INDEX

Summary

In the evolving landscape of online business, optimizing website performance is pivotal for attracting and retaining customers. An effective strategy often employed is redesigning the homepage to enhance user experience or accentuate new offerings. An A/B test was conducted to evaluate the correlation and causation between a new homepage design and revenue changes at GloBox, specifically, a new banner was incorporated on the homepage, highlighting the new F&B section. Post analysis, it was concluded that the findings were insufficient to confidently recommend moving forward, prompting the necessity for some revisions to the employed methodology and a possible redesign of the banner.

Context

The motivation behind conducting this A/B test stems from GloBox's initiative to bolster its Food and Beverages (F&B) segment. The company aims to divert more attention towards this sector, and the introduction of the new banner on the homepage is a part of this strategic move. The hypothesis test was designed to assess whether this visual change would have a significant effect on users' interaction and spending behaviour.

1. A/B Test

Our A/B test employed a controlled approach over a 2-week period. To mitigate any bias, two comparable groups were randomly assigned, following a blindness method to prevent awareness of group assignments. The control group experienced the existing homepage design, while the treatment group was exposed to the new homepage featuring the F&B banner.

2. Hypothesis Testing

The central goal was to evaluate the impact of the new homepage design on two metrics: users' response and average expenditure per user. Hypothesis testing was conducted with the following in mind:

Null Hypothesis (H0): There would be no statistically significant difference in results between the control group (existing homepage) and the treatment group (new homepage).

Alternative Hypothesis (Ha): There would be a statistically significant difference between the control group and the treatment group, indicating a causal impact of the new homepage design.

It is crucial to remember that any changes to the website aim to enhance the customer experience, drive user engagement, and ultimately, increase revenue from the F&B offerings. Through the statistical analysis explained in the next chapter, we hope to gather insights to inform future design and marketing decisions.

Statistical Analysis

To evaluate both hypotheses, a statistical analysis was conducted. The p-value, which measures the probability of obtaining results as extreme as the observed ones under the assumption of the null hypothesis, was used to determine statistical significance. A significance level of 5% was employed, where a p-value less than 0.05 would lead to the rejection of the null hypothesis.

In other words, if we find evidence in our data that makes us 95% confident that group A and group B behaved differently, we'll assume this was caused by the new banner.

Additionally, a confidence interval was calculated to estimate the range of possible differences in repeated or increased samples, i.e., our estimation of how much could the results vary if later tests were executed.

1- Conversion Rate

To assess if there is a difference in the conversion rate between Group A and Group B, the Proportion Hypothesis method was used (refer to *Appendix 2* for detailed steps and formulas).

The pooled proportion was determined to calculate the Standard Error. A standardized Z-Score statistic was then calculated to obtain the p-value, which was compared to the assigned significance level (α).

To ensure the reliability of the test for future analysis, a margin of error was calculated using the standard error and critical values table, enabling the presentation of a confidence interval for the results.

RESULT: *Since the p-value was smaller than the significance level, we can conclude with 95% confidence that there is a difference in conversion rate between the two groups, thus **Rejecting** the null hypothesis*

2- Average Spent

To test the Null Hypothesis that there is no difference in the average spent per user between both groups, a T-Test was performed using the t-distribution and a 5% significance level, assuming unequal variance and applying Welch's method (refer to *Appendix 3* for details).

After calculating the standard error, the degrees of freedom were adjusted according to Welche's Test. The t-test was conducted using Excel's built-in formula for a two-tailed, independent variable Type 3 test. Confidence calculations were performed to further assess the analysis.

RESULT: *Since the p-value was much larger than the confidence level, we cannot conclude with statistical significance that the groups differ in results. Therefore, we **Fail to Reject** the null hypothesis.*

Visualizing and further analyzing the results

Using Tableau, I present our findings for your consideration:

1. Test findings:

 - The charts below display the actual difference in conversion rate and average spending per user using a dual-axis bar chart for conversion rate and a reference line for average spending.

- Additionally, a stacked histogram divided into bins of $30 displays the count of users within each bin for both groups. By stacking both groups one can see how they behaved similarly.

Distribution of User Expenditure

Both Groups behaved similarly when looking at the amount spent

 Excluded from this chart:
- non-converted users: 46849
- some outliers above $210 (not enough data)

