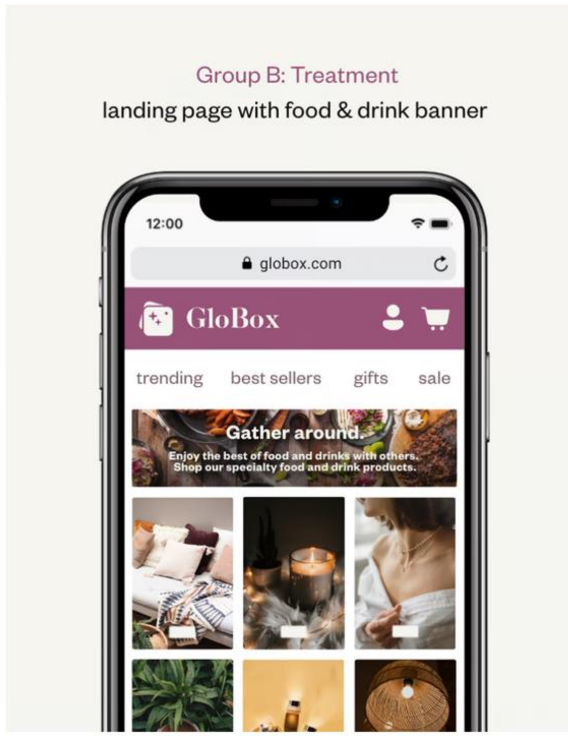
Globox A/B Test Analysis

By: Azat Beisheev



Summary

This report is presenting the result of A/B test to identify probability of success of new food and beverage banner in mobile Globox website. The test conducted by dividing into two groups “control = A” and “treatment = B”. Control group did not see the banner on website unlikely treatment group had seen the banner. The base metrics for a test were “conversion rate” and “average amount per user”. The test run between January 25 and February 6.

The analysis discovered statistically significant difference in conversion rate between treatment and control group. However, there was no significant difference in average amount spent per user.

Base on this, the recommendation is not to launch, conduct competitor analysis, divide food and beverage by categories and conduct more analysis with more samples.

Context

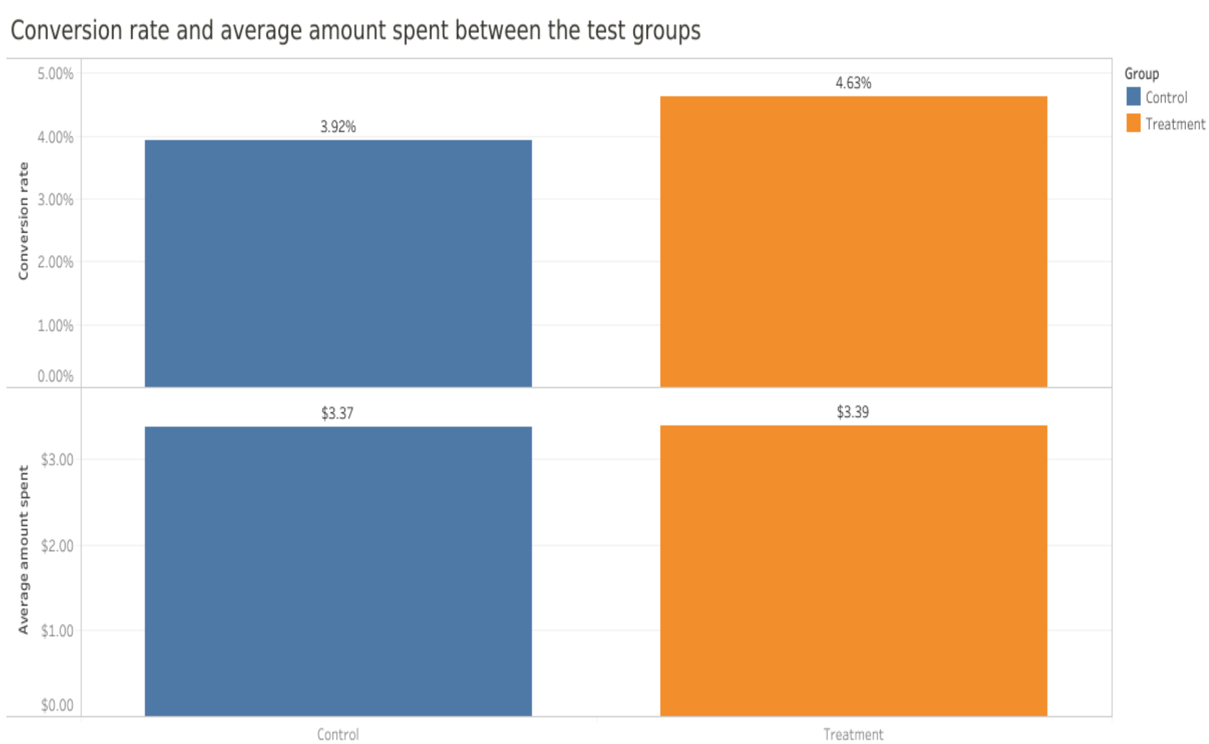
The test conducted only on mobile website from January 25 to February 6. Two group of users randomly assigned to Control 24,343 users and Treatment 24,600 users relatively. The Control group saw the website without banner and Treatment group saw the food and beverage banner. The main metrics of interest were Conversion rate and Average amount per user. The conversion rate's result was number of users conversion divided by all visited users. The average amount was the average of total spent. Both metrics were analyzed for each group separately. These metrics are very crucial because it will help to identify should we launch the new service or not.

