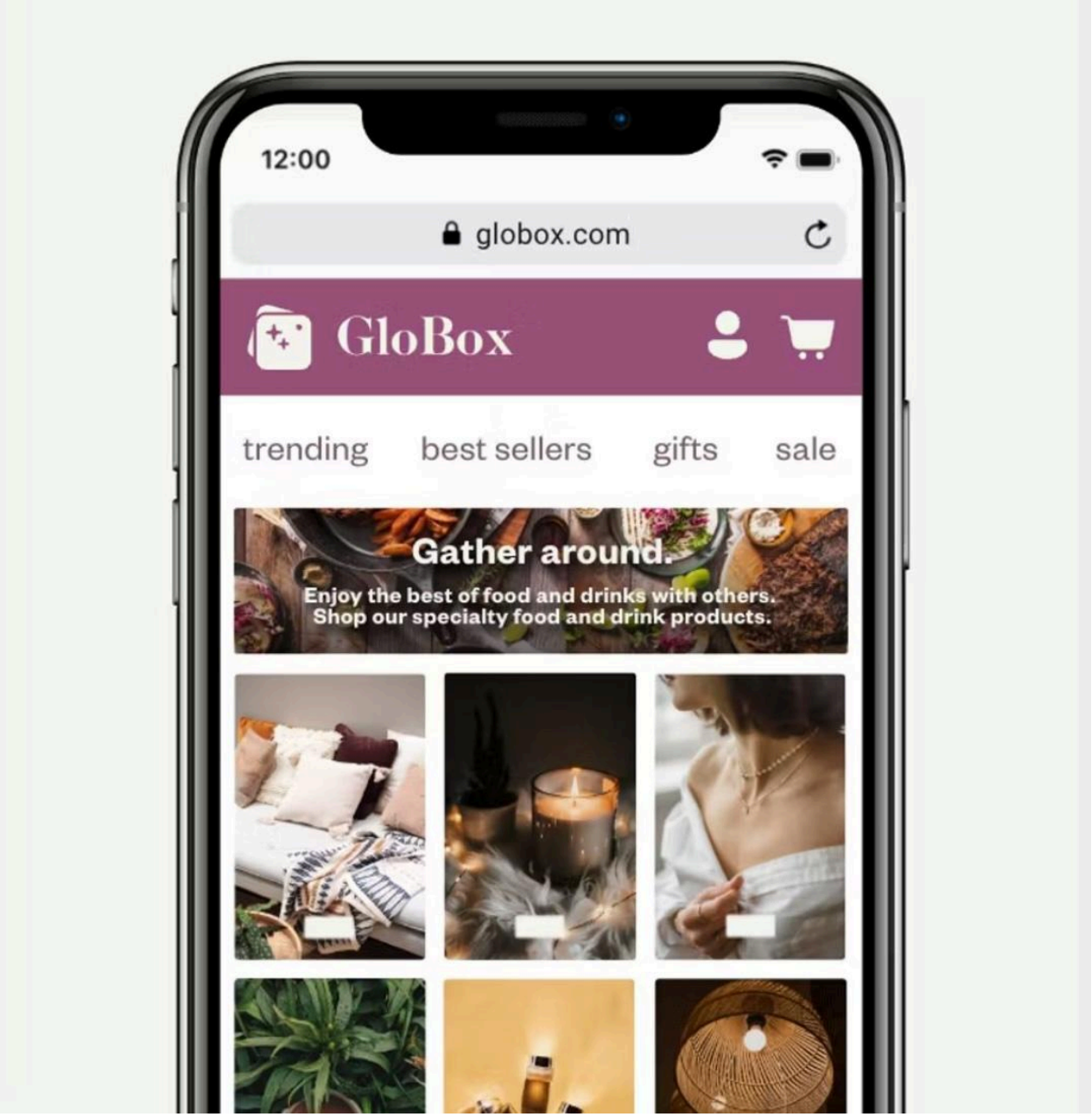
Analysis on the Impact of Banner Implementation on User Engagement and Spending



Introduction

GloBox recently conducted an A/B test to evaluate the effectiveness of a new banner promoting the food and drink category on its website. The goal was to determine whether the banner influenced user spending and conversion rates. This report presents the findings of the test, providing insights into user behavior and offering recommendations for future marketing strategies.

