, we anticipate an increase in the number of users making a transaction in that category, ultimately leading to higher conversion rates across the site.

Metrics and Data Collection

The key metrics we are using in our evaluation are Total User Conversion Rate and Total Revenue. The key metrics can be further aggregated for more in depth analysis of the user interactions such as Conversion Rate by Country and Average Spend Per User.

The data was collected from the GloBox mobile site and from users of the site between the dates of the experiment. We used SQL on the Database to filter out null values and perform some basic analysis. Once we had refined the data we took the output from our SQL query and downloaded it in a CSV format so that we can import it into Google Sheets for further analytics and into Tableau to visualize our findings.

We conducted additional analytics in Google Sheets which entailed running hypothesis tests and calculating confidence intervals around the resulting differences in conversion rate and average amount spent.

We had the following hypotheses before conducting our test;

- H_0: The implementation of the banner will have no effect on the conversion rate and average spend per user.
- H_1: The implementation of the banner will have an effect on the conversion rate and average spend per user.

The rationale behind this hypothesis is that a clear and prominent Food and Drinks banner can enhance the category's visibility and encourage engagement. By making it easier for users to find and click on the desired action, we anticipate an increase in the number of users making a transaction in that category, ultimately leading to higher conversion rates across the site.

We used a 95% confidence level for our tests and an Alpha of 5%. The 95% confidence level allows us a balance between having a precise estimate and having enough caution to not make any fake claims about the data. Have the Alpha set at 5%, we limit the risk of committing a Type 1 error/ a false positive in our conclusion.

For the results of the experiment see the results section of the report. For additional viewing of the calculations head to the Google Sheet attached in the appendix.

A few challenges we encountered during the data collection were Almost 7,000 users in the study had their gender status as ‘Unknown’. A question I would ask the business is why this is? Could this be because they signed up before it was compulsory to enter gender in their submission? Could this be an error in their database and Data Management?

Results

Conversion Rate Test Results

Alpha:	0.05
Count Of Group A Converted(X1):	955
Count Of Group B Converted(X2):	1139
Count Of Group A (N1):	24343
Count Of Group B (N2):	24600
Sample Proportion Group A (P1):	0.03923099043
Sample Proportion Group B (P2):	0.04630081301
Pooled Proportion(P):	0.04278446356
Test Statistic(Z):	-3.86429177
Confidence level	95%
Degrees of Freedom:	48941
P- Value:	0.0001114119853
Null Hypothesis Result	REJECT

