Project Title:

"Evaluation of A/B Testing of the Control and Conversion Group of the Mobile Homepage of the Food and Drink Product Category of Glo Box Online Company."

Author:

Ejikeme Justine Ekwem

Master school

A/B Testing Mastery Project

December 2023

Project Summary:

The project aimed to evaluate the influence of a new banner feature (Group B) in comparison to the current setup (Group A) on crucial metrics, notably the conversion rate and average user spending. Following a thorough analysis, the findings and recommendations have been identified and presented.

Table of Contents:

1.0. Introduction.

GloBox, an e-commerce company, specializes in curating unique, high-quality products from around the world. Their philosophy is to transform shopping into an adventure, aiming to make the global market accessible to consumers through mobile phones.

1.1. Problem statement.

GloBox, renowned for its boutique fashion and high-end decor, has witnessed significant growth in its food and drink offerings. To boost awareness and revenue for this expanding category, the company's growth team proposes an A/B test. The test involves showcasing key products from the food and drink category as a banner at the top of the website. The control group does not view the banner, while the treatment group is exposed to it.

1.2. A/B Test Setup

The setup of the A/B test is as follows:
1. The experiment is only being run on the mobile website.
2. A user visits the GloBox main page and is randomly assigned to either the control or test group.
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. However, if they do make one or more purchases, this is considered a "conversion".
5. The project has a completion date of one month from the start date.

1.3. Objective and motivation of the use A/B test for the project.

A/B testing, also known as split testing, is a method used by businesses to compare two or more versions of a webpage or advertisement to determine which one performs better in terms of achieving specific goals. The objective of the company to use A/B testing to launch a website page on a mobile phone is to create a mobile-optimized user experience that leads to higher conversion rates and better overall performance. The motivation for the test is to use data-driven insights to make informed design and content decisions that result in a more effective and user-friendly mobile website.

2.0. Extraction of the Datasets

The experiment utilized datasets comprising 48,943 records extracted from Beekeeper Studio's relational database using SQL. Customers were randomly assigned to control or treatment groups, and data from the company's database was utilized. Beekeeper Studio, an open-source SQL editor, facilitated easy data extraction. Control and treatment group sample sizes (24,343 and 24,600, respectively) were obtained through SQL.

Additional variables, such as user conversion rates and average spending, were extracted from Beekeeper Studio. After database extraction, datasets were saved in CSV format for statistical analysis in a spreadsheet and Tableau. Test metrics were applied, and visualizations were

created to produce a performance-oriented dashboard on Tableau. The objective was to enhance understanding and facilitate a comprehensive overview of the study.

3.0. Analysis of the A/B test results using statistical methods

3.1. A hypothesis test to see whether there is a difference in the conversion rate between the two groups.

The hypotheses are defined as below:

- The null hypothesis (H0): There is no significant difference in the conversion rates between the control and treatment group:

- The alternative hypothesis (Ha): There is a significant difference in the conversion rates between the control and treatment groups:

The significance level (α) for the test is set as 0.05, which corresponds to a 5% significance level. The test is conducted for a difference in proportions as a two-sample z-test with pooled proportion with formular below:
z= P1-P2/sqrt(P(1-p) (1/n1 + 1/n2))

Where:

P1 and P2 are the sample proportions of conversions in the control and treatment groups, respectively.

N1 and N2 are the sample sizes of the treatment and control groups.

p^ is the overall pooled sample proportion. Which is calculated as:

$$\hat{p} = \frac{\hat{p}_1 * n_1 + \hat{p}_2 * n_2}{n_1 + n_2}$$

P1 =4.63, P2=3.92, N1=24600, N2=24343, P=4.28

z= PT-PC/sqrt(P(1-P) (1/NT + 1/NC))

4.63-3.92/sqrt (4.28(1-4.28) (1/24600 + 1/24343))

0.71/sqrt (-14.04) (8.172)

0.71/sqrt (-114.73)

0.71/-10.71

z = -0.06

The p-value for a z-score of -0.06 at a 5% significance level (95% confidence level) is greater than 0.05. In other words, the observed z-score of -0.06 is not extreme enough to be considered statistically significant at the 0.05 significance level.