Methodology

Data Aggregation:

- The initial data extraction and aggregation were performed using SQL. This process involved querying multiple tables to retrieve relevant user information, including group assignment, spending behavior, and conversion events.
- Specific SQL queries were utilized to join user data with transaction data, ensuring a comprehensive dataset that included user group assignments, spending amounts, and conversion indicators.
- Data cleaning steps were applied during the SQL aggregation phase, where null values in the spending column were treated appropriately to maintain data integrity.

Statistical Analysis:

- After data aggregation, the dataset was imported into a Python environment for detailed statistical analysis.
- A T-test was conducted to compare the average spending between Group A (control) and Group B (test), providing insights into the banner's impact on spending behavior.

- A Z-test for two proportions was used to assess the difference in conversion rates between the two groups, determining the effectiveness of the banner in influencing user purchases.

A/B Test Design:

- Users were randomly assigned to Group A or Group B to ensure the test's validity.
- Group A users did not see the banner, while Group B users were exposed to it, allowing for a direct comparison of user behavior between the two groups.

Results

1. User Distribution

- Total users: 48,943
- Users in Group A: 24,343
- Users in Group B: 24,600

2. Spending Analysis

- The T-test did not reveal a statistically significant difference in average spending between Group A and Group B.
- Mean spending in Group A: $3.37
- Mean spending in Group B: $3.39

3. Conversion Rate Analysis

- Overall conversion rate: 4.28%
- Conversion rate for Group A: 3.92%
- Conversion rate for Group B: 4.63%
- The Z-test indicated a statistically significant difference in conversion rates, with Group B showing higher conversions.

Discussion

The analysis suggests that while the banner did not significantly impact average user spending, it did have a positive effect on the conversion rate. This indicates that the banner may have influenced more users in Group B to make a purchase, even if it didn't affect the amount spent per user.

Conclusions

- The banner's introduction is associated with a higher conversion rate but not with increased spending per user.
- These findings suggest that the banner effectively attracts user attention and encourages purchasing behavior, even though it doesn't lead to higher spending.

Recommendations

1. Implement the Banner: Consider adopting the banner across the site, given its positive impact on conversion rates.
2. Segmentation Analysis Perform deeper analysis on user segments to identify specific groups that may be more influenced by the banner.
3. Monitor Key Metrics: Continue monitoring spending and conversion rates to ensure the banner maintains its effectiveness over time.

Appendix

- Detailed statistical analysis output
- Data collection methodology
- Graphs and charts illustrating the findings

Appendix A: Detailed Statistical Analysis Output

1. T-Test for Comparing Means between Two Groups (Group A and Group B Spending)

- Objective: To determine if there is a statistically significant difference in the average spending between users who saw the banner (Group B) and those who did not (Group A).

- Test Statistics:
    - T-statistic: -0.0597
    - P-value: 0.9524
    - Degrees of Freedom: 48941 (calculated based on sample size minus two since we're comparing two independent samples)

- Interpretation: The high p-value (0.9524) indicates that we fail to reject the null hypothesis, suggesting no significant difference in average spending between the two groups.

2. Z-Test for Comparing Proportions (Conversion Rates between Group A and Group B)

- Objective: To assess whether the observed difference in conversion rates between the two groups is statistically significant.

- Test Statistics:
    - Z-score: -3.8643
    - P-value: 0.0001114

- Interpretation: The low p-value suggests that the difference in conversion rates between Group A and Group B is statistically significant. We reject the null hypothesis, indicating that the banner had a significant impact on conversion rates.