Conversion Rate and Average Spent per Group
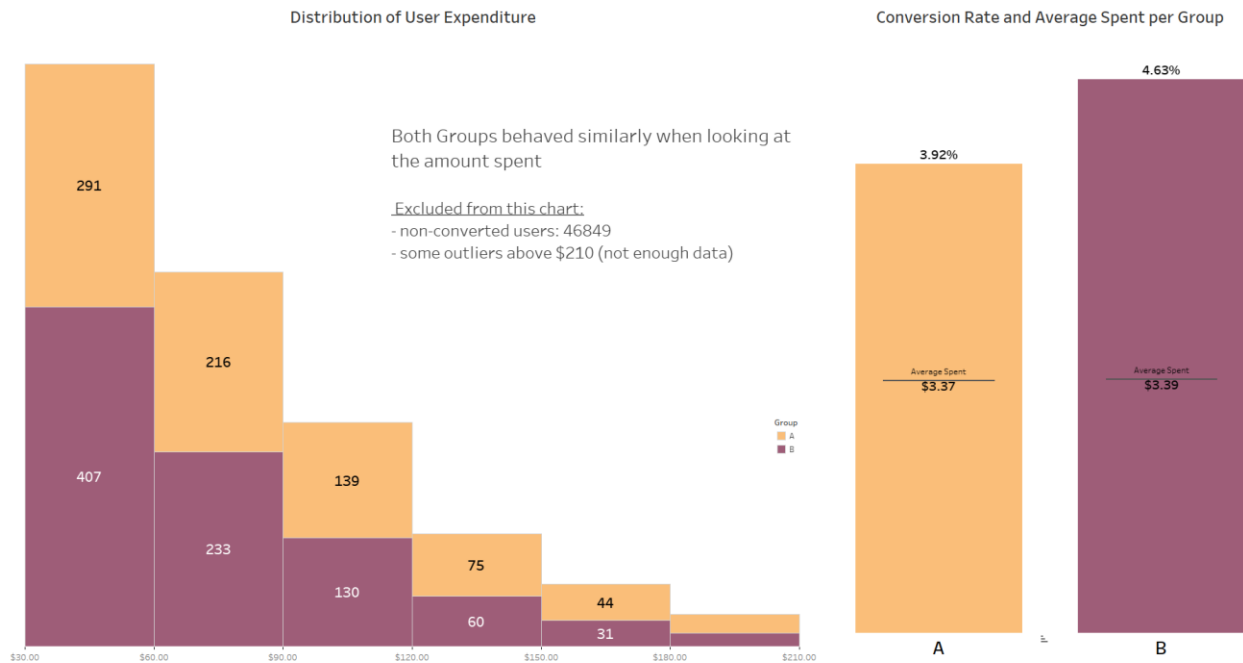
2. Demographics:

 - To analyze the participants of the A/B test, a single dashboard categorizes our users based on gender, country of login, and the operating system used (Android or iOS).
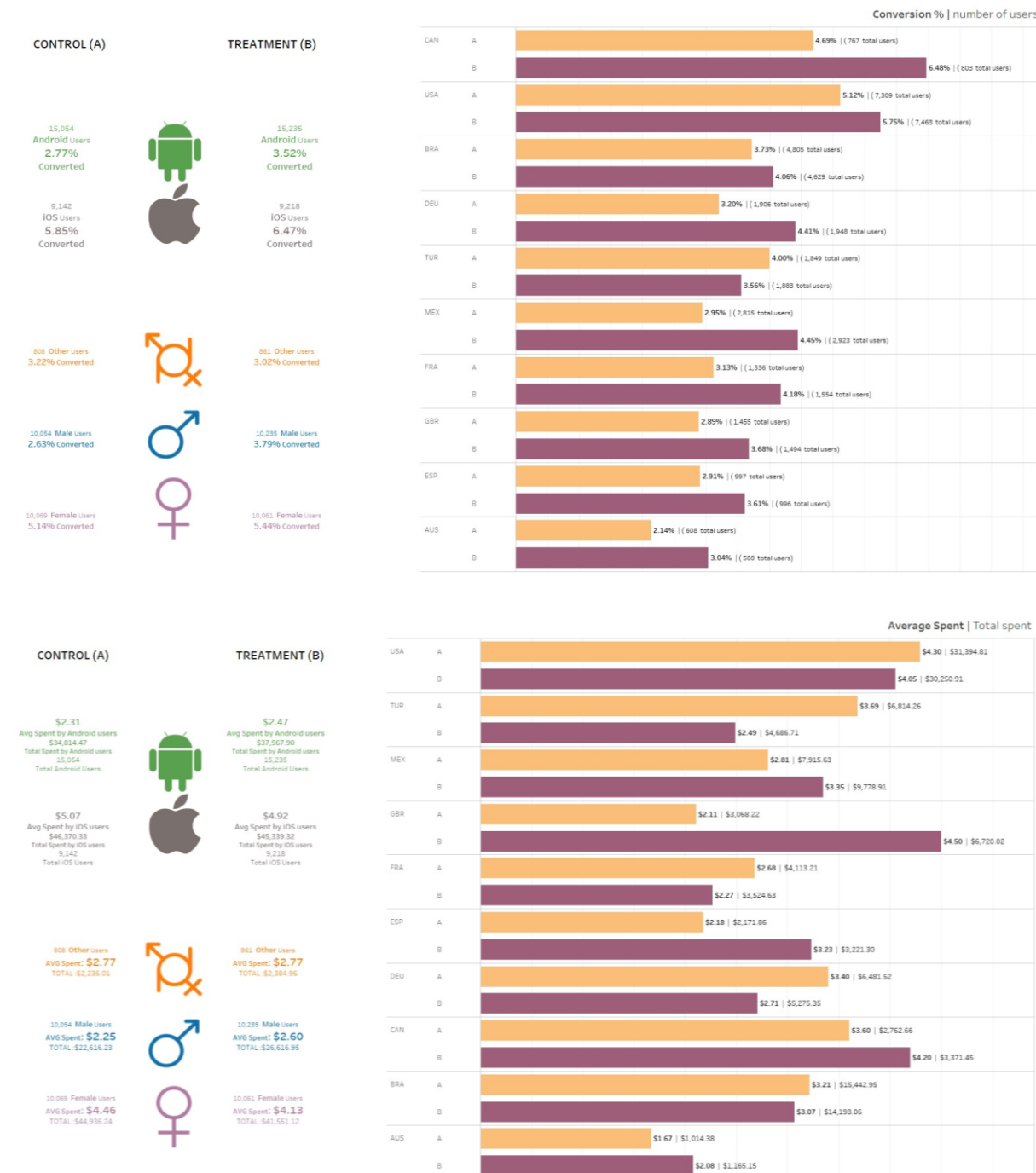
Some of the insights we can draw are that that Android users are well below the average for iOS users.