Take a decision:

Compare the p-value to the significance level (α).

- If p-value < α, at 5%, reject the null hypothesis.

- If p-value ≥ α, at 5%, fail to reject the null hypothesis.

Conclusion:

Since the p-value is greater than 0.05, we fail to reject the null hypothesis, which indicates that there is not enough evidence to reject the null hypothesis. In other words, we do not have strong statistical evidence to conclude that there is a significant difference between the two groups.

3.2. The confidence interval for a difference in proportion is calculated using Two-sample z- interval with unpooled proportions with the formular below:

CI= (PT-PC) ± Z (sqrt (pt (1-PT)/NT + PC (1-PC)/NC)

Where:
The conversion rate in the treatment group (PT) = 4.63%
The conversion rate in the control group (PC)= 3.92%
The sample size of the treatment group (NT)= 24600
The sample size of the control group (NC) =24343
Z is the critical value from the standard normal distribution corresponding to a 95% confidence level. For a 95% confidence level, Z =1.96

Sample statistics = (4.63-3.92) = 0.71%
Critical value (Z), for a 95% confidence level is set at: 1.96.
Standard Error = sqrt ((4.63(1-4.63)/24600) + 3.92(1-3.92)/24343)) =0.0339.
Margin of Error = critical value *standard error
Margin of Error= 1.96 * 0.0339 = 0.0664

Substituting the formular:
4.63-3.92 ± 1.96 *sqrt ((4.63(1-4.63)/24600 + 3.92(1-3.92)/24343))
0.71 ± 1.96 * sqrt ((-16.8069/24600) + (-11.4464/24343))
0.71 ± 1.96 * sqrt (-0.0006832) + (-0.00047)
0.71 ± 1.96 * sqrt (0.00115) =0.0339
0.71 ± (1.96 * 0.0339)
0.71 ± 0.0664
CI = 0.71 ± 0.0664
CI=0.7764 (Upper limit) and 0.6436 (Lower limit)

Conclusion: We conclude that with 95% confidence, we estimate that the true difference in the parameter (conversion rates) between the two groups lies somewhere between 0.6436 and 0.7764.

3.3. A hypothesis test to see whether there is a difference in the average amount spent per user between the two groups.

Define the Hypotheses:

- Null Hypothesis (*H0*): There is no significant difference in the average amount spent per user between the two groups.
- Alternative Hypothesis (*Ha*): There is significant difference in the average amount spent per user between the two groups.

The significance level (α) for the test is set at 0.05, which corresponds to a 5% significance level. The test is conducted as a two-sample t-test for the difference in means with unpooled variance with the formular below:

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:
X1 as XT= 3.391 is the sample means of the amount spent in the treatment group
X2 as XC = 3.375 is the sample means of the amount spent in the control group
S1 as ST= 100.24 is sample standard deviations of the amount spent in the treatment group
S2 as SC= 100.24 is sample standard deviations of the amount spent in the control group
n1 as nc=24600 is the sample size of the treatment group
n2 as nc=24343 is the sample size of the control group

Substitute the formular:

T = 3.391 – 3.375 /sqrt ((94.03*94.03)/24600) + (100.24*100.24)/24343))
0.016/ sqrt ((8470/24600) +(10048/24343))
    0.016 / sqrt (0.3443+0.4127)
0.016 / sqrt (0.757)
0.016/0.87
T = 0.02

Degrees of Freedom:

Using the formular for DF: min (n1-1, n2-1)
DF = (24600-1) =24599
      (24343-1) =24342
The minimum value is = 24342.

Level of significance:

The level of significance is estimated at 5%, for a 95% confidence level and degree of freedom.

Calculate the P-Value:

Given the t-score of 0.02 and DF as 24342, a t-distribution online calculator was used for a two-tailed sample with a significance level of 5% which resulted to a p-value of 0.98. p-value > 0.05

0.98 >0.05.

Take a decision:

- If the p-value is <than the chosen level of significance of 0.05, we reject the null hypothesis.
- If the p-value is > than or equal to the chosen level of significance of 0.05, we fail to reject the null hypothesis.

Conclusion:
Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This suggest there is not enough evidence to conclude that there is a significant difference in the average amount spent per user between the groups. This also indicates a high probability that the observed result could occur by random chance.

