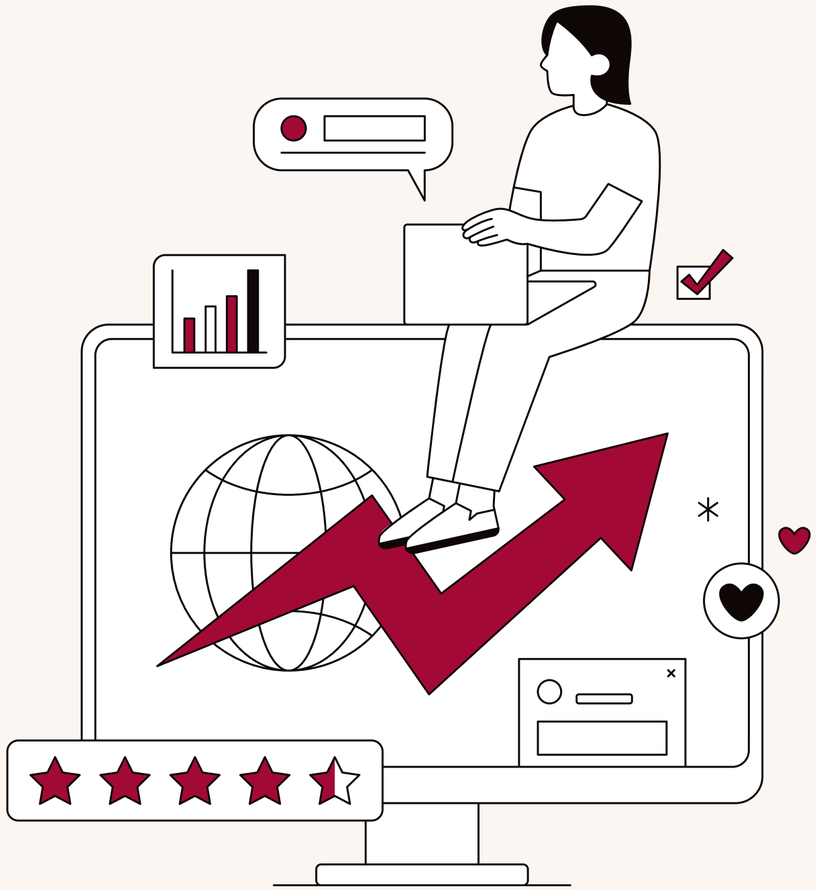
Analysis of results

A/B Test

Beatriz | Ivan | Lucas

Vanguard®



Vanguard & The Digital Challenge

- US-based investment management company
- Intuitive and modern interface
- Smoother online process

Did the new UI lead to higher completion rates?

Data Overview

Client Profiles

Clients demographic:

Age

Gender

Clients behavior:

Tenure

Engagement

Digital Footprints

Records of the clients digital interaction between the current platform and testing platform that has more intuitive and modern User Interface (UI).
It help us understand user behavior on digital platforms and evaluate the effectiveness of different process variations.

Experiment Roster

Focused on experiments showing the assignment of each client to specific variation.
There are two types of variation:

- **Control** - which is related to the group of users selected to use the old User Interface
- **Test** - which is the group of users selected to test the new User Interface

Data Cleaning

Removing Duplicates

We load each dataset from GitHub using raw URLs.
We checked for rows that were exact duplicates (using df.duplicated()) and dropped them if they were not valid or needed.
This ensures each record (row) in the dataset is unique and consistent.

Handling Missing Data

We identified columns with missing values and either filled them with placeholder labels or decided to remove those rows entirely.
This step prevents errors and bias from unaccounted-for blank entries.

Addressing Inconsistent Values & Data Types

We checked for strange codes (like "U" or "X" in gender) and standardized them
We ensured numeric columns if were correctly,
treated as numbers and that string/categorical columns had consistent labels.

Data Merging

We vertically concatenate (pd.concat) the two parts of the web data into a single df_web.

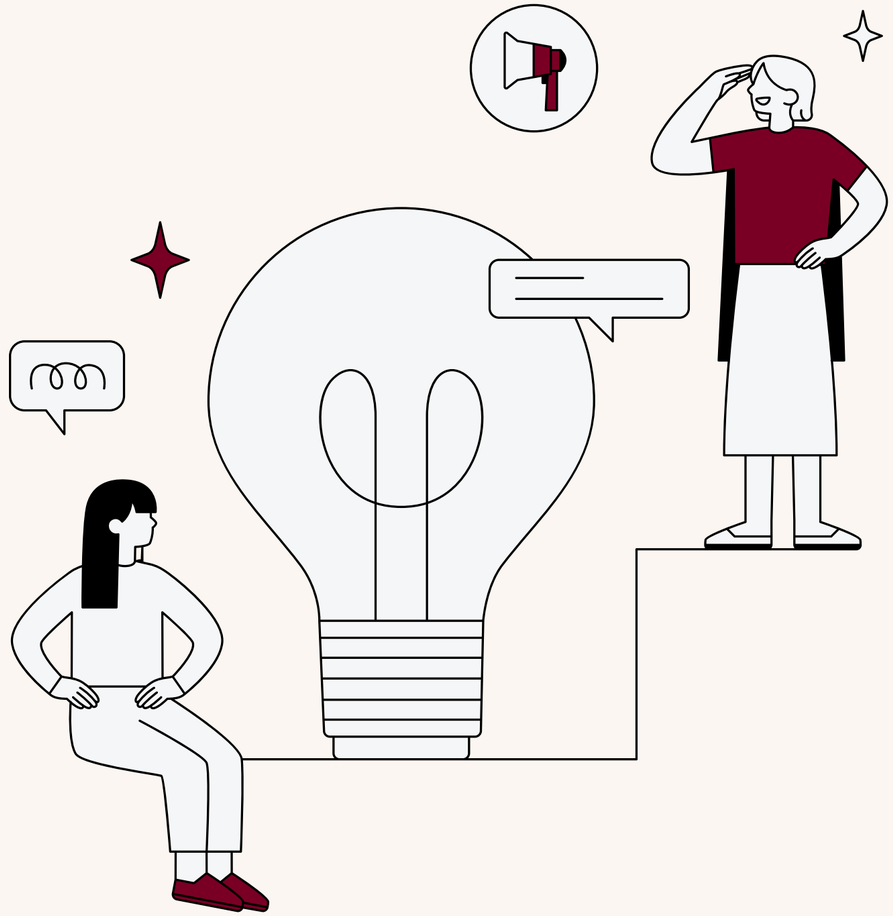
We left-join (pd.merge) that df_web with the demographics df_demo (so all web records remain).

