Analysis on Vanguard’s new user interface design (US-based investment management company)

-Govarthini G S

Introduction

Problem statement

The primary goals are to improve user engagement and increase the completion rate of the process. Specifically, Vanguard aims to determine if the new design leads to a higher completion rate and lower error rate while also considering the cost-effectiveness of the new design implementation.

Hypotheses

I have formulated several hypotheses to test various aspects of the new design's effectiveness

- Completion Rate Hypotheses
- Error Rate Hypotheses
- Average time spent on each step
- Cost-Effectiveness Hypotheses
- Demographics Hypotheses

Data Overview

Timeline

The experiment was conducted from March 15, 2017, to June 20, 2017.

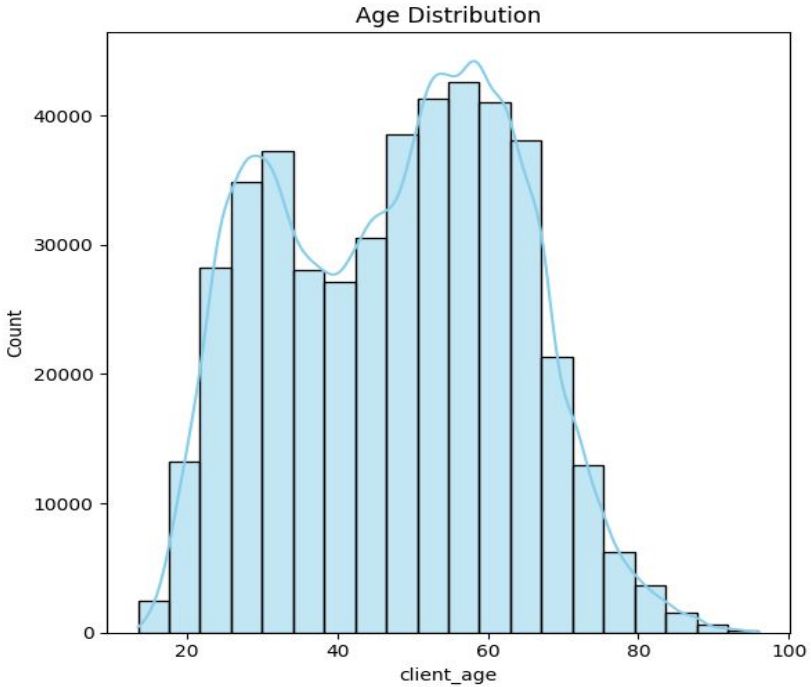
Key Components

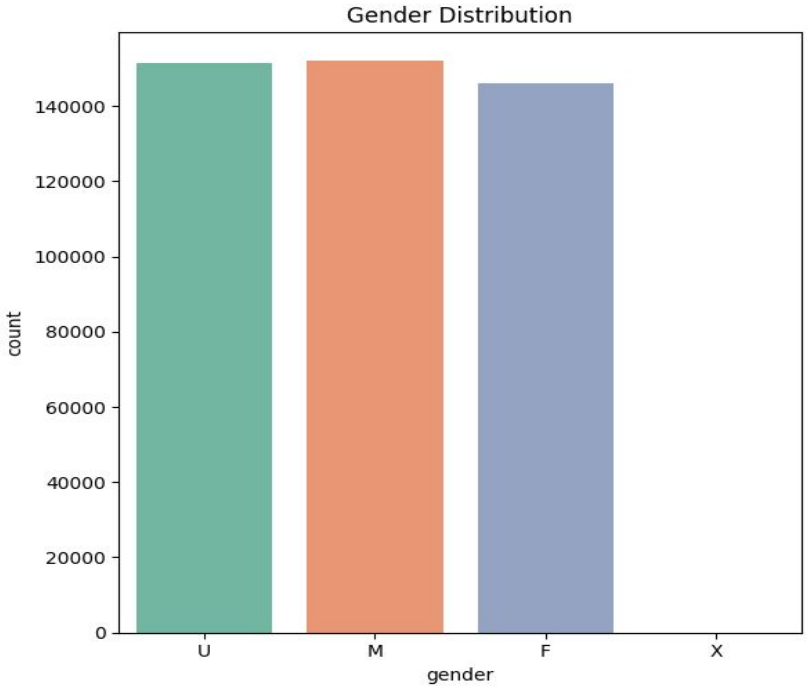
1. **Control Group**: This group interacted with the traditional online process.
2. **Test Group**: This group interacted with the new, enhanced digital interface.
3. **Process Sequence**:
    - Initial page
    - Step 1
    - Step 2
    - Step 3
    - Confirmation page (indicating process completion)

