

results

November 29, 2024

1 Data Analysis Results

1.0.1 Metadata

After cleaning the three raw datasets provided by Vanguard and merging them, we have the following data to begin with.

Rows:

- `client_id (int)`: Every client's unique ID.
- `visitor_id`: A unique ID for each client-device combination.
- `visit_id`: A unique ID for each web visit/session.
- `step`: Marks each step in the digital process.
- `date_time`: Timestamp of each web activity.
- `tenure_years (int)`: Represents how long the client has been with Vanguard, measured in years.
- `tenure_months (int)`: Further breaks down the client's tenure with Vanguard in months.
- `age (int)`: Indicates the age of the client.
- `gender`: Specifies the client's gender (four unique values: 'Female', 'Male', 'Other' and 'Unspecified').
- `accounts (int)`: Denotes the number of accounts the client holds with Vanguard.
- `balance (float)`: Gives the total balance spread across all accounts for a particular client.
- `calls_last_6_months (int)`: Records the number of times the client reached out over a call in the past six months.
- `logons_last_6_months (int)`: Reflects the frequency with which the client logged onto Vanguard's platform over the last six months.
- `variation (object)`: Indicates if a client was part of the experiment (three unique values: 'Test', 'Contro' and 'Unknown').

2 EDA (Exploratory Data Analysis)

Initial examination of the dataset to summarize the main characteristics in order to gain insights into the data, uncover patterns, spot anomalies and check assumptions before applying further statistical methods.

2.0.1 Client behavior analysis

Answering the following questions about demographics:

- Who are the primary clients using this online process?

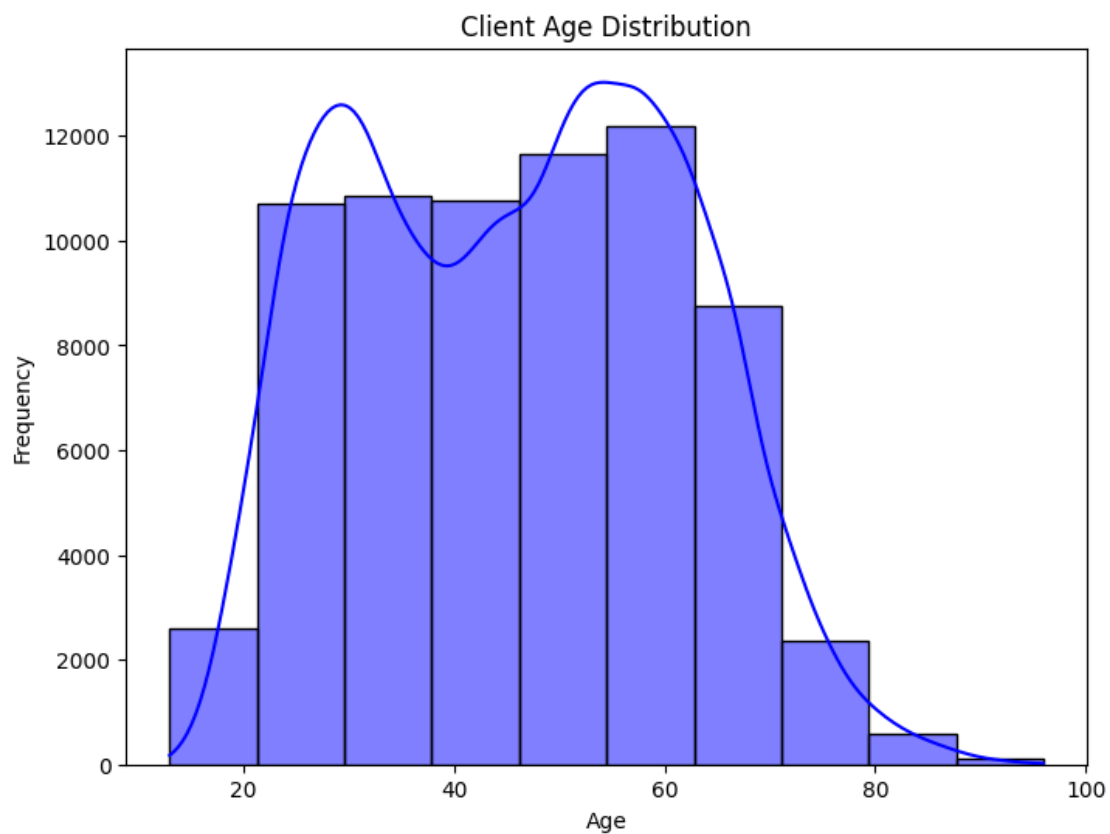
- Are the primary clients younger or older, new or long-standing?
- Next, carry out a client behaviour analysis to answer any additional relevant questions you think are important.

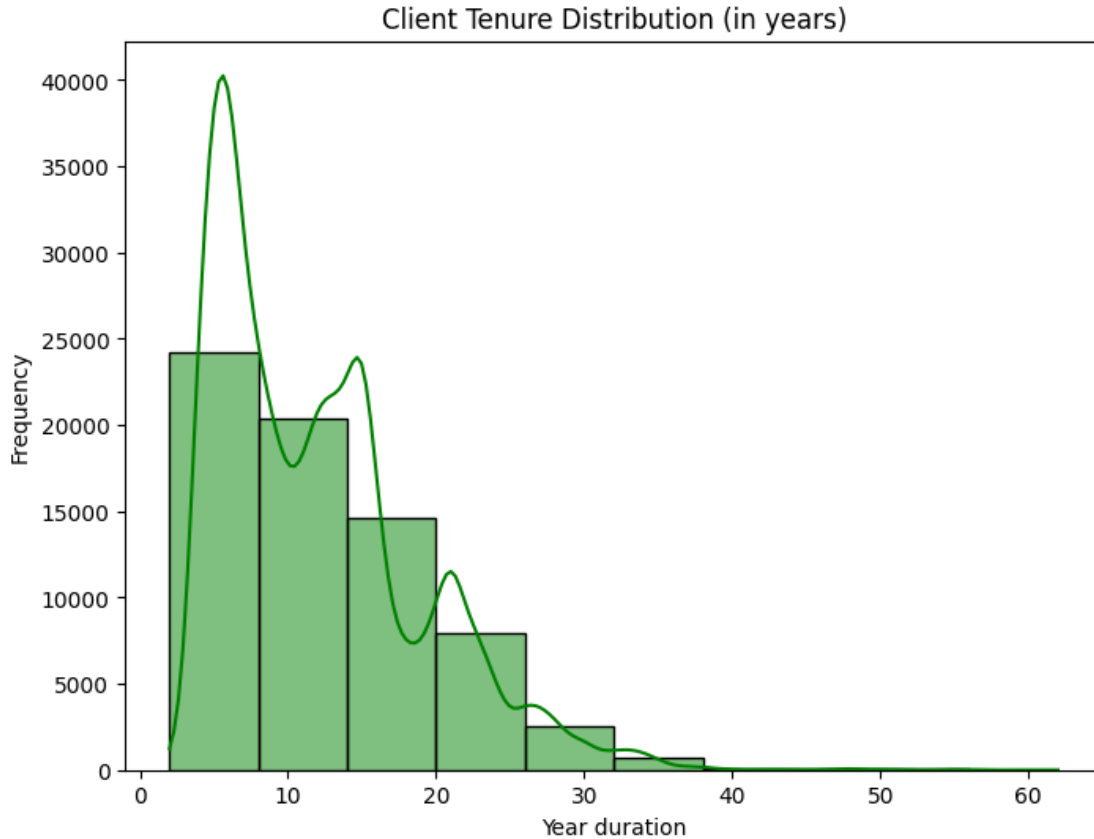
2.0.2 Active clients (higher logons and accounts):

- Average client age: 46.18 years
- Average client tenure: 12.05 years

2.0.3 Conclusion:

- The primary clients are *generally older* (above 40 years old).
- The primary clients are *generally long-standing* (over 3 years).





2.0.4 Performance Metrics

Success Indicators

Discovering what key performance indicators (KPIs) will determine the success of the new design. We calculated the completion rate, time spent on each step and error rates.

2.1 KPIs - Completion Rate

The proportion of users who reach the final `confirm` step.

2.1.1 Results

Summary:

Out of 70,594 total clients, **47,787** successfully completed the process (reached step 4), resulting in a completion rate of approximately **67.7%**.

Key Insights:

- **Clients Who Finished (47,787):**
These clients reached the final step (step 4), successfully completing the process.
- **Total Unique Clients (70,594):**
This is the total number of distinct clients in the dataset, calculated using `nunique()` on `client_id`.