Another, female users are not only double the male users but also spend almost twice as much per user.

Conversion % | number of users

	CONTROL (A)	TREATMENT (B)
Android Users	15,054 / 2.77% Converted	15,235 / 3.52% Converted
iOS Users	9,142 / 5.85% Converted	9,218 / 6.47% Converted
Other Users	808 / 3.22% Converted	861 / 3.02% Converted
Male Users	10,054 / 2.63% Converted	10,235 / 3.79% Converted
Female Users	10,069 / 5.14% Converted	10,061 / 5.44% Converted

Country	A	B
CAN	4.69%	(767 total users)	6.48%	(803 total users)
USA	5.12%	(7,309 total users)	5.75%	(7,463 total users)
BRA	3.73%	(4,805 total users)	4.06%	(4,629 total users)
DEU	3.20%	(1,906 total users)	4.41%	(1,948 total users)
TUR	4.00%	(1,849 total users)	3.56%	(1,883 total users)
MEX	2.95%	(2,815 total users)	4.45%	(2,923 total users)
FRA	3.13%	(1,536 total users)	4.18%	(1,554 total users)
GBR	2.89%	(1,455 total users)	3.68%	(1,494 total users)
ESP	2.91%	(997 total users)	3.61%	(996 total users)
AUS	2.14%	(608 total users)	3.04%	(560 total users)

Average Spent | Total spent

	CONTROL (A)	TREATMENT (B)
Avg Spent by Android users	$2.31	$2.47
Total Spent by Android users	$34,814.47	$37,567.90
Total Android Users	15,054	15,235
Avg Spent by iOS users	$5.07	$4.92
Total Spent by iOS users	$46,370.33	$45,339.32
Total iOS Users	9,142	9,218
Other Users	808 / AVG Spent: $2.77 / TOTAL $2,236.01	861 / AVG Spent: $2.77 / TOTAL $2,384.96
Male Users	10,054 / AVG Spent: $2.25 / TOTAL $22,616.23	10,235 / AVG Spent: $2.60 / TOTAL $26,616.95
Female Users	10,069 / AVG Spent: $4.46 / TOTAL $44,936.24	10,061 / AVG Spent: $4.13 / TOTAL $41,551.12

Country	A	B
USA	$4.30	$31,394.81	$4.05	$30,250.91
TUR	$3.69	$6,814.26	$2.49	$4,686.71
MEX	$2.81	$7,915.63	$3.35	$9,778.91
GBR	$2.11	$3,068.22	$4.50	$6,720.02
FRA	$2.68	$4,113.21	$2.27	$3,524.63
ESP	$2.18	$2,171.86	$3.23	$3,221.30
DEU	$3.40	$6,481.52	$2.71	$5,275.35
CAN	$3.60	$2,762.66	$4.20	$3,371.45
BRA	$3.21	$15,442.95	$3.07	$14,193.06
AUS	$1.67	$1,014.38	$2.08	$1,165.15

3. Confidence and Power

- Visualized 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.

The graph clearly shows us that there is virtually no chance of the difference in conversion being null, which gives confidence to the findings that if repeating this test, Group B would always turn higher conversions. The same cannot be said for the difference in avg spent, where the bottom line shows us that with repeated samples comes the possibility of the difference in spending being null or even negative (Group A spending more).

