Data analytics - Bootcamp June-2024

— Vanguard’s experiment

João Sousa

Introduction

Vanguard is one of the world's largest investment management companies, known for its wide range of mutual funds and exchange-traded funds (ETFs).

Recently, the company invested in improving the website layout

Did the new UI lead to higher completion rates?

Data Overview

- **Collected User Information:** Client ID, Gender, Age, Number of Accounts per Client, Date and Time of Interaction, Visit ID, Balance, Client Tenure, Process Steps and more.

Experiment design:

Initiated an A/B test from March 15, 2017, to June 20, 2017.

- **Control Group:** Clients interacted with Vanguard's traditional online process.
- **Test Group:** Clients experienced the new, spruced-up digital interface.

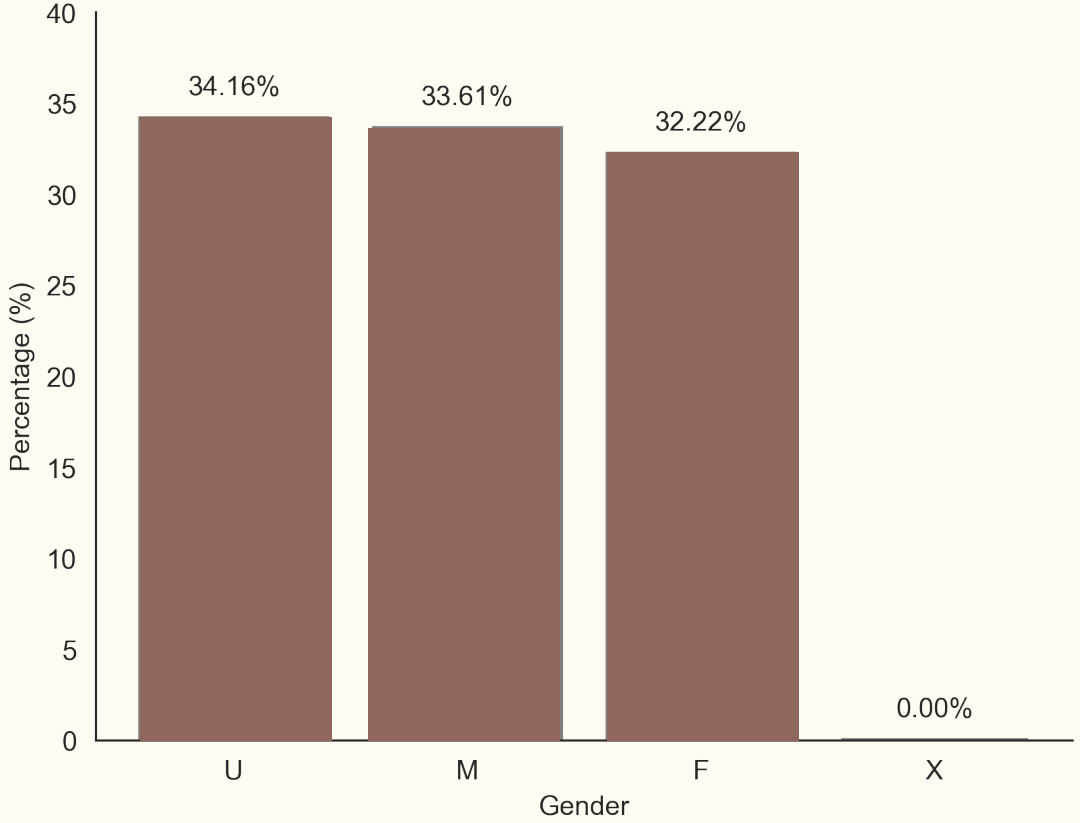
Both groups followed the same sequence: an initial page, three steps, and a final confirmation page

The goal is to determine if the new design improves user experience and increases process completion rates