We then left-join the experiment roster df_experiment to identify which group each client is in.

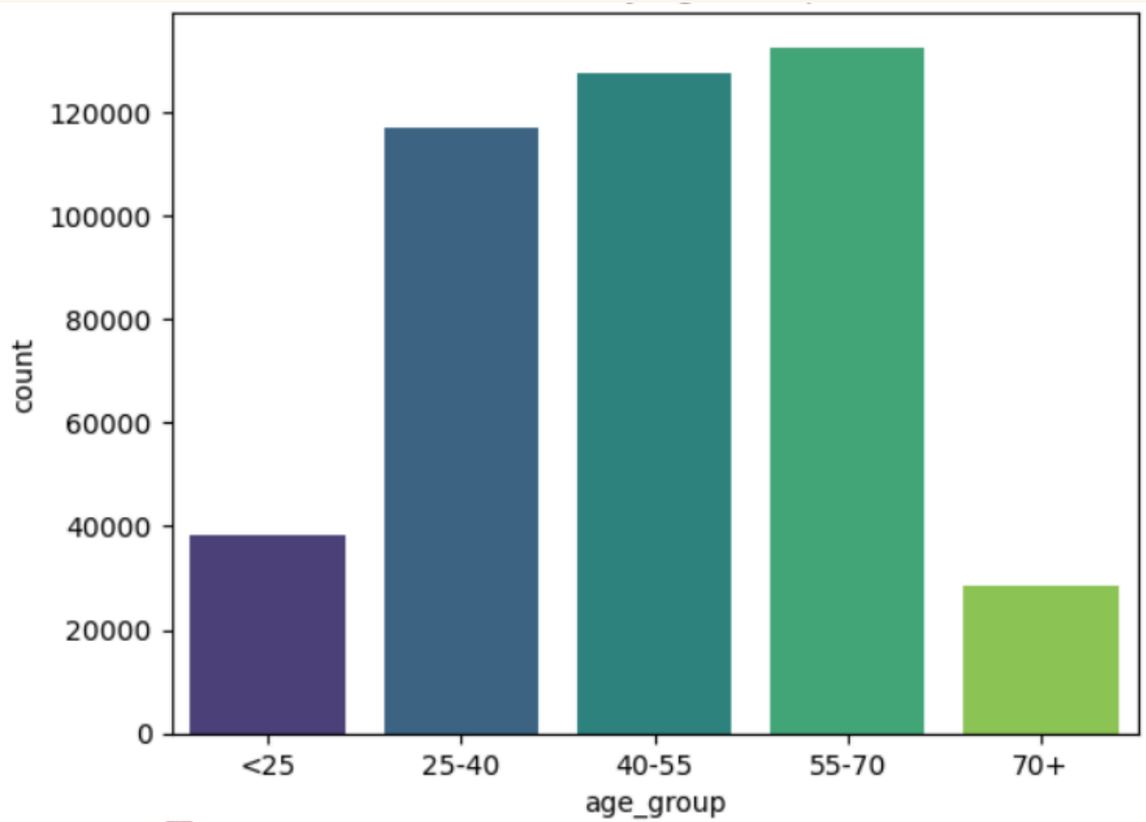
EDA

EDA
Key Demographics and Behaviors

- Age
- Gender & Tenure (Years with Vanguard)
- Process Steps / Online Visits
- Calls (calls_6_mnth)
- Logins (logons_6_mnth)



Exploratory Data Analysis



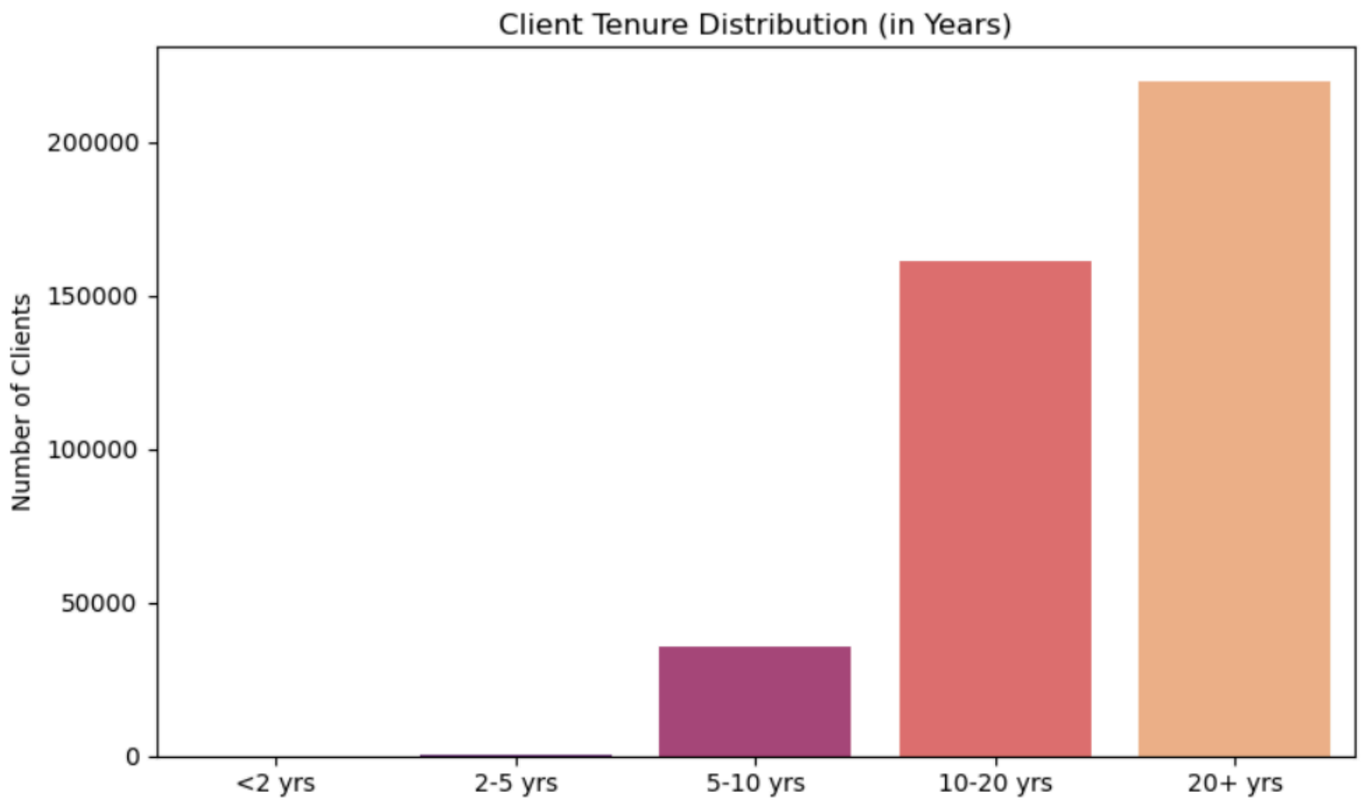
What it shows: ?