- Completion Rate: $\text{Completion Rate} = (47,787 / 70,594) \times 100 = 67.7\%$
- Client Drop-off (32.3%): A total of 22,807 clients ($70,594 - 47,787$) did not finish the process. This represents a drop-off or abandonment rate.

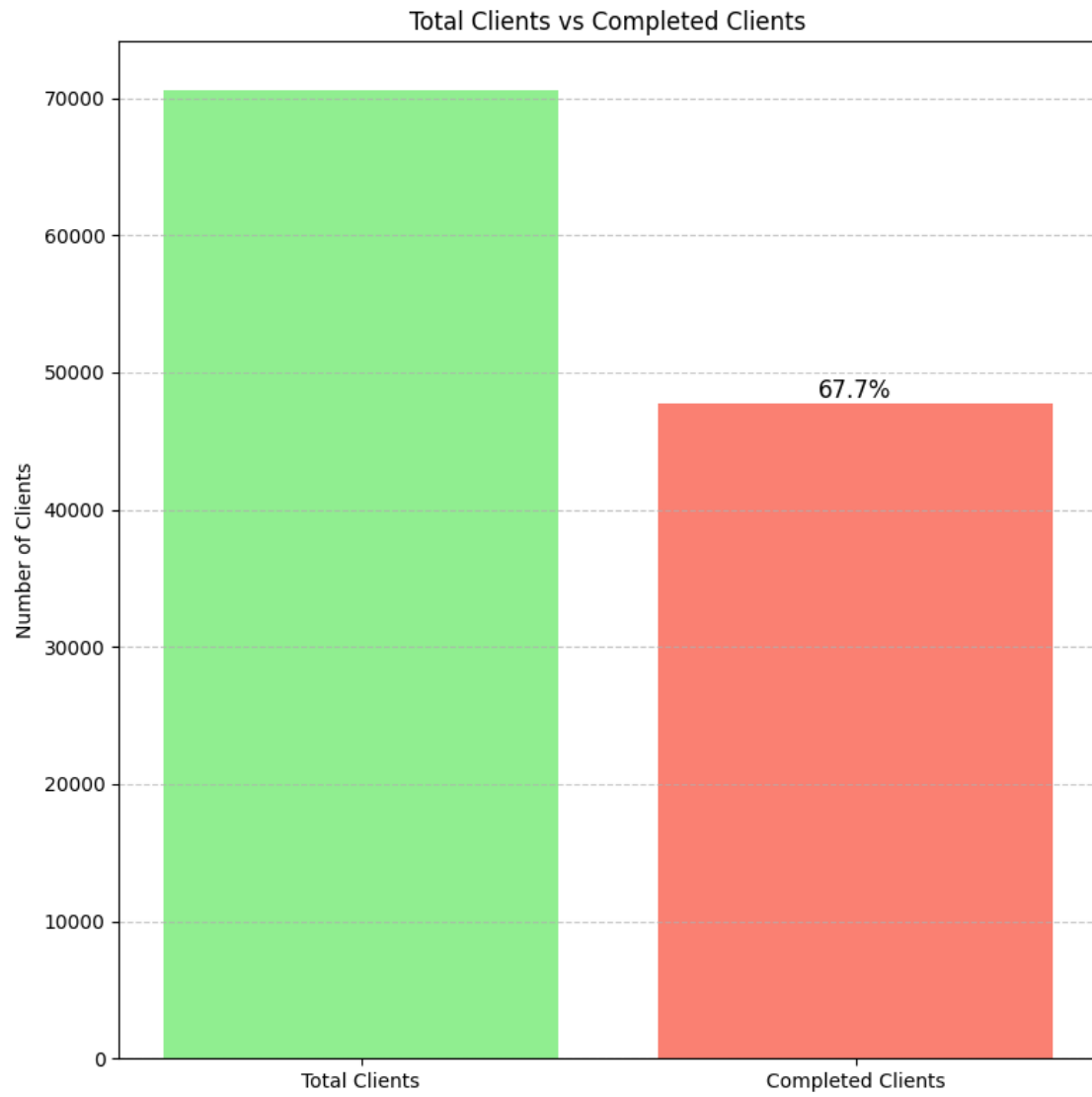
Implications:

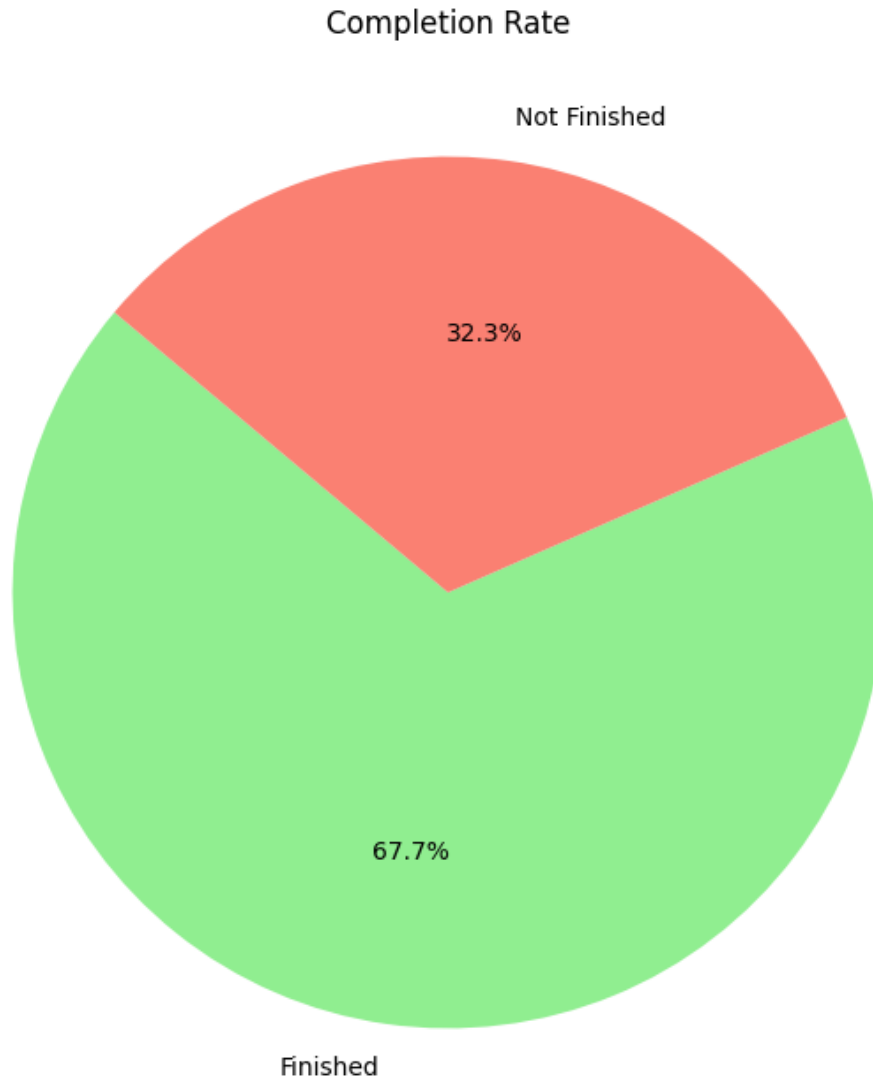
- Process Effectiveness: A completion rate of 67.7% may indicate good engagement, but there is room for improvement depending on project goals.
- Potential Goals: If 80% or higher is the target, interventions may be needed to reduce drop-offs and improve completion rates.

Follow-up Suggestions:

1. Analyze Drop-off Points: Identify where most clients drop off (e.g., step 2 or step 3). This can help locate friction points.
2. Segment Clients: Study client characteristics (e.g., demographics, behavior) to see if certain groups are more likely to drop off or complete the process.
3. Repeat Completions: Investigate whether some clients completed the process multiple times and consider focusing on unique client completions.
4. Retention Link: Explore whether completing the process correlates with retention or customer loyalty to gauge its impact on long-term success.

Conclusion: With a **67.7% completion rate**, the process appears moderately successful but highlights opportunities to reduce the **32.3% drop-off rate**. This data provides a strong foundation for refining the process and setting achievable performance targets.





2.2 KPIs - Time Spent on Each Step

The average duration users spend on each step.