Appendix B: SQL Query for Data Aggregation and Conversion Calculation

This appendix details the SQL query used to extract and prepare the data for the A/B test analysis. The query retrieves user demographic information, group assignments, and spending activity, which are essential for evaluating the impact of the banner on user behavior.

```
SELECT
   u.id,
   u.country,
   u.gender,
   g.group,
   g.join_dt AS join_date,
   g.device,
   a.dt AS spent_date,
   a.spent
FROM
   users u
   INNER JOIN groups g ON u.id = g.uid
   LEFT JOIN activity a ON u.id = a.uid
ORDER BY
   a.dt, u.id;
```

Explanation of the Query

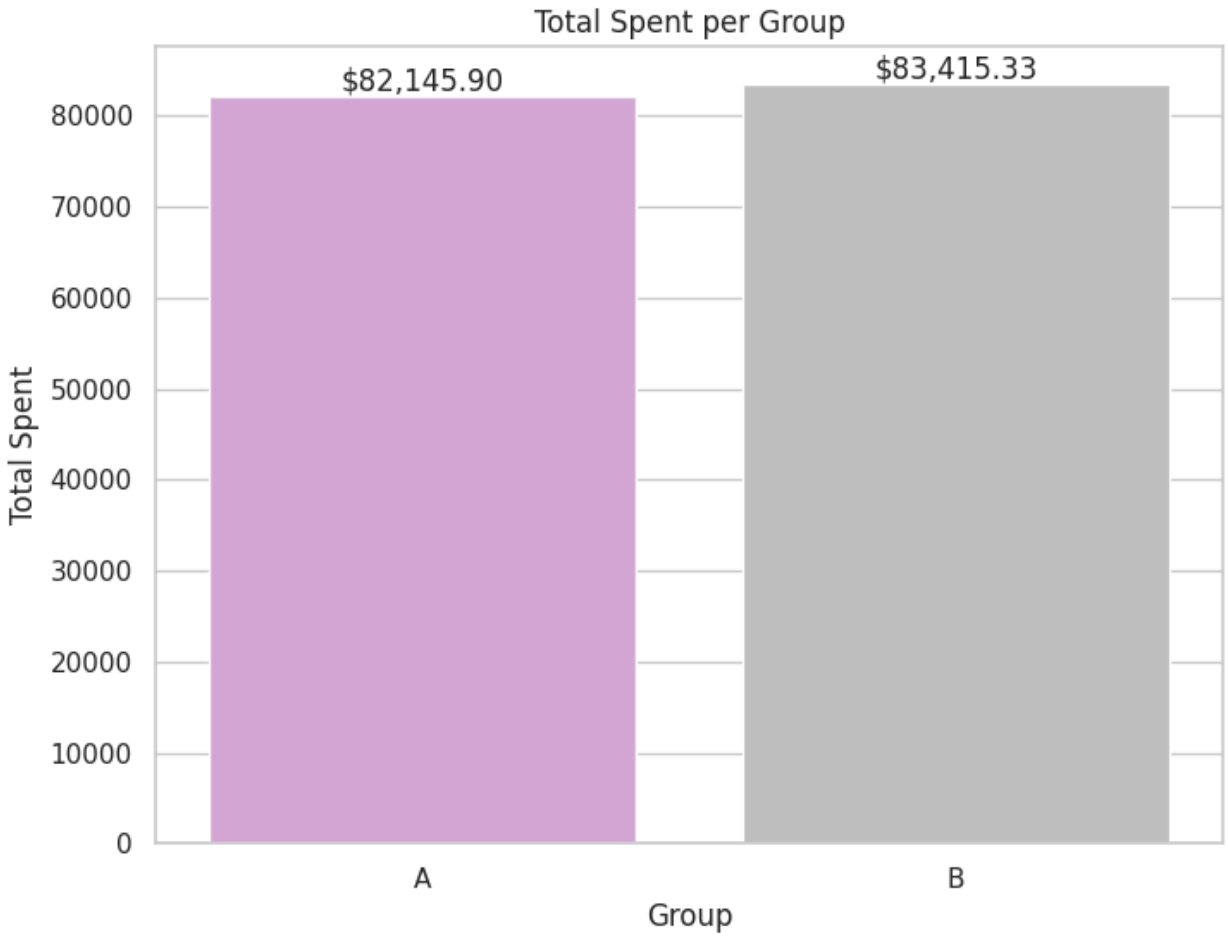
- FROM Clause: The query combines data from three tables—users, groups, and activity. The users table contains demographic data about the users. The groups table has information on which group (control or test) the users belong to, along with their joining date and the device they use. The activity table tracks user spending activity.
- JOIN Clauses: An INNER JOIN is used to ensure that only users present in both users and groups tables are included. A LEFT JOIN with the activity table is used to include all user spending activity, which ensures that even if users did not spend anything (no matching record in activity), they are still included in the results with NULL values for spending and spending date.
- SELECT Clause: Specific columns are selected from each table to provide a complete dataset that includes user identification, demographics, group assignment, and spending behavior. The AS keyword is used to rename some columns for better clarity in the resulting dataset.
- ORDER BY Clause: The result set is ordered first by the spending date (a.dt) to organize the data chronologically, and then by user ID (u.id) to maintain a consistent order within each date.