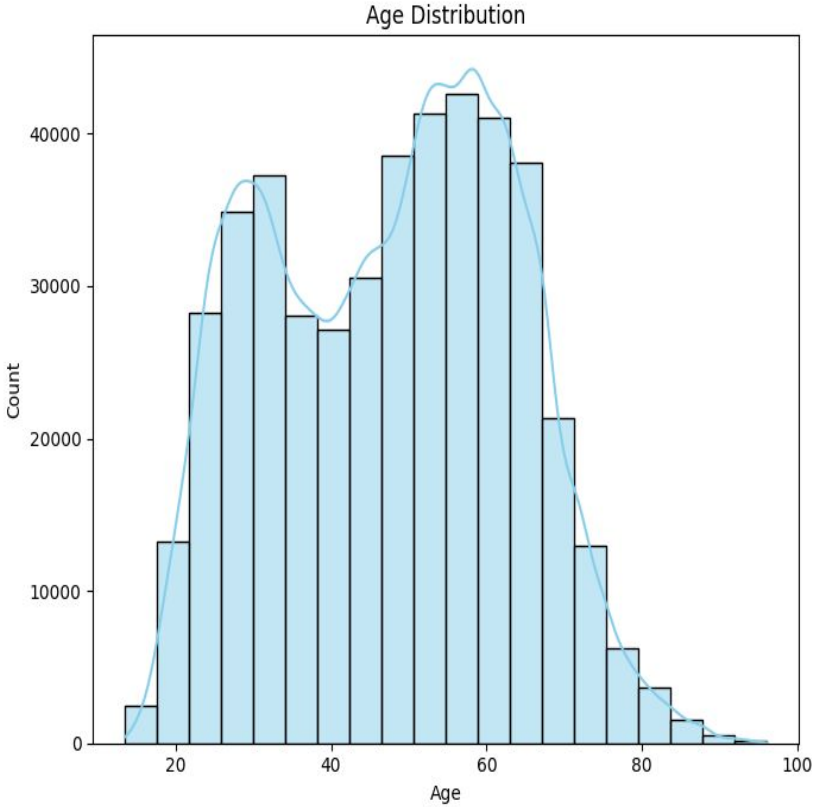
Data Wrangling

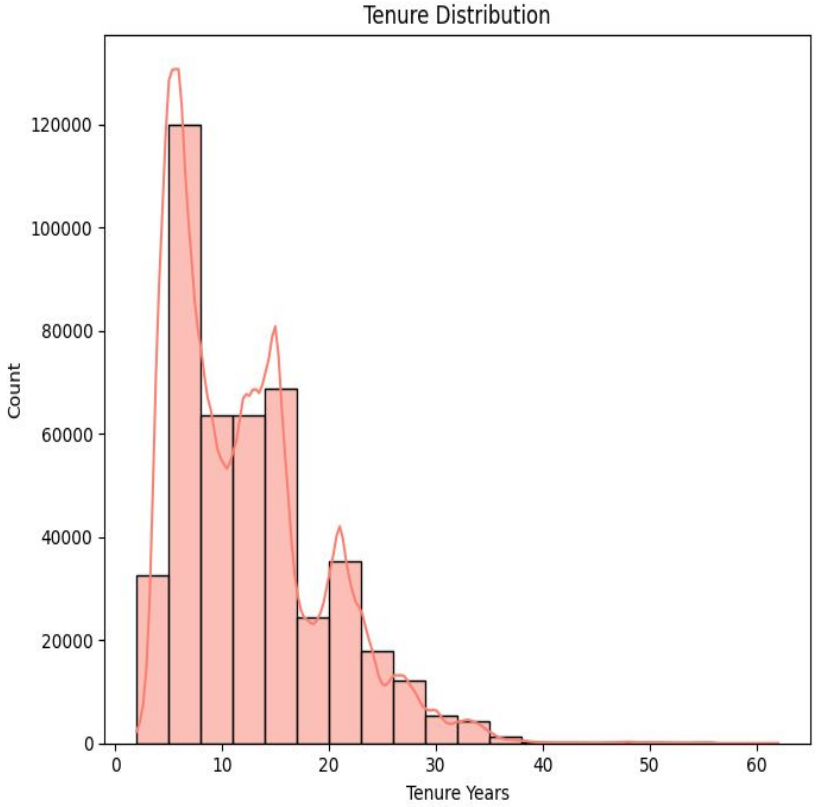
- **Checking for null values** : df.isnull().sum()
- **Handling missing values** : df= fill na ()
- **Checking for duplicates:** df.duplicated()
- **String formatting**: df.rename()

