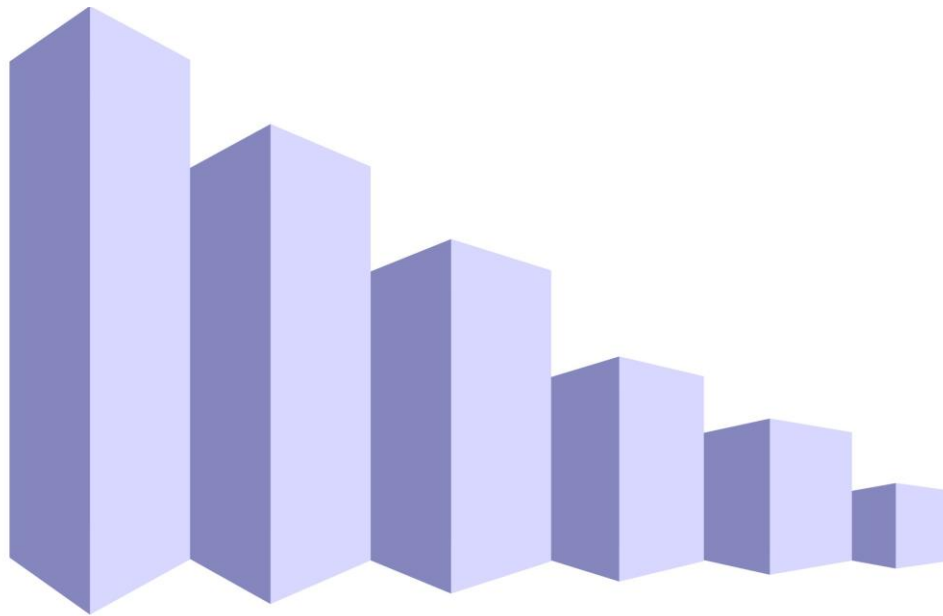


A/B Test Report: New User Interface Conversion Rate Analysis

Date: *17/09/2023*

Prepared by: *Samuel Obadan*

Project: New Conversion rate/spend rate analysis for the *Glowbox* cooperation



Executive Summary:

This report presents the results of an AB test conducted to assess the impact of a new user interface on the conversion rate of a digital platform. The objective of the test was to determine whether the new user interface leads to a statistically significant improvement in the conversion rate compared to the old interface.

[Link to tableau dashboard](#)

<https://public.tableau.com/app/profile/samuel.obadan/viz/ABtestsample/GlowboxABtest>

Methodology:

Test Duration: The AB test was conducted over a period of two weeks, starting on [25th Jan 2023] and ending on [6th Feb 2023].

Participants: A random sample of [Number of Participants] users was selected for the test, with approximately [50%] exposed to the new user interface (Group B- Treatment group in orange (fig.1.1)) and approximately [50%] using the old interface (Group A – Control group in blue (fig.1.1))

Metric: The primary metric of interest was the conversion rate, which measures the percentage of users who completed the desired action (e.g., signing up, making a purchase) on the platform.

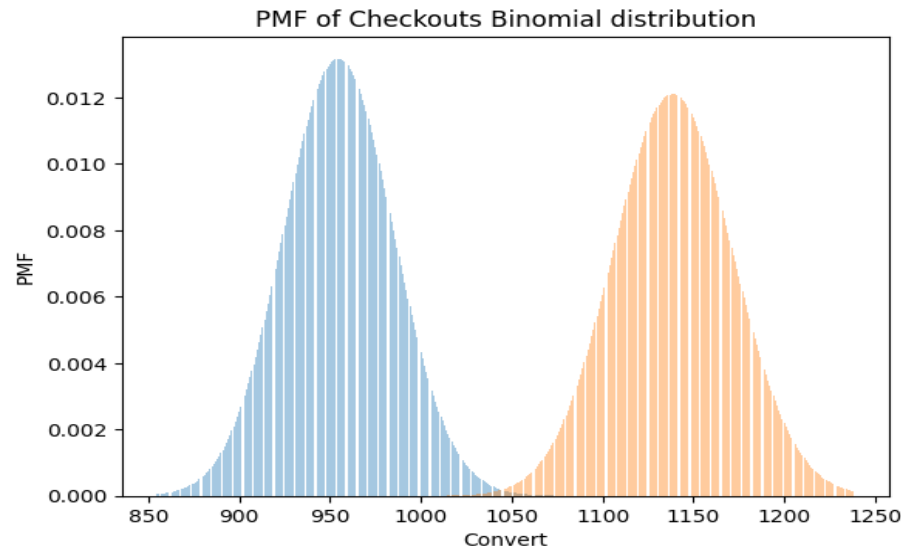


Fig.1.1 Binomial distribution for Group A –control group (blue) and Group B-treatment group (orange)