Conversion Difference Confidence Interval

0.35%	1.07%
0.71%	
0.0	

Avg.Spent Difference Confidence Interval

$-0.44	$0.47
$0.02	
0	
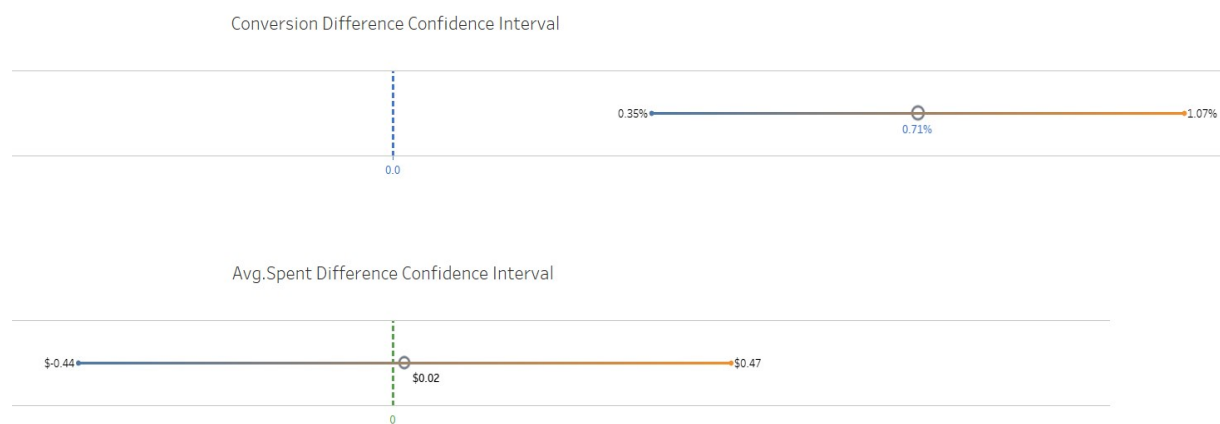
- Conducted further analysis to determine the potential presence of a novelty effect (this was inconclusive due to the duration and methodology of the A/B Test- more on this in the following chapter): visualized a timeline of Group B users converting and compared it with a bar chart of daily number of users joining the experiment. I further presented a categorization of the count of users who converted by the number of days It took them to convert (refer to *Appendix 4*).

- Performed a Power Analysis to ascertain the validity of our decision to reject and the statistical reliability of our findings. The conclusions here are that a minimum sample size to have a validity power of 90% was achieved when Rejecting H0 (refer to *Appendix 5*).

While crucial to strengthen the confidence in our findings, the advanced analysis described above makes certain assumptions about the baseline and general user behavior that we cannot confidently ascertain, while also not considering several other factors, discussed in the next chapter (Considerations).

Considerations

<u>Novelty Effect</u>: Checking for the novelty effect is crucial in decision-making, as it is a common error that can impact the accuracy of our conclusions.

To avoid the novelty effect when designing an A/B test, certain steps should be taken. These steps may include carefully selecting the duration of the **test** to ensure users have enough time to adapt to any changes and tracking user behaviour over a longer period to assess any sustained effects.

In our case, I strongly recommend categorizing our Treatment Sample into Returning Users vs New users: simply put, the idea is that if we find a larger conversion rate in returning users when compared to that of new users, chances are those "usual visitors" are merely reacting to the new feature on the website: this novelty effect will be short-lived and should not impact our decision to implement in the long run. Alternatively, if both user groups display similar conversion rates and spending, we can be confident that these results will be representative of the bottom-line results when choosing to keep the banner.

For visual checks and insights into a potential novel effect, please see appendix 4.

<u>Seasonality</u>: Seasonal variations in data can introduce patterns or fluctuations that need to be accounted for in power analyses. For example, consumer purchasing behaviour might be influenced by seasonal trends, and failing to consider these variations can lead to an underpowered study.

Given the timeline of the experiment, there is a possibility that the low average spent and the almost null difference between both groups be a result of low purchasing ability, typical of the month of January, and consequent larger sensitivity to pricing.

However, this is the opposite mindset of luxury goods consumers, which is our target market. More on this paradox, in the final "Recommendations" chapter. Nonetheless, to be able to completely exclude seasonality one would have to collect data over multiple seasons or adjust for seasonal effects in statistical models.

<u>Metrics</u>: While the suggested homepage change is designed to attract users to the new F&B offerings, the measurements of this A/B Test are not specific to this section of our business. By testing for across-the-board proportions and behaviour, the question of whether the banner has made a difference in actual revenue from F&B is left unanswered.

Lastly, not having historic values from before the experiment limits our comparison and conclusion making.

Conclusions

In light of our recent analysis on GloBox's homepage banner initiative for the Food & Beverages (F&B) sector, we have arrived at certain insights that call for a considered approach. While the aim to enhance user interaction and bolster the F&B segment is commendable, the current findings urge caution and further investigation.