The intention of this query is to organize and retrieve a dataset that allows for a detailed analysis of how group assignment (control group vs. test group) may influence user behavior in terms of engagement and spending on the GloBox platform.

Appendix C: Graphs and Charts

This appendix provides further details and descriptions of the graphs and charts used in the analysis report. Each graph/chart is accompanied by a brief description and insight that it aims to convey.

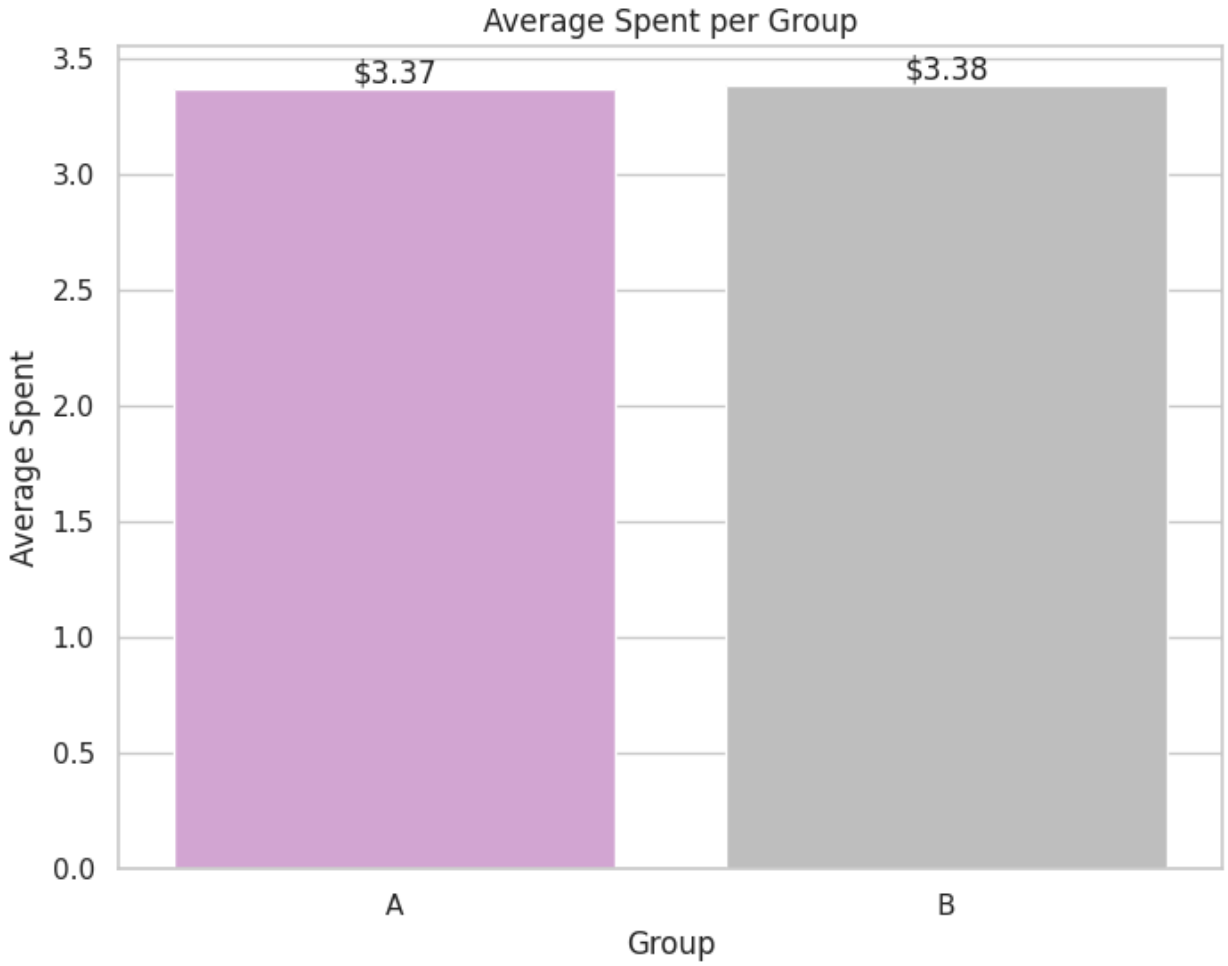
Total Spent per Group:



- Bar chart comparing the total spending between Group A and Group B.

- Illustrates the overall impact of the banner on total spending for each group as nearly identical, though Group B is slightly higher.

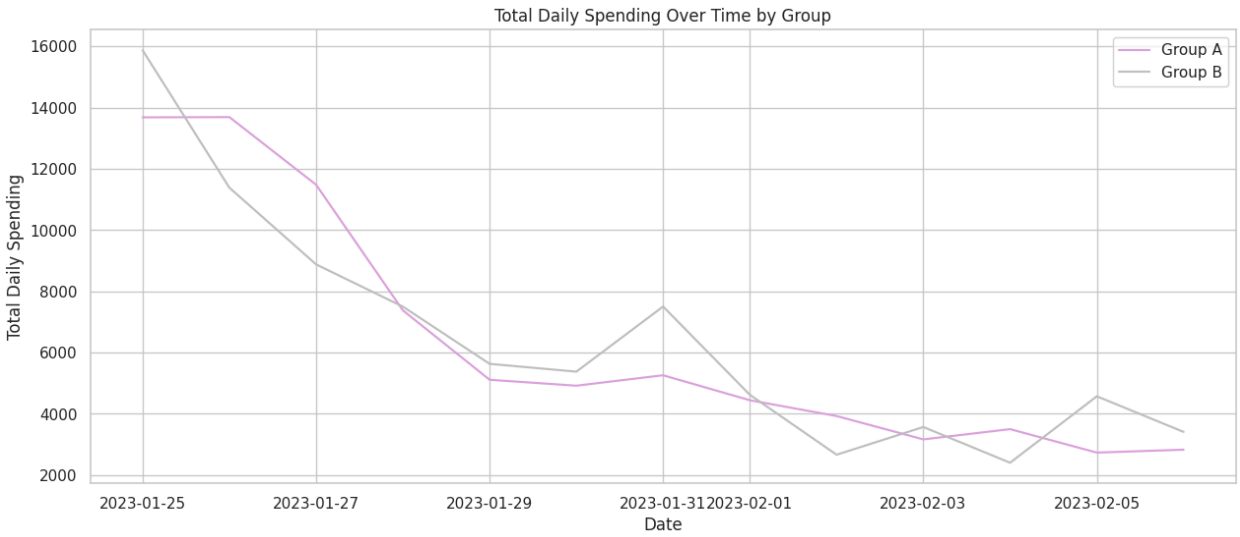
Average Spent per Group:



- Bar chart showing the average amount spent per user in each group.

