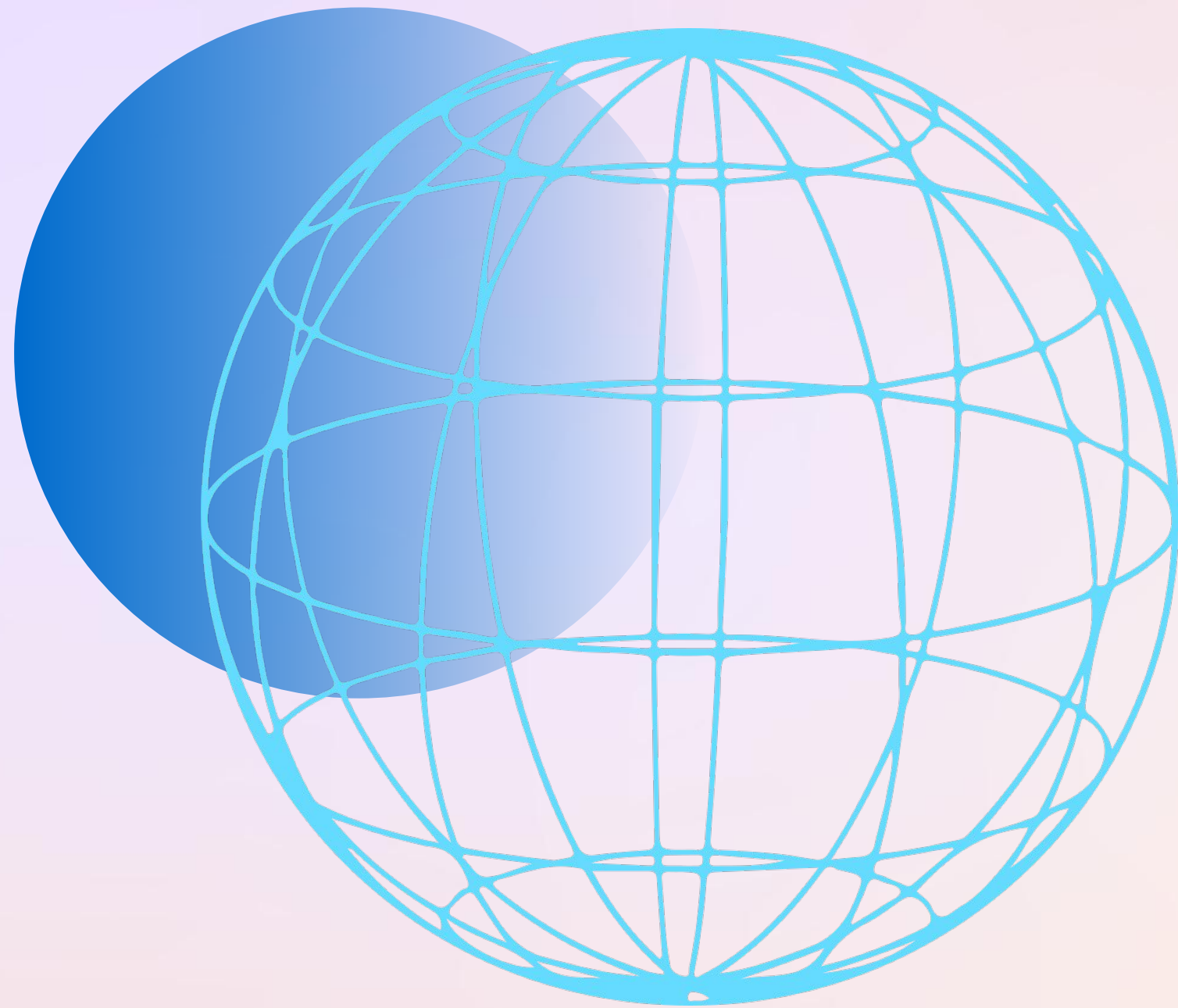


VANGUARD Company

# VANGUARD

A/B TEST

Presented by  
Esra & Theophilus



---

# Introduction

## Vanguard's Challenge:

Improve user completion rates through a redesigned digital process.

## Main Question:

Did the new UI lead to significantly higher process completion rates compared to the old design?

# Datasets preparation process

## Dataset

- Client Profiles: demographic and account info
- Digital Footprints: timestamped user interactions across steps
- Experiment Roster: test/control group assignments

## Cleaning & Merging:

Merged all sources into a single DataFrame

### Modular Cleaning Pipeline

- Created a separate Python script, containing reusable functions
- Functions applied across datasets via consistent logic

### Key Functions Used

- `read_function()` → Loads and parses raw data
- `conv_datetime()` → Converts date fields to datetime format
- `quick_data_report()` → Provides a snapshot of shape, types, and missing values
- `fill_categorical_with_mode()` → Handles missing categorical data
- `remove_duplicates()` → Ensures unique records per client/session