A bar chart categorizing by clients into age ranges: <25, 25–40, 40–55, 55–70, and 70+.

Interpretation:

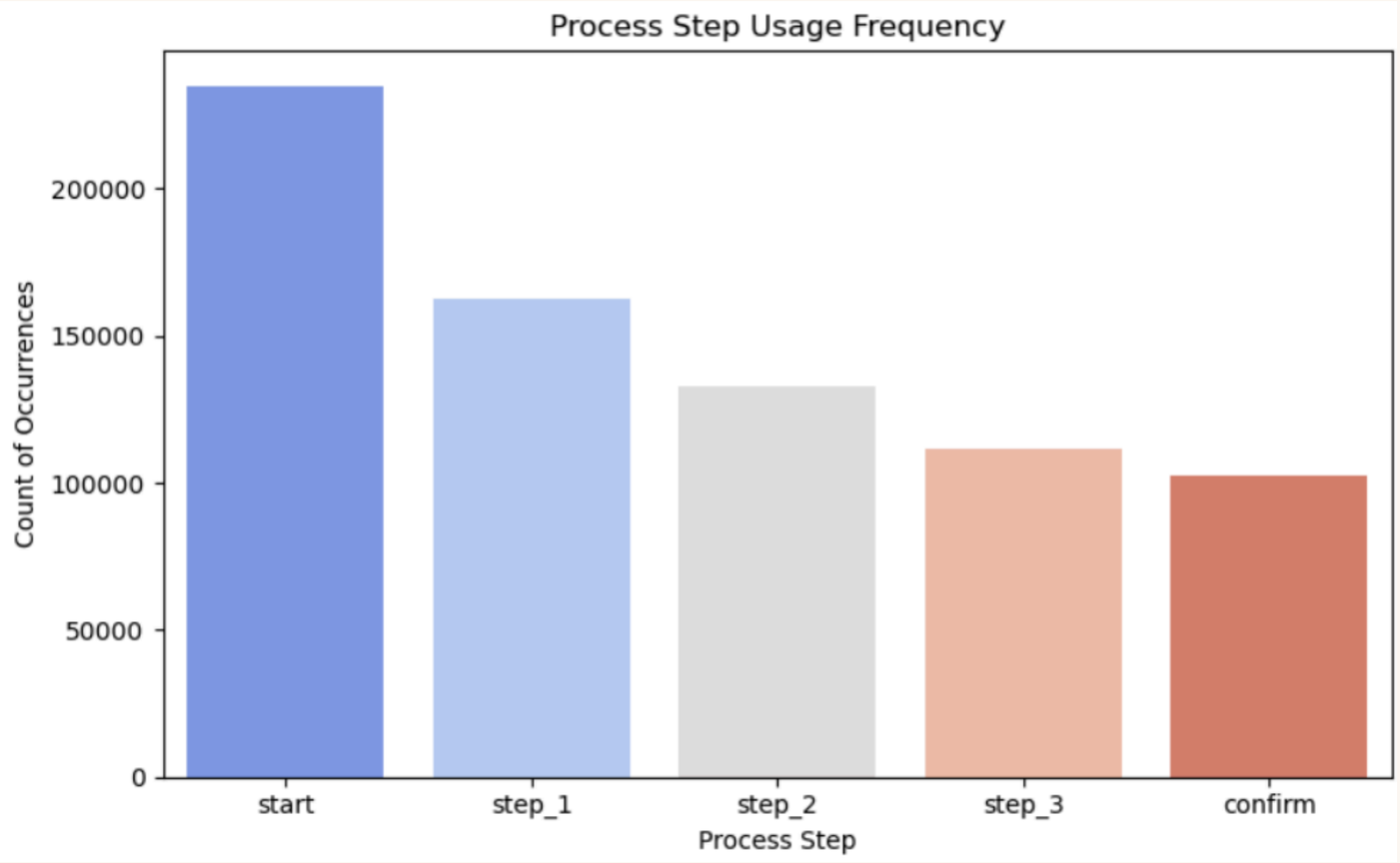
Most clients are in the 40–55 and 55–70 range, followed by 25–40, indicating a mature audience. The <25 and 70+ groups have the fewest clients.

Exploratory Data Analysis



Clients are binned into tenure ranges:
(<2 yrs, 2–5 yrs, 5–10 yrs, 10–20 yrs, and 20+ yrs).
The bar indicate how many
clients have been in for that length of time.
Here, we can also see an overwhelming
majority in the 20+ yrs and 10–20 yrs categories.
This suggests many clients have been with the
organization for a long time,
with relatively few newer
(<2 or 5–10 years) clients.