Results:

The following table summarizes the results of the AB test:

GROUP	USERS	CONVERSIONS	CONVERSION RATE
A (CONTROL GROUP)	24,343	955	3.92%
B (TREATMENT GROUP)	24,600	1139	4.63%

GROUP	USERS	Spent (Purchase)	Average Purchase Rate
A (CONTROL GROUP)	24,343	\$82,145.90	\$3.37
B (TREATMENT GROUP)	24,600	\$83,415.32	\$3.39

T-test Result for conversion Rate

index	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-3.866406	48712.312188	two-sided	0.000111	[-0.01 -0.0]	0.03494	17.919	0.971595

T-test Result for spend/ purchase mean

index	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-0.070425	48894.497311	two-sided	0.943856	[-0.47 0.44]	0.000637	0.01	0.050568

Power analysis:

The minimum population for both A and B groups for a possible 1% increase in conversion rate is 5104. With a population of approximately 24,000 for each group, we thus satisfy the minimum requirement.

Sanity check:

With a P-value > 0.01 , we conclude that a ‘Sample Ratio Mismatch’ (SRM) is unlikely. Thus, there exist a fine balance between the control group population and the treatment group population.

```
Power divergence Result (statistic=1.3495086120589257, p-value=0.24536404027139797)
SRM likely not present
```

Checking for Simpson’s paradox:

Table 1.1

Group	Device	Rate
A	Andriod	2.31
	Iphone	50.7
	others	6.54
B	Android	2.47
	Iphone	4.92
	others	3.46

Tableau Snapshot



Fig 1.2 Tableau KPI dashboard and conversion rate

We have detected Simpson's paradox even when the counts between the two groups and their subgroups are proportional. Simpson's paradox is a statistical phenomenon in which a trend appears in different groups of data, but disappears or reverses when these groups are combined. I suspect there is a confounding variable that affects the “**others**” **devices** that needs to be taken into account. We find the average spent (purchase) to be twice as much in the control group than in the treatment group as shown in table 1.1, and \$5.00 as shown in figure 1.2

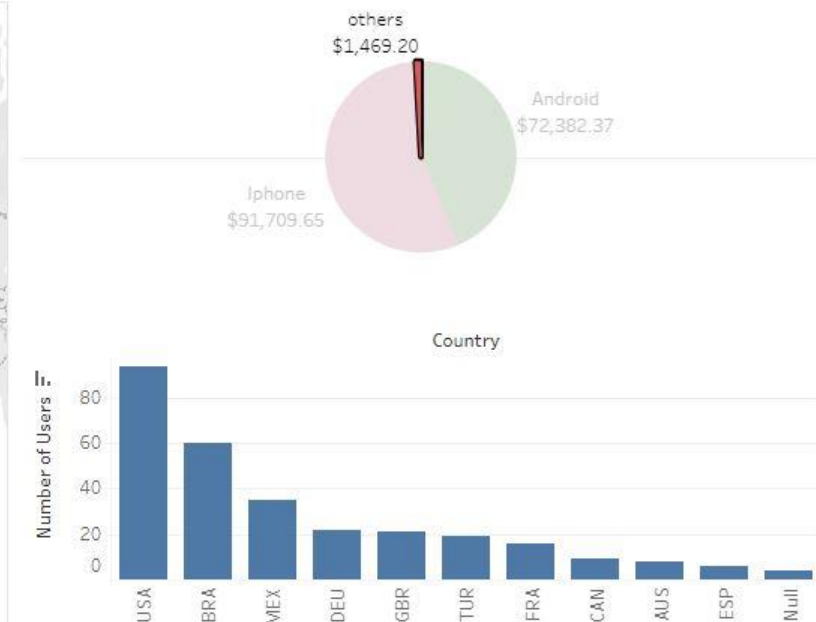


Fig 1.3 Tableau Maps and Graphs for devices and countries

From the fig 1.3, most of this purchase has come from Mexico (The darkest green dot on the map), which also is the country with the third highest number of users. They may exist confounding variables such as events in Mexico during the run of the simulation which may need further investigation especially between 26th of January to 28th of January.