# Exploratory Data Analysis EDA

## Client Demographics:

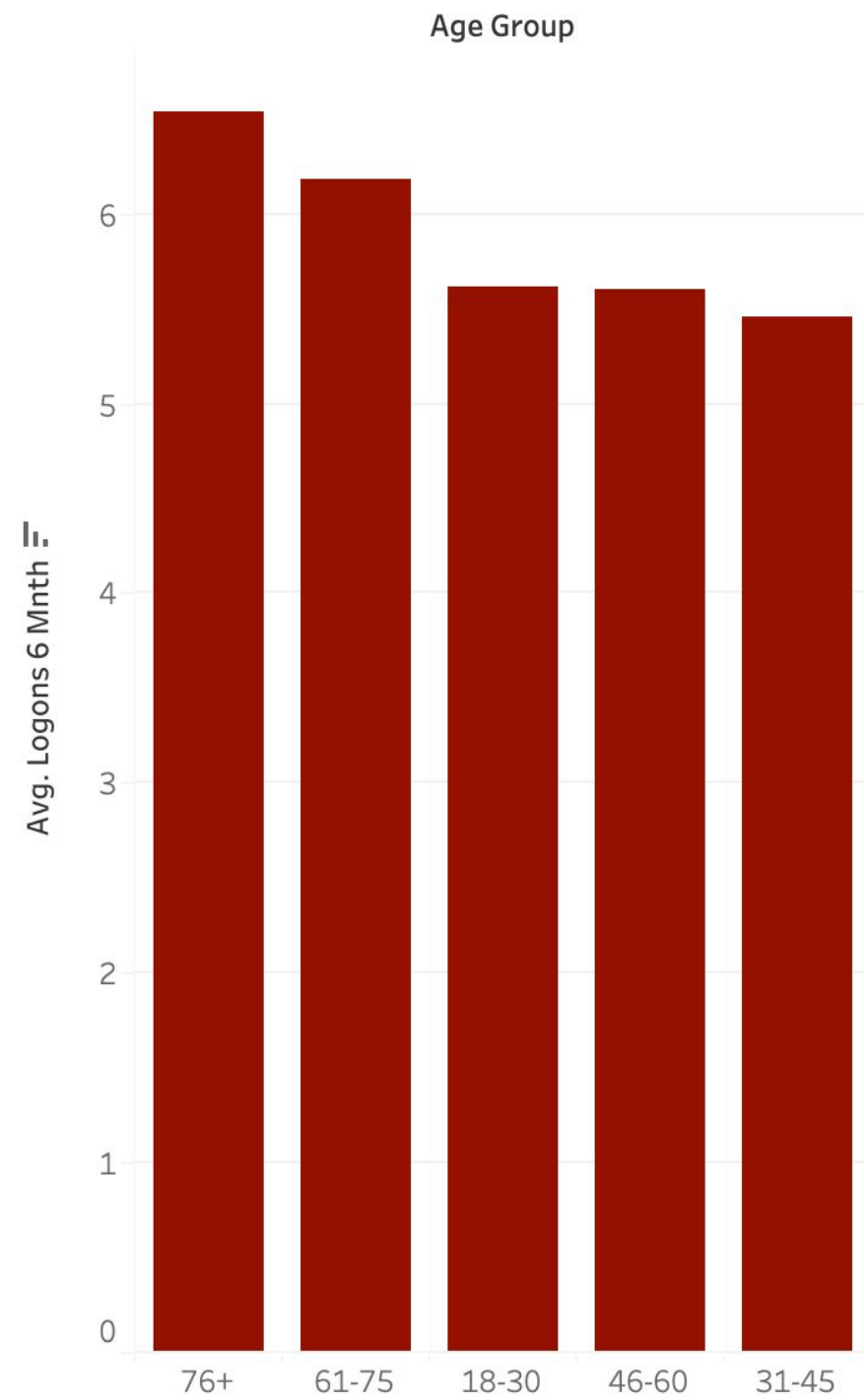
- Majority are aged between 35–55
- Slight male majority
- Wide range of tenure with Vanguard