- Provides a comparison of the spending behavior per user between the groups, highlighting the direct influence of the banner on individual spending as nearly identical.

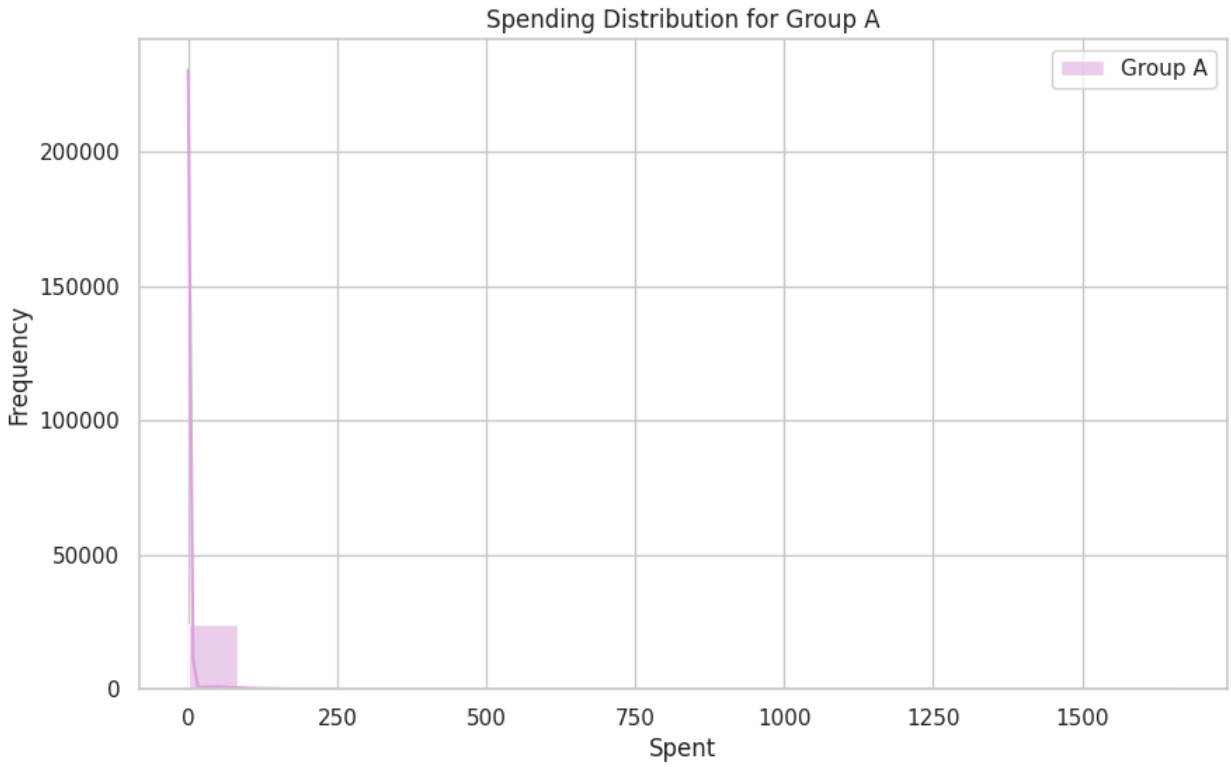
Total Daily Spending Over Time by Group:



- Both groups started with a certain level of spending on the first day, with Group A beginning slightly higher than Group B. Immediately on the following day, there's a sharp drop in spending for both groups, which is typical as initial interest or activity may wane after a new feature's launch or the start of a campaign.
- Over the subsequent days, there are fluctuations in the spending patterns of both groups. Such fluctuations could be influenced by various factors including weekday versus weekend behavior.