Checking for Novelty Effect:

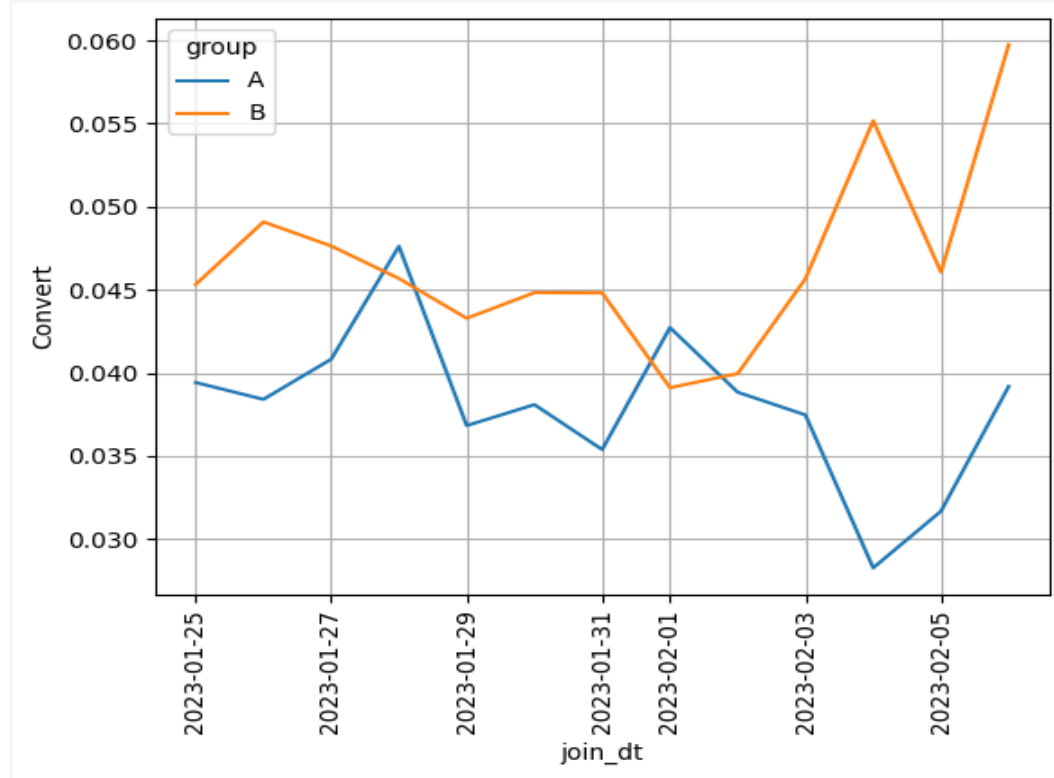


Fig 1.4 Conversion rate vs Duration in weeks

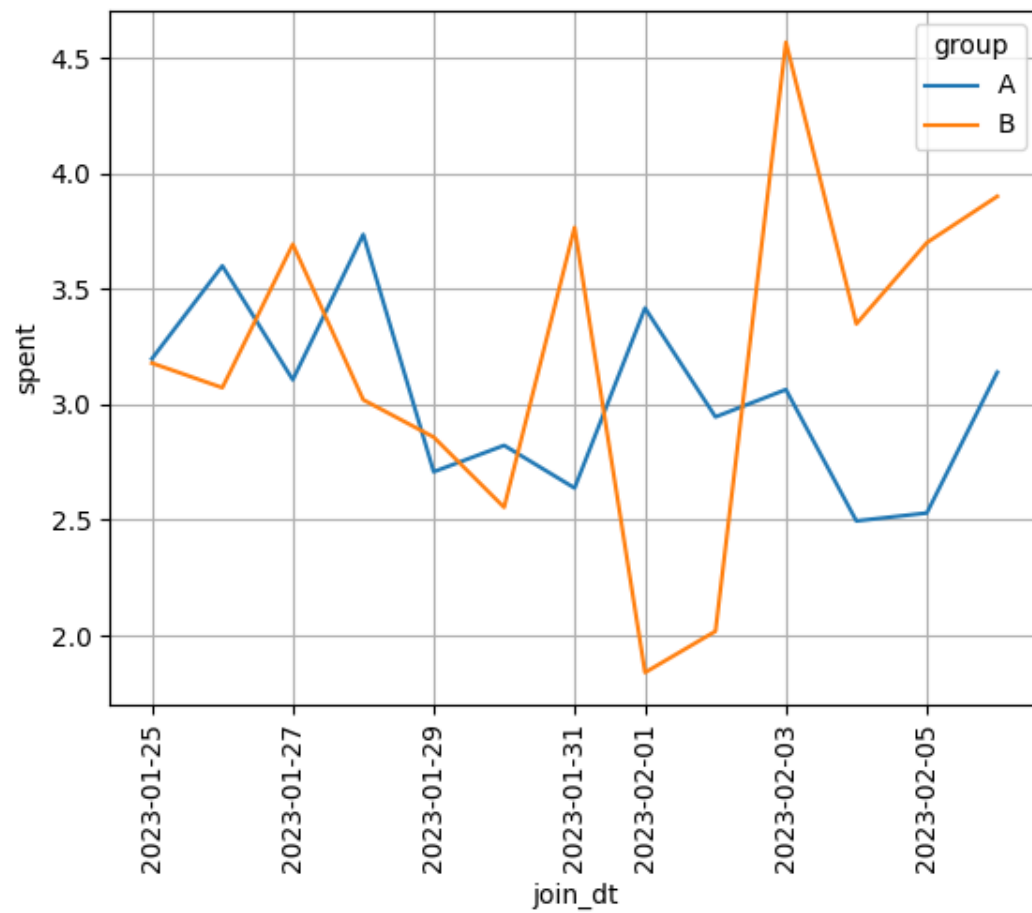


Fig 1.5 Average spent vs duration in weeks

While the randomness among both groups suggest that there are no novelty effects in both conversion rate the average purchase (spent) amount (fig 1.4, and fig 1.5 