

# GLOBOX A/B TESTING PROJECT

**By: Ninya Narayanan**

**Sep 12, 2023**

**SUMMARY**

The A/B test revealed a significant divergence in the conversion rates between the control and treatment groups, favoring the treatment group. However, there was no notable difference in the average amount spent per user across both groups. Consequently, our recommendation is to withhold the launch of the new feature to all users. Further examination is necessary to gauge its impact on revenue, empowering stakeholders to make informed decisions regarding its broader implementation.

**CONTEXT**

**Project Background**

GloBox, renowned for its boutique fashion and premium decor, has recently expanded its food and drink offerings. Seeking to bolster revenue from this category, the Growth team initiated an A/B test. In this test, a banner showcasing select food and drink items is displayed to the test group at the website's top, while the control group does not encounter this banner. The aim is to enhance awareness and engagement with these new offerings.

**The setup of the A/B test is as follows**:

* The experiment is only being run on the mobile website.
* A user visits the GloBox main page and is randomly assigned to either the control or test group. This is the join date for the user.
* 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.
* 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. If they do make one or more purchases, this is considered a “conversion”.

**Task**

Analyze the results of the A/B test and provide a recommendation to the stakeholders about whether Globox should launch the experience to all users.

**Stakeholders**

Growth Product & Engineering Team: This is the team that you work with at GloBox. The team is made up of a product manager, a user experience designer, an engineering manager, and several software engineers, and you, the data analyst. The team develops features for the GloBox website that drive growth in users and revenue.

Leila Al-Farsi, Product Manager, Growth: Leila is the product manager for the Growth product and engineering team. Alongside Alejandro, she leads the Growth team by deciding their goals and projects, measuring their success against defined KPIs, and communicating results to other company leaders like Mei.

Alejandro Gonzalez, User Experience Designer, Growth: Alejandro is the designer for the Growth product and engineering team. He conducts user research and designs the experience that the A/B test is evaluating.

Mei Kim, Head of Marketing: Mei oversees the Marketing team, which works on targeting audiences with effective marketing campaigns to drive customers to the GloBox website. She collaborates frequently with Leila and Alejandro to design website experiences that will align well with the current marketing efforts.

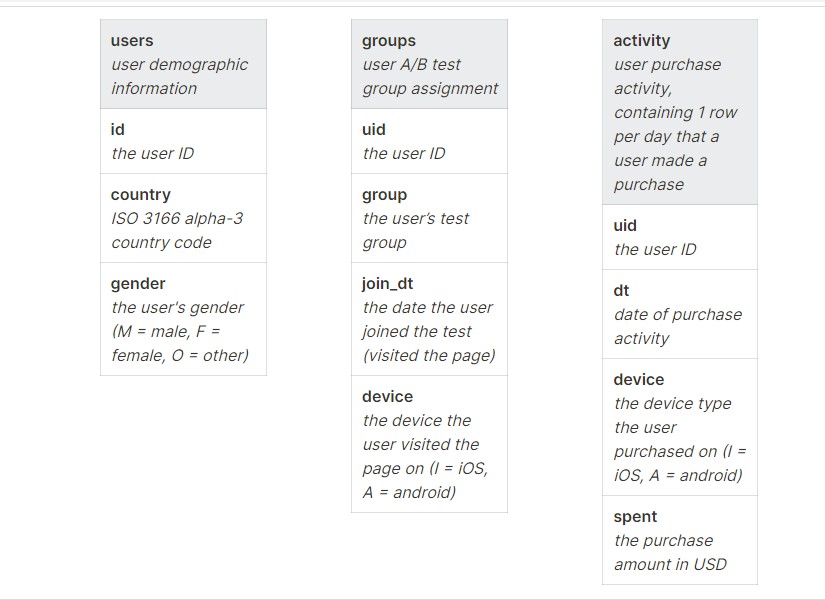
**The Dataset**

Globox stores its data in a relational database, which was extracted for this test using SQL queries.

To consider:

All users should be assigned to one A/B test group.

Not all users make a purchase.



Together, Leila, Alejandro, and Mei will decide whether or not to launch the experiment based on the results.