2.2.1 Results

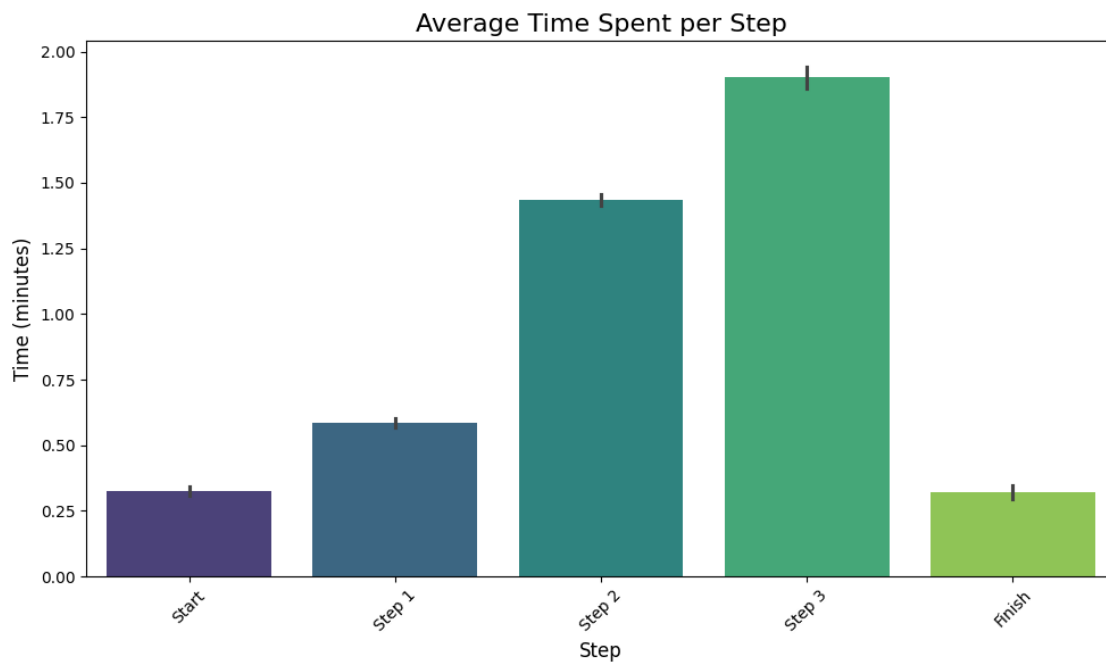
Step	Average time (minutes)
Start	0.673287
Step 1	0.813043
Step 2	1.563080
Step 3	2.144761
Finish	0.320263

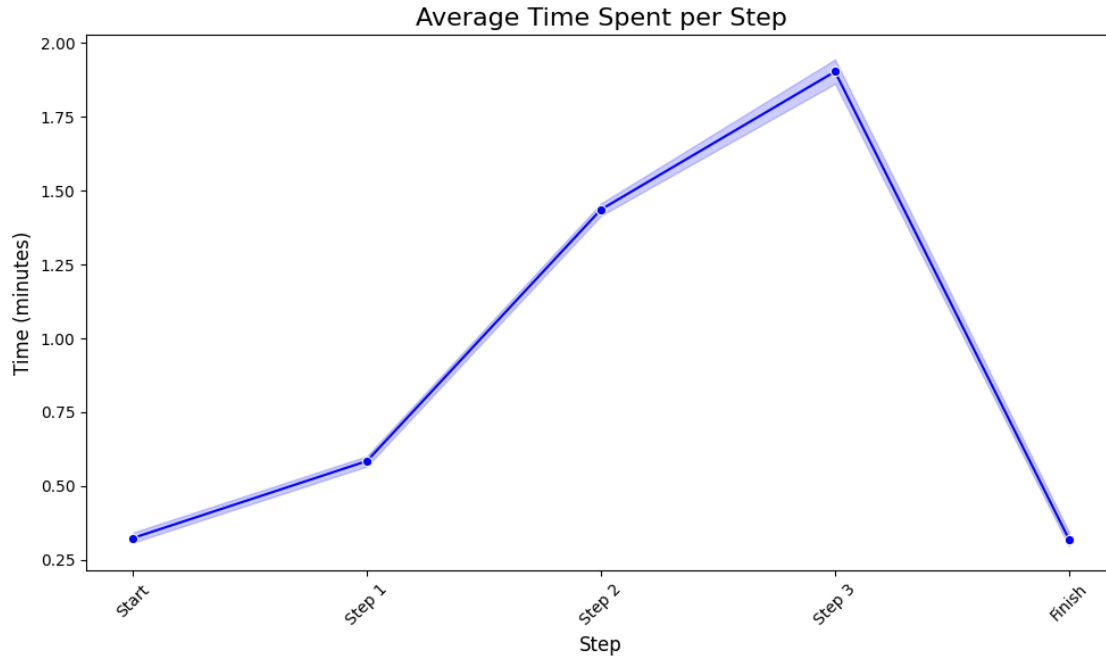
2.2.2 Conclusion:

The average time analysis shows that Steps 2 and 3 require more client attention or effort, while Step 4 takes less time. This decrease at Step 4, the final step, could indicate areas for improvement in user experience or process completion, but further investigation is needed to understand what happens at this stage.

Optimizing Steps 2 and 3, which take the longest, may improve client engagement and overall process efficiency.

It's important to note that defining what constitutes a “step” is subjective and can influence the analysis.





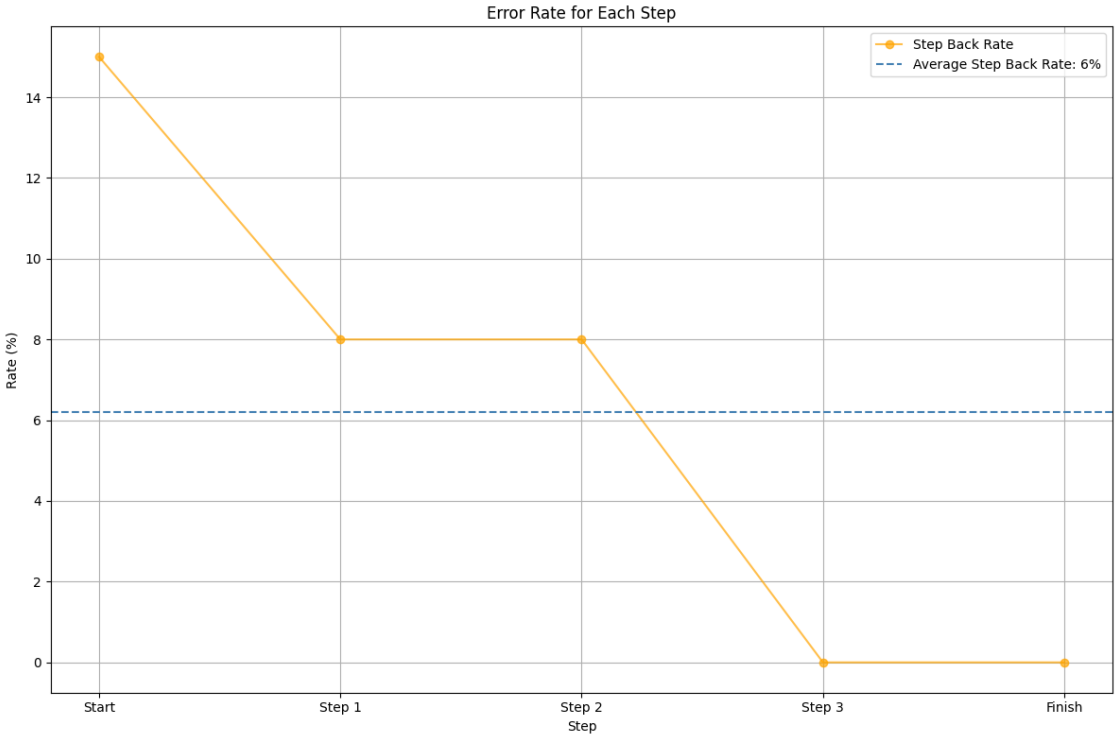
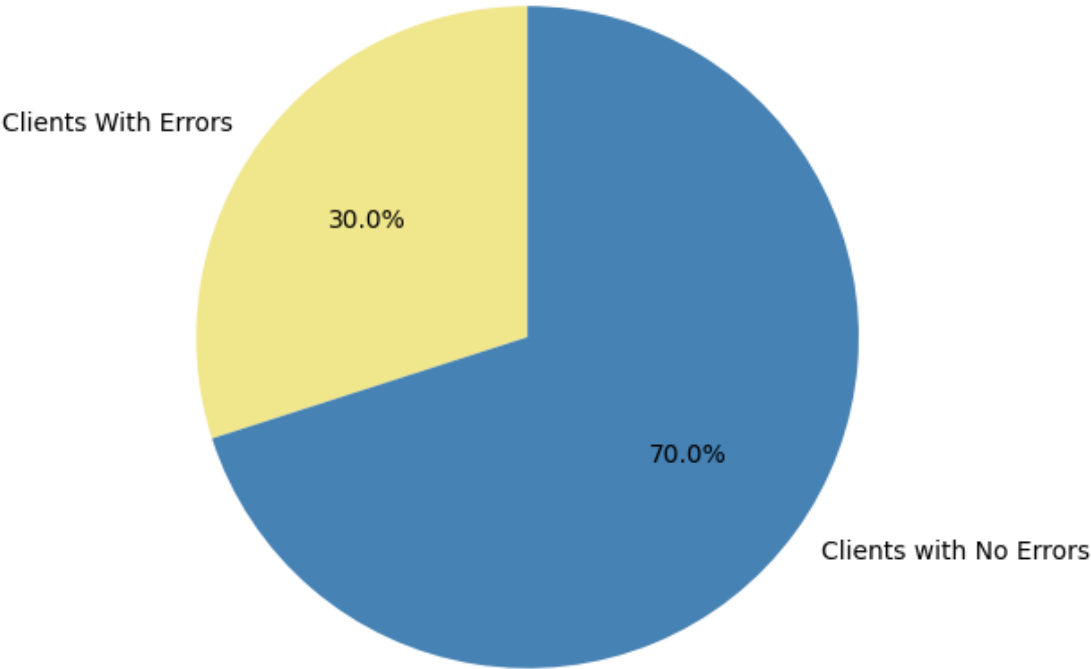
2.3 KPIs - Error Rates