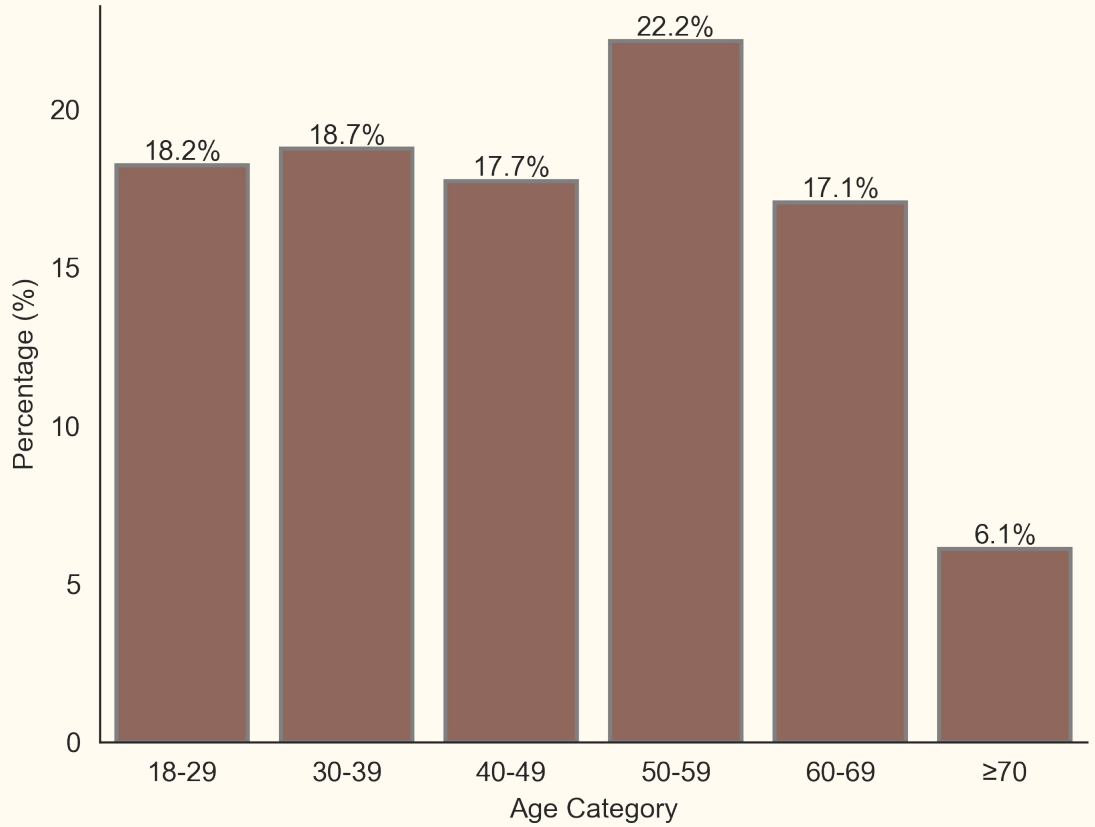
Data handling

- **Data wrangling**
  - Combining datasets.
  - Cleaning data.
- **Exploratory analysis**
  - Exploring data before diving in A/B test.
- **Performance metrics**
  - Define the KPIs to evaluate the new design's performance.
- **Hypothesis tests and Analysis**
  - Define hypothesis tests.
  - Data visualization of A/B test.