3.4. The confidence interval for a difference in Means is calculated using Two-sample t- interval with unpooled variance with the formular below:

$$\text{Confidence Interval} = (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\text{df}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- X1 is the sample mean of the treatment group (average amount spent).
- X2 is the sample mean of the control group (average amount spent).
- S1 is the sample standard deviation of the treatment group.
- S2 is the sample standard deviation of the control group.
- n1 is the sample size of the treatment group.
- n2 is the sample size of the control group.
- α/2 is the significance level 0f 5% divided by 2 to account for the two tails for a 95% confidence interval.
- Df is the degrees of freedom for the distribution, which can be calculated using the formular for unequal variances:

CI= (3.391 − 3.375) ± $t\alpha$/2,df sqrt (94.03*94.03)/24600) + (100.24*100.24)/24343)

CI = 0.016 ± 1.96 sqrt (94.03*94.03)/24600) + (100.24*100.24)/24343
CI =0.016 ± 1.96 sqrt (8842/24600) + (10048/24343)
Sqrt (0.3594 +0.4128) = 0. 8787
CI =0.016 ± 1.9600 (0.8787)
CI = 0.016± 1.72
CI = -1.704(lower limit) and 1.736 (upper limit).

Conclusion:
In view of the result, we are 95% confident that the true difference in the average amount spent per user between the treatment and control groups falls between -1.704 and 1.736.

NOTE:

Critical value =t ($\alpha/2$, df)
  t (0.05/2, 24342) = (0.025,24342)
  However, due to the volume of the datasets, the df of 24342 cannot be accessed on the t-table and on the online calculator, hence a two-tailed 95% confidence interval was used at ±1.96 for a 95% confidence level.

Sample Statistics = Average amount spent (X1) - Average amount spent (X2)
  3.391 – 3.375 = 0.016

Standard Error Formula:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

SE = Sqrt (0.3594 +0.4128) = 0.88

4.0. Data Visualization/Chart.

Tableau was used to visualise the results of the datasets on a performance-based dashboard. The application was used because of its powerful data visualisation capabilities, which can help create interactive and insightful visualisations to better understand the test results, identify trends, and communicate findings to stakeholders effectively. In addition, its flexibility to connect to various data sources, including databases and spreadsheets, allows for easy import and integration of the data collected during A/B tests, making it more accessible for analysis. In the test, five visualisations and a dashboard were created on tableau application for statistical analysis, as captured below.

4.1. Conversion rate and average amount spent between the test groups.
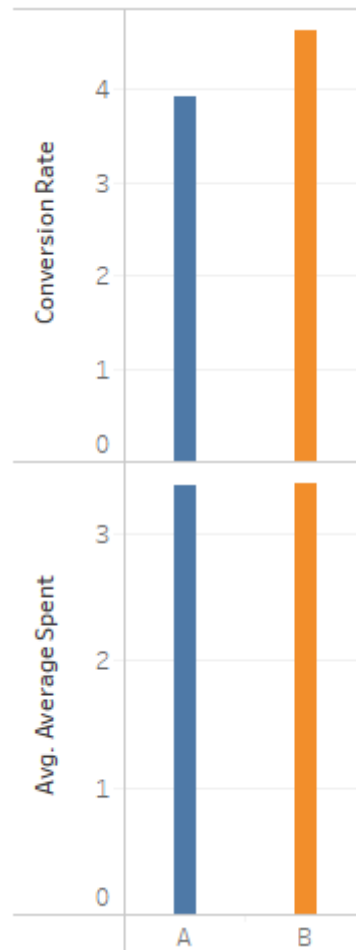

Fig. 4.1. Own Source: Tableau.

Figure 4.1 is a bar chart illustrating the conversion rate and average spending for the groups. The chart highlights that the treatment group spent an average of $3.391, while the control group spent $3.375, indicating a $0.01 difference. Additionally, the user conversion rate for the treatment group is 4.63%, compared to 3.92% for the control group, showing a difference of 0.71%.