If there's a step where users go back to a previous step, it may indicate confusion or an error. You should consider moving from a later step to an earlier one as an error.

2.3.1 Results

- Error rate (clients with errors): 29.95% of clients have at least one backward step during their visit. This means that nearly 30% of the clients in our dataset encountered some form of issue, where they moved backward in the process. This is a relatively high error rate, suggesting that a significant portion of clients are experiencing difficulties during their journey, which may require further investigation to improve the process.
- Error rate (steps with errors): 8.13% of all steps across all visits are backward steps (errors). This shows that a small but noticeable percentage of the steps in the process are errors, meaning that clients are not always following the expected forward progression. While this percentage is not overwhelmingly high, it still represents a non-negligible proportion of the entire process, and further investigation could help optimize the flow.
- Error rate per step: This is identical to the "steps with errors" rate because each step is either an error (backward movement) or not. Essentially, this shows that 8.13% of all individual steps across all visits are backward steps. This rate gives an overall picture of how often clients make errors during the entire process, not just at specific stages.

Error Rate (Clients with Steps Back)



2.4 KPIs for each Variation Group

Confirm if the difference in completion rate of the the new design and the old design is statistically significant. Given the data and KPIs you have explored discussed, one interesting hypothesis to test is related to the completion rate between the Test and Control groups. Since the new design (Test group) had a higher completion rate compared to the old design (Control group), you are required to confirm if this difference is statistically significant.

2.4.1 Results

Control Variation Completion Rate: Clients who finished the process: 65.6% (15428 out of 23526).

Total number of errors across all clients: 9576

Median time spent across all steps: 0.88 minutes

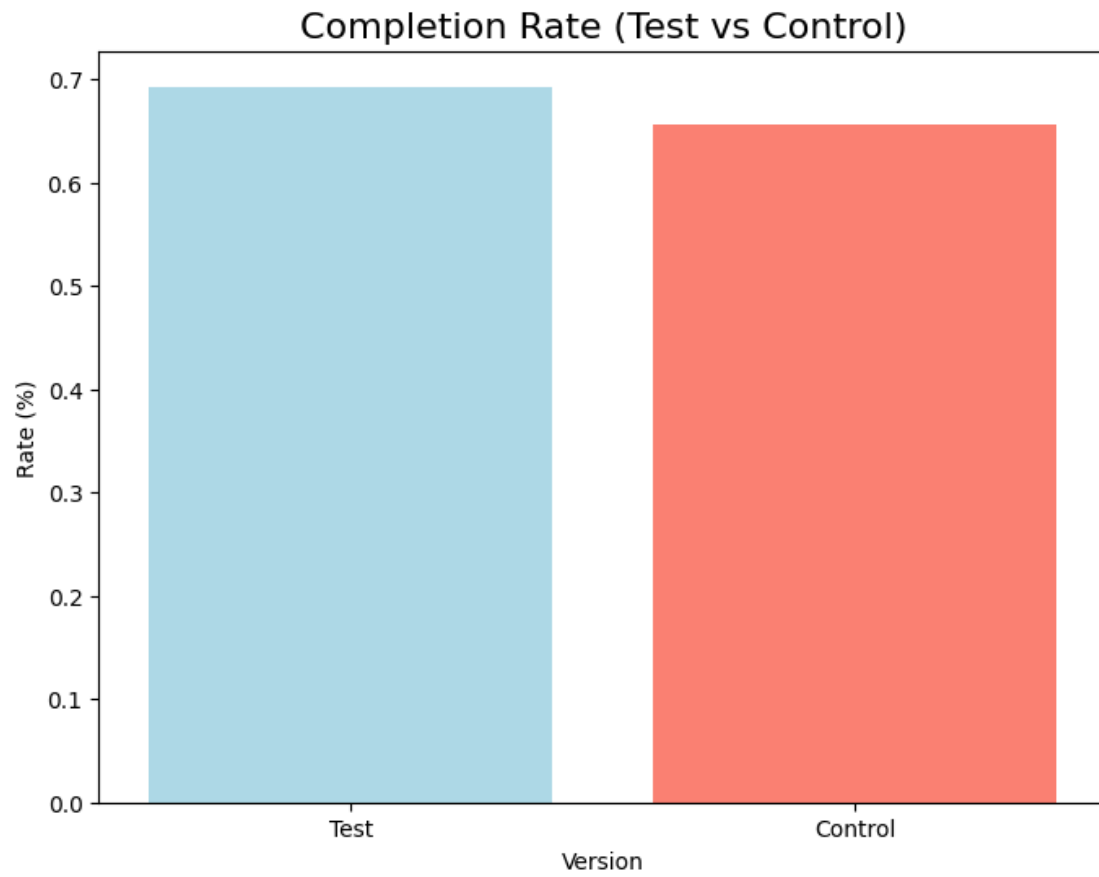
Step	Average Time (Minutes)
0	0.88
1	0.76
2	1.45
3	2.38
4	0.37

Test Variation Completion Rate: Clients who finished the process: 69.3% (18682 out of 26961).

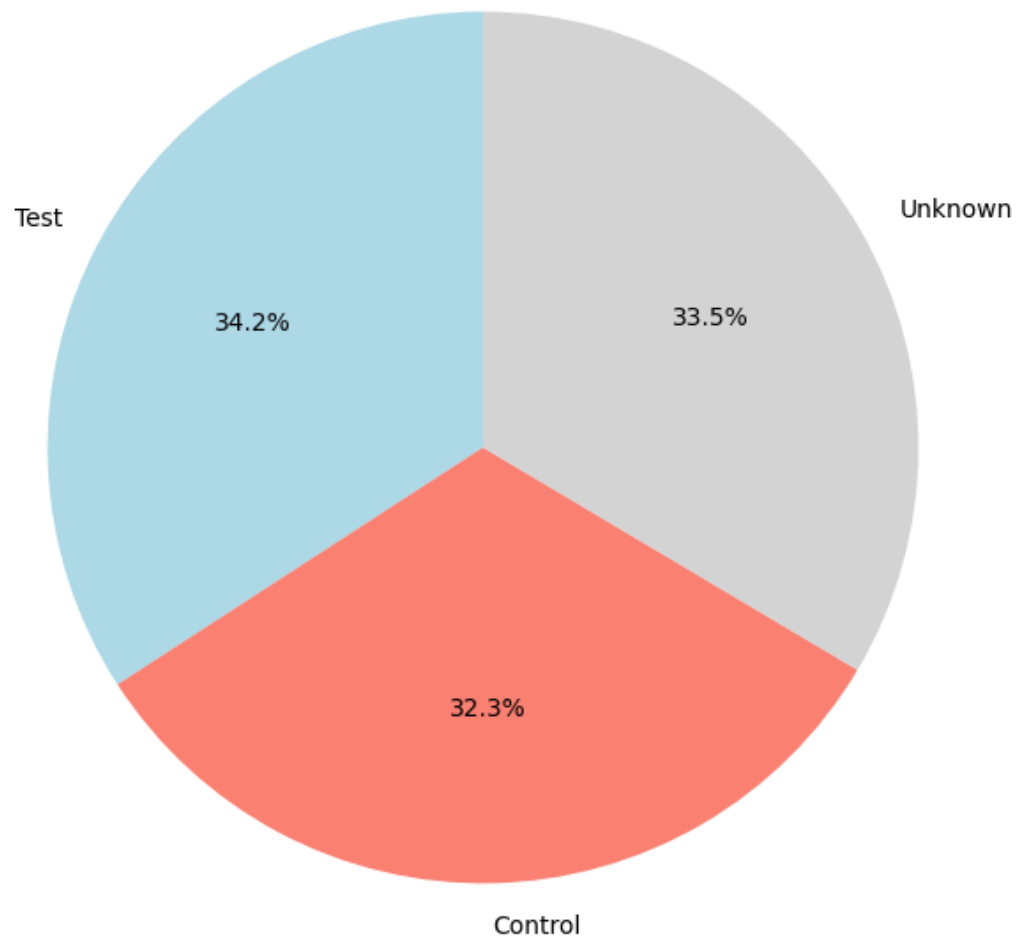
Total number of errors across all clients: 16229