## Behavioral Insights:

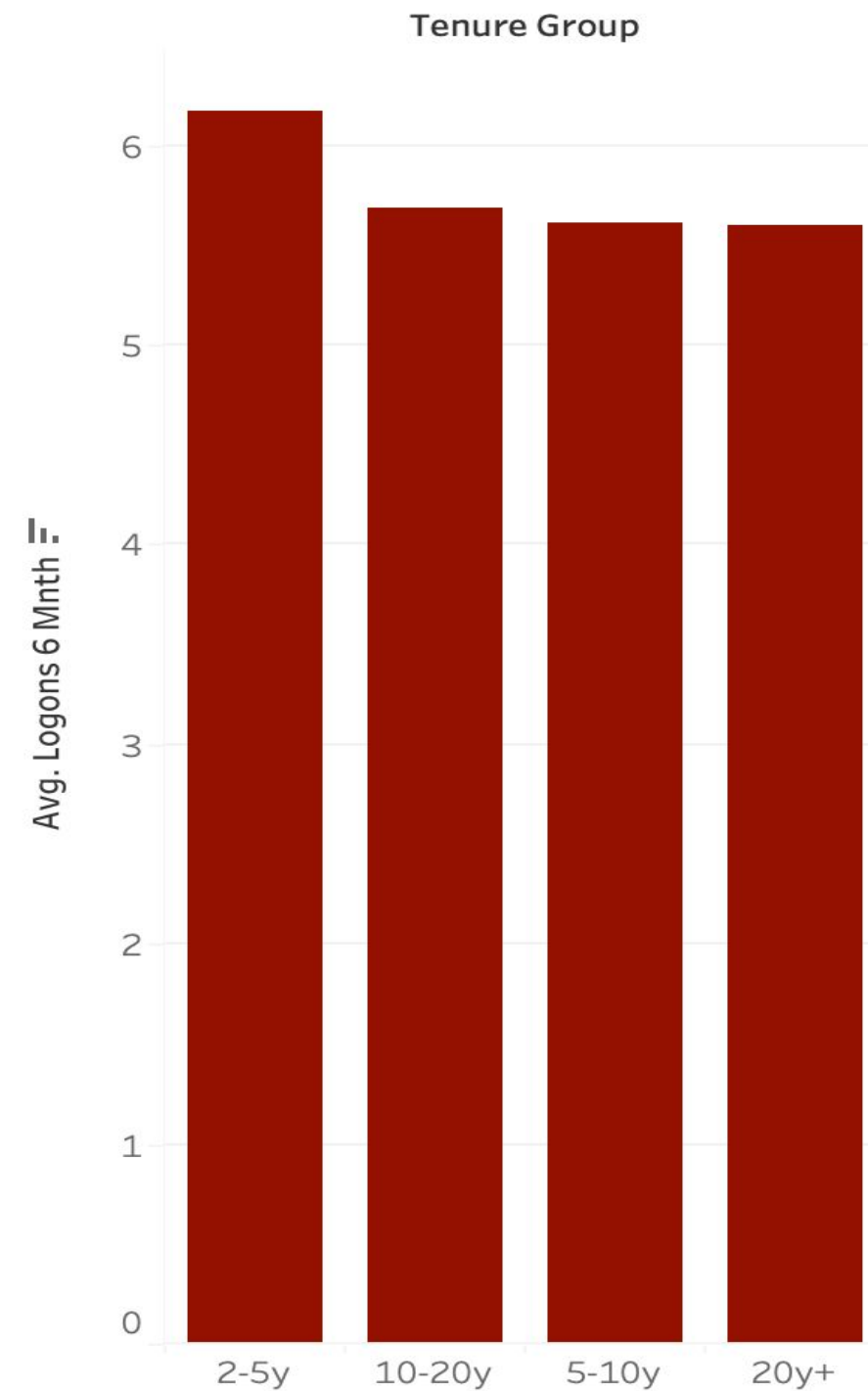
- Older clients tend to log in more often
- Clients with longer tenure engage more fully in multi-step flows
- No major imbalance between Test and Control groups pre-funnel