# Metric

* Revenue
* Conversion Rate = #converted / #converted + # notconverted

## Statistical Hypothesis

* Ho : CRcon = CRexp
* Ha : CRcon ≠ CRexp
* Ho : μcon = μexp
* Ha: μcon ≠ μexp

## Power Analysis

In this instance, power analysis was already conducted and the test data was readily available and extracted from the database. Nonetheless, I have established the confidence level and the significance level for the study.

* **Confidence Level**: 95%
* **Significance Level**: 5% or 0.05

## Minimum Detectable Effect

The targeted objective is to discern the magnitude of change between the new and old versions, as a basis for deciding whether to launch the new feature. Although the stakeholders did not provide a specific percentage, we have set a benchmark of a 2% change in conversion rate and revenue for this analysis, which will be assessed based on the average amount spent per user.

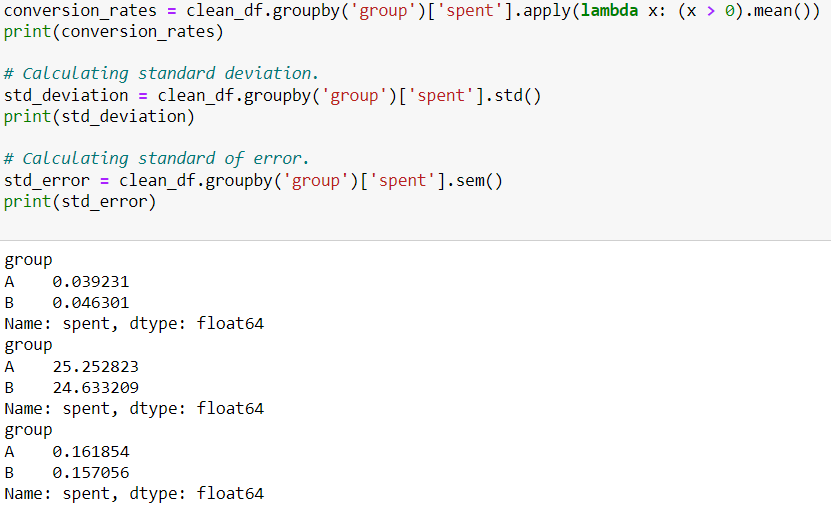
# Analyzing A/B test results

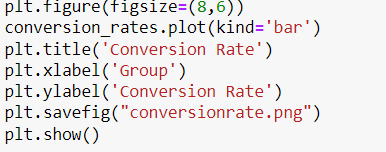


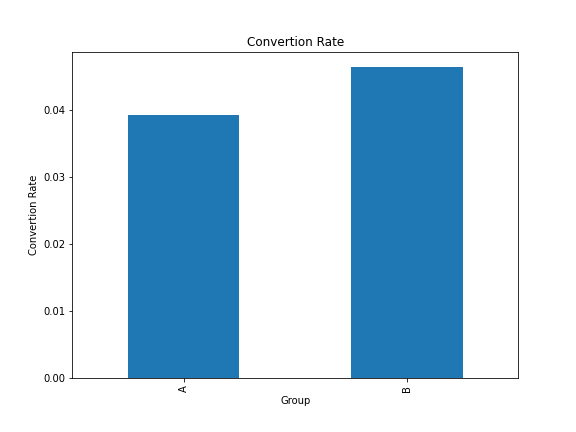
## After identifying 139 duplicates in the dataset, it was determined that the impact of their removal would not significantly affect the results. Therefore, I proceed with deleting the duplicates from the dataset.

Approximately 139 duplicates were found and removed, reducing the dataset from 49,082 to 48,943 entries. As these duplicates didn't significantly impact the analysis, their removal aimed to ensure a more accurate count of unique converting users. The dataset, after being read into a Pandas data frame, underwent thorough checks for

# Conversion Rate







From the graph above, we can observe that the treatment group or “B” performs slightly better with approximately 4.63% conversion rate than control group or “A” with 3.92% conversion rate. However, we cannot draw conclusions yet without identifying if the results or difference is statistically significant.