Median time spent across all steps: 1.17 minutes

Step	Average Time (Minutes)
0	0.65
1	1.00
2	1.50
3	2.35
4	1.17



Completion Rate (Test, Control and Unknown Variation)





2.5 Completion Rate with a Cost-Effectiveness Threshold

2.5.1 Hypotheses:

- **Null Hypothesis (H0):** The completion rate for the new design (Test group) is **equal to or better than** the completion rate for the old design (Control group), with a 5% improvement.
- **Alternative Hypothesis (Ha):** The completion rate for the new design (Test group) is **worse than** the completion rate for the old design (Control group), even with the 5% improvement.

2.5.2 Z-Statistic Formula:

The **Z-statistic** is used to measure how much the difference between the Test group and the Control group is, adjusted for a 5% improvement. It helps us determine if the difference is big enough to be real and not just random.

$$Z = \frac{(p_1 - p_2 - 0.05)}{\sqrt{p \cdot (1 - p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

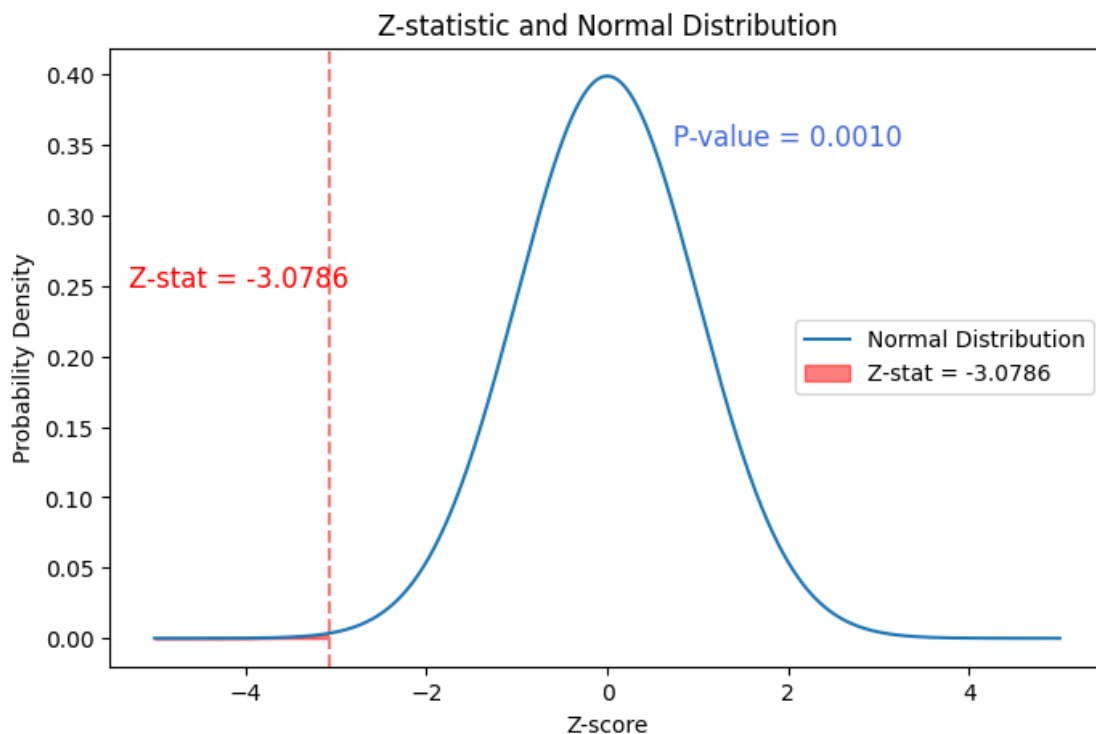
2.6 Hypothesis Test Results

- **Completion Rates:**

- The completion rate for the Control group (old design) is 0.6558, which means about 65.58% of clients completed the process.
- The completion rate for the Test group (new design) is 0.6929, which means about 69.29% of clients completed the process.
- **Z-statistic:**
 - The Z-statistic of -3.0786 tells us how much the difference between the Test and Control groups stands out from what we would expect if both designs were the same.
 - A negative value means **the Test group has a lower completion rate**, which is what we are testing in this case. The more extreme (negative or positive) the value, the stronger the evidence against the null hypothesis.
- **P-value:**
 - The P-value of 0.0010 is very small (less than 0.05), meaning there is strong evidence to reject the null hypothesis.
 - This suggests that **the difference between the Test and Control group is statistically significant**, and the Test group has a significantly lower completion rate than the Control group after adjusting for the 5% improvement.

Conclusion

- The new design (Test group) has a lower completion rate than the old design (Control group), and this result is statistically significant.
- The P-value of 0.0010 tells us this difference is unlikely to have happened by chance.



2.7 Design Effectiveness Analysis

2.7.1 Key Questions:

- Was the experiment well-structured?
- Were appropriate methods used to test the hypothesis?
- Were customers randomly assigned to the control (old design) and test (new design) groups to ensure fairness? (Client Allocation)
- Were there any factors that could have influenced the allocation or results unfairly? (Bias check)

3 ANOVA Results

We performed an ANOVA to compare the balance values across three groups: *Control*, *Test*, and *Unknown*.

What Are “Balance Values”? Balance values represent the amount of something (like money or points) that people in different groups (Control, Test, Unknown) have. We compare these amounts to see if one group has significantly more or less than the others.

3.0.1 Results

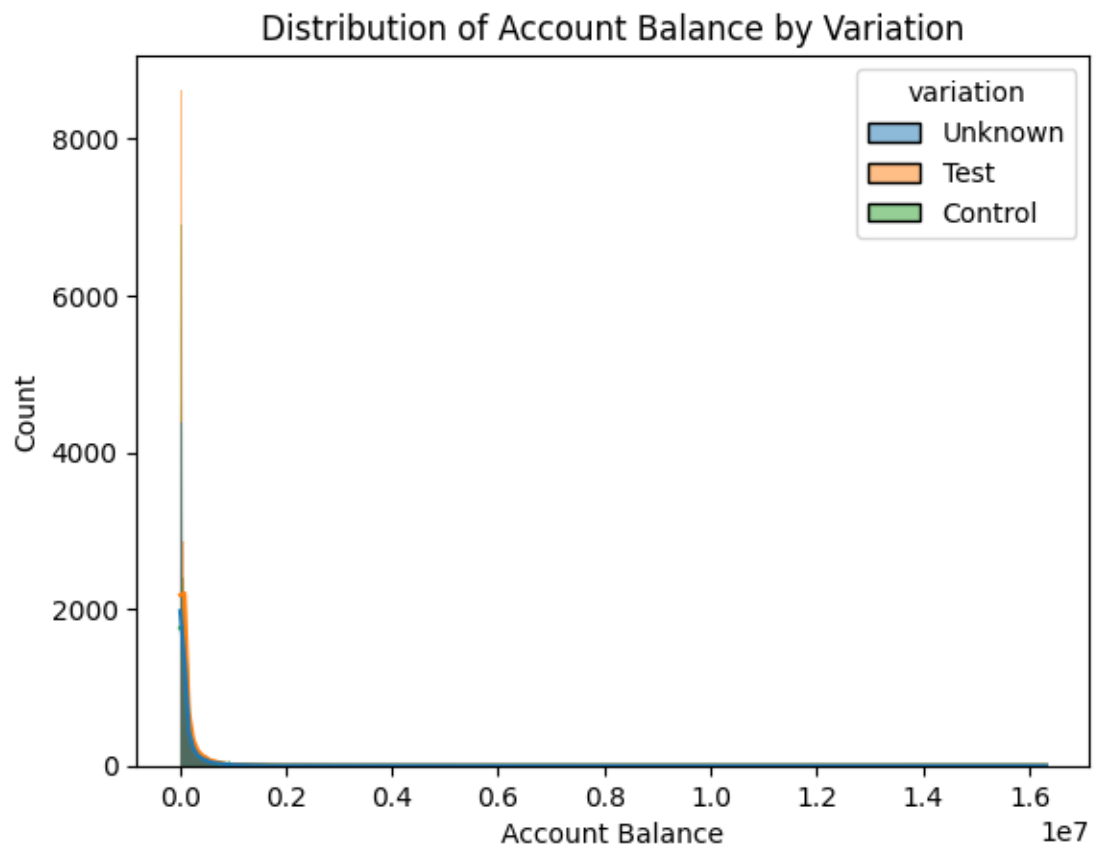
- **F-statistic:** 24.348
- **p-value:** 2.67e-11