respectively) for both groups, we however observe a novelty effect trend in both the control and treatment groups when taking into account the total amount of money spent (purchases) during the experiment, for the period as shown in the fig 1.6 below

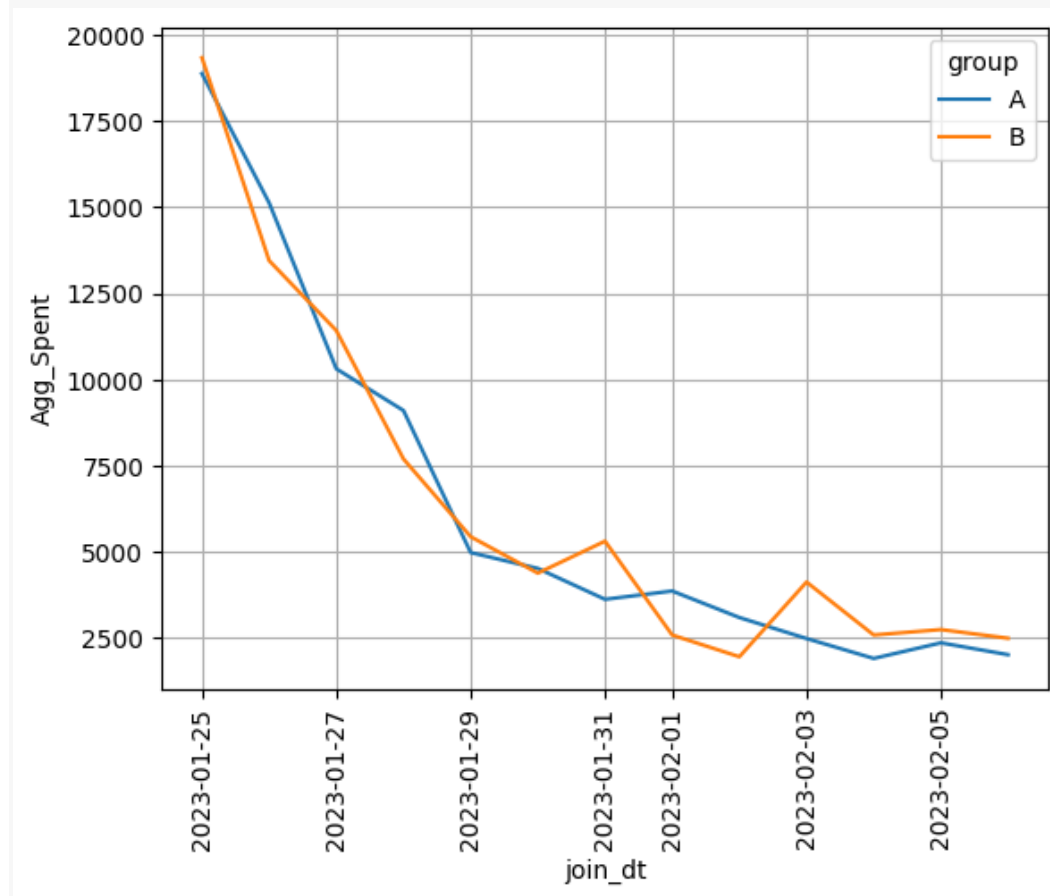


Fig 1.6 Aggregated amount spent (Agg_spent) vs duration in weeks

Statistical Analysis:

Conversion rate:

A two-sample t-test was performed to determine the statistical significance of the difference in conversion rates between the two groups. The results indicated a [p-value: 0.000111], which is [less than] the significance level of [$\alpha = 0.05$]. Therefore, we [reject] the null hypothesis that there is no difference in conversion rates between the two user interface groups.

Purchase rate:

A two-sample t-test was also performed to determine the statistical significance of the difference in purchase rates between the two groups. The results indicated a [p-value: 0.943856], which is [greater than] the significance level of [$\alpha = 0.05$]. Therefore, we [fail to reject] the null hypothesis that there is no difference in average amount spent (purchase) between the two user interface groups.

Conclusion:

Conversion Rate

Based on the results of the AB test, the new user interface led to a [statistically significant] improvement in the conversion rate compared to the old interface. The conversion rate for the new user interface group was [4.63] %, while the conversion rate for the old interface group was [3.92] %. This suggests that the new user interface positively impacted user engagement and interaction with the platform.

Purchase rate

Based on the results of the AB test, the new user interface led to a [non-statistically significant] improvement in the purchase mean compared to the old interface. The conversion rate for the new user interface group was [\$3.37], while the purchase mean for the old interface group was [\$3.39]. Consequently, we cannot conclude at this moment that there exist a significant difference in the spending (purchase means) between the control and treatment groups.

Recommendation:

Despite the overall improvements in conversion rate, with the presence of a possible novelty effect and Simpsons' paradox in the observations, it is recommended that we delay the implementation of the new interface subject to further investigations.

1. **Long-Term Analysis:** Towards the end of the second week, the novelty effect seems to begin to plateau. Thus it is recommended that we analyse the experiment over an extended period to determine if the initial increase is temporary or a lasting effect.

2. **Explore Other Variables:** In order to find a possible cause for the Simpson's paradox, we need to consider whether there are other variables or factors at play especially in Mexico that could be influencing the increase in spending average especially in the controlled group A. It's important to examine other potential explanations for the observed change in this behaviour within the 26th of January to the 28th of January which showed a lower conversion rate in the control group A but significantly higher amount of aggregated sum of money spent (purchases) as indicated by the thickness of the blue line in fig 1.2.

Addendum:

SQL code for data extraction

```
SELECT g.uid,  
       g.group,  
       join_dt,  
       g.device,  
       country,  
       gender,  
       COALESCE(spent, 0) AS spent  
FROM   groups g  
       JOIN users u  
         ON u.id = g.uid  
       LEFT JOIN activity a  
         ON a.uid = g.uid
```

Data Cleansing and wrangling using Python code

Extracted SQL code saved as ABtesData.csv loaded into data frame (df)

```
df = pd.read_csv('ABtestData.csv')
```

Aggregate sum of all money spent(purchases) on the uid of each user and store in data frame (df2)

```
df2 = df.groupby(['uid'], as_index=False) ['spent'].agg({'Spent': 'sum'})
```

Drop duplicate uid records from the original (df) data frame

```
df.drop_duplicates('uid', inplace=True)
```

Create a new data frame (df3) via a pandas join on uid between df and df2 using an inner join

```
df3=df2.merge(df, on =['uid'], how = 'inner')
```

Drop the old spent column as the new created Spent column in step 2 has the needed aggregated sum of purchases

```
df3.drop('spent',axis=1, inplace=True)
```

Create a new column 'Convert' using a pandas inline function to place a 1 where there was a purchase and 0 otherwise

```
df3['Convert'] = df3['Spent'].map(lambda x: 1 if x>0 else 0)
```

replace null values with the word 'others' in the device column.

```
df3.device.fillna('others', inplace=True)
```

replace null values with 'not specified' in the gender column

```
df3.gender.fillna('not specified', inplace=True)
```

optional, but for best practice, replace null values for where the country was not specified with a 'NULL' string

```
df3.country.fillna('NULL', inplace=True)
```

Set the date column (join_dt) to a date data type to enable plotting and date computation else it would be treated as a string.

```
df3['join_dt'] = pd.to_datetime(df3['join_dt']).dt.date
```