Exploratory Data Analysis



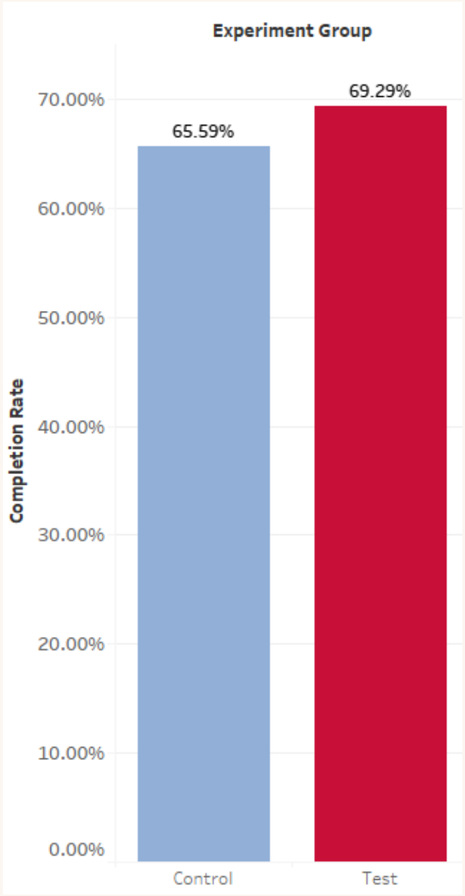
Here we can see the frequence of each “process step” occurs .
(e.g., start, step_1, step_2, step_3, confirm).
Think of it like a funnel: most users enter at the “start,” and fewer continue through each step.
The largest bar is “start,” followed by progressively smaller bars for step_1, step_2, etc.
This indicates drop-offs at each stage of the process.
The final “confirm” step has fewer users,
showing not everyone completes all steps.