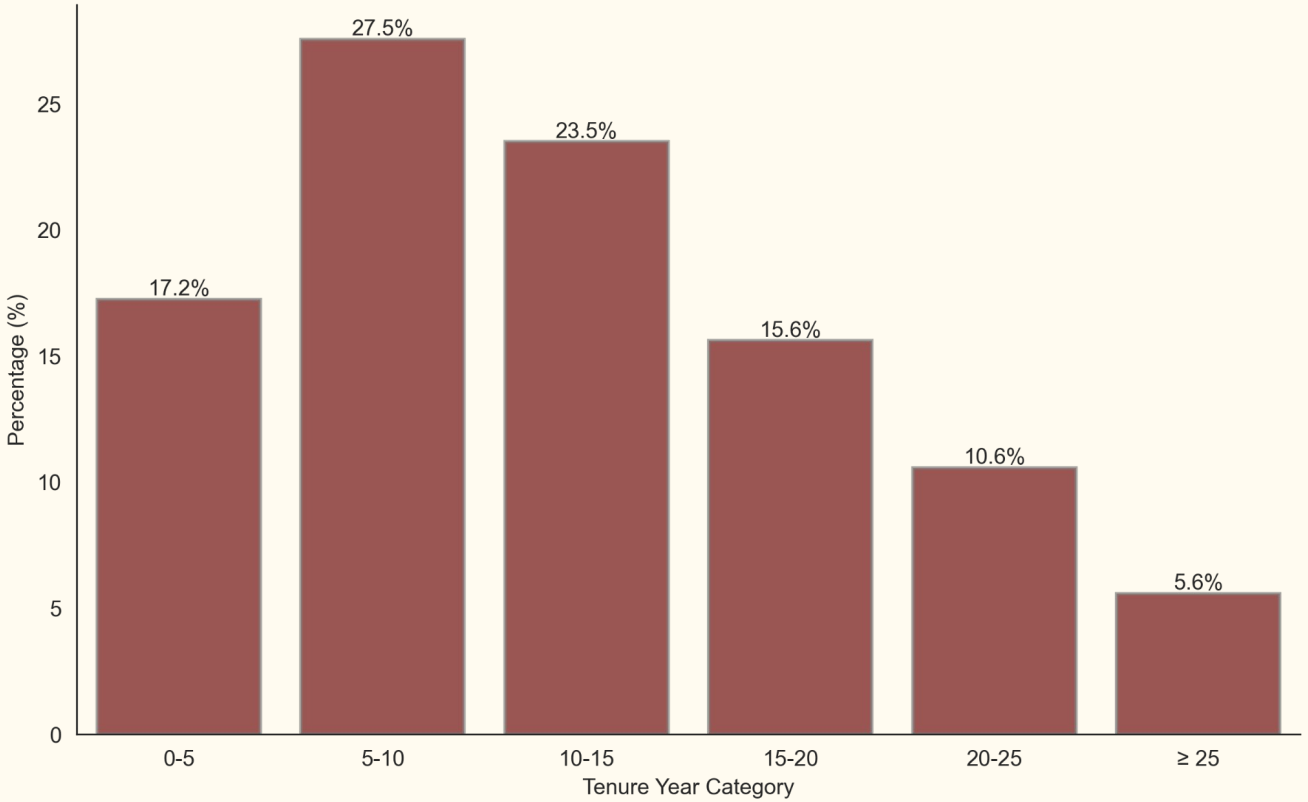
Exploratory analysis: Gender by percentage(%)



Exploratory analysis: Age of the clients per category



Exploratory analysis: Tenure of the clients



Performance metrics

KPI's:

- **Time Spent on Each Step:** The average duration users spend on each step.
- **Completion Rate**: The proportion of users who reach the final 'confirm' step.
- **Error Rates:** If users go back to a previous step, it may indicate confusion or an error. Transitioning from a later step to an earlier one should be considered as an error.