# EDA Visualizations

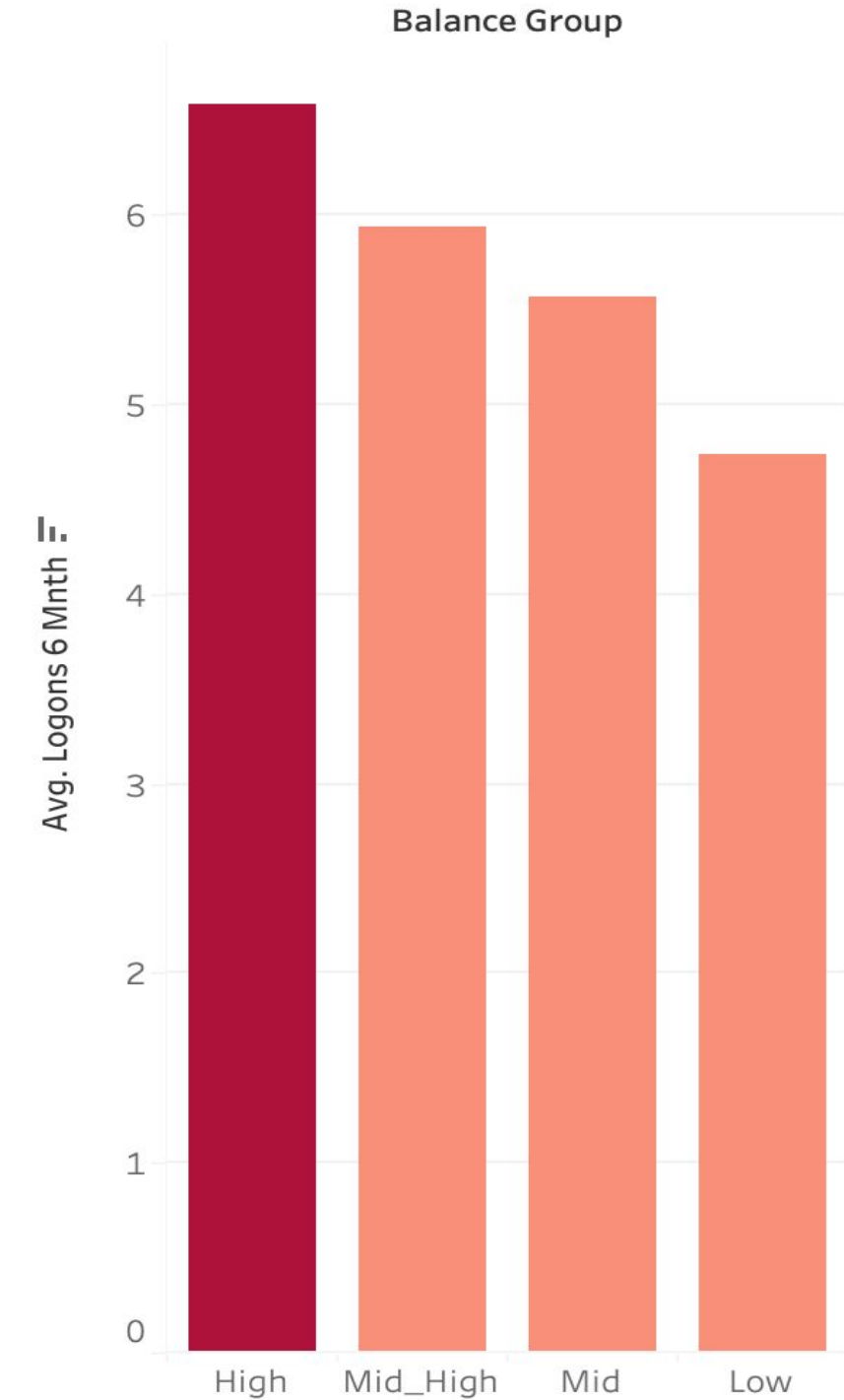
Average Logons by Age Group



Average Logons by Tenure Years



Average Logons by Balance Level

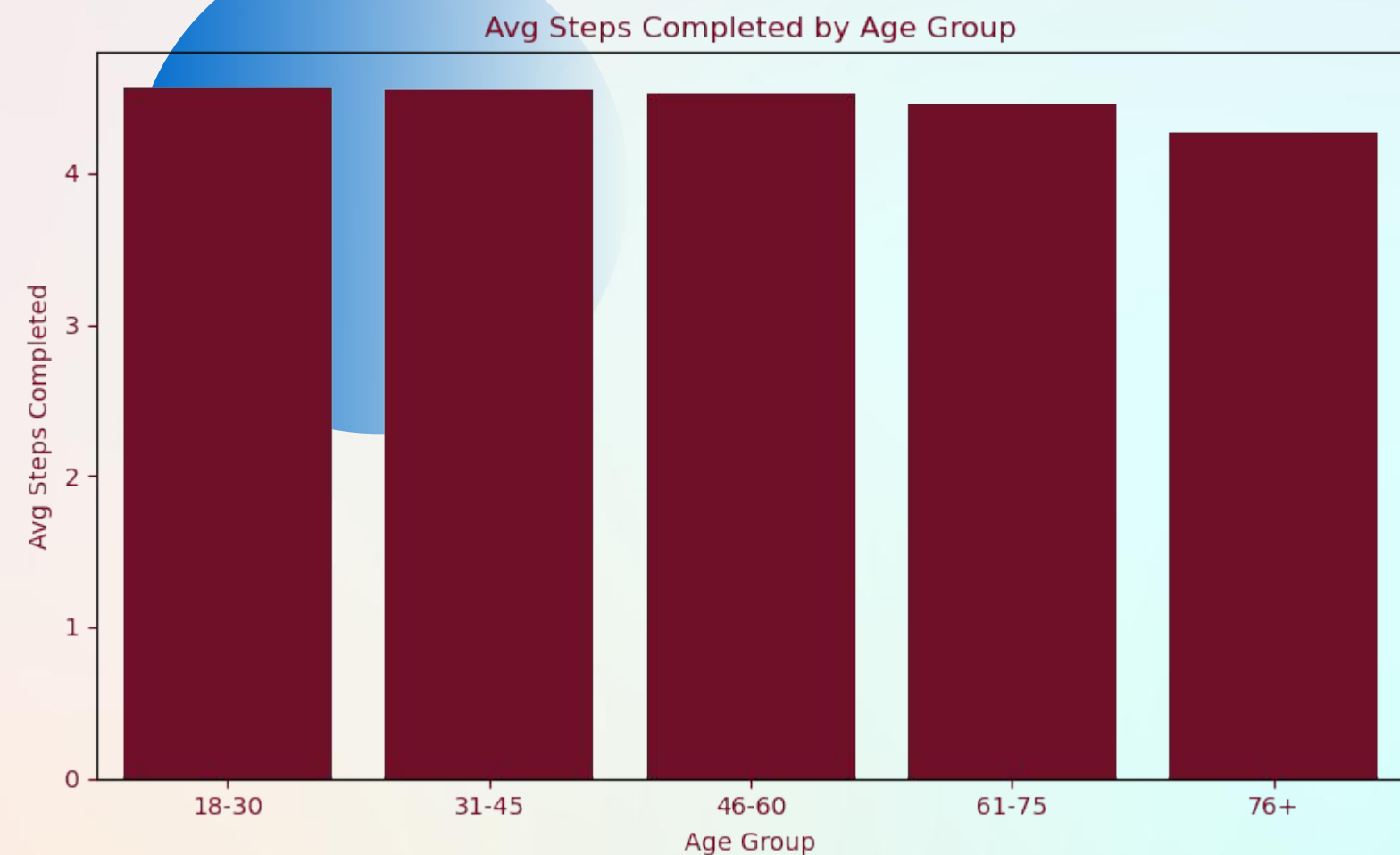


# Performance Metrics

1- Completion Rate: The proportion of users who reach the final 'confirm' step.