Exploratory Data Analysis(EDA)

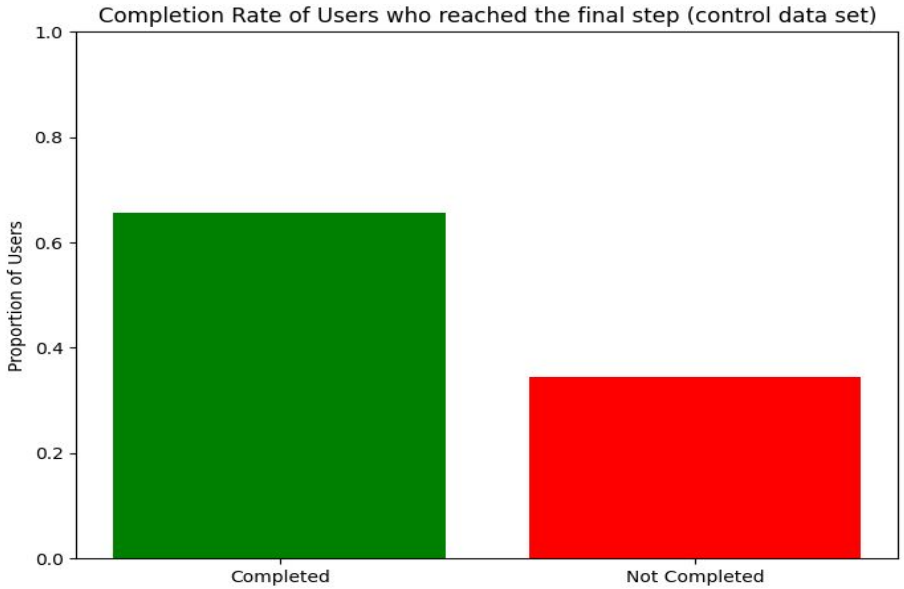


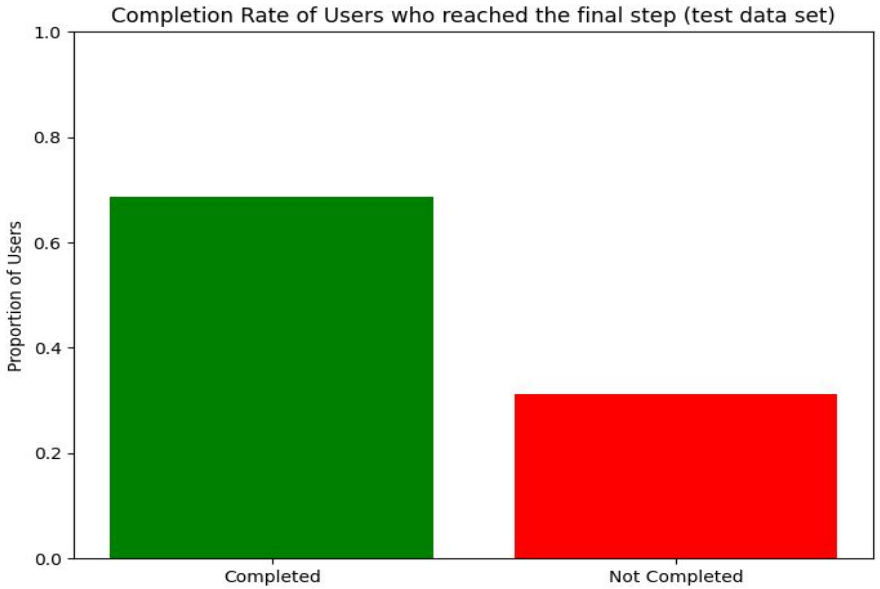



Age Distribution
Count
Age
0
10000
20000
30000
40000
20
40
60
80
100


Tenure Distribution
Count
Tenure Years
0
20000
40000
60000
80000
100000
120000
0
10
20
30
40
50
60