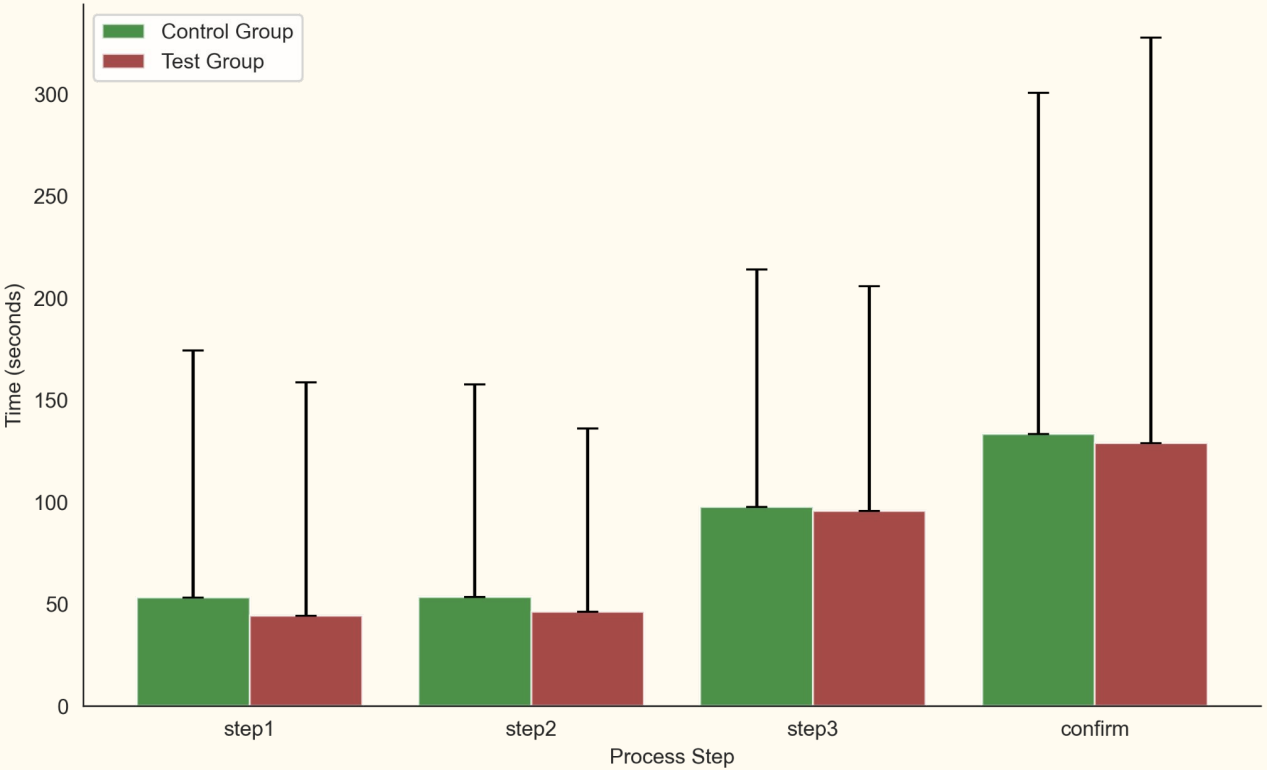
Hypothesis tests

- **In these analysis, users who spent more than 30 minutes on any step in either group were excluded.**
- **Time Spent on Each Step :**
  - Test-t - determine if there is a significant difference between the means of two groups.
  - Null hypothesis (H_0) : Don't exist differences between each step in control and test.
  - Alternative hypothesis(H_a): Exist differences between each step in control and test.

Hypothesis tests

- **Completion Rate:**
  - Test-Z - comparing proportions or percentages between two groups.
  - Null hypothesis (H_0) : Don't exist differences between "confirm" step in control and test.
  - Alternative hypothesis(H_a): Exist differences between "confirm" step in control and test.
- **Error Rates:**
  - Test-Z - comparing proportions or percentages between two groups.
  - Null hypothesis (H_0) : Don't exist differences between each step in control and test.
  - Alternative hypothesis(H_a): Exist differences between each step in control and test.