Based on the results, a small positive correlation is seen in user response to the treatment when it comes to conversion rates, but no significant correlation was verified regarding average spending amounts per user.

This indicates to:

a) no evidence that the new homepage design improves revenue per user.

b) no certainty that the increase in conversion rate is representative of future long-term results.

Here are my recommendations:

Reevaluate the Data: Our initial tests have shown promise, but the evidence is not robust enough to support a full-scale launch. There may be novelty effects influencing the results. We recommend further subdividing the test groups to gauge the long-term effects of the new homepage design more accurately.

Consider Seasonal Effects: Our consumer base appears less sensitive to price fluctuations, but the timing of our experiment may not reflect typical behavior throughout the year. Repeating the experiment at different times is of low cost to our company, and could provide a more balanced perspective, free from potential seasonal biases.

Explore Historical Comparisons: Our ability to draw meaningful conclusions is hindered by a lack of historical records. Going forward, consistent record-keeping and benchmarking, or repeating tests across multiple years, will enable more confident decision-making.

Personalize and Optimize the Design: Opportunities exist to tailor the banner to specific audiences, such as female users, and to enhance its compatibility with different operating systems. These refinements could contribute to a more targeted and effective design.

Strategic Pause for Reflection: Given the above, we propose a hold on the launch. This time can be used to consider a thoughtful redesign of both the banner's user interface and the testing format, with a view to re-launching at a more opportune time.

Insights for stakeholders:

Alignment with Strategic Goals: Our recommendations are aligned with the company's broader goals of innovation and customer-centric growth. We are not abandoning the initiative but advocating for a more informed and strategic approach.

Risk Mitigation: By heeding these insights, we act prudently, potentially avoiding unforeseen setbacks that could affect shareholder value.

Investment in Future Success: The proposed additional testing and refinements are not merely expenses but investments in the future success of the company. They can lead to more informed decisions and a stronger impact from our F&B segment.

Transparency and Diligence: This process demonstrates our commitment to transparency and diligence in decision-making, affirming our responsibility to act in the best interests of our customers and stakeholders.

In conclusion, our recommendation is not a step back but a strategic pause. It reflects our commitment to making well-informed decisions that align with our long-term objectives, uphold our values, and maximize stakeholders' confidence in us. By taking the time to refine and reassess, we lay the groundwork for a more successful and resilient landing page.

Appendix 1.1

Data Prep

```
 GloBox_Prep.sql
 1       -- create tables
 2     create table users(
 3                    id numeric,
 4                    country varchar,
 5                    gender varchar
 6     );
 7     create table groups(
 8                    uid numeric,
 9                    "group" varchar,
10                    join_dt date,
11                    device varchar
12     );
13     create table activity(
14                    uid numeric,
15                    dt date,
16                    device varchar,
17                    spent numeric
18     );
19
20    /*import values from csv, after saving each page separately as .csv
21     (after 1st line, replace all ",," with "") | click: tables > Import/Export data */
22
23     -- round all SPENT values to 2 decimals
24     UPDATE activity
25     SET spent = ROUND(spent::numeric, 2);
```

```
56
57     --8.    What was the conversion rate for each group?
58     SELECT
59         ROUND(100 * COUNT(DISTINCT(a.uid)) FILTER (WHERE "group" = 'A') * 1.0 / COUNT(DISTINCT(g.uid))
60         FILTER (WHERE "group" = 'A'),2) || '%' AS conversion_A,
61         ROUND(100 * COUNT(DISTINCT(a.uid)) FILTER (WHERE "group" = 'B') * 1.0 / COUNT(DISTINCT(g.uid))
62         FILTER (WHERE "group" = 'B'),2) || '%' AS conversion_B
63     FROM groups g
64     FULL JOIN activity a USING(uid);
65
66     --9.    What is the average amount spent per user for each group, including users who did not convert?
67     SELECT
68         "group",
69         ROUND(AVG(SUM(spent::numeric)) OVER(PARTITION BY "group") / COUNT(DISTINCT g.uid),2) AS avg_spent
70     FROM groups g
71     FULL JOIN activity a USING(uid)
72     GROUP BY "group";
73     --OR
74     SELECT groups.group,
75     ROUND(SUM(activity.spent::numeric) / Count(DISTINCT(groups.uid)),2) As avg_spntA
76     from groups
77     left join activity on groups.uid=activity.uid
78     group by groups.group;
79
```