The dataset totally includes 48,943 users who visited website during the A/B Test. The data included information about did user purchased or not, amount spent, group, gender, country and device.

During the data exploration it was found that columns Country, Gender and Device had many null values. Null values were replaced to 'N/A' using coalesce function. In addition, used "Case When" conditional statement on SQL to filter by assigning "True" to users who made purchase and "False" to users who did not. SQL query can be reviewed on SQL query page.

Results

The Control group's conversion rate was 3.92%, which means users who didn't see the food and beverage banner on mobile website while the Treatment group's conversion rate was 4.63%. The average amount spent for Control group was 3.37$ and for Treatment group was 3.39$.



Hypothesis test for Conversion rate

In other words, we are testing to see if is in there a statistically significant difference between Control and Treatment group. Google Spreadsheet and Excel was used to complete these calculations. Significance level for identification of the power of our evidence is 0.05.

H0: p1 = p2 (The conversion rates in the control and treatment groups are equal)

HA: p1 ≠ p2 (The conversion rates in the control and treatment groups are not equal)

Control group = A

Treatment group = B

Test: Two Sample Z test

Distribution: Normal Distribution

Sample Size: A = 24343, B = 24,600

Conversion rate: A = 0.03923099, B = 0.046300813

Pooled proportion: (0.03923099 * 24343 + 0.046300813 * 24600) / (24343+ 24600) = 0.042784464

Standard Error: √(0.042784464 * (1-0.042784464) * (1 / 24343 + 1 / 24600) = 0.001829526

Test Statistic: (0.03923099 - 0.046300813) / 0.001829526 = 3.86429177

P-Value: 2 * (1 – NORM.S.DIST(3.86429177, TRUE)) = 0.000111

P-Value 0.001829526 is less than significance level 0.05 and that's why we rejected Null Hypothesis. There's a statistically significant difference in conversion rate between two groups.

95% Confidence Interval for Conversion Rate

The 95% Confidence interval for the difference between two groups [0.0035, 0.01065].

Control group = A

Treatment group = B

Critical Value: NORM.S.INV(1-(aplha/2) = 1.959963985

Standard Error: √(0.042784464 * (1-0.042784464) * ((1/24343) + (1/24600)))

Margin of Error: 1.959963985 * 0.001829526 = 0.003585805

Lower Bound: (0.046300813 - 0.03923099) - 0.003585805=0.003484017

Upper Bound: (0.046300813 - 0.03923099) + 0.003585805=0.010655628

Point Estimate: 0.010655628 - 0.003484017 = 0.00717161

Hypothesis test average amount per user

Hypothesis test was conducted to identify a difference in average amount per user between two groups. Spreadsheet and Excel was used for this test.