Data Visualization: Time spend on each step

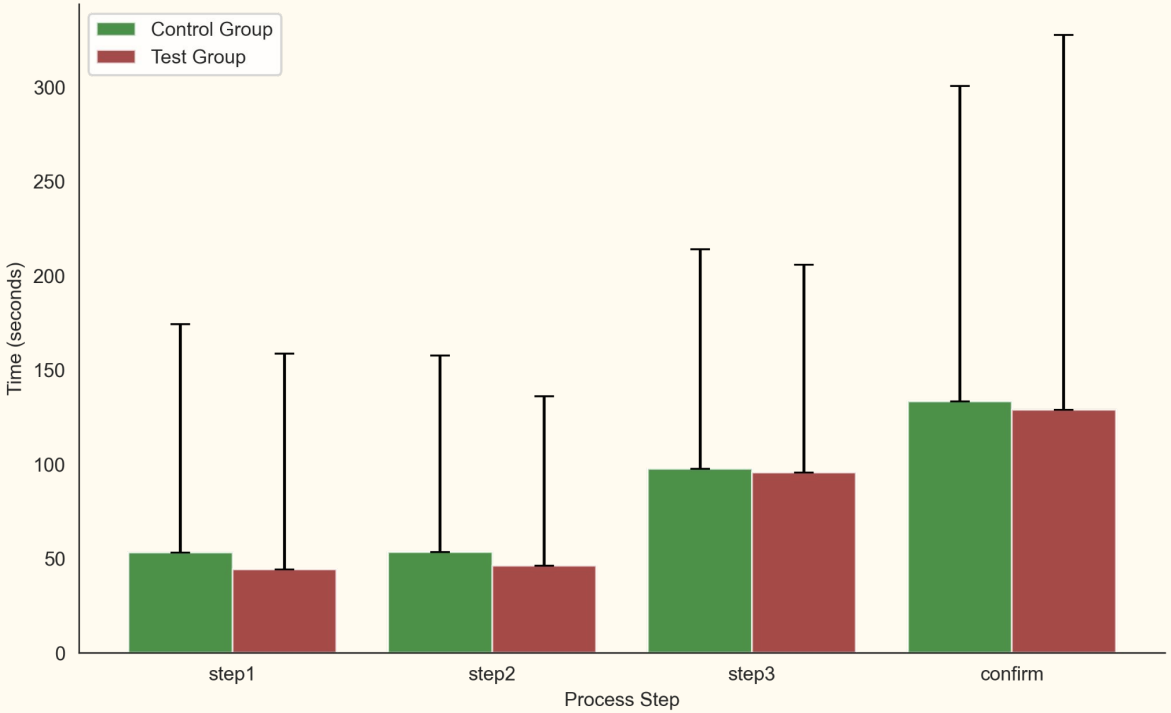


Time spend on each step : Results test-t

	Step 1	**Step 2**	**Step 3**	**Confirm**
T-statistic	7.42	6.9	1.86	2.14
P-value:	1.17e-13	3.55e-12	0.06	0.03