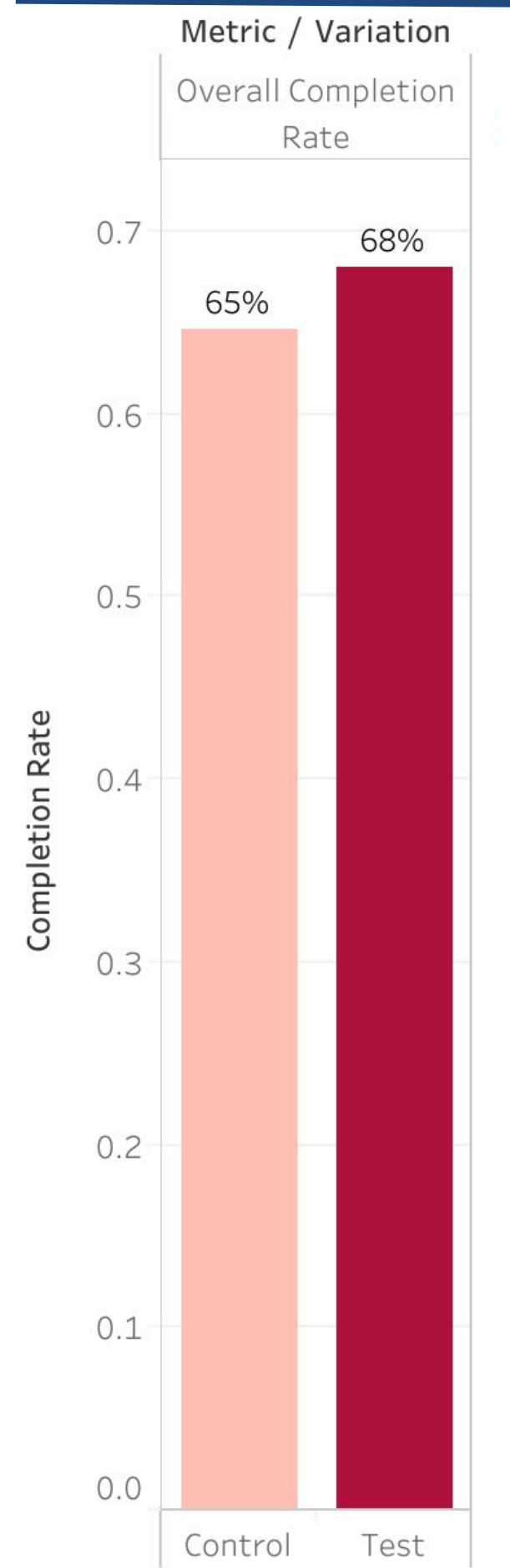
2- Time Spent on Each Step: The average duration users spend on each step.

3- 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.