H0: $\mu1 = \mu2$ (The average amount spent per user in the control and treatment groups is equal)

HA: $\mu1 \neq \mu2$ (The average amount spent per user in the control and treatment groups is not equal)

Control group = A

Treatment group = B

Test: Two sample Z test

Distribution: Normal distribution

Significance level: 0.05

Sample means (avg spent): X-bar A = 3.374518468, X-bar B = 3.390866946

Sample Standard Deviation: A = 25.93639056, B = 25.4141096

Sample size: A = 24343 , B = 24600

Degrees of Freedom: MIN((24343-1), (24600-1)) = 24342

Standard Error: SQRT((25.4141096)^2/ 24600+(25.93639056)^2/ 24343) = 0.232140559

Test Statistic: ((3.390866946- 3.374518468)-0)/ 0.232140559 = 0.07042491

P-Value: 0.943856044 is more than 0.05 significance level and that's why we fail to reject null hypothesis. There's a statistically insignificant difference in conversion rate between two groups

95% Confidential Interval in average amount between two group

The 95% Confidence interval for the difference between two groups [-0.4386, 0.4713].

Control group = A

Treatment group = B

Test: Two sample Z test

Distribution: Normal Distribution

Significance level: 0.05

Sample means: x- bar A = 3.374518468, x-bar B = 3.390866946

Sample Standard Deviation: A = 25.93639056, B = 25.4141096

Sample size: A = 24343 , B = 24600

Degrees of Freedom: MIN((24343-1), (24600-1))= 24342

Critical Value: TINV(0.05, 24342) = 1.960061445

Standard Error: SQRT((25.4141096)^2/ 24600+(25.93639056)^2/ 24343)= 0.232140559

Margin of Error: (1.960061445* 0.232140559) = 0.455009759

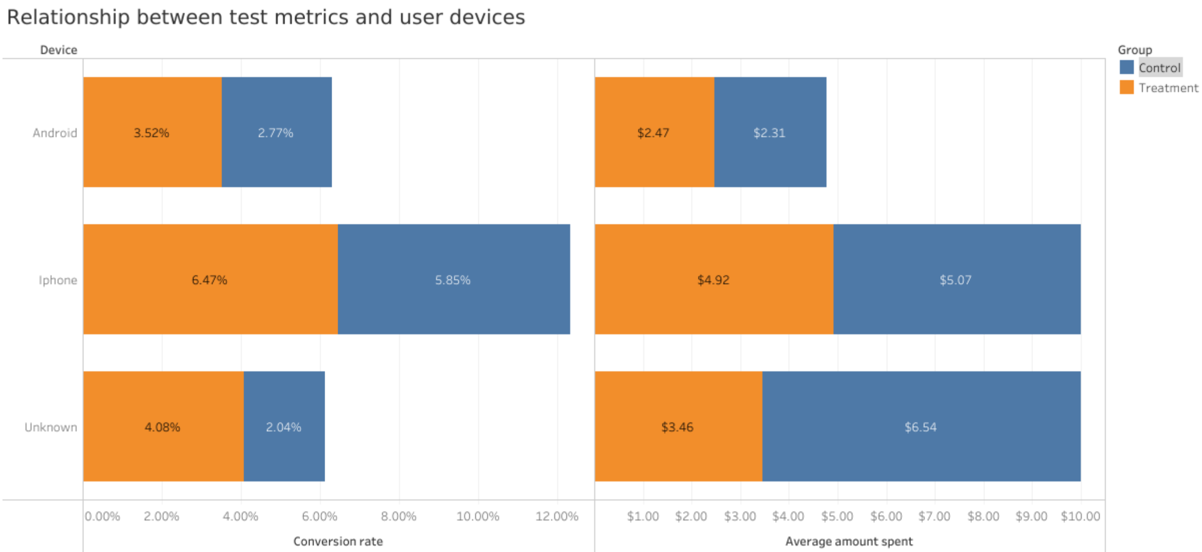
Lower Bound: (3.390866946- 3.374518468) - 0.455009759 = -0.438661281

Upper Bound: (3.390866946- 3.374518468) + 0.455009759 = 0.471358237

Point Estimate: 0.471358237 - 0.438661281 = 0.032696956

Exploratory Data Analysis

The relationship between test metrics and user's device.



Andoid

Conversion rate for Android users treatment group was 3.52% and for control group 2.77%. For average amount spent the treatment group 2.47$ was also a little bit higher than control group 2.41$.

IOS

The conversion rate for IOS device for treatment group is 6.47% and control group 5.85% however in average amount spent the treatment group 4.92$ is lower than control group 5.07$.

Unknown

For users with unknown device the conversion rate for treatment group is 4.08% and 2.04% for a control group. Here the conversion rate of treatment group is higher than control group.

The average amount spent in treatment group 3.46$ is lower than control group 6.54$.

