Vanguard - A/B Test Analysis

House Tully
By Anna & Katya

Introduction

Vanguard Context

Vanguard, a US-based investment management company, conducted a digital experiment over a duration of three months, which required a thorough analysis of the results.

Digital Challenge

Vanguard believed that a more intuitive, modern UI, combined with timely in-context prompts (like cues and instructions), could streamline the online process for clients.

Analysis question: Did the new UI lead to higher completion rates?

Data Overview

Client Profiles

Demographic data on Vanguard's clients, including gender, age, tenure and interactions with the company.

Digital Footprints

Summary of interactions with the Vanguard UI for each client, divided into two files.

Experiment Roster

Distribution of clients between Test and Control groups.

Data Merging and Cleaning

Used client_id field to merge all data sets and perform initial EDA with Pandas, Seaborn and Matplotlib.

01

02

Removed empty rows and duplicates.

Created additional fields such as step indexes and errors to calculate and visualize the KPIs.

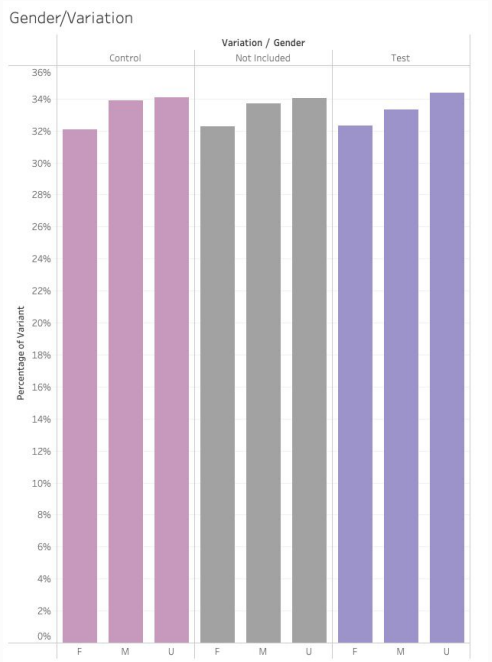
03

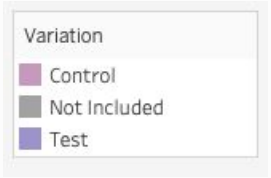
Used IQR approach to remove outliers for analysis of account balance and time on step.

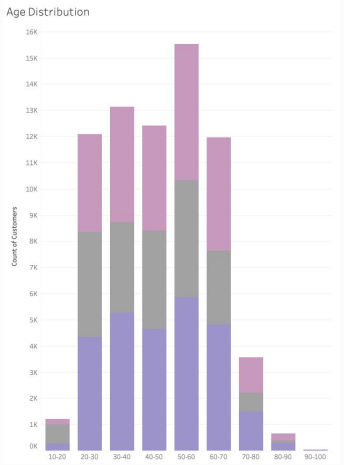
04

Customer Demographics

- Gender distribution is even across all groups but unknown for ⅓ of customers.
- Most Vanguard customers are aged 50-60.
- Very few customers are younger than 20 or older than 80.