Appendix 1.2

Spreadsheet migration

```sql
80   -- create a query for next fase of the project and dowload as CSV
81   WITH summing AS
82       (SELECT DISTINCT(uid),
83                    SUM(spent) AS expenditure
84   FROM activity
85   GROUP BY uid
86   HAVING SUM(spent) > 0)
87
88   SELECT U.*,
89           g.device,
90           g.group,
91           CASE WHEN expenditure > 0 THEN 1 ELSE 0 END AS converted,
92           COALESCE(expenditure, 0.00) AS total_spent
93   FROM users u
94   LEFT JOIN groups g ON u.id = g.uid
95   LEFT JOIN summing ON u.id = summing.uid;
96
97   -- date query for Novelty vizz in Tableau
98   SELECT id,
99           "group",
100          join_dt,
101          dt AS conversion_dt,
102          dt - join_dt AS days_to_convert,
103          COALESCE(a.spent, 0.00) AS spent
104  FROM users u
105  RIGHT JOIN activity a ON u.id = a.uid
106  LEFT JOIN groups g USING(uid)
107  INNER JOIN (
108      SELECT uid, MIN(dt) AS first_purchase_date
109      FROM activity
110      GROUP BY uid
111  ) subq ON a.uid = subq.uid AND a.dt = subq.first_purchase_date
112  ORDER BY u.id;
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	id	country	gender	device	group	converted	total_spent								Diff Conversion	diff Avg Spent	Conversion LB	Conversion UB	AvgSpent LB	AvgSpent UB
2	1000000	CAN	M	I	B	NO	0.00								0.0071	0.0163	0.0035	0.0107	-0.438649909	0.471346865
3	1000001	BRA	M	A	A	NO	0.00			User Count	Converted	Avg Spent								
4	1000002	FRA	M	A	A	NO	0.00		Total	48943	2094	3.38								
5	1000003	BRA	M	I	B	NO	0.00													
6	1000004	DEU	F	A	A	NO	0.00													
7	1000005	GBR	F	A	B	NO	0.00													
8	1000006	ESP	M	A	B	NO	0.00													
9	1000007	BRA	F	A	A	NO	0.00													
10	1000008	BRA	F	A	A	NO	0.00		Group A	User Count	Converted	Avg Spent	St Dev							
11	1000009	USA		A	A	NO	0.00			24343	955	3.37	25.9364							
12	1000010	BRA	M	I	B	NO	0.00													
13	1000012	USA	M	A	B	NO	0.00													
14	1000013	GBR	F	A	A	NO	0.00													
15	1000014	USA	M	A	A	NO	0.00			User Count	Converted	Avg Spent	St Dev							
16	1000015	AUS	F	A	A	NO	0.00		Group B	24600	1139	3.39	25.4141							
17	1000016	DEU	F	I	A	NO	0.00													