The relationship between metrics and users gender

Relationship between test metrics and users gender

Females

Conversion rate for females in treatment group was 5.44% which is slightly higher than control group 5.14%. Average amount spent in treatment group was 4.13$ which is comparatively lower than control group 4.46$.

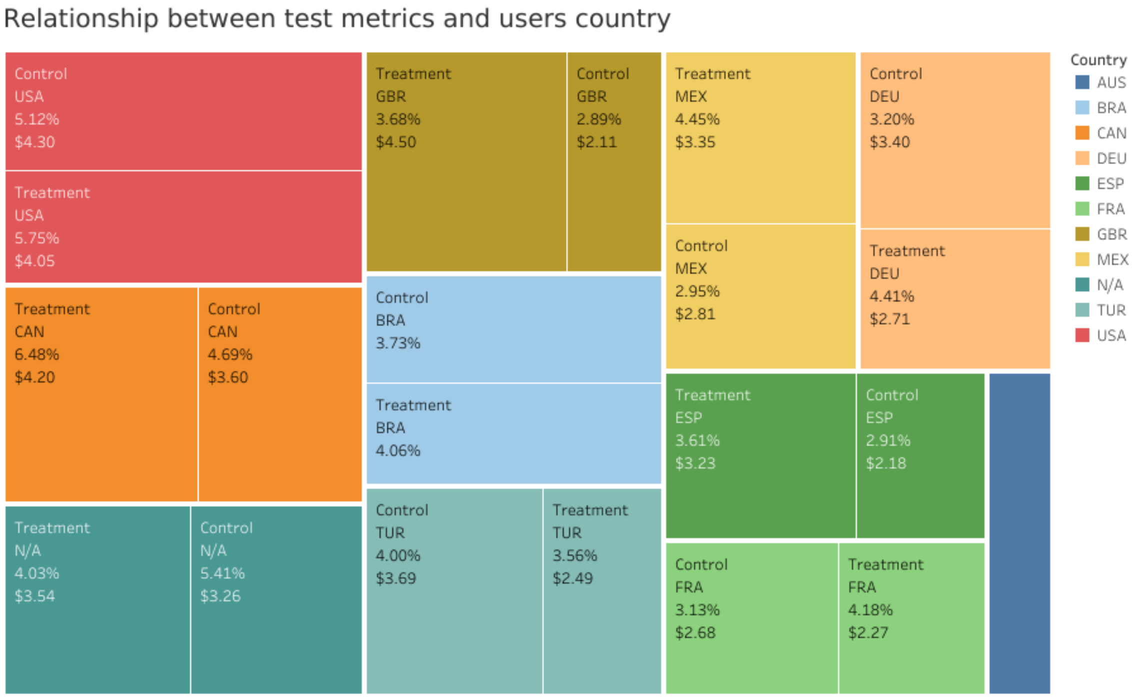
Males

Conversion rate for males in the treatment group was 3.79% which is higher than in the control group 2.63%. The average amount was also higher in treatment group 2.60$ than control group 2.25$.

Other genders

For other users, conversion rate in the treatment group was 3.02% which slightly lower than control group 3.22%. And average amount has no significant difference 2.77$ between two groups.

Relationship between test metrics and users country



USA

Conversion rate in control group was 5.12% and treatment group has higher rate 5.75%. For average amount in control group was 4.30$ and 4.05$ for treatment group.

England

Conversion rate in control group was 3.68% which is higher than treatment group 2.89%. However average amount has significant difference with 4.50$ in treatment group and 2.11$ in control group.

Turkey

Conversion rate in control group was 4.00% and treatment group has slightly lower 3.56% rate. For average amount spent in control group is 3.69$ and treatment group has lower 2.49$ spent average.