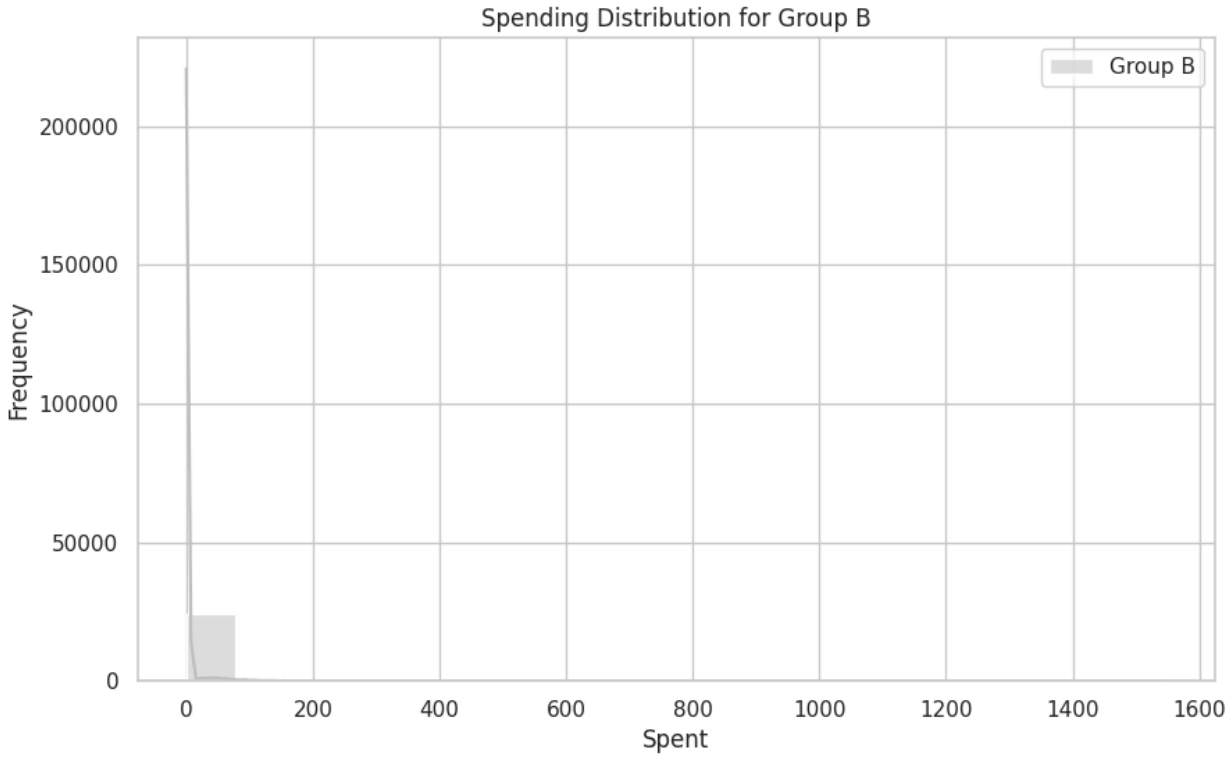
- Around 2023-01-31, Group B shows a significant peak in spending that is not mirrored in Group A. This could suggest a particular response to the banner by Group B on that day, or it could be the result of an external factor not controlled for in the test.
- After the peak, spending in Group B drops back down and the trends for both groups begin to converge, suggesting that the initial impact observed may be leveling off.
- Throughout the period, Group B generally spends more than Group A, except on the first and last days shown. This could suggest that the presence of the banner in Group B had a positive impact on spending, though not consistently day-to-day.