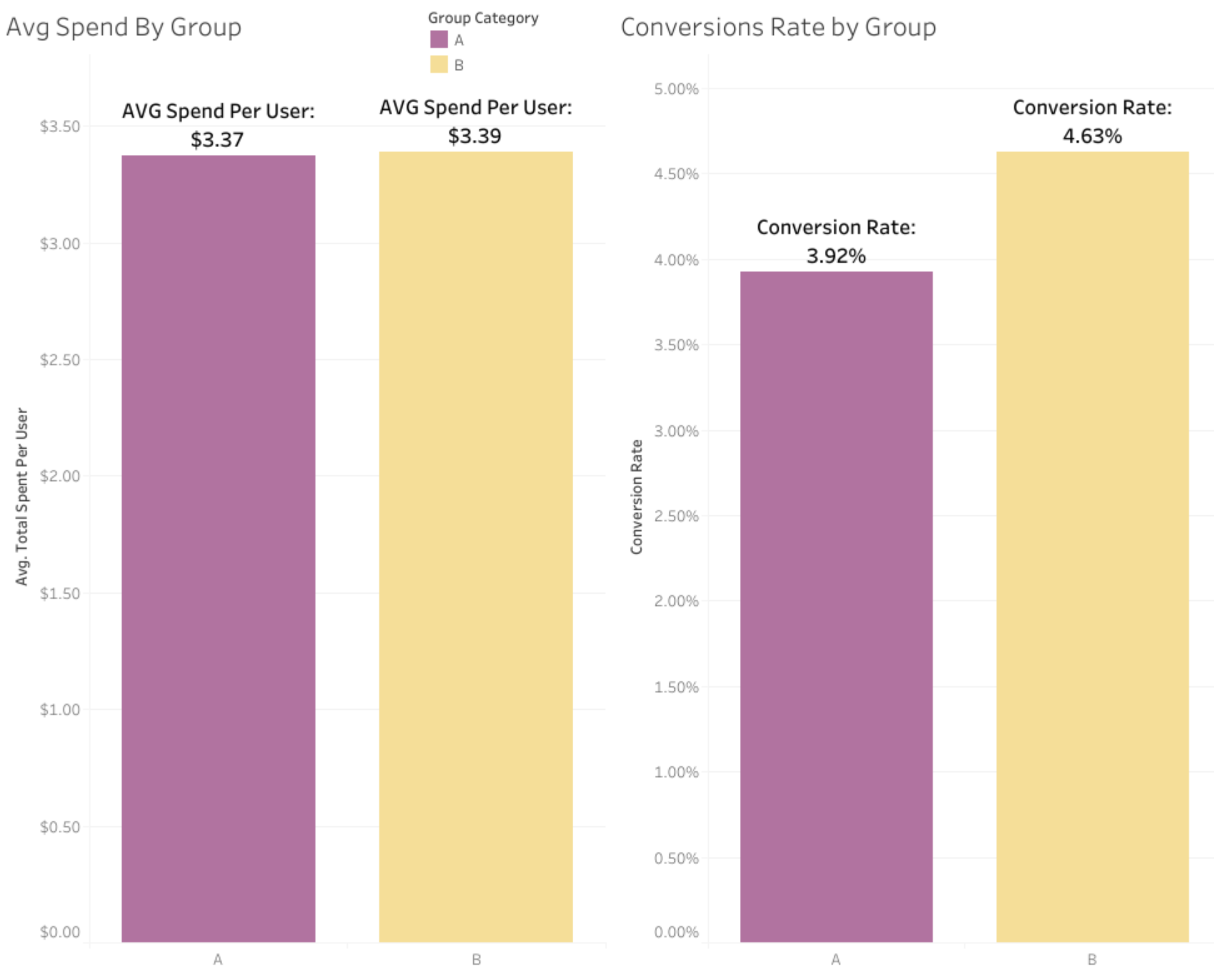
Average Spend Test Results

Average Spend Group A(X1):		3.374518468
Average Spend Group B(X2):		3.390866946
Group A Sample STEDV(S1):		25.93639056
Group B Sample STEDV(S2):		25.4141096
Count of Average Spend Group A(N1):		24343
Count of Average Spend Group B(N2):		24600
Pooled Standard Deviation (SP):		0.2321405588
T:		-0.07042491
Degree of Freedom:		48941
P-Value:		0.9438557529
Confidence Level:		95%
Null Hypothesis Response:		FAIL TO REJECT

These results are from our hypothesis test. The next section shows visualisations regarding the key metrics of **Average Spend** and **Conversion Rate**.

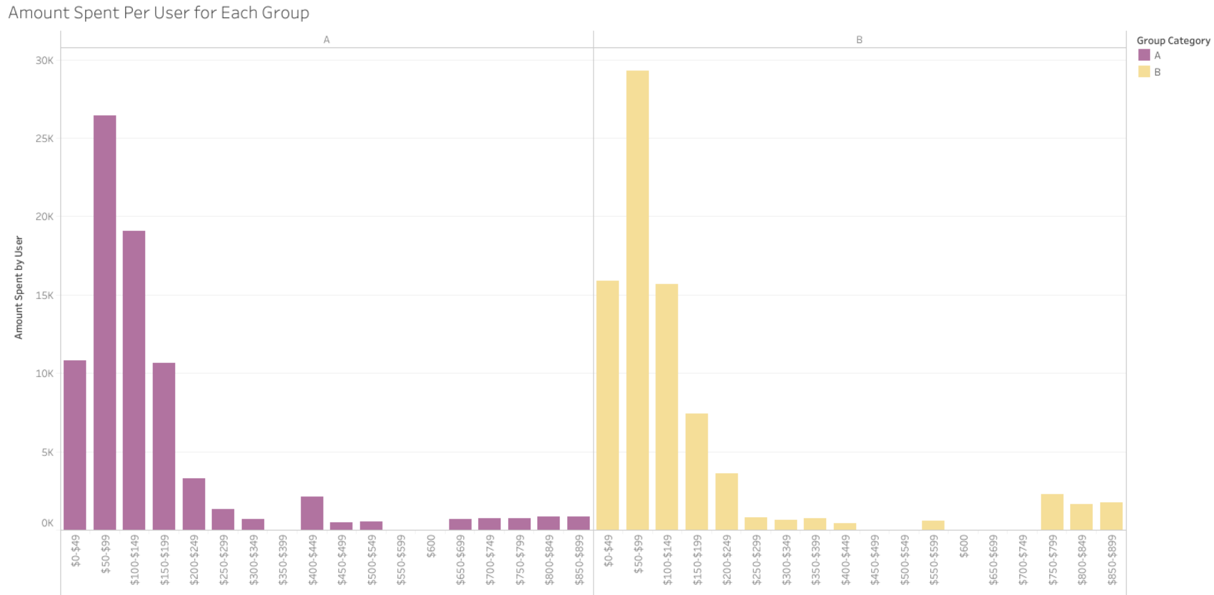
I visualised our findings using the data visualisation tool, Tableau;