Canada

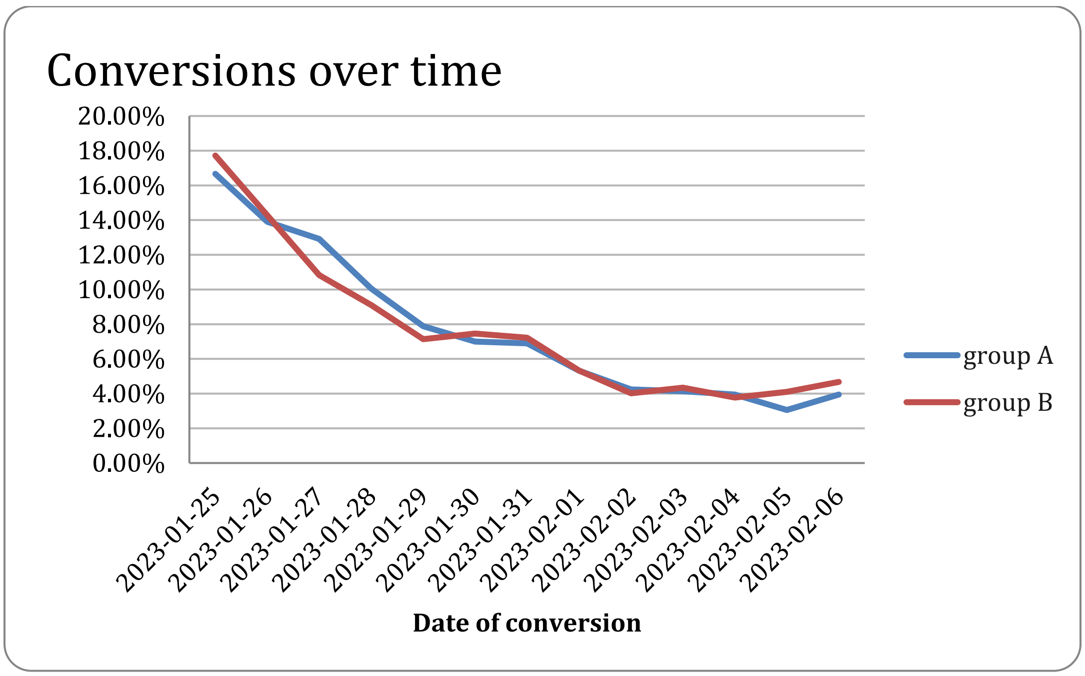
Canada has highest conversion rate among other countries for a treatment group 6.48% higher than control group.

‘N/A’

There are unknown countries with conversion rate in control group 5.41% which is much higher than 4.03% treatment group rate. In average amount there’s a slight different with 3.54$ in treatment group and 3.26$ in control group.

Advanced analysis

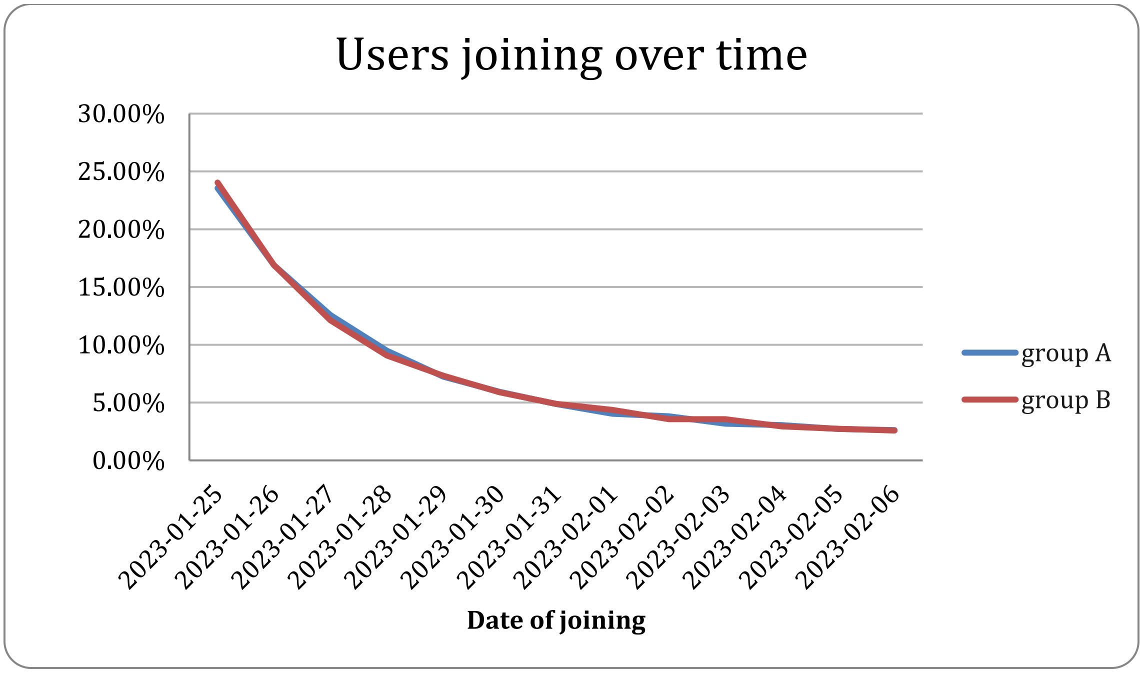
Novelty Test



A novelty test, also known as an anomaly detection or outlier detection test, is a statistical technique used to identify unusual or unexpected observations in a dataset. The goal of novelty test is to detect the tendency of users to respond more effective changes to a new experiment, object, service or any information than to the old one.