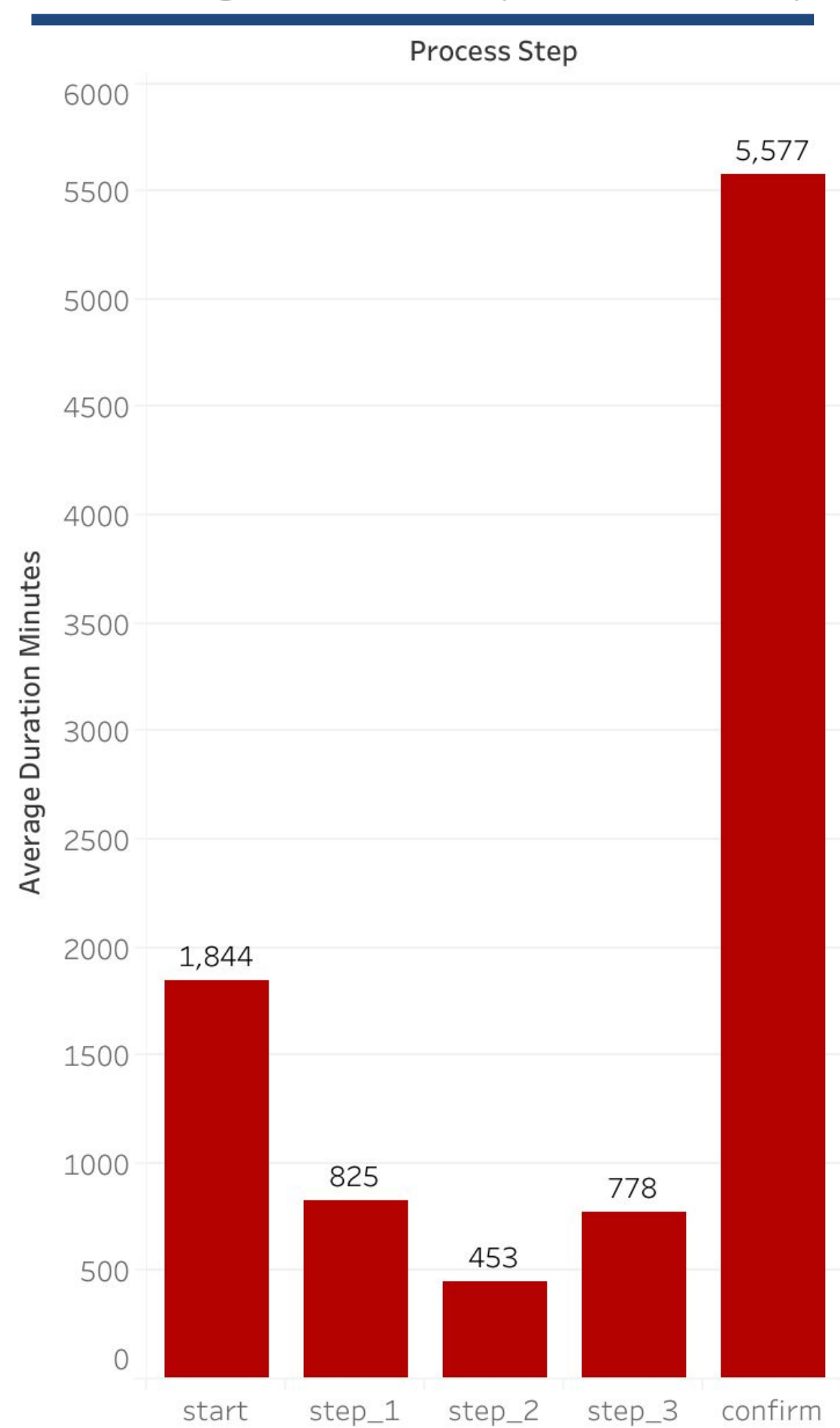




Overall  
Completion Rate

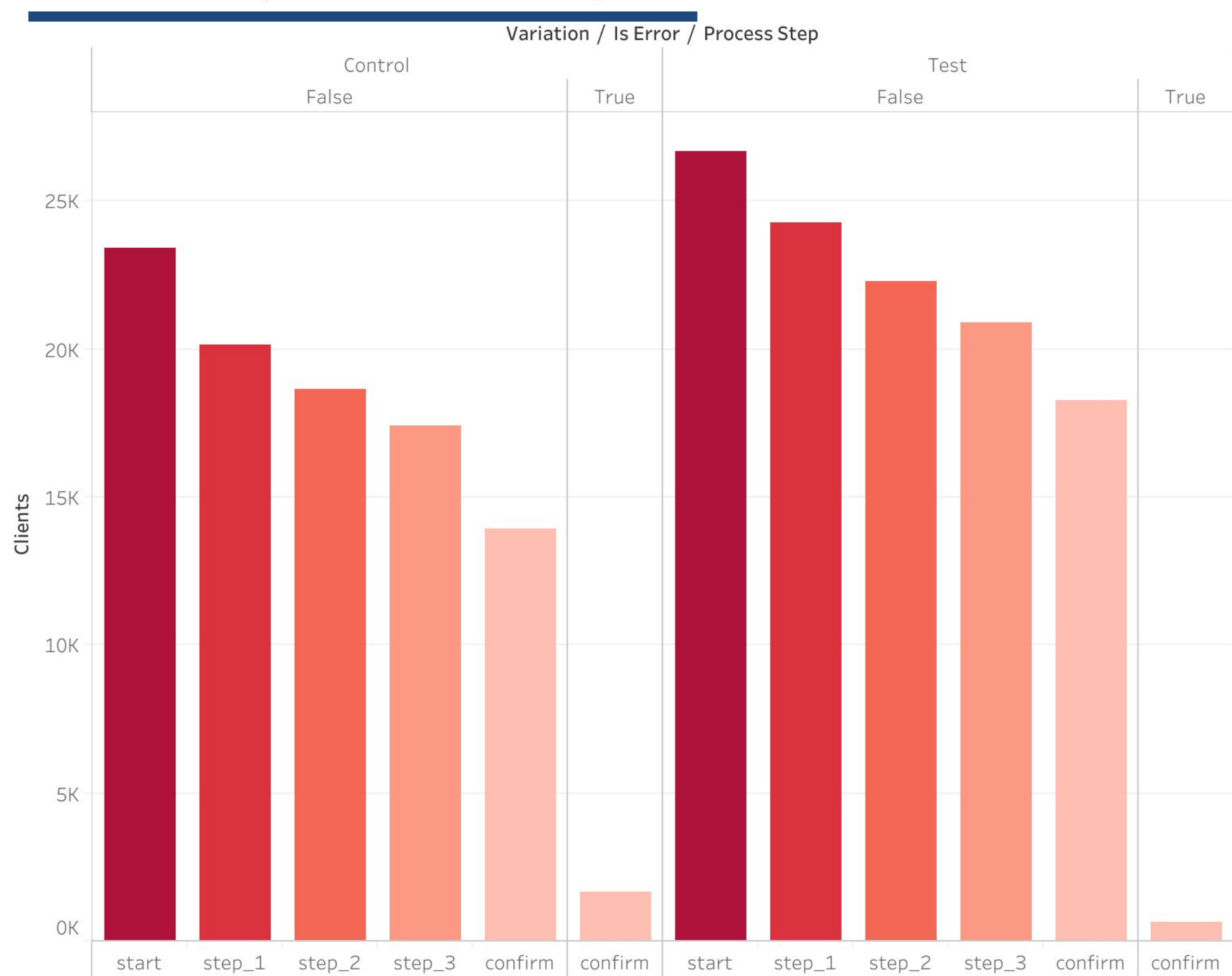


The average duration spent each step



Metrics

step regressions (users going backward)



*Is the new design led to more clients completing the process compared to the existing design*

# Hypothesis Testing Rationale

## Why This Structure Matters

By framing our test this way, we avoid bias and ensure that any observed improvement is genuinely supported by the data, not assumed in advance.