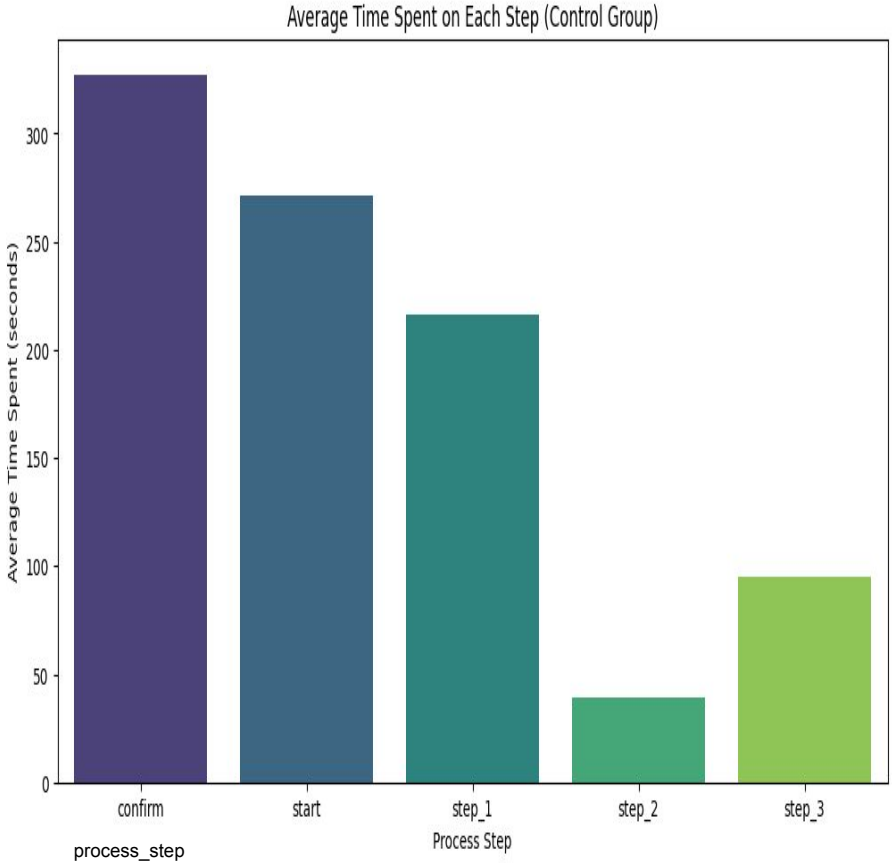
Performance Metrics



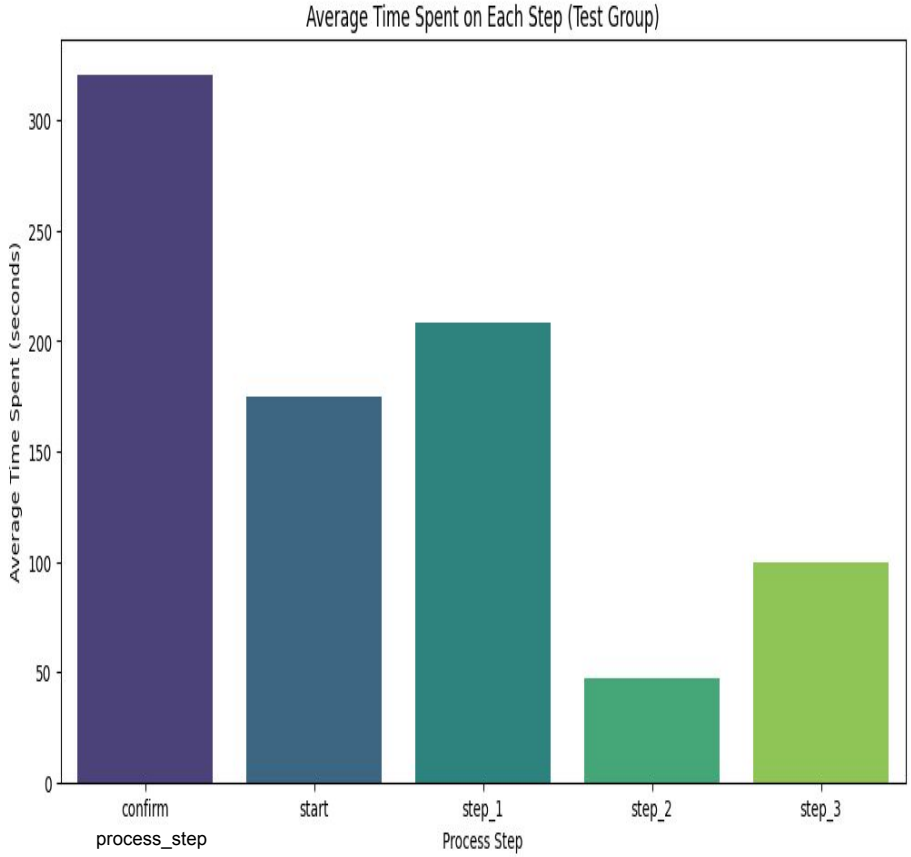


Completion Rate: 0.6558728539860615

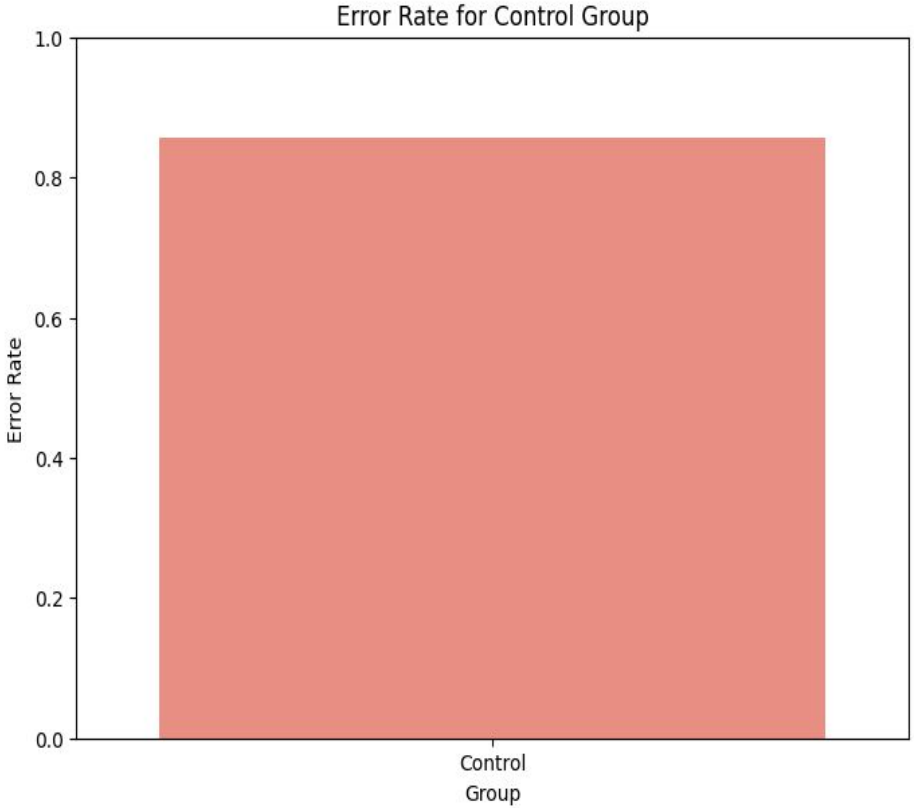
Completion Rate: 0.6875119485098881

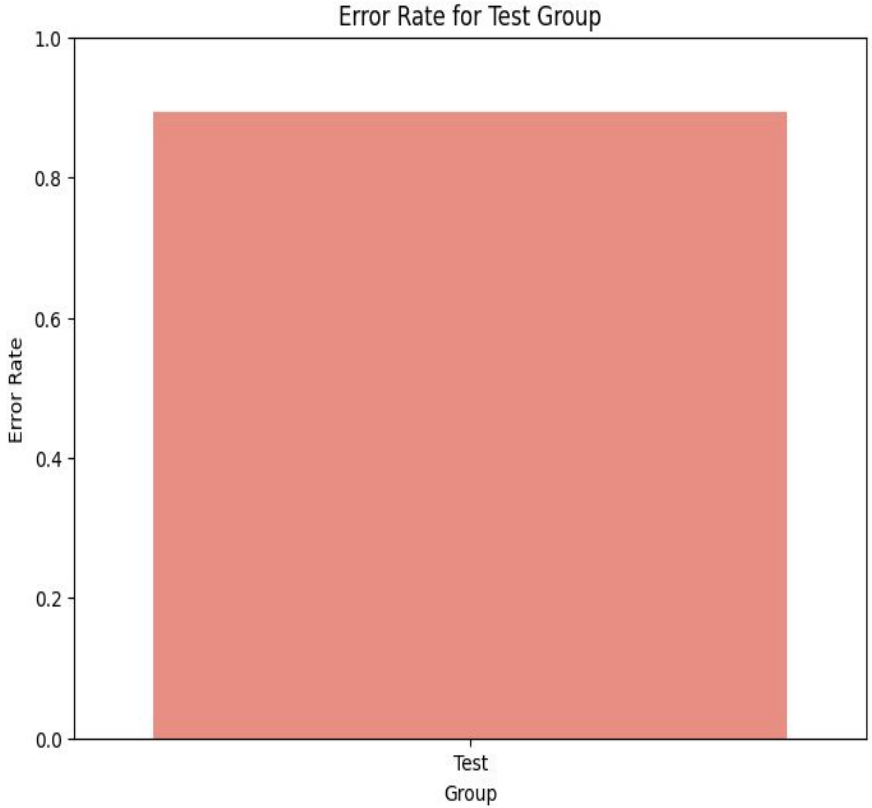


process_step
confirm 327.280340
start 271.015422
step_1 215.995650
step_2 39.298378
step_3 95.373610
Name: time_diff_seconds, dtype: float64



process_step
confirm 320.820497
start 175.007696
step_1 208.431814
step_2 47.324164
step_3 99.779869
Name: time_diff_seconds, dtype: float64