4.2. Distribution of the amount spent per user for each group.
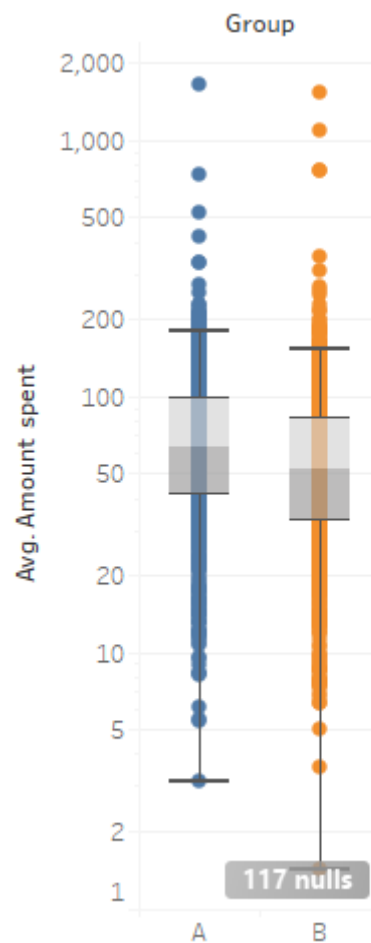

Fig. 4.2. Own Source: Tableau.

Figure 4.2 presents boxplots offering insights into the distribution of amount spent for both Group A and Group B. Notably, both groups exhibit similar median spending levels ($64 in Group A, $52 in Group B). However, it's important to highlight that Group A has a wider interquartile range (IQR) compared to Group B. The presence of outliers in both groups raises the potential for distortion and bias in the distribution results. Furthermore, it's crucial to acknowledge the impact of 117 filtered null values, which could have influenced the visualizations for both groups.

4.3. Relationship between the test metrics (conversion rate and average amount spent) and the user's device.
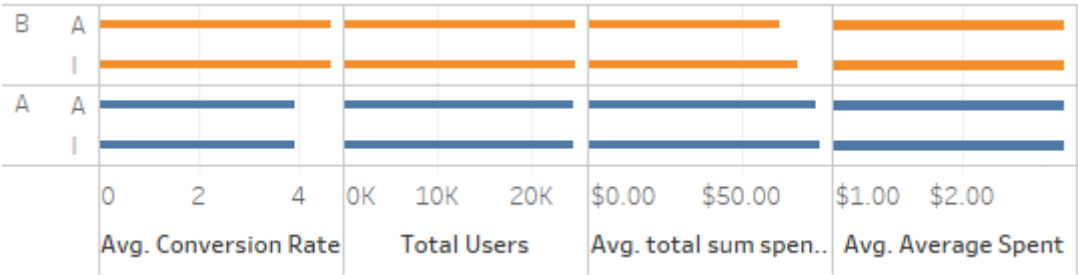

Fig. 4.3. Own Source: Tableau.

Figure 4.3 showcases a horizontal bar chart, providing a clear comparison of the datasets. Notably, the chart reveals that the highest average total spending in Group A is $75.71, particularly associated with the use of an iPhone. In contrast, Group B shows its highest average total spending at $68.22, also associated with iPhone usage. This visual representation offers a concise and impactful overview of spending patterns across different devices in both groups.

4.4. Relationship between the test metrics (conversion rate and average amount spent) and the user's gender.
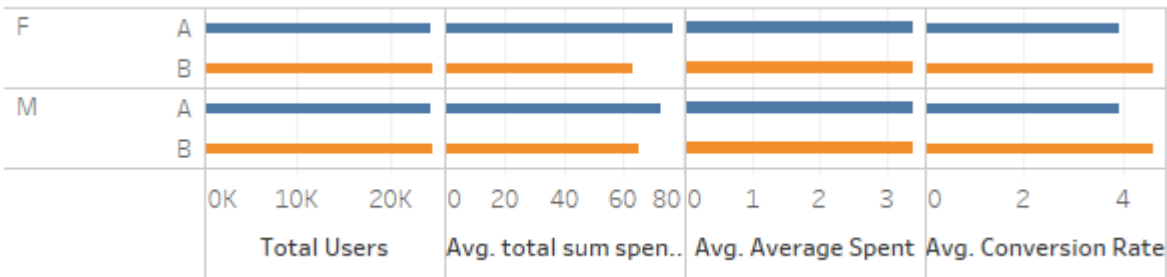

Fig. 4.4. Own Source: Tableau.