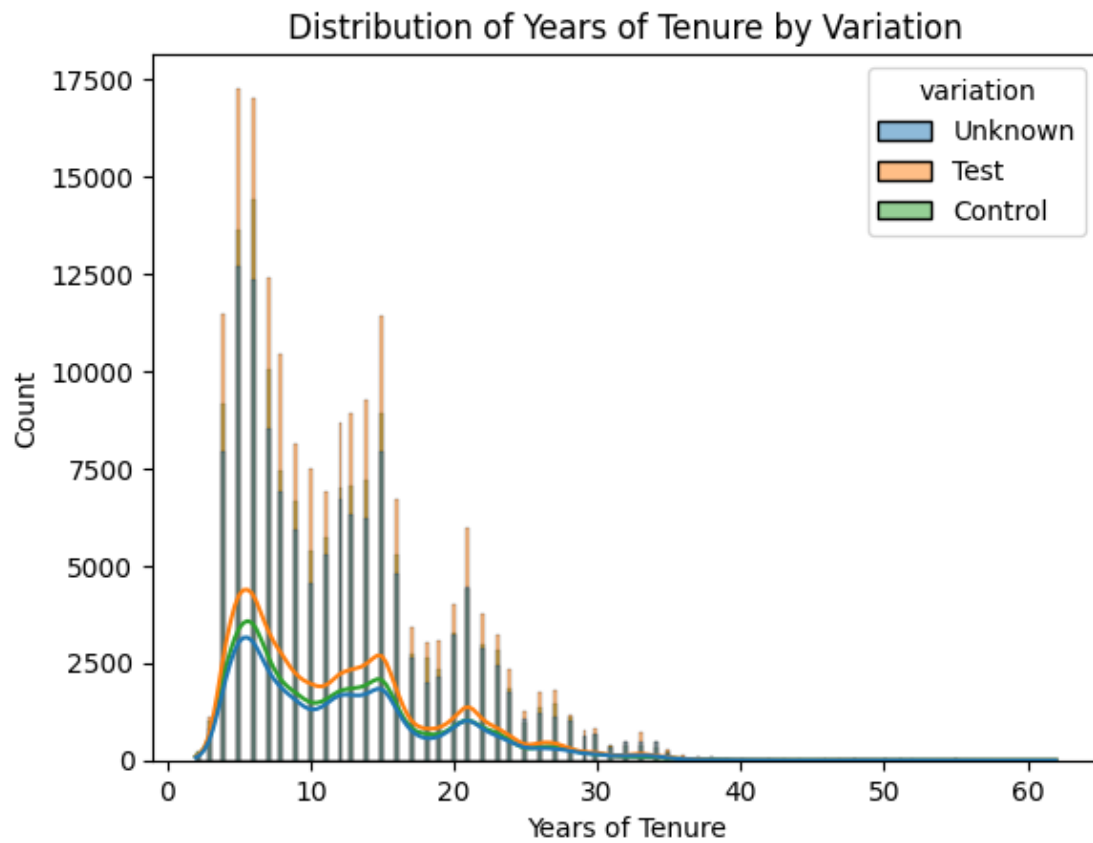
Interpretation

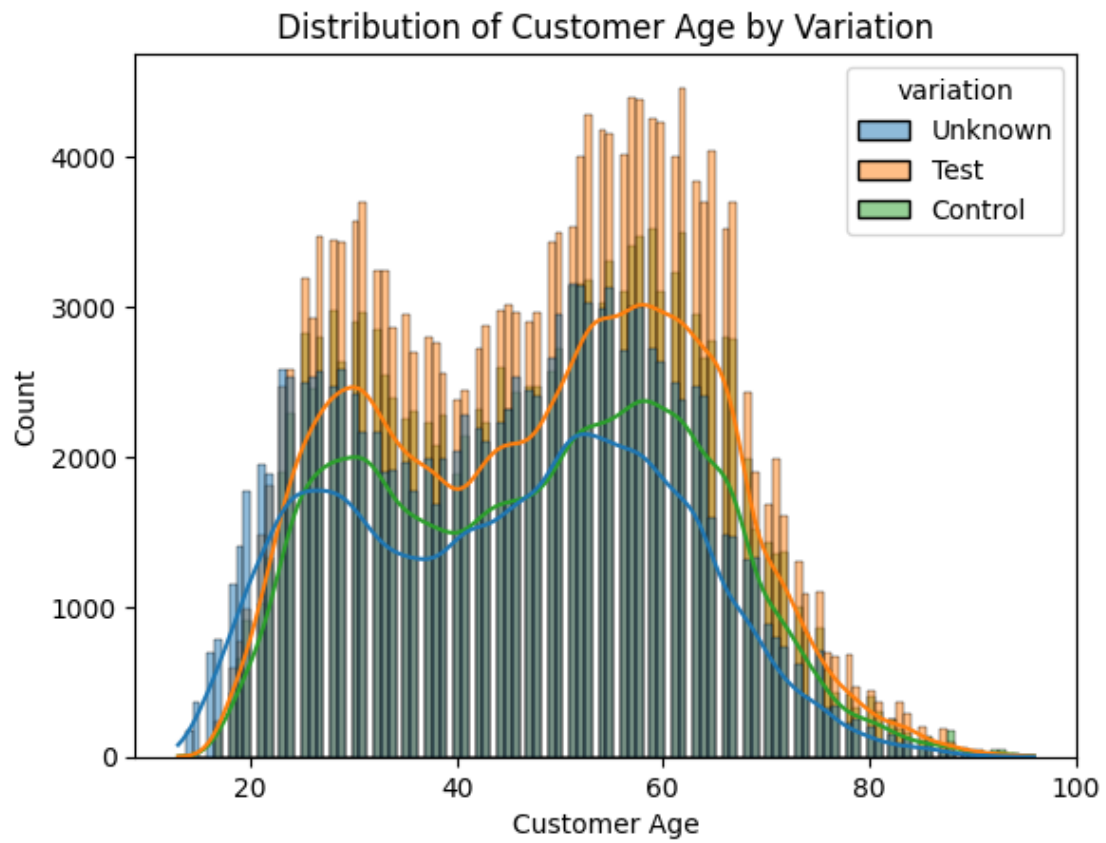
- The **F-statistic** measures how much the group means differ relative to the variability within the groups. A higher value indicates larger differences between groups.
- The very small **p-value** (< 0.05) suggests that there is a statistically significant difference in “balance” between at least two groups.