18	1000017	MEX	O	A	B	NO	0.00													
19	1000018	USA	M	I	B	NO	0.00													
20	1000019	MEX	F	A	B	NO	0.00													
21	1000020	AUS	M	A	A	NO	0.00													
22	1000021	GBR	M	A	B	NO	0.00													
23	1000022	DEU	F	A	B	NO	0.00													
24	1000023	BRA	M	I	B	NO	0.00													
25	1000024	CAN	M	A	A	NO	0.00													
26	1000025	USA	M	I	A	NO	0.00													
27	1000026	DEU	F	A	B	NO	0.00													
28	1000027	BRA		I	B	NO	0.00													
29	1000028	BRA	F	A	A	NO	0.00													
30	1000029	MEX	M	A	A	NO	0.00													
31	1000030	FRA	F	I	A	NO	0.00													
32	1000031	GBR	F	I	B	NO	0.00													
33	1000032	GBR		I	A	NO	0.00													
34	1000033	FRA		A	A	NO	0.00													

Appendix 2

Proportions Hypothesis Theory

H0	There is **no** difference in conversion rate between the 2 groups	pA = pB
H1	There is a difference in conversion rate between the 2 groups	pA ≠ pB

xA	955	Conv_Rate A	**pA =** 0.039231
xB	1139	Conv_Rate B	**pB =** 0.046301
nA	24343	conv_rate_diff	**Pdiff =** 0.007070
nB	24600		
α	0.05	H0 RESULT	REJECTED

pooled proportion ($p\hat{}$)	$(xA+xB) / (nA+nB)$	$p\hat{}$**pooled =**	0.04278
standard error *(SE)*	sqrt $(p\hat{} * (1-p\hat{}) * ((1 /\ nA)+ (1 /\ nB)))$	**S E =**	0.00183
test statistic *(Z-score)*	$(conv\_rateB - conv\_rateA) / SE$	**z =**	3.86429
critical value *(crit)*	(for α=0.05)	**crit =**	1.96
margin of error *(ME)*	*critical value * standard error*	**M E =**	0.00359
P-value	*(using NORM.S.DIST)*	**p-value =**	0.00011
confidence level *(CL)*	1 - α	**C L =**	95%
confidence interval Lower Bound	Pdiff - ME	**LB =**	0.0035
confidence interval Upper Bound	Pdiff + ME	**UB =**	0.0107

Appendix 3

Averages Hypothesis Theory

H0	There is **no** difference in average spent between the 2 groups	μA = μB
H1	There is a difference in average spent between the 2 groups	μA ≠ μB

μA	**3.37**	Sample StDev A	**sA =**	**25.9364**
μB	**3.39**	Sample StDev B	**sB =**	**25.4141**
nA	**24343**			
nB	**24600**			
α	**0.05**	H0 RESULT	FAILED TO REJECT	

standard error *(SE)*	sqrt((sA^2 / nA) + (sB^2 / nB))	S E =	0.23214
degrees of freedom *(df)*	(nA-1) +(nB-1)	d f =	48941
critical value (crit)	*using TINV(1-a/2, df)*	crit	1.96001