KPIs

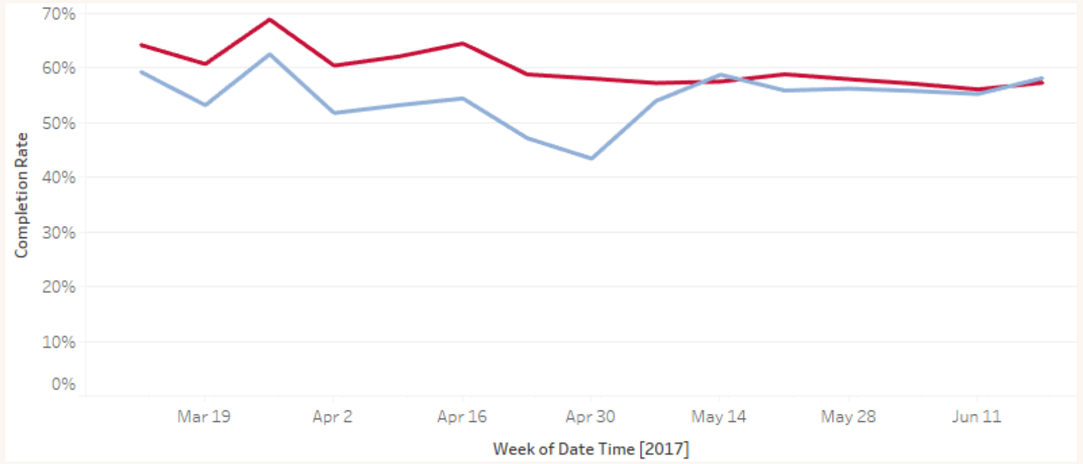
Completion Rate

Control
Test

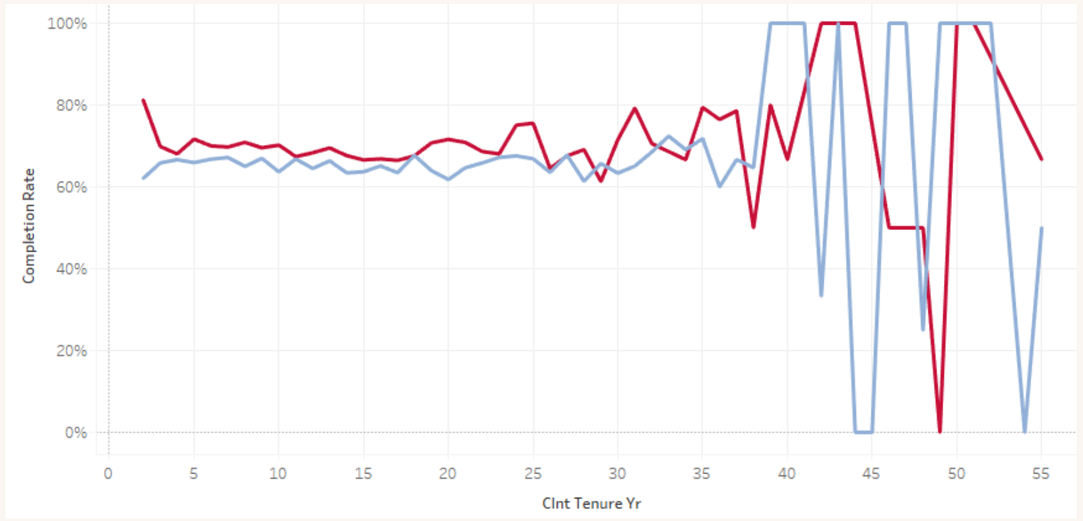
Completion Rate



Completion Rate Over Time



Completion Rate Per Tenure

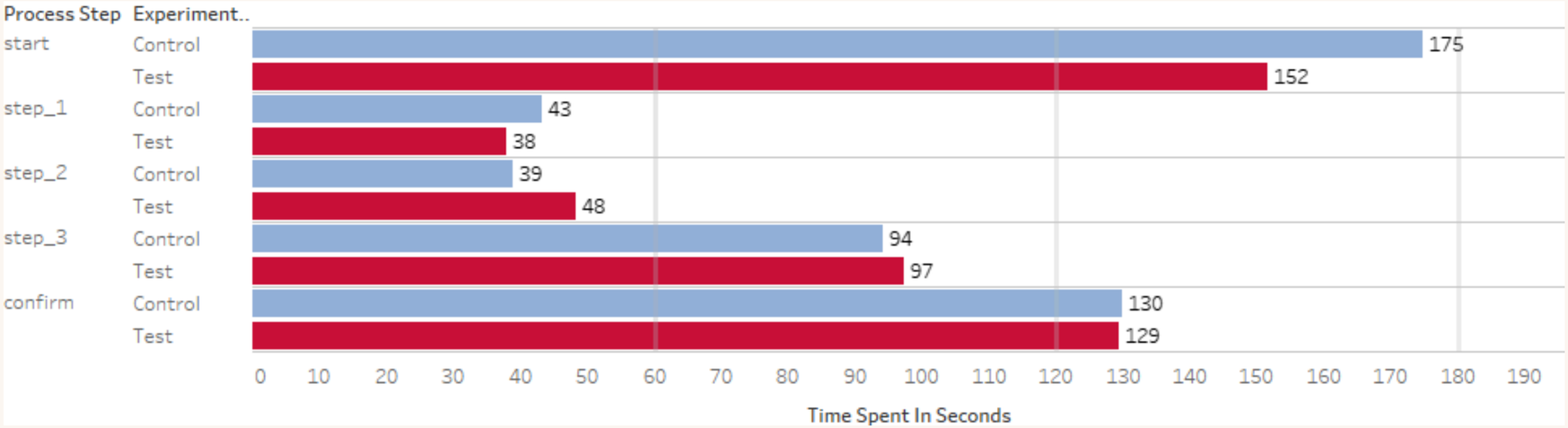


- There is a **difference of 3.7%** in terms of Completion Rate between both groups, which indicates that the **Test group** had a **positive impact** for this KPI.
- The **Test group** consistently **outperforms** the Control group over time. Both **Completion Rates tend to stabilize** as time goes by.
- The **Test group** maintains **higher Completion Rates in earlier tenures**. The variability for later tenures is most likely related to the amount of users analyzed.

Time Spent on Each Step

Control
Test

Average Time Spent on Each Step

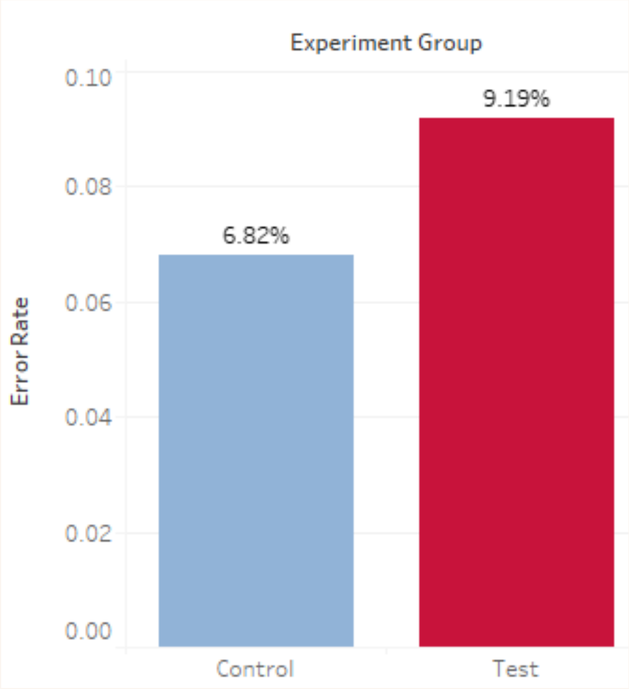


- The Test group shows significant improvements when compared to the Control group for both the Start and Step 1.
- The increased time spent on Step 2 by the Test group could indicate more complexity on this step or require more user interaction. To prove this we would need more information in regards to this step to further analyze it.
- The **differences in time for Step 3 and Confirm are minimal**, which does not indicate a significant impact.

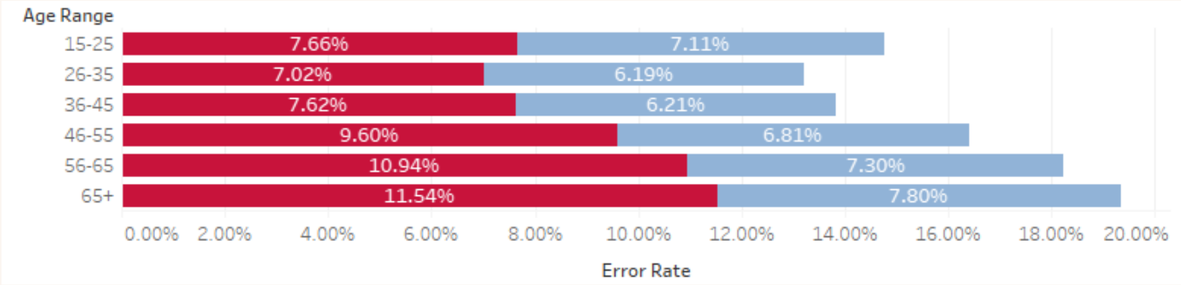
Error Rates

Control
Test

Error Rate

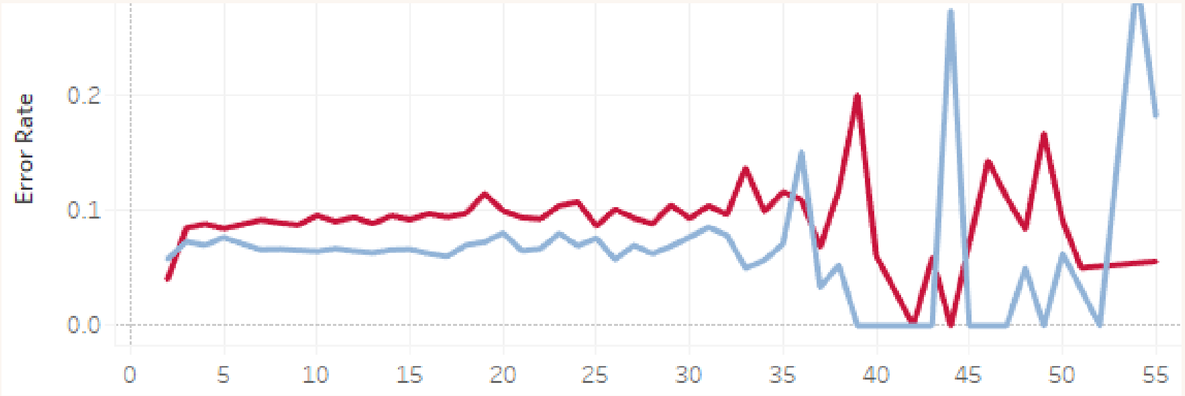


Error Rate Per Age Range



As we can see **older users** tend to be **particularly affected by the new UI**, which leads us to question the complexity of the features introduced to the Test group.