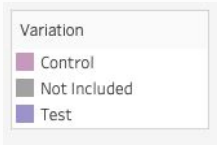


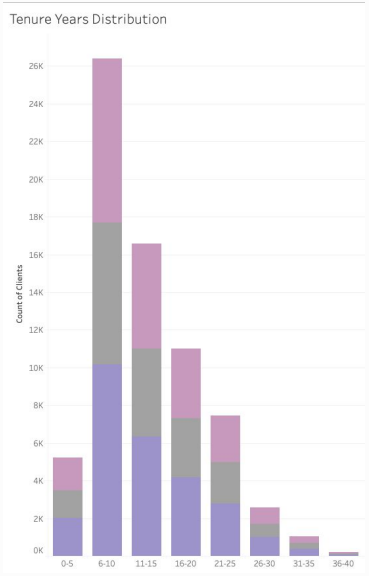


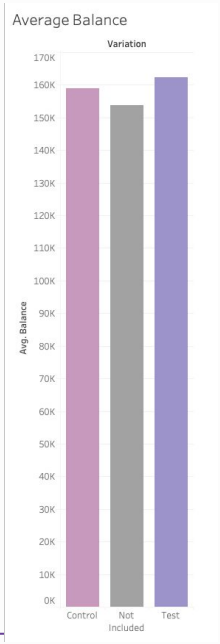


Customer Demographics

- **Customer Tenure:** Most customers have been with the company for 5-6 years.
  However, the share of new customers is relatively low.
- **Average balance**: around $150k.
- **Outliers**: balances exceeding $12M and $16M.
- **Experiment Exclusions:** Customers excluded from the experiment have slightly lower average balances.
- **Additional Characteristics**: The number of accounts, logins, and calls are evenly distributed across groups.

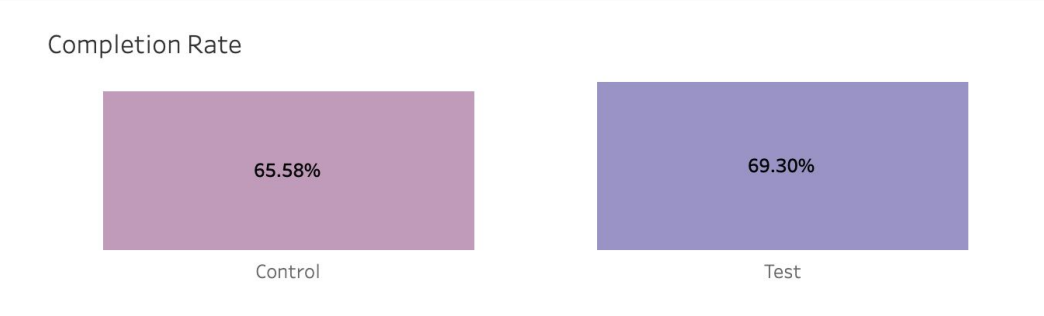






KPI: Completion Rate

- **Completion Rate:** The proportion of users who reach the final ‘confirm’ step.
- **CR =** (Number of unique customers who reached the confirmation step) / (Total number of unique customers)



Completion Rate Hypothesis 1

- **H0:** The **CR** is not significantly different between the Test and Control groups.
- **H1:** The **CR** of the Test group is significantly higher than that of the Control group.

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed proportions z-test Alpha = 0.05	• Z-Statistic: 8.89 • P-Value: < 0.0001	Since the p-value is much lower than 0.05, we reject the null hypothesis, **concluding that the Test group has a significantly higher CR than the Control group.**

Completion Rate Hypothesis 2

- **H0:** The **CR** for the Test group is equal to or less than the CR for the Control group **increased by 5%.**
- **H1:** The **CR** for the Test group is greater than the CR for the Control group **increased by 5%.**

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed proportions z-test Alpha = 0.05	• Z-Statistic: -160.09 • P-Value: 1.0	Based on the p-value, we failed to reject the null hypothesis, **concluding that the Test group's CR did not exceed the Control group's by more than 5%.**

KPI: Error Rate

- **Error** = Any action that is not in the expected sequence of steps; can be multiple for one customer ID.
- **Error Rate = Count of errors / Total count of actions.**
- The result for both groups was around 22%.