Figure 4.4 presents a visually appealing horizontal bar chart highlighting key data variables. Notably, the chart emphasizes that the highest average total spending occurred in Group A. Specifically, females in Group A had an average total spending of $76.75, while males in the same group had an average total spending of $72.82. This clear and concise visualization effectively conveys spending patterns based on gender in Group A.

4.5. Relationship between the test metrics (conversion rate and average amount spent) and the user's country.


Fig. 4.5. Own Source: Tableau.

Figure 4.5 presents a visually impactful horizontal column chart illustrating the highlighted data variable. Notably, the chart emphasizes that the highest average total sum spent was observed in Group A, specifically in Germany (DEU), amounting to $95.32. In contrast, for Group B, the highest average amount spent occurred in the United Kingdom (GBR) at $95.27. The chart effectively communicates the range of spending across different groups and regions, with the lowest amounts being $53.79 in Group B and $64.52 in Group A.

5.0. GloBox Dashboard System

Below is the dashboard visualization of the variables, providing a comprehensive overview of key metrics and trends. The dashboard system enhances accessibility and allows for a more interactive exploration of the experiment results.
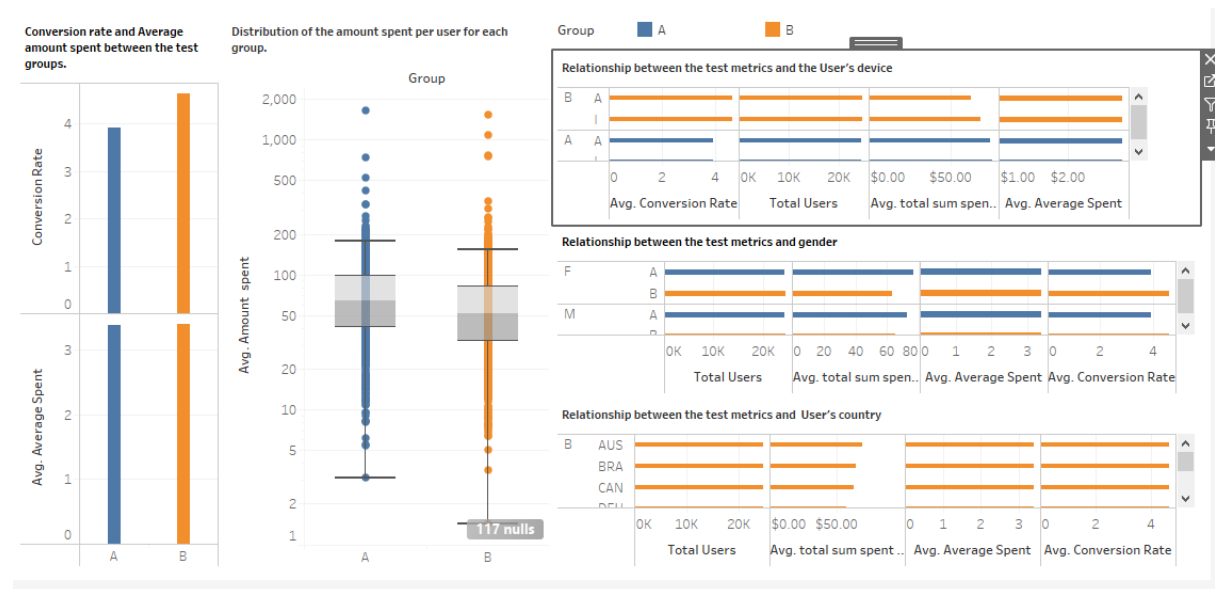
6.0. Findings

The statistical test results are presented below:

1.Conversion Rate: The results of our hypothesis testing indicate that there is not enough evidence to conclude a significant difference in the conversion rate between Group A and Group B. This suggests that the observed variations in conversion rates could likely occur by random chance.

2.Average Amount Spent: Similarly, our analysis did not reveal a statistically significant difference in the average amount spent per user between the two groups. The observed differences in spending may also be attributed to random chance.