Conversion Rate and Average Spend



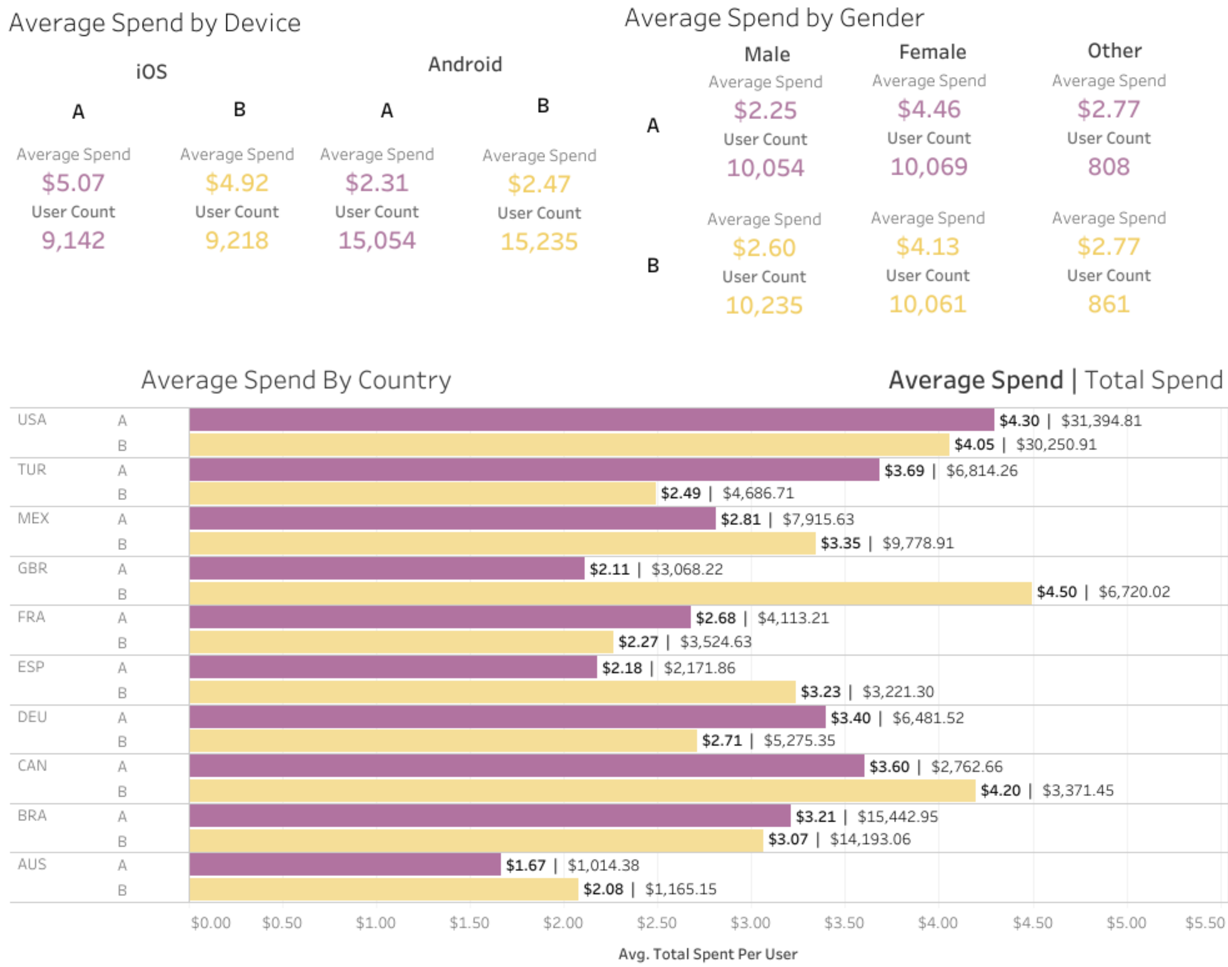
The charts illustrate the difference in conversion rate and average spending of the two groups . Group A refers to our control group and Group B refers to the test group. As we can see there is a **miniscule difference** in **Average Spend** between the two groups. This is reflected by how we **fail to reject the null hypothesis** for the Average Spend.