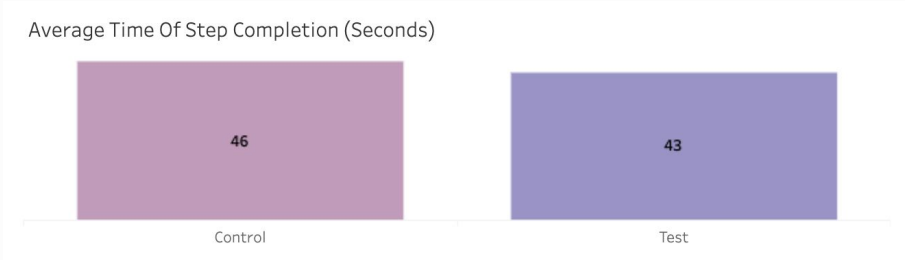
Error Rate Hypothesis

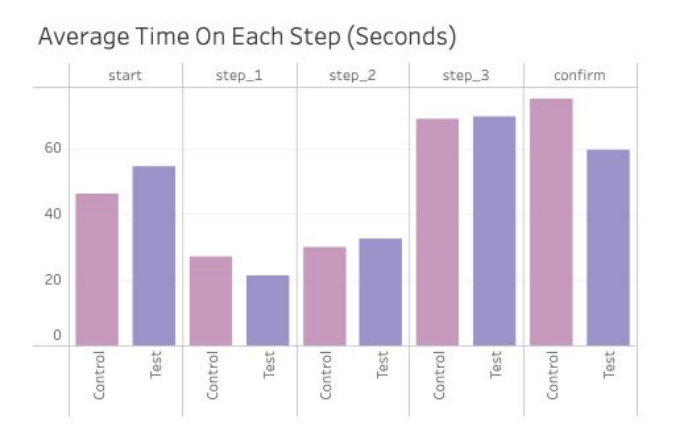
- **H0: The error rate** between the Test and Control groups **is the same.**
- **H1: The error rate** of the Test group is **lower than the Control group.**

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed t-test Alpha = 0.05	• Z-Statistic: 0.68 • P-Value: 0.75	We cannot reject the null hypothesis, concluding that **the difference in error rates** between the Test and Control groups is **not significant.**

KPI: Time on Step

- **Time on Step = The difference between timestamp of current step of the flow and the timestamp of the previous step, expressed in seconds.**
- We excluded all errors (backward steps) and outliers.
- The average Time on Step was 46 seconds for the Control group and 43 seconds for Test, but the results for each step varied.





Time on Step Hypothesis

- **H0: The average time** spent on a step by clients in the test group is **equal** to that of clients in the control group.
- **H1: The average time s**pent on a step by clients in the test group is **less than** that of clients in the control group**.**

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed t-test Alpha = 0.05	• Z-Statistic: 0.68 • P-Value: 0.75	Since the p-value is much lower than 0.05, we reject the null hypothesis, concluding that the **Test group spent significantly less time on steps compared to the Control group.**

Average Age Hypothesis

- **H0:** The **average age** of clients engaging with the new process is the same as those engaging with the old.
- **H1:** The **average age** of clients engaging with the new process is lower than as those engaging with the old.

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed t-test Alpha = 0.05	• Z-Statistic: 7.83 • P-Value: ~1.0	Based on the p-value, we **failed to reject** the null hypothesis**, concluding that the average age of clients engaging with the new process is not lower than those engaging with the old process.**

Client Tenure Hypothesis

- **H0:** The average **client tenure** of those engaging with the new process is the same as those engaging with the old process.
- **H1**: The average **client tenure** of those engaging with the new process is lower than those engaging with the old process.

HYPOTHESIS TEST	RESULTS	OUTCOME
One-tailed t-test Alpha = 0.05	• Z-Statistic: 0.57 • P-Value: 0.716	Based on the p-value, we **failed to reject** the null hypothesis, concluding that the average client tenure of **those engaging with the new process is not significantly lower** than those engaging with the old process.

Experiment Evaluation

Experiment Design

The experiment was well-structured, but variable balance suggests a slight bias toward higher-profile accounts.