margin of error *(ME)*	*critical value * standard error*	M E =	0.45500
P-value	*using T.Test, 2tailed, type 3(unequal variance))*	p-value =	0.94386
confidence level *(CL)*	1 - α	C L =	0.95
confidence interval Lower Bound	(μB - μA) - margin_of_error	LB =	-0.43865
confidence interval Upper Bound	(μB - μA) + margin_of_error	UB =	0.47135

Appendix 4

Novelty Visualizations (inconclusive)

Cumulative Conversions by Group (A/B)

Cumulative Total Spending by Group (A/B)

Distribution of Days to Convert (Group A)

Distribution of Days to Convert (Group B)
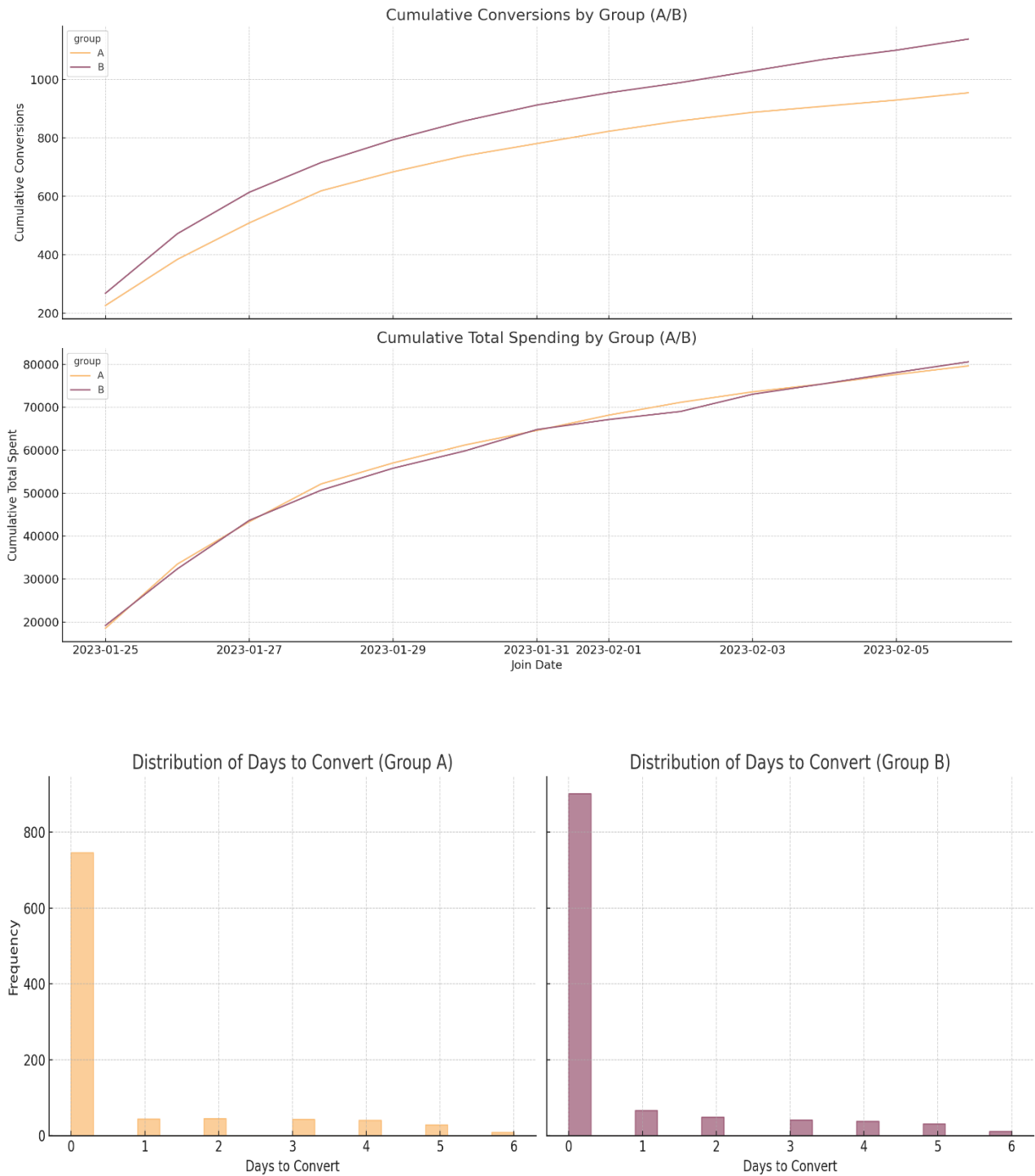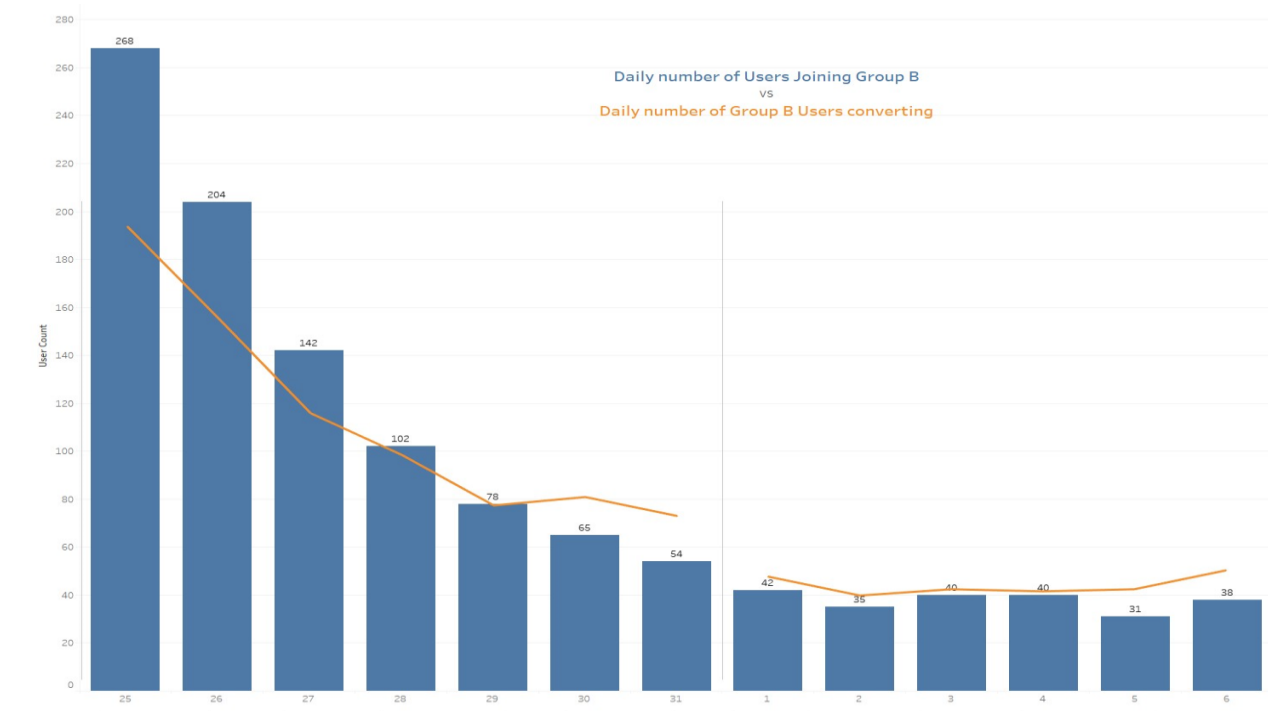
Based on the visualizations and analyses we've conducted, here are some conclusions about the potential novelty effect:

Cumulative Conversions and Spending: We didn't observe a distinct and sharp spike in either cumulative conversions or cumulative spending for group B, followed by a more gradual incline. Such a pattern might have indicated a novelty effect.

Change in Conversion Rate: The plot of the conversion rate over time did not show a pronounced temporary increase for group B, which might have been a sign of a novelty effect.

Spending per Conversion: Similarly, the spending per conversion over time did not reveal a significant temporary increase in spending for group B.

Days to Convert: The distribution of days to convert didn't show a clear difference between the two groups.

Given these observations, the visualizations do not provide strong evidence of a novelty effect at play in this dataset. However, we cannot correctly ascertain the novelty effect in this case due to several limitations in the available data

Appendix 5

Power Analysis

	Power Analysis for 2-sample Z-test		
pooled proportion ($p^$)	$(x\text{A}+x\text{B}) / (n\text{A}+n\text{B})$	$p^\textbf{pooled =}$	0.0428
complement of $p^$	$1 - p^$	$p^\textbf{ comp =}$	0.9572
Critical Value (95% confidence)	*Check table*	**crit =**	1.9600
Min. detectable effect	*assigned (high certainty)*	**Δ =**	0.1000
Z- value (90% power, 2tailed)	*Check table*	**zvalue =**	1.2800

Test sample	$nA + nB$		48943
Power Analysis	*2*p^pooled * p^comp * (crit+zvalue)^2 / (Δ*p^pooled)^2*		29422