However we can clearly see that there is a **significant change in Conversion Rate** between the two groups. We can see a **positive growth in Conversion Rate** when the user group are **exposed to the new banner**.

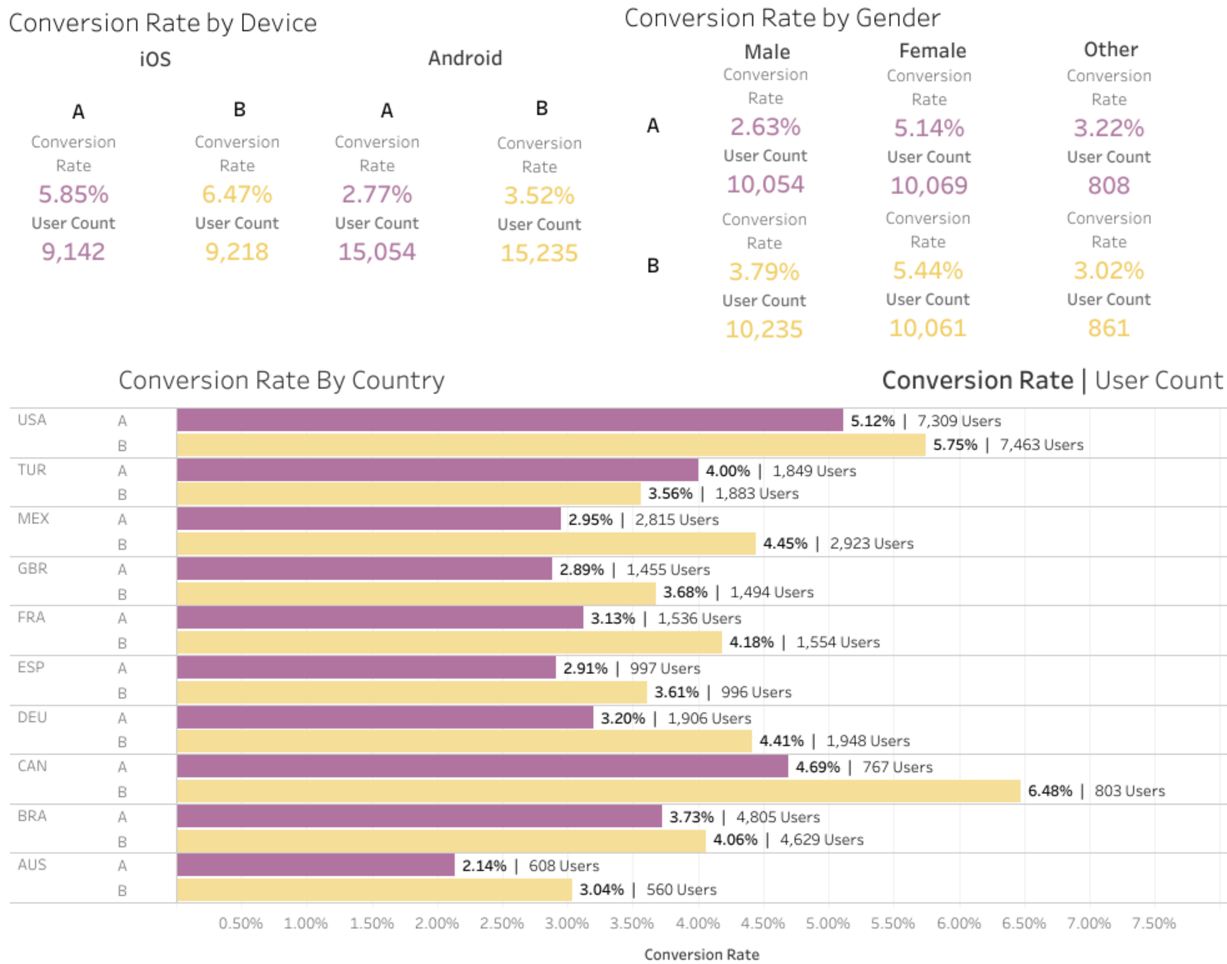


The above graph, divides the checkout totals into bins of $50, the distribution shows where the majority of the transactions took place in terms of price point. On Tableau when you go into the visual, the tool tip will display the User ID and how much they spent exactly. The basket shows the **most amount of activity** took place between **$50-$99.99**. This could infer more people added extras to their baskets from the **Food and Beverages** sections. However to see if this is the case we would **need more specific itemised data** so we could see which products/ categories were in their baskets.