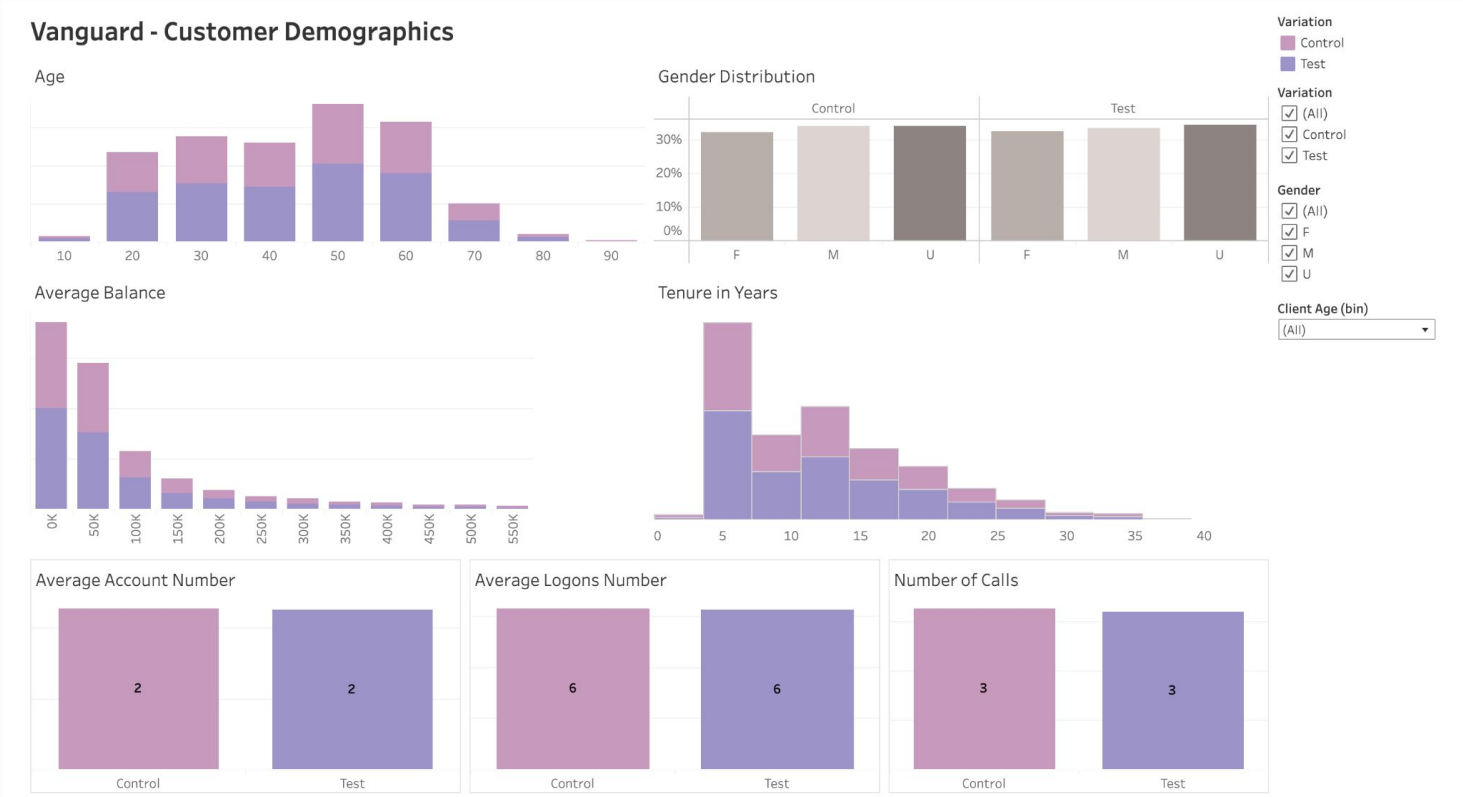
Duration

The timeframe was sufficient for analyzing the Completion Rate and Time on Step KPIs. Based on the power analysis of the Error Rate, extending the experiment likely wouldn't have led to more significant results.