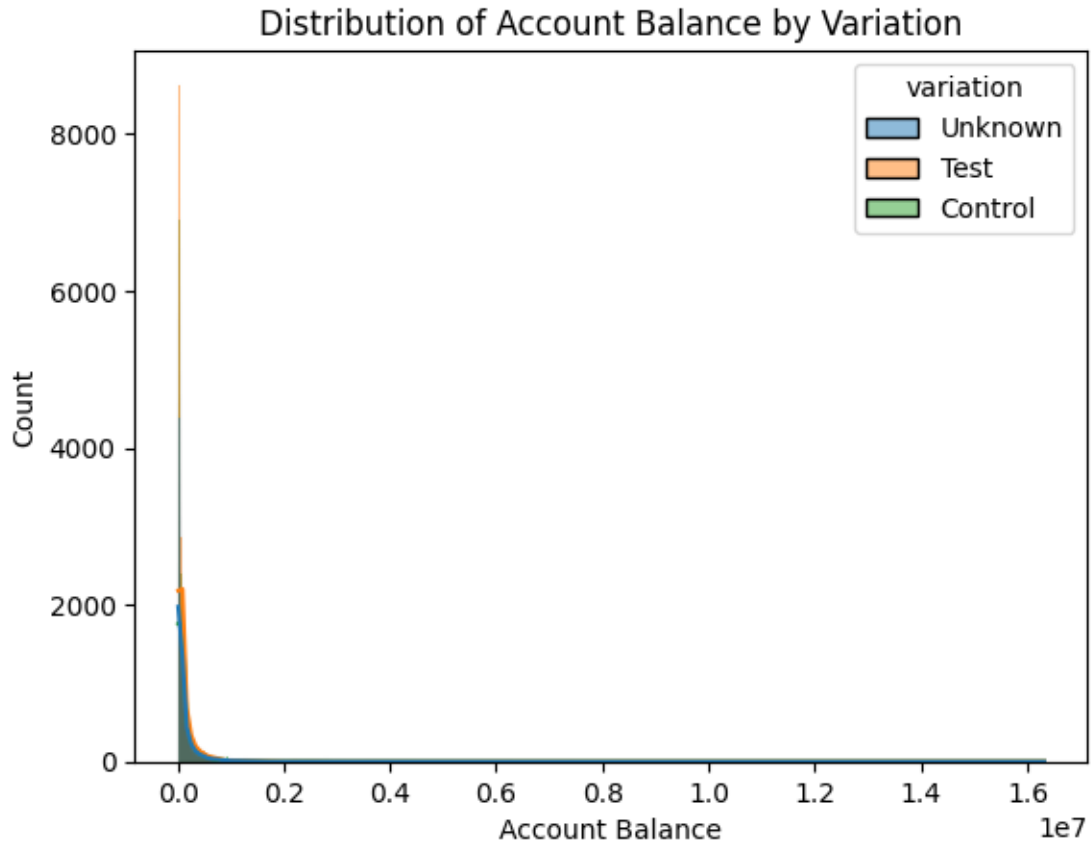
ANOVA shows that differences exist, but it doesn’t tell us which groups differ. To find out, the next step is to perform a **Tukey’s HSD test** to compare all pairs of groups.

4 Descriptive Statistics by Group









5 Tukey HSD Test Results

The Tukey HSD test was performed to compare the balance values between the three groups: **Control**, **Test**, and **Unknown**. This will identify the specific pairs with significant differences and provide p-values for each, and to pinpoint the exact group comparisons driving the significant ANOVA result.

Here's a summary of the findings:

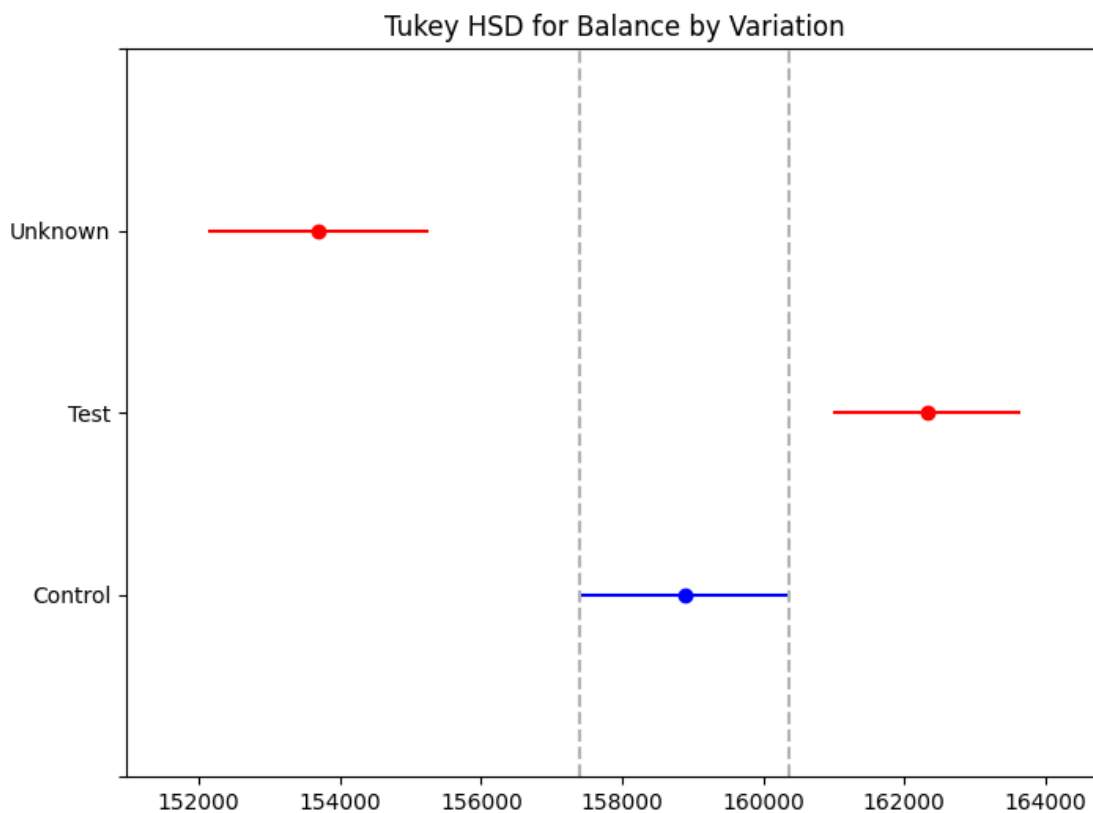
Results

Group 1	Group 2	Mean Difference	p-value	95% Confidence Interval	Significant?
Control	Test	3432.54	0.0118	[618.30, 6246.77]	Yes
Control	Unknown	-5195.64	0.0002	[-8246.02, -2145.27]	Yes
Test	Unknown	-8628.18	0.0000	[-11526.76, -5729.60]	Yes

5.0.1 Interpretation

- Control vs. Test:
 - The mean difference is **3432.54**, with a p-value of **0.0118**.

- The confidence interval does not include 0, and the result is statistically significant.
 - This means there is a significant difference in balance between the **Control** and **Test** groups.
2. Control vs. Unknown:
- The mean difference is **-5195.64**, with a p-value of **0.0002**.
 - The confidence interval does not include 0, and the result is statistically significant.
 - This shows a significant difference in balance between the **Control** and **Unknown** groups.
3. Test vs. Unknown:
- The mean difference is **-8628.18**, with a p-value of **0.0000**.
 - The confidence interval does not include 0, and the result is statistically significant.
 - This indicates a significant difference in balance between the **Test** and **Unknown** groups.



5.0.2 Conclusion 1:

All pairwise comparisons show statistically significant differences in balance between the groups: - Control vs. Test - Control vs. Unknown - Test vs. Unknown

This confirms that balance values vary significantly between all groups.

5.0.3 Conclusion 2:

- Significant Differences Between Groups:
 - ANOVA results show statistically significant differences in average balances among groups (p-value $2.67e-11$).
 - Tukey’s post-hoc test identifies the specific group comparisons with significant differences:
 - * **Control vs Test:** The Test group has significantly higher average balances compared to the Control group (meandiff 3432.54).
 - * **Control vs Unknown:** The Unknown group has significantly lower average balances compared to the Control group (meandiff -5195.64).
 - * **Test vs Unknown:** The Unknown group has significantly lower average balances compared to the Test group (meandiff -8628.18).
- Visualization Insights:
 - Tukey’s plot shows that the Test group has the highest average balance, the Control group is in the middle, and the Unknown group has the lowest average.
 - Error bars do not overlap significantly, supporting the statistical findings.

Interpretation:

- The analysis indicates that the groups are not balanced in terms of their average initial balances. This imbalance, particularly the significantly lower average for the Unknown group, may affect the fairness of the experiment.
- Differences in initial balances might influence participant behavior, potentially impacting the experimental outcomes.

Implications for the Experiment:

- Experimental Design:
 - The disparity in initial balances could bias results and should be accounted for in future analyses.
 - Controlling or adjusting for these differences (e.g., through statistical techniques like ANCOVA) is recommended to ensure a clearer understanding of the experimental effects.
- Follow-up Suggestions:
 - Investigate if other variables (e.g., customer tenure or number of accounts) also show similar imbalances.
 - Consider redesigning the experiment to ensure more homogeneous group assignments or apply statistical adjustments for existing disparities.

Summary: While the results highlight significant differences between groups, the initial imbalances in average balances must be considered to accurately interpret the experimental findings. Adjustments or redesign may be necessary to ensure robust conclusions.

5.1 A/B Test

When analyzing the results of an A/B test (where you compare two versions of a product or design), you’re required to ensure that the new design (Test) improves the completion rate by at least 5%

compared to the current version (Control).

We performed an analysis to compare the completion rates between a Test Group and a Control Group using a Z-test for proportions.

5.1.1 Calculate Completion Rates

The code calculates the completion rate for each group. Both rates are printed as percentages, rounded to two decimal places.