Demographics



The dashboard above shows how the average spend differs between each metric. This offers insights such as in the **US, Turkey, France, Germany and Brazil** the **average spend** is actually **lower than the control groups**. When exposed to the new banner the **Average Spend drops** in the **Female** identifying population. This could become a problem as the **Female** demographic are the highest spending Gender group.



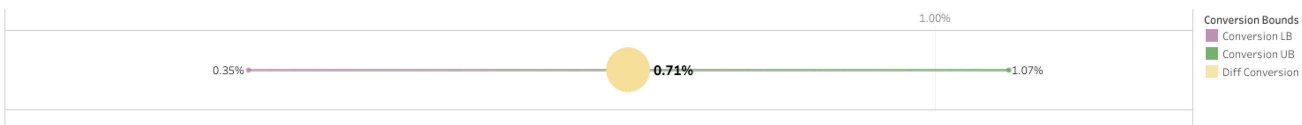
The dashboard above shows how the conversion rates vary between each metric. For instance we can see how **people who identify outside of the binary genders** have a **slightly lower conversion rate when exposed to the banner**.

Another point to note, in **Turkey not only did the average spend retract when exposed to the new banner but the conversion rate was reduced as well**. This suggests that the Turkish market did not respond well to the banner implementation and perhaps the **User Experience team will need to rework** it to suit this market more effectively.

Confidence Intervals

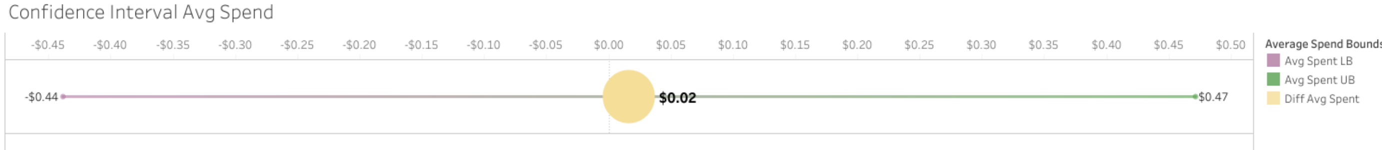
I visualised the confidence intervals for each metric, by adding a new data source in conjunction with calculated fields. I charted double axis for differences in metric (as a circle) and upper/lower bounds as a line

Confidence Interval Conversion



A 90% confidence interval means that if we take a large amount of samples then compute for each sample group an interval estimate, we would expect the true population parameter to appear in about 90% of the samples. However this does not mean it implies a 90% chance of our hypothesis being correct. However it provides a range of potential values for the parameter with an indication to the precision of the estimate.

Confidence Interval Avg Spend



When we conducted the tests we set out with a 95% confidence level, meaning we wanted to say with 95% confidence that the average would fall between these two bounds.

Recommendation

We've observed enough evidence that makes us confident there is a statistically significant positive effect on the conversion rate due to the introduction of the banner. The banner overall had a **positive impact** on the overall **conversion rate** although with a **18% improvement** when only implementing one banner. This shows potential for the addition of additional banners across the page to help increase revenue and that further A/B tests could be conducted for the additional banners.

I failed to reject the null hypothesis for our averages as the values were so close between both groups and there isn't enough statistical evidence to be able to reject the null.

Moreover, we could not reject the possibility that the change in conversion rate may be merely due to a novelty effect of the user noticing the change if they are a returning customer and having the novelty wear off after the first few exposures.

This indicates to:

a) No concrete evidence to support the new homepage raises the average spend per user overall.

b) Positive evidence that the increase in conversion rate is reflective of future results with the banner implementation.

Furthermore, there are opportunities to adapt the banner to be more specific to certain markets and demographics. In particular **female users** as they had a **higher conversion rate when exposed to the banner** and had the **largest total spend** out of all the Gender groups. However we would need to tailor where it directs them to help increase their average spend when on the site.