In order to test for a statistically significant difference between control and treatment group, we used a two-sided z-test for proportion, since is ideal to determinate a statistical significance positive or negative results, or if there is not statistically significant between the proportions.

**Hypothesis**

The established hypotheses for the conversion rate and differences in proportions between the treatment and control groups are as follows:

Conversion Rate Hypotheses:

Null Hypothesis (H0): There is no difference in conversion rates between the treatment and control groups.

Alternative Hypothesis (H1): There exists a difference in conversion rates between the treatment and control groups.

Differences in Proportions Hypotheses:

Null Hypothesis (H0): The conversion rate for the treatment group minus the control group's rate equals zero.

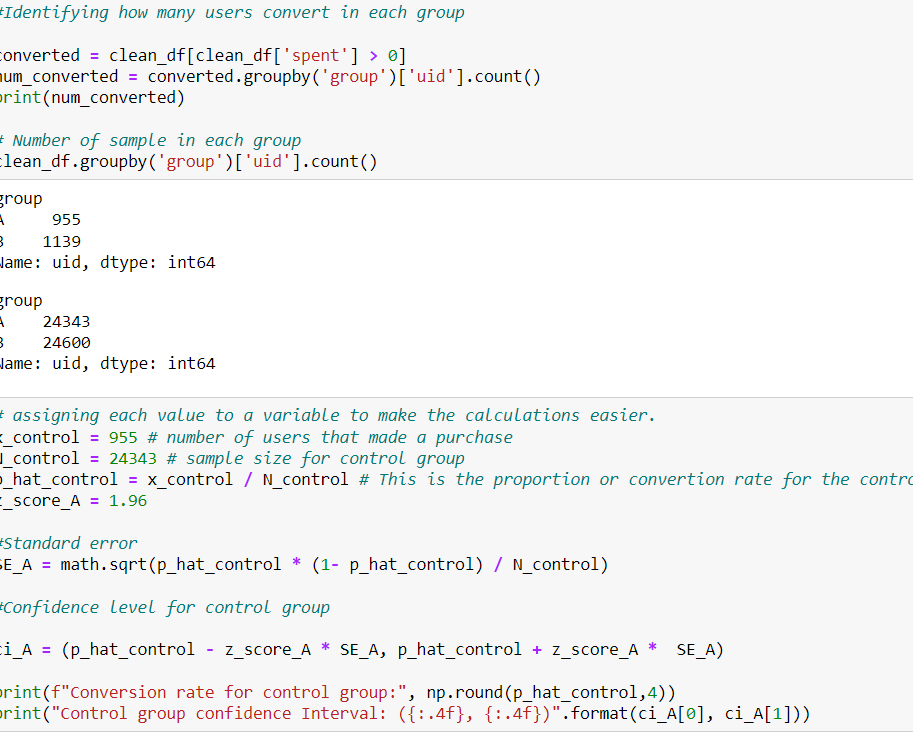
Alternative Hypothesis (H1): The difference in conversion rates between the treatment and control groups is not zero.

Ho : CRcon = CRexp

Ha : CRcon ≠ CRexp

Ho: CRexp - CRcon= 0

Ha : CRexp- CRcon ≠ 0



Confidence intervals were computed for both control and treatment groups. Control group's conversion rate: 3.92% (CI: 0.0368 - 0.0417), and treatment group's rate: 4.63% (CI: 0.0437 - 0.0489). Their values within these intervals suggest the sample's representation of the actual population.