Here we tested if users started to buy products because of the new service has launched in website and that we can count as first impression but it does not mean that they really liked the product or service.

For a test I used spreadsheet analyzing by conversions over time and users joining over time.



Almost half of the conversions happened at first three days regardless of users saw the banner or not. And around 50% of users also join at first three days of A/B Test. It looks like there's no Novelty Effect because the number of conversions almost similar to number of users joining per day and there is no significant differences between treatment group and control group. Another interesting fact we can see here that joining date and number of conversions started to decrease for both groups at the end of the test.

Power Analysis

Power analysis in statistics is a method used to determine the sample size needed for a study to detect a statistically significant effect. It helps researchers plan their studies by estimating the minimum sample size required to achieve meaningful results.

Power analysis was conducted by using Python programming.

The result show that in order to get meaningful result we need to have 38,443 users per group for a conversion rate. And 92,733 users per group for an average amount per user. Which are significantly higher than we had in our A/B test dataset.

	Required Sample Size
Metric	
Conversion Rate	38443.800962
Average Amount Spent	92733.481872

Recommendation

Not to launch

As the conclusion, we conducted many different analysis where only conversion rate metrics show a strong positive result. However, average amount spent did not show big difference between two groups. Generally, Exploratory data analysis didn't indicate strong negative or positive result because if the treatment group was better in conversion rate the average amount spent result was opposite or vice versa.

Another test is a Power analysis where we found that in order to get more precise and accurate result, we need to get higher number of users. 38,443 users per group for conversion rate and 92,733 users for average amount spent which is significantly higher than we had during this analysis.

In my opinion, a novelty test was one of the important part of the analysis. About 50% of users joined the test and did conversion on first 3 days of the A/B test but the number of conversions almost similar to number of users joining per day and there's no significant differences between two groups. Additionally, I have noticed negative tendency on both metrics which probably means that the number of users and number of conversions will decrease in the future. May be there are other competitors in the market which mainly focused on food and beverage products and of course they will provide better offers to the customers. Competitors analysis would help us to understand this situation.

In the end, as company wants to bring awareness to this product category to **increase revenue** I would like to recommend not to launch. For example: You can make a purchase a pen 15 times for 1$ which you will spend only 15$. However, you can buy 1 time a book for 50$. The conversion rate for purchasing pen will be higher but the revenue will be much lower than a book's revenue. Conversion rate is not that important if your revenue is not high.

Appendix

1.Can a user show up more than once in the activity table? Yes or no, and why?

Yes, same customer can make many purchases in activity table.

select

count(distinct uid)

from activity

Select

count(uid)

from activity

2.What type of join should we use to join the users table to the activity table?

Left join

Select

count(distinct u.id)

From users u

Left join activity a on u.id=a.uid

3.What SQL function can we use to fill in NULL values?

Coalesce

4.What are the start and end dates of the experiment?

start date

"2023-01-25"

select

min(join_dt)

from groups

end date

"2023-02-06"

select

max(join_dt)

from groups

5.How many total users were in the experiment?

"48943"

```
select
count(users)
from users
```

6.How many users were in the control and treatment groups?

```
select
groups.device,
count(*)
from groups
group by 1
```

7.What was the conversion rate of all users?

```
SELECT Count(*) AS user_counter, conversion
FROM (SELECT u.id,
 u.country,
 u.gender,
 g.group,
 g.device,
 CASE
 WHEN a.spent > 0 THEN 'True'
 ELSE 'False'
 END AS Conversion,
 Sum(a.spent) AS total_spent
 FROM users u
 LEFT JOIN groups g
 ON g.uid = u.id
 LEFT JOIN activity a
 ON a.uid = g.uid
 GROUP BY 1, 2, 3, 4, 5,
6) t1
GROUP BY 2
```