Error Rate Per Tenure



The results in terms of **longer tenure users are impacted by the low amount** of these users. However, this can also indicate that **changes in familiar workflows create confusion**.

Hypothesis

Hypothesis Testing
Completion Rate

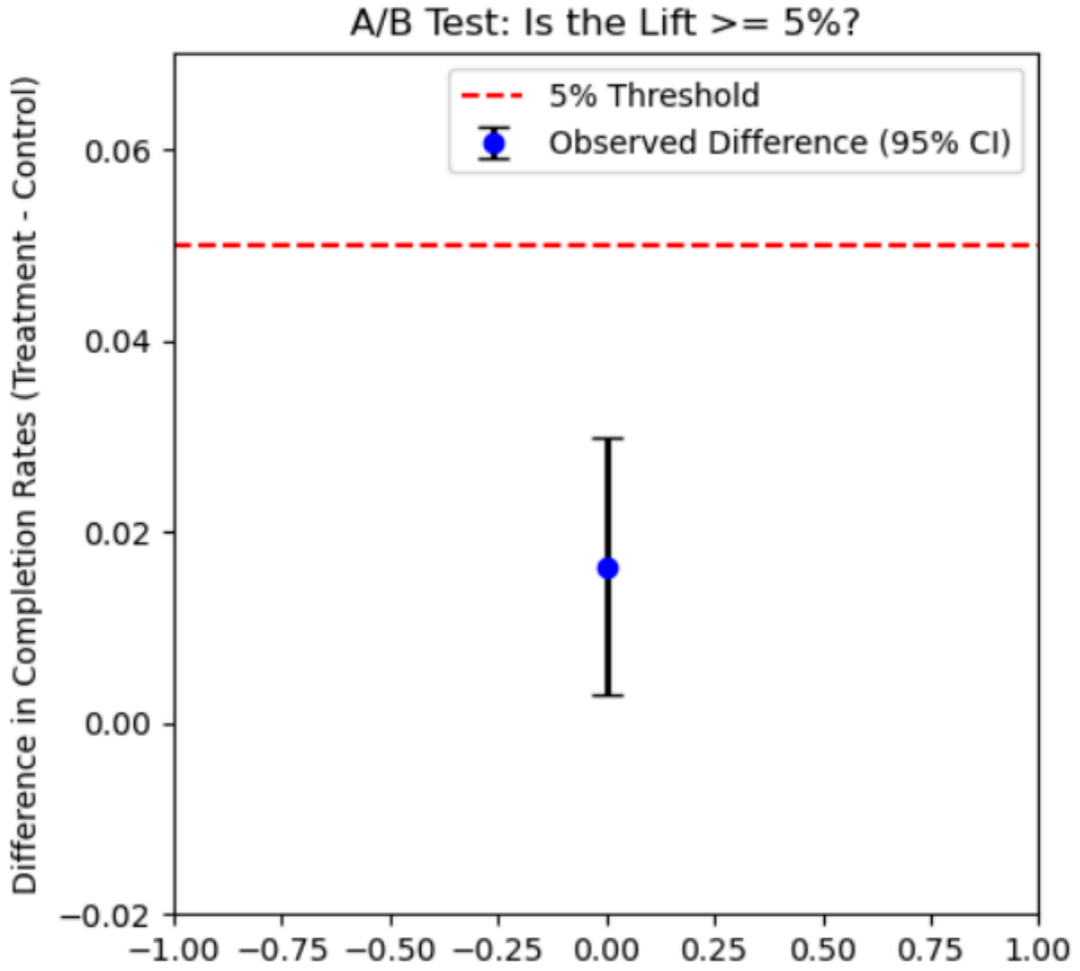
- **Null Hypothesis:** The completion rate for the Test group (new design) is equal to the completion rate for the Control group (old design).
- **Alternative Hypothesis:** The completion rate for the Test group (new design) is not equal to the completion rate for the Control group (old design).

Significance Level (α)	Z-Statistic	P-Value
0.05	8.875	0.0

- **Test used:** Z-Test for Two Proportions, which is the most appropriate when comparing two different proportions.
- As we can observe, the P-Value is 0, which indicates a significant difference between the test and control rates. As we can attest by the previously analyzed KPI.

Control Completion Rate	Test Completion Rate
65.59%	69.29%

Hypothesis Testing
Cost Effectiveness Threshold



Blue Dot, Shows how much higher (or lower)
the Treatment completion rate is compared to Control.

Black Error Bars (95% CI),
Indicate the range
of plausible values for that difference.

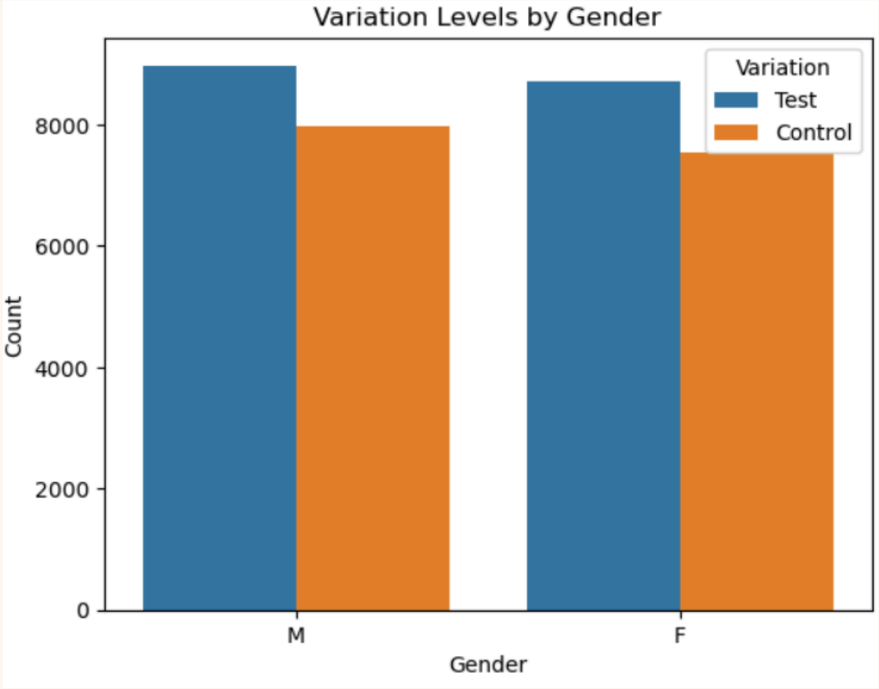
Red Dashed Line (5% Threshold):
Marks the minimum improvement needed for cost-effectiveness.
In this example, the dot and confidence interval lie below 5%,
suggesting the Treatment does not meet the 5% improvement target.