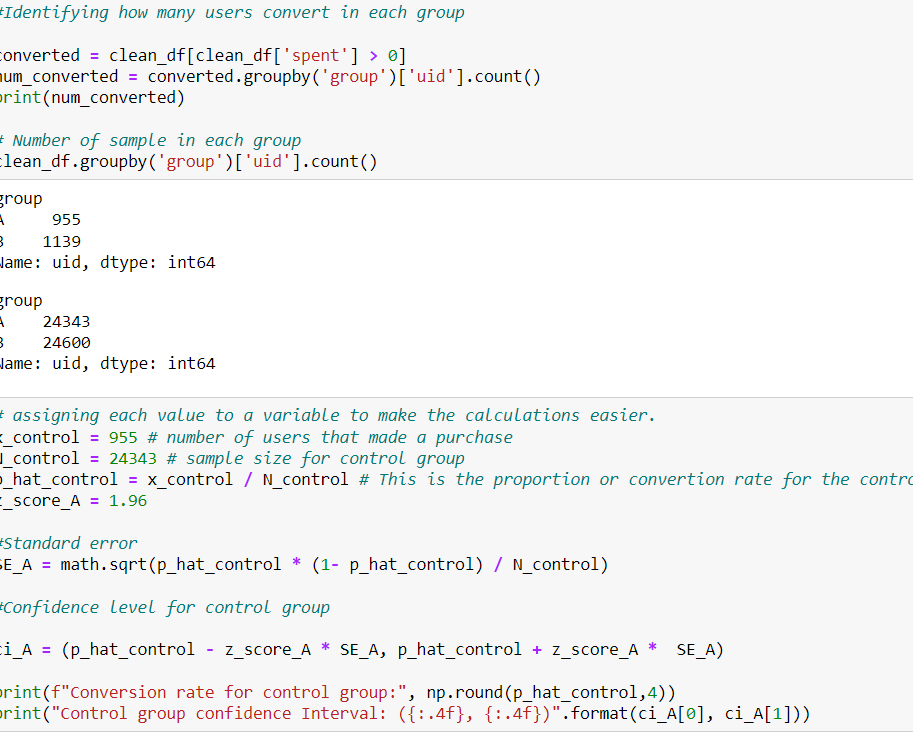
Top of Form

.

The analysis revealed a p-value of 0.001, surpassing our set significance level of 0.005. As a result, we rejected the null hypothesis, implying a significant distinction in proportions between the control and treatment groups. Additionally, the confidence interval for the difference in proportions, (0.0035, 0.0107), excluding zero, further supports this conclusion of a significant difference between the groups.

After analyzing the conversion rate results and rejecting the null hypothesis, we proceeded to examine the revenue, or the average amount spent per user to determine the statistical significance of launching the new feature as part of our A/B testing objectives.

**Revenue confidence Interval**



We conducted a hypothesis test and confidence interval analysis for the average revenue per user using the original dataset, inclusive of duplicates to encompass all user purchases during the test. The treatment group showed a slightly higher average revenue of 3.391 compared to the control group's 3.375. However, without executing the hypothesis test and confidence interval, no definitive conclusions could be drawn from these observations.

Hypothesis

Ho : μcon = μexp

Ha : μcon ≠ μexp

Ho: μexp - μcon = 0

Ha : μexp - μcon≠ 0

In our analysis of the average amount spent per user, the control group's sample mean (3.37) was slightly lower than the treatment group's. The control group's confidence interval (3.048, 3.685) suggested no statistically significant difference, as our sample mean lay outside this interval. Conversely, the treatment group's interval (3.071, 3.689) included our sample mean (3.380), indicating potential significance. To evaluate the mean difference's statistical significance, we computed the confidence interval and its corresponding p-value.

The calculated p-value (0.952) surpassed the set significance level (0.05), indicating inadequate evidence to dismiss the null hypothesis. Hence, we couldn't reject the hypothesis that the mean difference between the groups equals the true population mean. This implies that observed mean differences might be due to chance, not a significant distinction between the groups. Importantly, a failure to reject the null hypothesis doesn't confirm its truth; it suggests insufficient evidence to support the alternative hypothesis.

# Calculating the confidence interval for the difference of mean using unequal variance and t-distribution

**CONCLUSION AND RECOMMENDATION**

The A/B test showed a significant boost in the treatment group's conversion rate (4.63%) compared to the control group (3.92%). Although meeting the 2% conversion rate goal, the new feature's effect on revenue remains uncertain due to no significant variance in the average amount spent per useTop of Form

**Recommendation**

Further analysis is recommended to comprehend the new feature's influence on revenue, involving an in-depth study of customer behavior and engagement. These insights will enable stakeholders to make informed decisions regarding the potential launch of the experience to all users.