In conclusion I recommend holding off on launching the banner to all regions, especially Turkey, until further tests and changes have been done so that we can conduct more tests in order to see positive growth in Average Spend Per User and a higher Conversion Rates across all regions.

Appendix

-“Big Spend on the Weekend.” *https://www.jpmorganchase.com/institute/research/cities-local-communities/insight-big-spend-on-the-weekend*, 2016. Accessed 05 07 2023.

Excel Sheets:

https://docs.google.com/spreadsheets/d/1Tm6KDoJPuXHF7pu56Nalyv3P1ev3wuOA/edit?usp=sharing&ouid=115044881614916536862&rtpof=true&sd=true

Tableau Workbook:

https://public.tableau.com/app/profile/william.brooker/viz/ABTestingAnalysis_16984961351770/ABTestUserBehaviourAnalysis

Presentation:

Video
https://www.loom.com/share/9390b30815474a8ab76e95460d5ce21b?sid=bebd34e9-99b6-4575-8f8d-674b1287ead8

https://docs.google.com/presentation/d/1ox0W6zGiEpKc8KJywn68ZVYJ43BYVNfY/edit?usp=sharing&ouid=115044881614916536862&rtpof=true&sd=true

SQL Code:

1 -

```
SELECT uid, COUNT(uid)
FROM activity
GROUP BY uid
HAVING COUNT(uid) > 1;
```

2 -

```
SELECT *

FROM activity

INNER JOIN users

ON users.id = activity.uid;
```

3 -

By using the ISNULL function where can scan to see if any values in a column are empty/null. By doing the following I was able to realise that there are not Null data points in the spent column on the activity table.

```
SELECT *

FROM activity

WHERE spent ISNULL;


The Function to fill in NULL values would be;


COALESCE(column, fill_value)
```

4 -

```
SELECT dt

FROM activity

ORDER BY dt ASC;

SELECT dt

FROM activity

ORDER BY dt DESC;
```

5 -

```
SELECT DISTINCT COUNT(id)

FROM users;
```

6 -

```
SELECT "group", COUNT(*)

FROM groups

WHERE "group" = 'A'

GROUP BY "group";

SELECT "group", COUNT(*)

FROM groups

WHERE "group" = 'B'
```

```
GROUP BY "group";
```

7 -

```
SELECT

    COUNT(DISTINCT id) AS total_users,

    COUNT(DISTINCT activity.uid) AS total_converted,

    ROUND((COUNT(DISTINCT activity.uid)*1.0 / COUNT(DISTINCT id) * 1.0 *
100),2) || '%' AS conversion_rate

FROM users

LEFT JOIN activity ON users.id = activity.uid

LEFT JOIN groups on activity.uid = groups.uid;
```

8 -

```
WITH converted_users AS (

    SELECT g.group, COUNT(DISTINCT a.uid) AS activity_user_count

    FROM activity AS a

    JOIN groups AS g ON g.uid = a.uid

    GROUP BY g.group

),


all_users AS (

    SELECT "group", COUNT (DISTINCT uid) AS group_user_count

    FROM groups

    GROUP BY "group"

)


SELECT converted_users.group,

       converted_users.activity_user_count,

       all_users.group_user_count,
```

```
       (converted_users.activity_user_count * 1.0) /
(all_users.group_user_count * 1.0) * 100 AS conversion_ratio

FROM converted_users

JOIN all_users ON all_users.group = converted_users.group;
```

9 -

```
SELECT

  "group",

  ROUND(AVG(SUM(spent::numeric)) OVER(PARTITION BY "group") / COUNT(DISTINCT
g.uid),2) AS avg_spent

FROM groups g

FULL JOIN activity a USING(uid)

GROUP BY "group";
```

10 -

```
SELECT

 COALESCE(u.id, 0) AS user_id,

 COALESCE(a.uid, 0) AS activity_uid,

 COALESCE(g.uid, 0) AS group_uid,

 COALESCE(u.country, '') AS country,

 COALESCE(u.gender, '') AS gender,

 COALESCE(g.device, '') AS device,

 COALESCE(g.group, '') AS group_name,

 COALESCE(SUM(a.spent), 0) AS total_spent

FROM users u
```

```
FULL OUTER JOIN activity a ON a.uid = u.id

FULL OUTER JOIN groups g ON g.uid = u.id

GROUP BY 1,2,3,4,5,6,7

order by 1 ASC;
```