5.1.2 Check for Valid Completion Rates

The code checks if either completion rate is 0 or 1, as this would indicate that all participants either completed or did not complete the action. A rate of 0 or 1 results in a standard error of 0, making the Z-test invalid. If this occurs, the code prints a message and skips the Z-test due to invalid completion rates.

5.1.3 Perform Z-Test for Proportions

The Z-test checks if the difference in completion rates between the Test and Control Groups is **statistically significant** (i.e., not just due to random chance).

Here's the process:

1. **Combine the Groups:**

Calculate the overall completion rate for both groups combined (the “pooled proportion”).

2. **Calculate Standard Error (SE):**

The SE measures how much completion rates can vary. A small SE indicates similar groups, while a larger SE suggests more difference.

3. **Validate SE:**

If the SE can't be calculated (due to extreme data), the test is skipped and an error is shown.

4. **Calculate Z-Score:**

The Z-score compares the difference between the two groups to the SE, showing how significant the difference is.

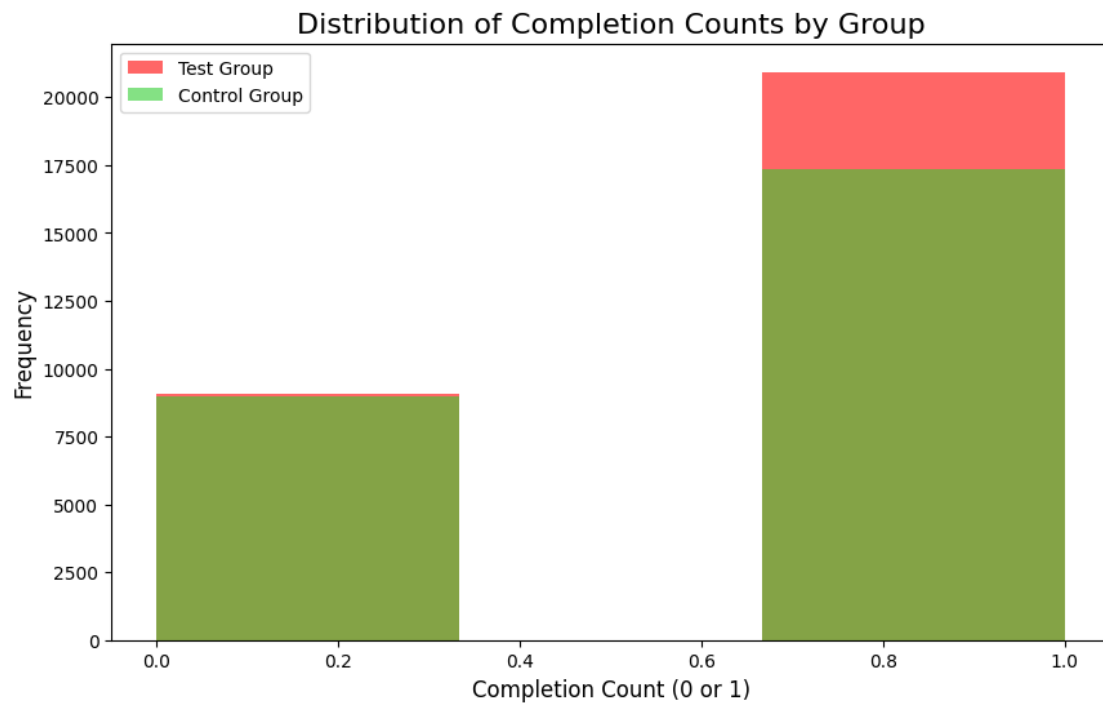
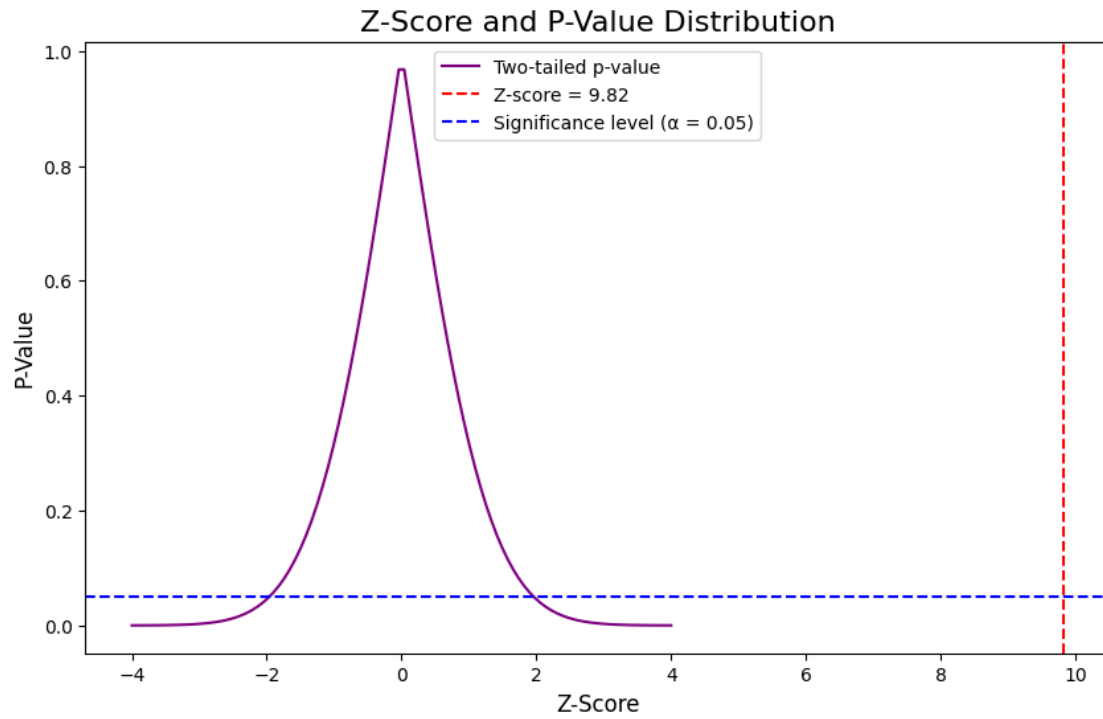
5. **Find P-Value:**

The p-value shows the likelihood that the observed difference is due to chance. If the p-value is below 0.05, the difference is considered **statistically significant**.

In summary, the Z-test determines if the difference between the groups is large enough to be considered meaningful and not random.

5.1.4 Check Statistical Significance

The code compares the p-value to the significance level ($\alpha = 0.05$, or 95% confidence). If the p-value is less than 0.05, the difference between the completion rates is considered statistically significant, and the code prints a message confirming this. If the p-value is greater than 0.05, the difference is not statistically significant, and the code prints a message indicating this.



5.2 Conclusions

Completion Rate Comparison:

- **Test Group:** 69.77% completion rate

- **Control Group:** 65.90% completion rate

The **test group** (new design) outperformed the **control group** (old design), suggesting the new design encourages more completions.

Statistical Significance:

- **Z-score:** 9.82
- **P-value:** 0.0000

The **Z-score** is high, and the **p-value** is very low, indicating the difference is **statistically significant** and not due to chance.

Hypothesis Testing:

- The **null hypothesis** (no difference) is **rejected**.
- The **alternative hypothesis** (test group performs better) is **accepted**.

Practical Implications:

- The new design appears to be **more effective** and is likely the better option.
- The large difference suggests the result is **reliable** and not random.

Confidence Interval: The 95% confidence interval shows the true difference is unlikely to be zero, confirming the result's reliability.

Next Steps:

1. **Implement the new design:** Given the positive results, consider adopting it.
2. **Evaluate costs:** Assess whether the improved completion rate justifies any additional costs.
3. **Continue testing:** Perform further tests to ensure consistent results across different users or situations.

Final Takeaway: The new design leads to more completions and the result is statistically significant. If the benefits outweigh the costs, proceed with implementing it.

5.3 Hypotheses

- H_0 - Null Hypothesis: The average tenure of clients in the Test group equals the average in the Control group.
- H_1 - Alternative Hypothesis: The average tenure of clients in the Test group is not equal to the average in the Control group.

$$H_0 : mean_{test} = mean_{control}$$

$$H_1 : mean_{test} \neq mean_{control}$$

Data Summary

Group	Mean (Years)	Std Dev (Years)	Sample Size
Control	12.088	6.878	23,526
Test	11.983	6.845	26,961

Conclusion

- H_0 - Null Hypothesis is sustained: The average client tenure is similar for both groups.
- H_1 - Alternative Hypothesis is not sustained: No significant difference in average tenure between the groups.

6 More Insights

More graphics showing further analysis and insights can be found on our [Tableau Dashboard](#)

For further questions do not hesitate to contact us!

Constanza Tabora and Brugera Herrero