8.What is the user conversion rate for the control and treatment groups?

```
with t1 as (
select
u.id,
u.country,
u.gender,
a.device,
Coalesce(g.group, 'N/A') as group,
Case when coalesce(a.spent, 0) > 0 Then 'Converted' ELSE 'Not Converted'
END as Conversion,
sum(a.spent) total_spent
from users u
left join activity a on u.id = a.uid
left join groups g on a.uid = g.uid
group by
1, 2,3,4,5,6)
select
count(*) as user_count,
t1.group,
conversion
from t1
where t1.conversion = 'Converted'
group by 2,3
```

SQL extract to Tableu

```
WITH t1
 AS (SELECT u.id,
 u.country,
 u.gender,
 g.group,
 g.device,
 CASE
 WHEN COALESCE(a.spent, 0) > 0 THEN 'True'
 ELSE 'False'
 END AS Conversion,
 Sum(a.spent) AS total_spent
 FROM users u
 LEFT JOIN groups g
 ON g.uid = u.id
 LEFT JOIN activity a
 ON a.uid = g.uid
 GROUP BY 1,
 2, 3, 4, 5, 6)
SELECT id,
 COALESCE(country, 'N/A') country,
 COALESCE(gender, 'N/A') gender,
 t1.group,
 COALESCE (device, 'N/A') device,
 conversion,
 COALESCE (total_spent, 0) total_spent
FROM t1
ORDER BY 4
```

SQL query for Novelty Test

```
SELECT u.id as user_id, coalesce(cast(a.dt as varchar), 'N/A') as date, g.group, g.join_dt as
join_date,
   CASE WHEN a.spent > 0 THEN 'True'
   ELSE 'False' END as Conversion,
   SUM(coalesce(a.spent, 0)) as total_spent
FROM users u

LEFT JOIN groups g on g.uid = u.id
LEFT JOIN activity a on a.uid = g.uid
GROUP BY
   u.id, a.spent, g.group, g.join_dt, a.dt
```

Python code for Power analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.power import TTestIndPower, NormalIndPower
csv_path = '/Globox A B test CSV.csv'
globox=pd.read_csv(csv_path)
globox.info()
globox.head()
# Segregating the data into control and test groups
control_group = globox[globox['group'] == 'A']
treatment_group = globox[globox['group'] == 'B']
# Calculating the baseline metrics for the control group
control_conv_rate = control_group['conversion'].mean()
control_avg_amount_spent = control_group['total_spent'].mean()
# Calculating the metrics for the test group
treatment_conversion_rate = treatment_group['conversion'].mean()
```

```
treatment_avg_amount_spent = treatment_group['total_spent'].mean()


# Preparing a summary of the metrics
metrics_summary = pd.DataFrame({
    "Metric": ["Conversion Rate", "Average Amount Spent"],
    "Group A: Control": [control_conv_rate, control_avg_amount_spent],
    "Group B: Test": [treatment_conversion_rate, treatment_avg_amount_spent]})
metrics_summary.set_index("Metric", inplace=True)
metrics_summary
# Parameters for the power analysis
alpha = 0.05  # Significance level
power = 0.80  # Statistical power
mde_relative = 0.10  # Minimum Detectable Effect (10% relative change)
conversion = 1
# Calculating the absolute MDE for conversion rate and average amount spent
mde_conversion_rate = control_conv_rate * mde_relative
mde_avg_amount_spent = control_avg_amount_spent * mde_relative
# Power analysis for conversion rate (two-sample proportion test)
effect_size_conversion_rate = NormalIndPower().solve_power(
    effect_size=(mde_conversion_rate / np.sqrt(control_conv_rate * (1 - control_conv_rate))),
    alpha=alpha,
    power=power,
    ratio=1  # Equal size for control and test groups)
# Power analysis for average amount spent (two-sample t-test)
effect_size_avg_amount_spent = TTestIndPower().solve_power(
    effect_size=(mde_avg_amount_spent / control_group['total_spent'].std()),
    alpha=alpha,
    power=power,
    ratio=1,  # Equal size for control and test groups
    alternative='two-sided')
required_sample_size = pd.DataFrame({
    "Metric": ["Conversion Rate", "Average Amount Spent"],
    "Required Sample Size": [effect_size_conversion_rate, effect_size_avg_amount_spent]})
```

```
required_sample_size.set_index("Metric", inplace=True)
required_sample_size
```

Hypothesis test, Power analysis and Novelty test documents will be attached with the report.