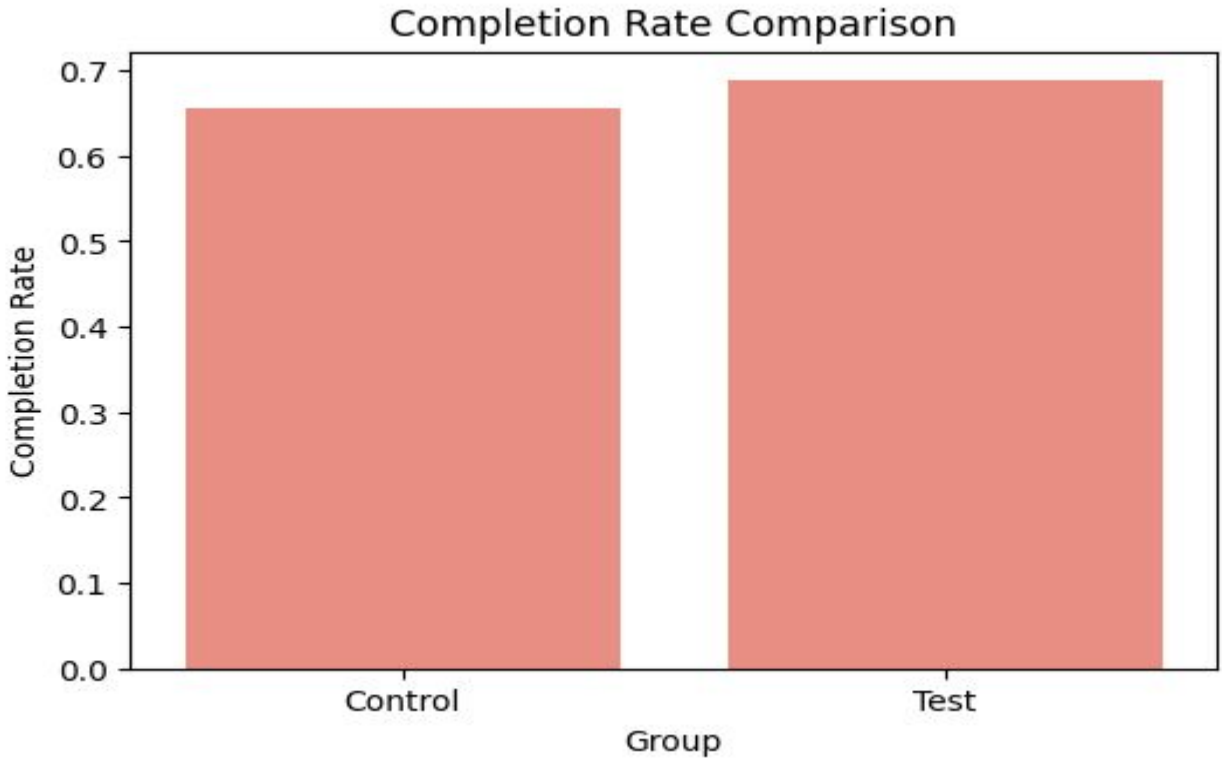




Control Group Error Rate: 85.75%

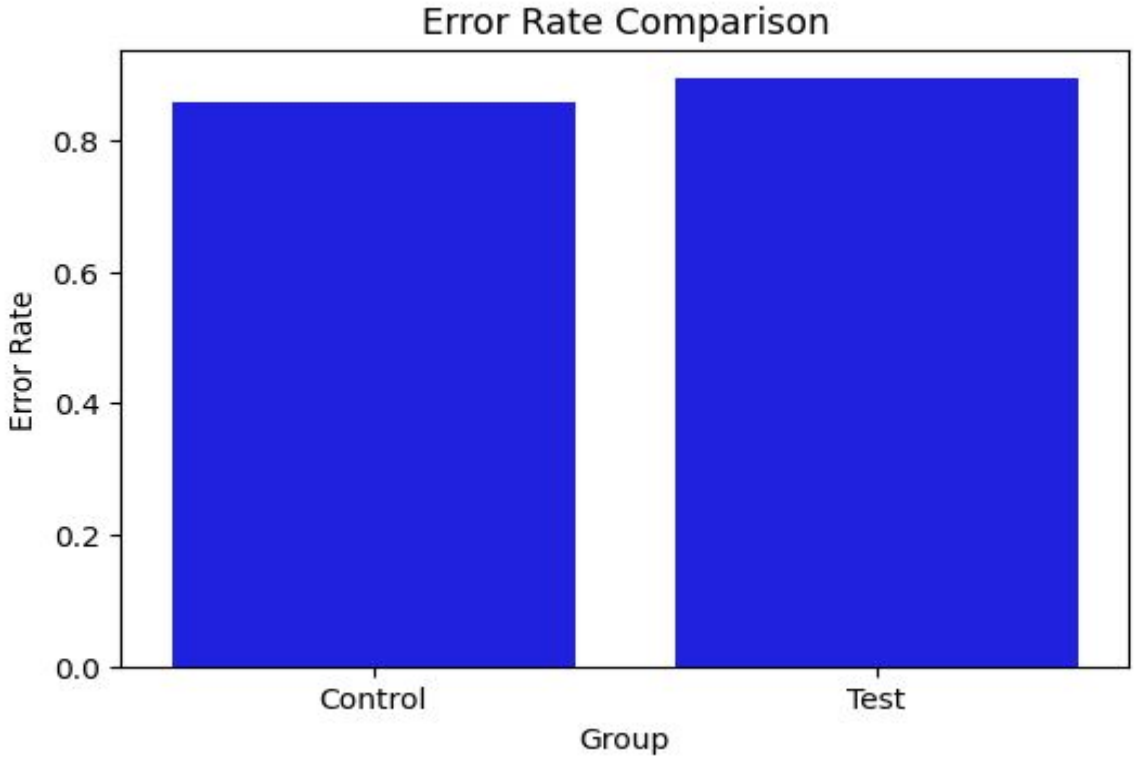
Test Group Error Rate: 89.28%

Performing Chi-Squared Tests for comparing Completion rate



- p-value = 0.00
- H0 is not equal to H1
- Rejecting the null hypothesis

Performing Chi-Squared Tests for comparing error rate



- p-value = 0.00
- H0 is not equal to H1
- Rejecting the null hypothesis

Completion rate with a Cost-Effectiveness threshold

:

- Control Completion Rate: 65.59%

  Test Completion Rate: 68.75%

  Completion Rate Difference: 3.16%

  threshold = 0.05(5%)

- The increase in completion rate (3.16%) does not meet the 5% threshold.