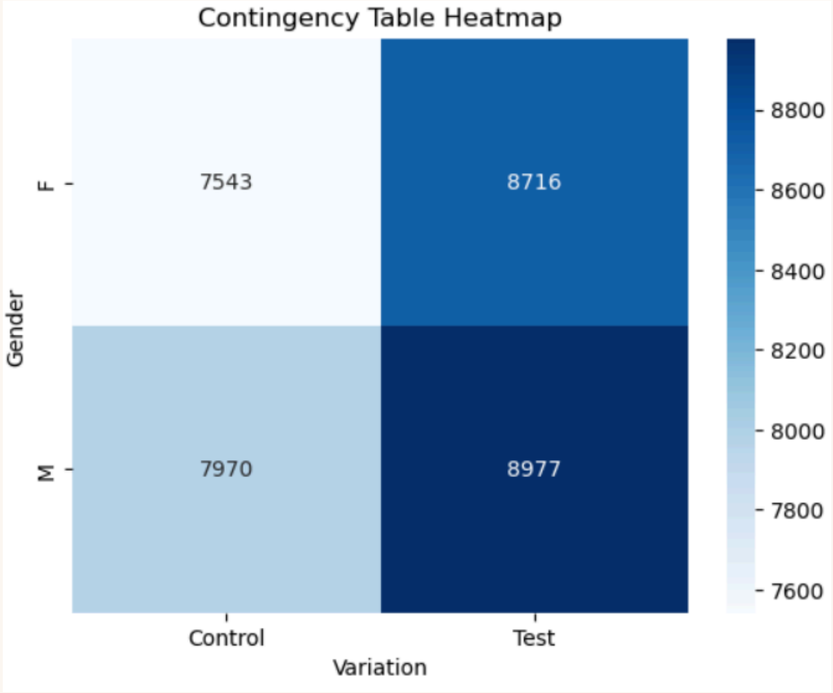
01

Hypothesis Testing
Gender Impact on Process Engagement

H0: Gender differences does not affect engaging with the new or old process
H1: Gender differences is affecting engaging with the new or old process

Based on the Chi-square test, with a Chi2 statistic of 1.324 and a p-value of 0.25, we cannot reject the null hypothesis. This suggests that gender differences do not significantly affect whether individuals are included in the new or old process. The data visualization further supports this conclusion by showing an even distribution.





Experiment Evaluation

- The experiment was well structured, featuring well-defined tasks and a clear concise hypothesis
- Good client distribution in terms of age and gender as well as experiment group.
- The timeframe was enough to gather insights but in order to stabilize the Error and Completion Rate we would need more time.
- The experiment might be biased due to having many participants who've been clients for a long time. This could skew results towards the habits of long-term clients, missing newer clients' perspectives
- Incorporating geographic information, device usage details (desktop vs. mobile), and customer feedback could enhance our analysis by providing deeper insights into client behavior and satisfaction.

Conclusion

Based on our analysis we came to this key findings:

1. **Competition Rate:**
- Old UI design is 65.59%
- New UI design is 69.29%(increase of 3.7%.)

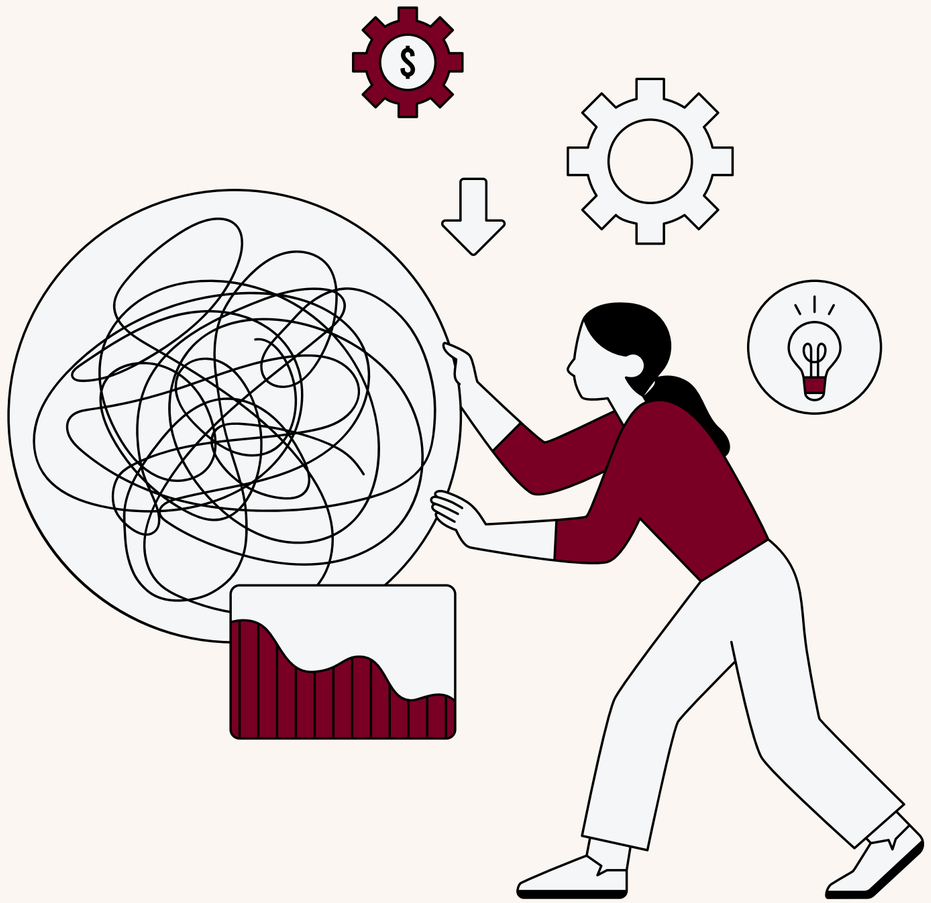
2. **Error rate:**
- Old UI design is 6.82%
- New UI design is 9.19% (increase of 2.37%)

3. **The average time users spent on each step:**
- The Test group did better than the Control group in the Start and Step 1. However, they spent more time on Step 2 and Step 3. This suggests these steps could be more complicated or require additional user interaction. We need further data to verify this..