3.Visualization Results: The visualization of the test metrics showed that Group A exhibited higher values compared to Group B in various aspects. This further supports the notion that the observed results could be a result of random variation.

7.0. Recommendations

Based on the findings and insights derived from the A/B test, we recommend against the launch of the new banner feature. Our analysis did not demonstrate a substantial improvement in our key success metrics. Therefore, it does not appear justified to release the feature to all users.

The cost-benefit analysis indicates that the perceived cost of launching the feature outweighs the potential benefits, especially considering the lack of significant improvements in conversion rates and average spending. We believe that further refinement or reconsideration of the feature may be necessary before proceeding with a full-scale launch.

8.0. Limitation of the Experiment

While visualizing the data, the presence of outliers in both Group A and Group B poses a risk of distorting and biasing the distribution results. Furthermore, the decision to filter out and ignore nulls in the datasets could potentially impact our analysis negatively. This approach not only introduces changes but also raises concerns about data loss and the potential for bias in our findings. Acknowledging and addressing these limitations is crucial for a comprehensive understanding of the experiment results.

9.0. Conclusion

In conclusion, while the A/B testing experiment provided valuable insights, the data did not support the implementation of the new banner feature at this time. We suggest revisiting the feature design and conducting additional tests to explore alternative strategies for achieving our desired outcomes.

10.0. References

Control group (Group A)

```
SELECT *
FROM users as u
JOIN groups as a
ON u.id = a.uid
WHERE a.group = 'A';
```

Treatment group (Group B)

```
SELECT *
FROM users as u
JOIN groups as b
ON u.id = b.uid
WHERE b.group = 'B';
```

User conversion rate for the control and treatment groups

```
SELECT "group", COUNT (DISTINCT u.id) As total users,
Round ((Count (DISTINCT a.uid) *100.0)/
Count (DISTINCT u.id),2) As conversion_rate
FROM users as u
LEFT JOIN groups as g
ON u.id=g.uid
LEFT JOIN activity as a
ON u.id=a.uid
GROUP BY "group";
```

Conversion rate of all users

```
SELECT
COUNT (DISTINCT a.uid) AS converted_users,
COUNT (DISTINCT u.id) AS total_users,
(COUNT (DISTINCT a.uid) * 100.0 / COUNT (DISTINCT u.id)) AS conversion_rate
FROM users as u
LEFT JOIN activity as a
ON u.id = a.uid
ORDER BY conversion_rate;
```

Average amount spent per user for the control and treatment groups, including users who did not convert.

```
SELECT "group",
COALESCE(SUM(a.spent),0)As total_spent,
Count(u.id) As total_transactions,
Round (COALESCE (SUM(a.spent), 0)/COUNT(DISTINCT u.id),3) As average_spent
```

```
FROM users as u
LEFT JOIN groups as g
ON u.id=g.uid
LEFT JOIN activity as a
ON u.id=a.uid
GROUP BY "group";
```

Aggregated database

```
SELECT u.id,sum(a.spent)as total_sum_spent,u.country,u.gender,g.device,g.group
FROM users as u
JOIN activity as a
ON u.id=a.uid
JOIN groups as g
ON a.uid = g.uid
WHERE spent >'0'
GROUP BY u.id,u.country,u.gender,g.device,g.group
ORDER BY total_sum_spent;
```

11.0. Appendix

Tableau charts

Q.1.https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_169463089206
10/conversionrateandaverageamountspentbetweenthetestgroups_?publish=yes

Q.2.https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_169463089206
10/Distributionoftheamountspentperuserforeachgroup?publish=yes

Q3.https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_1694630892061
0/RelationshipbetweenthetestmetricsandtheUsersdevice?publish=yes

Q4.https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_1694630892061
0/RelationshipbetweenthetestmetricsandtheUsersgender?publish=yes

Q5.https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_1694630892061
0/RelationshipbetweenthetestmetricsandtheUserscountry?publish=yes

Dashboard

https://public.tableau.com/app/profile/ejikeme.justine.ekwem/viz/Book2_16946308920610/D
ashboard1?publish=yes

https://www.omnicalculator.com/statistics/critical-value#t-critical-values