Spending Distribution for Group A:



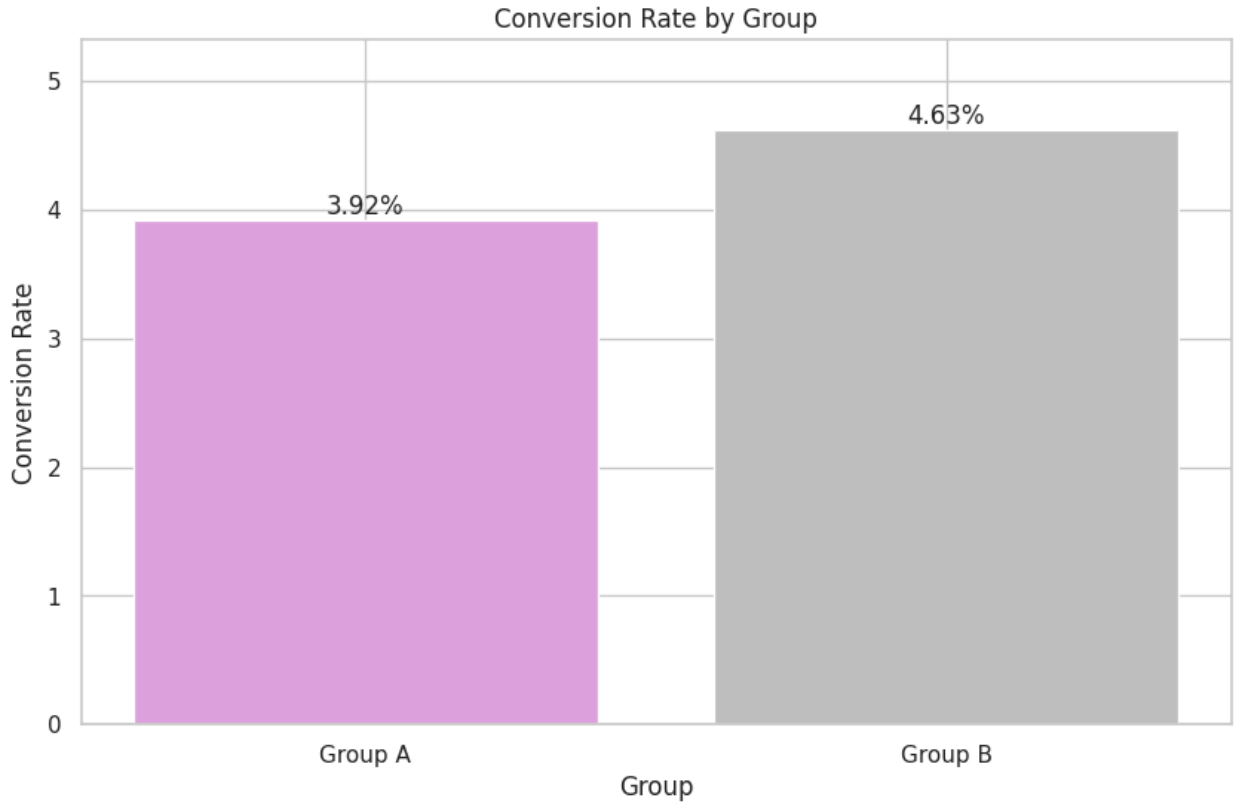
- The distribution is highly right-skewed, meaning there are a few instances of high spending that are not typical of the overall spending behavior.
- Most of the spending amounts are low, with a frequency that quickly tapers off as the amount spent increases.
- The absence of bars as the amount spent increases suggests there might be outliers or a long tail—just a few instances of high spending.

Spending Distribution for Group B:



- The single bar at the beginning indicates that almost all spending occurrences fall into the lowest spending bracket shown.
- The distribution would be considered right-skewed, which is typical for spending data, as there are usually more small transactions than large ones.
- Similar to Group A, there's a rapid drop-off in frequency as the spending amount increases, indicating few high spending amounts.

Conversion Rate by Group:



- Group A has a conversion rate of 3.92%
- Group B has a higher conversion rate of 4.63%
- Group B is converting at a higher percentage than Group A by ~0.71%. The banner appears to be more effective, leading to more users converting to sales.

++ The figures referenced above are intended to provide a visual summary and comparison of key metrics analyzed in the report. The color coding has been chosen to maintain consistency and aid in the quick identification of each group's data.