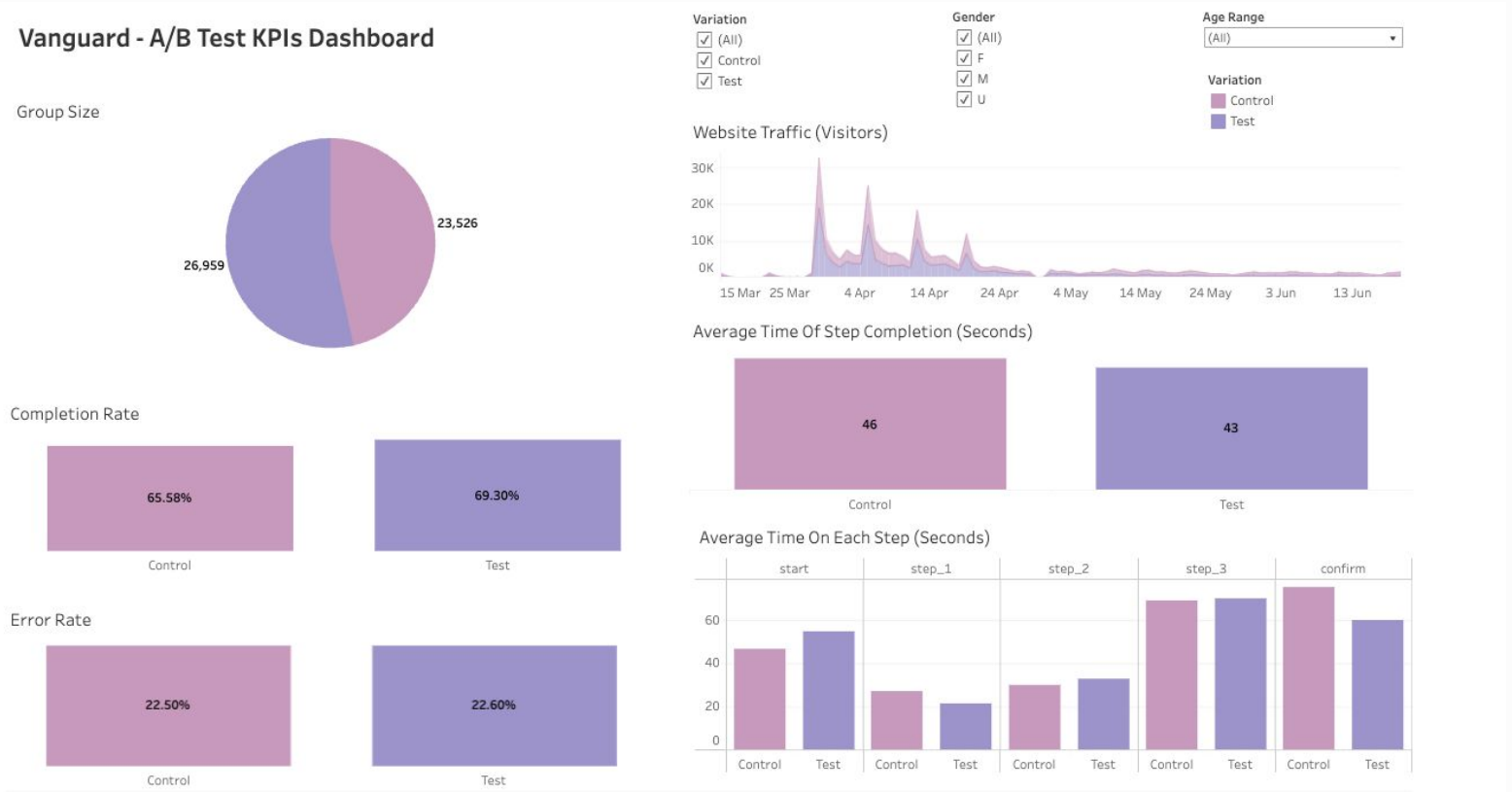
Additional Data Needs

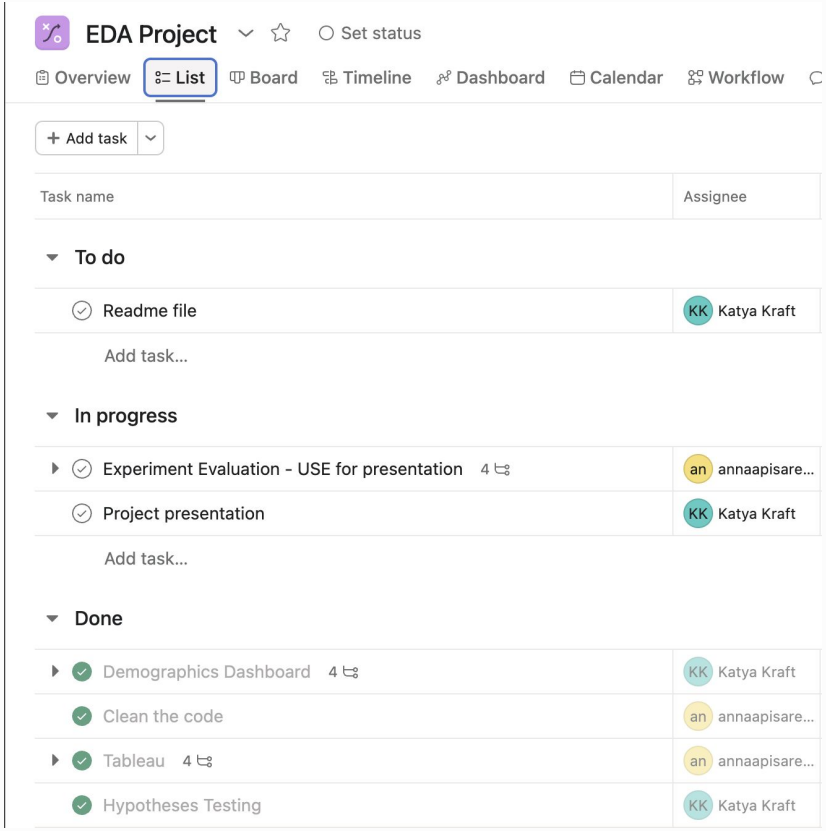
More gender data and insights on device usage (desktop vs. mobile) would have enhanced the analysis.

Customer Demographics Dashboard



KPIs Dashboard





Teamwork & Project Management

- We used an **Asana** board from the beginning, which helped **structure our workflow** and provided a clear overview of each project step.
- While we **collaborated on the same code**, we divided the creation of the **two dashboards**, as working on the same Tableau file simultaneously wasn't feasible.
  This ensured smooth progress without conflicts.

Challenges & Learnings

Time on Step KPI

Calculating this metric was complex and required additional research to implement correctly.

Sharing Large Data Files

We resolved this by using Google Drive for easier collaboration.

KPI Definitions and Outliers

Defining KPIs and handling outliers was challenging, so we experimented with various approaches before finalizing the solutions.

Conclusions

Completion Rate

The Test group's completion rate was higher than the Control's, but **did not meet the 5% threshold to justify the new UI's cost.**

Time Spent on Steps

The Test group spent significantly less time on steps, showing better efficiency with the new UI.

Error Rate

There was no significant difference in error rates, indicating the new UI didn't reduce errors.

Average Age and Tenure

No significant difference in age or tenure between the groups, meaning the new UI didn't attract different demographics.

Recommendations

Expand the Testing Scope

Test the interface with a larger or more diverse group of users to see if certain groups benefit more.

Review the UI changes

Focus on improving the features that help users complete tasks and reduce mistakes.

Iterative Design

Consider a step-by-step redesign to address key issues without the high cost of a full overhaul.

Thank you!

House Tully
By Anna & Katya