### Null Hypothesis ( $H_0$ )

There is no difference in completion rates between the Control and Test groups.

### Alternative Hypothesis ( $H_1$ )

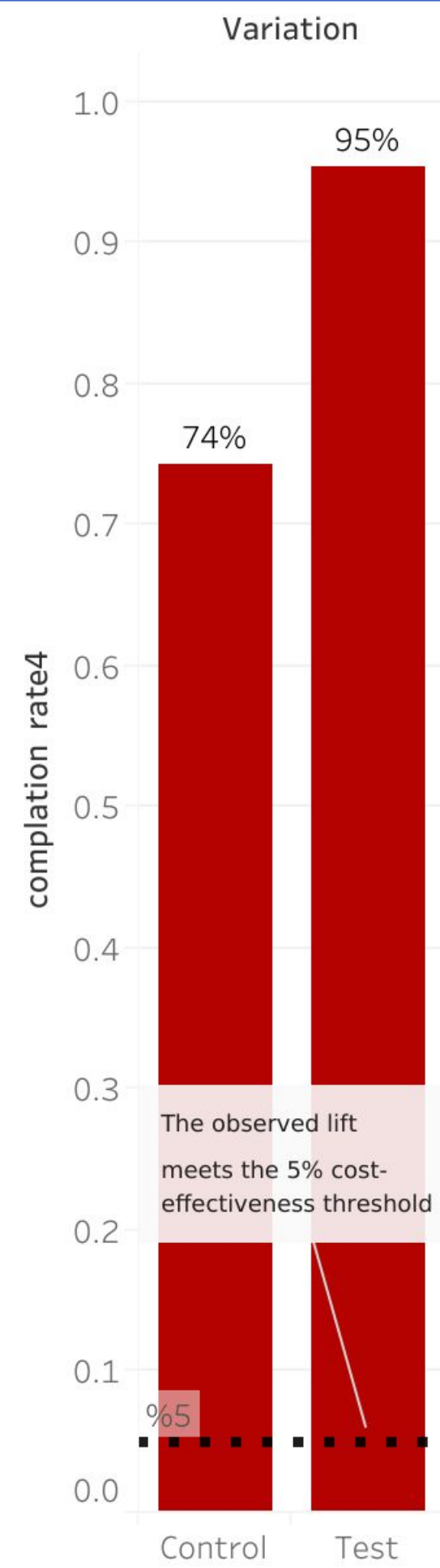
The Test group has a higher completion rate than the Control group.

### Objective

To evaluate whether the new design (Test group) led to more clients completing the process compared to the existing design (Control group)







### Hypothesis Testing Summary

- Z-statistic: -67.05
- P-value: 0.00

### Interpretation:

- A z-test comparing these two proportions would result in a very small p-value ( $p < 0.001$ ).
- →Reject the null hypothesis — the Test group has a significantly higher completion rate.

### Business Impact Summary

- Control Completion Rate: 74.38%
- Test Completion Rate: 95.38%
- Absolute Lift: +21.00%

### Result:

- The observed improvement exceeds Vanguard's 5% cost-effectiveness threshold, suggesting that—regardless of statistical test outcomes—the new UI may be highly effective in practice.

### Key Insight:

- The observed +21% lift exceeds Vanguard's required 5% cost-effectiveness threshold.
- →Strong business impact

The new UI leads to a statistically and practically significant improvement in completion rates.

*Is the conversion rate from 'start' to 'step\_1' is equal for Control and Test groups*

# Hypothesis Testing 2

## Null Hypothesis ( $H_0$ )

The conversion rate from 'start' to 'step\_1' is equal for Control and Test groups

$$\rightarrow p_1 = p_2$$

## Alternative Hypothesis ( $H_1$ )

The Test group has a higher conversion rate

$$\rightarrow p_2 > p_1$$

Test Type:

One-tailed Z-test for two independent proportions

## Objective

To determine whether the new user interface (Test group) leads to a higher conversion rate from the initial 'start' step to 'step\_1' compared to the original design (Control group)

# Hypothesis Testing 2



- *The Test group achieved a +4.89% improvement over the Control group in early funnel conversion.*
- *While this gain falls just short of Vanguard's 5% cost-effectiveness threshold, it signals a promising early impact from the new UI design — particularly at the top of the funnel.*
- *With further refinement, extended testing, or improved targeting, this uplift could potentially exceed the required threshold and justify broader implementation.*

## Findings & Conclusion

### Test Results

Z-statistic: 17

P-value: 0.00

### Statistical Result:

→ Reject the null hypothesis — the Test group has a significantly higher conversion rate from 'start' to 'step\_1'.



## What Is a Funnel Test?

A funnel test evaluates how users progress through a multi-step process, identifying where drop-offs occur and whether a new experience (Test group) improves flow through the funnel compared to an existing version (Control group).