H_0 is rejected when $p < 0.05$

Data Visualization: Time spend on each step



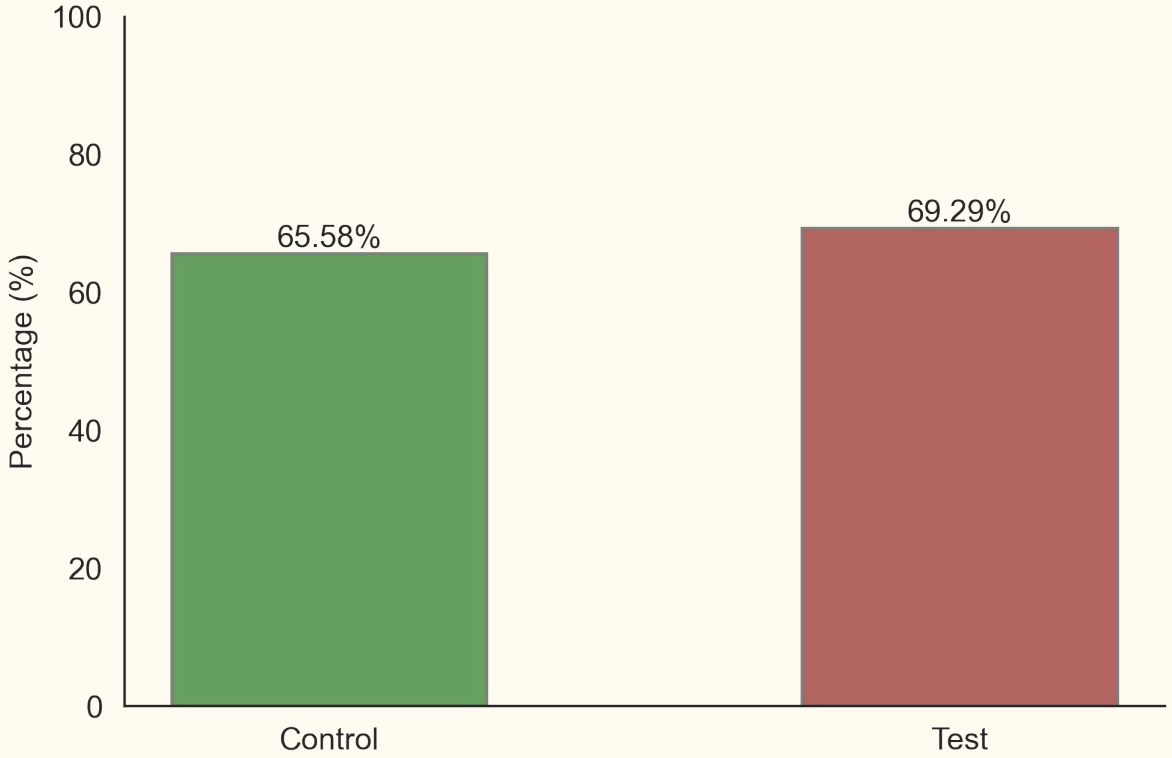
Step 1: Reject null hypothesis

Step 2: Reject null hypothesis

Step 3: Accept null hypothesis

Confirm: Reject null hypothesis

Data Visualization: Completion rate

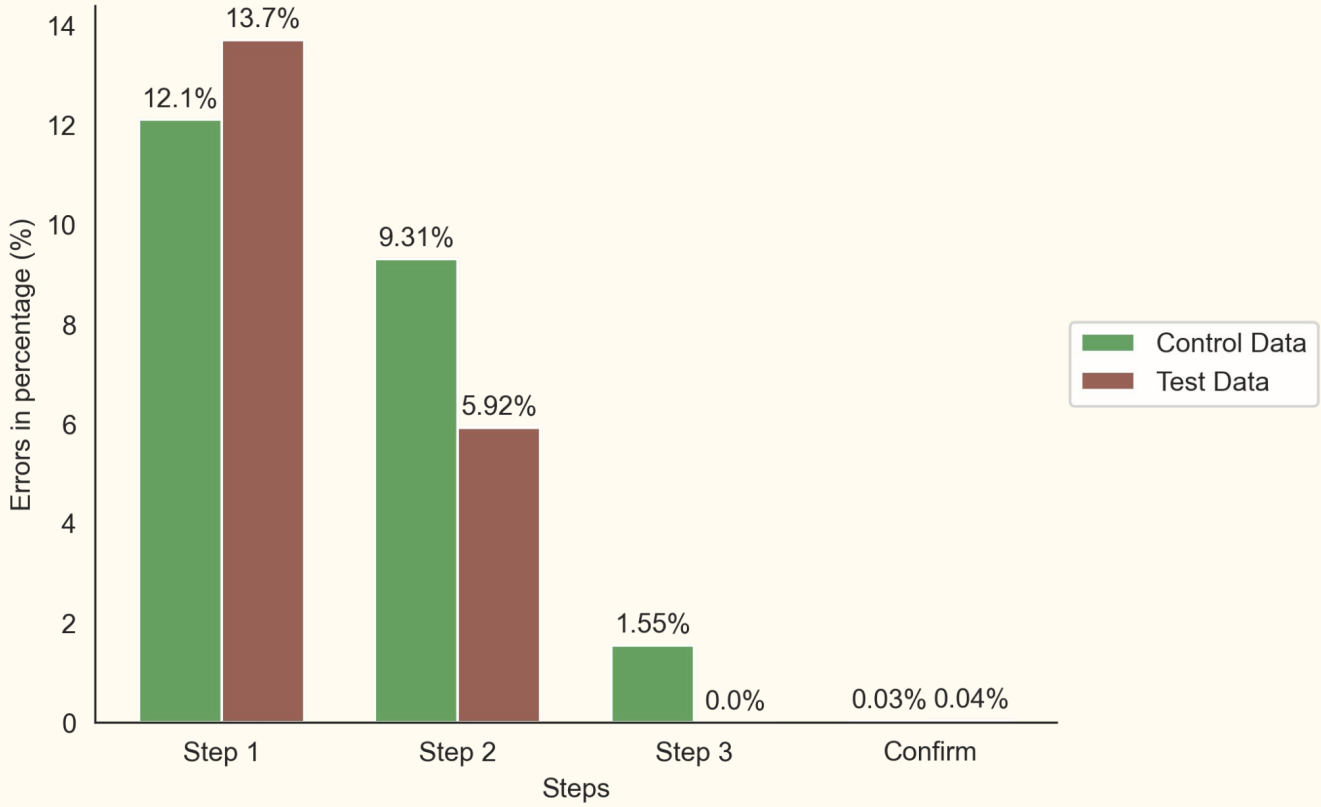


Completion rate: Result test-z

Test statistic (z):	-5.89
P-value	3.75e-09

H_0 is rejected when $p < 0.05$

Data Visualization: Error rates

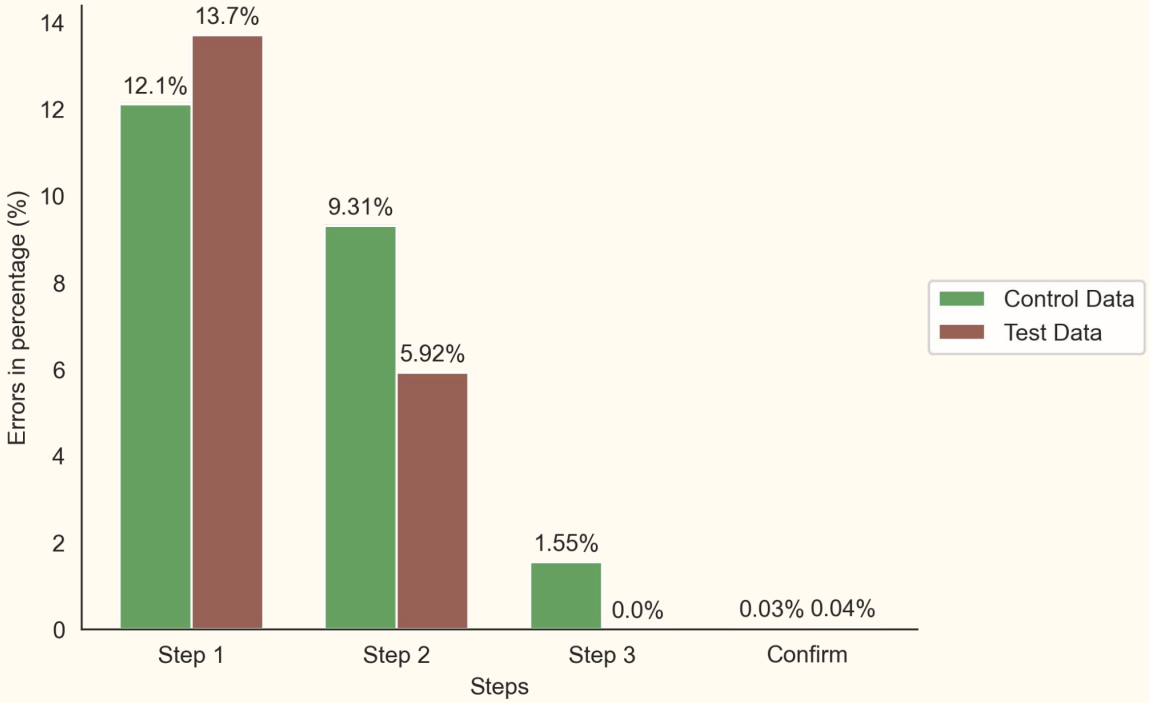


Erros rates: Result test-z

	Step 1	**Step 2**	**Step 3**	**Confirm**
T-statistic	-4.47	11.86	17.07	-0.50
P-value:	7.48e-06	0.0	0.0	0.61

H_0 is rejected when $p < 0.05$

Data Visualization: Erros rates



Step 1: Reject null hypothesis.

Step 2: Reject null hypothesis.

Step 3: Reject null hypothesis.

Confirm: Accept null hypothesis.

Results

- **Time Spent on Each Step**
  - Statistical differences exist between control and test groups in the average time spent on steps 1, 2, and confirm, with the test group spending less time. However, no differences were found in step 3.
- **Completion Rate**
  - Statistical differences exist between control and test groups in the proportion reaching the confirmation step, with a 3.71% increase observed in the test group.
- **Error Rates**
  - Statistical differences exist between control and test groups in error proportions for steps 1, 2, and 3, but not in the confirmation step. Errors increased in step 1 but decreased in steps 2 and 3 for the test group.

Conclusions

- **Time Spent on Each Step**
  - Observed improvements include reduced time spent on each step in the test group.
- **Completion Rate**
  - It does not meet the company's threshold, as they expected a minimum 5% increase to justify the investment.
- **Error Rates**
  - Overall, errors decrease throughout the process, but the first step has a higher error rate than the control, potentially causing users to abandon the session early.

Despite some progress, the outcomes don't meet company standards and the investment isn't justified.

Recommendations

Design a new test incorporating all the improvements from this analysis. We observed an increased ratio of reaching the final steps. With some adjustments to the previous test, it will definitely surpass the company's defined threshold.

Vanguard®

Tableau Visualisations

Thank you!