The Confirm step showed little time difference, indicating no major impact there.

Conclusion

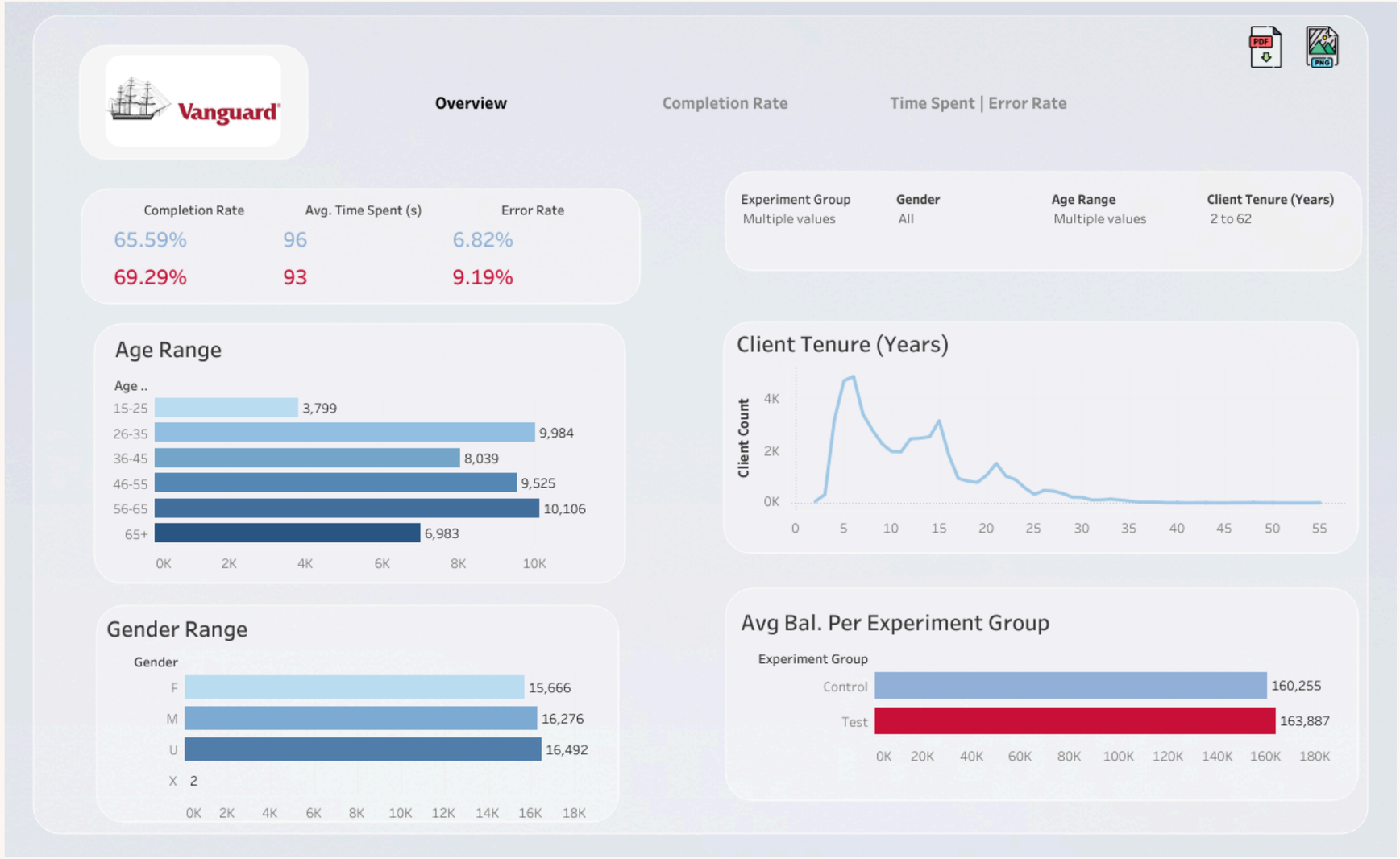
The new interface is definitely making changes! We've seen a solid 3.7% boost in completion rates, and users are getting through faster with a 3-second cut in their average time spent.
But, as expected with any fresh design, there's been a slight hiccup of 2.37% rise in error rates. Overall, the new look is working well, and there's plenty of potential to smooth out the errors and make things even better.



Offer recommendations for Vanguard:

1. Geographical Analysis: Study user behavior by region to find unique patterns and challenges.
2. Customer Feedback: Gather and analyze feedback to identify and fix design issues.
3. Error Rate Mitigation: Investigate and address the causes of increased errors to enhance user satisfaction.

Tableau


Vanguard
Overview
Completion Rate
Time Spent | Error Rate
Completion Rate
Avg. Time Spent (s)
Error Rate
65.59%
96
6.82%
69.29%
93
9.19%
Experiment Group
Multiple values
Gender
All
Age Range
Multiple values
Client Tenure (Years)
2 to 62
Age Range
Age ..
15-25 3,799
26-35 9,984
36-45 8,039
46-55 9,525
56-65 10,106
65+ 6,983
0K 2K 4K 6K 8K 10K
Client Tenure (Years)
Client Count
4K
2K
0K
0 5 10 15 20 25 30 35 40 45 50 55
Gender Range
Gender
F 15,666
M 16,276
U 16,492
X 2
0K 2K 4K 6K 8K 10K 12K 14K 16K 18K
Avg Bal. Per Experiment Group
Experiment Group
Control 160,255
Test 163,887
0K 20K 40K 60K 80K 100K 120K 140K 160K 180K

Thank you very much!