In this case, the funnel includes:

start → step\_1 → step\_2 → step\_3 → confirm

# Key Insights

## Funnel Analysis

**The biggest improvement occurs at the top of the funnel:**

→ Test group outperforms Control by +4.89% from 'start' → 'step\_1'

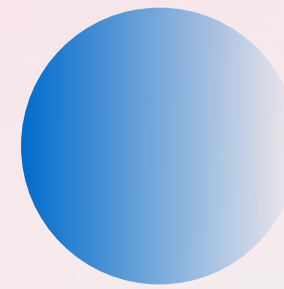
Middle and late funnel steps (Step 2 → Final) are strong and stable in both groups

→ Completion rates remain above 87% in all cases

The new UI helps more users enter and stay in the funnel, which is critical for improving total completions

Small differences in lower steps suggest existing UI is already effective beyond step\_1

# Hypothesis Testing 3



## Hypothesis Testing: Gender and Completion Rate

### Null Hypothesis ( $H_0$ )

$H_0$ : There is no association between gender and completion rate.

### Alternative Hypothesis ( $H_1$ )

$H_1$ : There is an association between gender and completion rate.

**Test Type:** Chi-Square Test of Independence Variables:

Significant

Gender: Male (M), Female (F)

Outcome: Completion (0 = Incomplete, 1 = Complete)

### Control Group:

-F: 2763 incomplete, 4780 complete

-M: 2673 incomplete, 5297 complete

-Chi-square = 16.14,  $p = 0.0001$  -> Significant

Reject  $H_0$ : Significant relationship between gender and completion in this group.

### Test Group:

-F: 2888 incomplete, 5828 complete

-M: 2627 incomplete, 6350

-Chi-Square = 30.70,  $p < 0.0001$  -> Significant

Reject  $H_0$ : Significant relationship between gender and completion in this group.



# Experiment Evaluation



t

## Design Effectiveness

### Structure

- The experiment was well-structured with clearly defined Test and Control groups
- A consistent funnel and outcome framework was used to measure success

## Potential Biases

Client type bias (e.g., older vs. newer users)

Self-selection bias by device, time of day, or entry point (e.g., mobile vs desktop)



# Experiment Evaluation



## ● Duration Assessment

### Experiment Timeline

Start: March 15, 2017

End: June 20, 2017

Duration: ~3 months

## ● Adequacy

Captures a full quarter's worth of client behavior

Includes weekdays and weekends

Minimal seasonal distortion

Timeframe sufficient for statistical insights, assuming reasonable sample size per group

# Experiment

# Evaluation

## Additional Data Needs

To enhance the depth and reliability of the analysis, the following data would be useful:

### Client segmentation

Demographics, risk profile, investment intent

### Traffic source / entry channel

Email, mobile app, referral, direct — to trace origin-based behavior

### Device/browser data

To detect performance or UX differences across platforms

### Qualitative feedback

Client-reported confusion, satisfaction, or usability issues

### Session granularity

Timestamps to analyze pause length, drop-off timing, or multitasking behavior





We used constant update with each other through meetings and slack channel.

Divided work into data prep, EDA, Tableau visuals into 2 parts.

Daily syncs helped ensure smooth collaboration



# Teamwork & Project Management

## 1. Data Visualization Challenges

Visualizing multi-step funnel data required careful structuring to avoid misinterpretation

Selecting the right chart types to convey comparisons and trends took iteration

Tableau's limitations (e.g., reference lines, transparency, exports) created minor workflow friction

## 2. Importance of Focused Hypotheses

Crucial to limit the number of hypotheses and focus on those most aligned with the business question

Too many tests risk diluting insights; choosing the right tests helped us stay efficient and impactful

We prioritized hypotheses that revealed real user behavior differences, not just statistical variations

# Challenges & Learnings


## Learnings:

- Hands-on experience with A/B test evaluation
- Importance of visual storytelling in data products
- The real value of analysis lies in asking the right questions, not just applying methods.
- By aligning our tests with user behavior and business goals, we turned raw data into clear, actionable insights.



## Key Findings

- Full-Funnel Completion Rates
- Control: 74.38% | Test: 95.38%
- +21% Absolute Lift — both statistically and practically significant
- $P < 0.001$  → Reject the null hypothesis

- 
- Top-of-Funnel (Start → Step 1)
  - Control: 85.84% | Test: 90.73%
  - +4.89% Absolute Lift — not statistically significant
  - $P = 0.11$  → Fail to reject the null hypothesis



## Recommendations for Vanguard

### Optimize Early Funnel

Focus improvements on the critical 'Start → Step 1' transition.

### Rerun A/B Test

Improve randomization, timing, and sample balance to strengthen results.

### Gather User Feedback

Use surveys or interviews to uncover early-stage usability issues.

### Segment & Personalize

Analyze by user type (age, device, tenure) to refine the experience.

# Conclusion





# Conclusion

## Verdict

The new UI shows strong overall impact, especially in driving full completion.  
Early-stage gains are promising, though they fall just below Vanguard's 5% cost-effectiveness threshold.

**Verdict:** Promising, but not yet conclusive at top-of-funnel.



---

**Thank You.**  
**Thank You.**  
**Thank You.**