  # Define hypotheses

  # H0: The increase in completion rate is less than 5%

  # H1: The increase in completion rate is at least 5%

  Z-statistic: -4.91

  P-value: 0.99

- **Fail to reject the null hypothesis: The increase in completion rate does not meet the 5% threshold.**

Additional Hypothesis testing

- Tested whether the average age of clients engaging with the new process is the same as those engaging with the old process

  T-statistic: -19.12

  P-value: 0.00

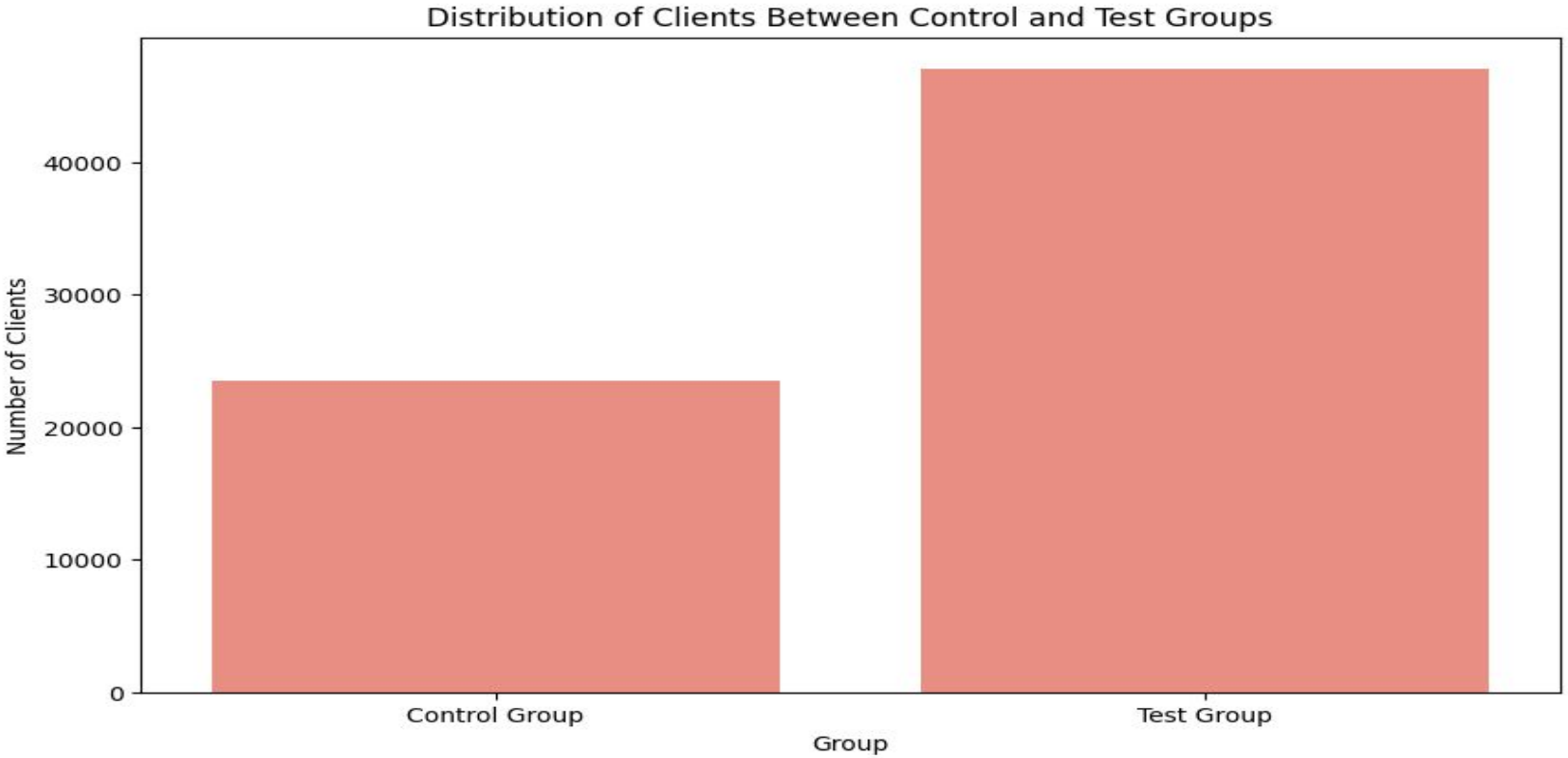
  # H0: The average age of clients in the Test group is not significantly different from the average of clients in the control group

  # H1: The average age of clients in the Test group is significantly different from the average of clients in the control group

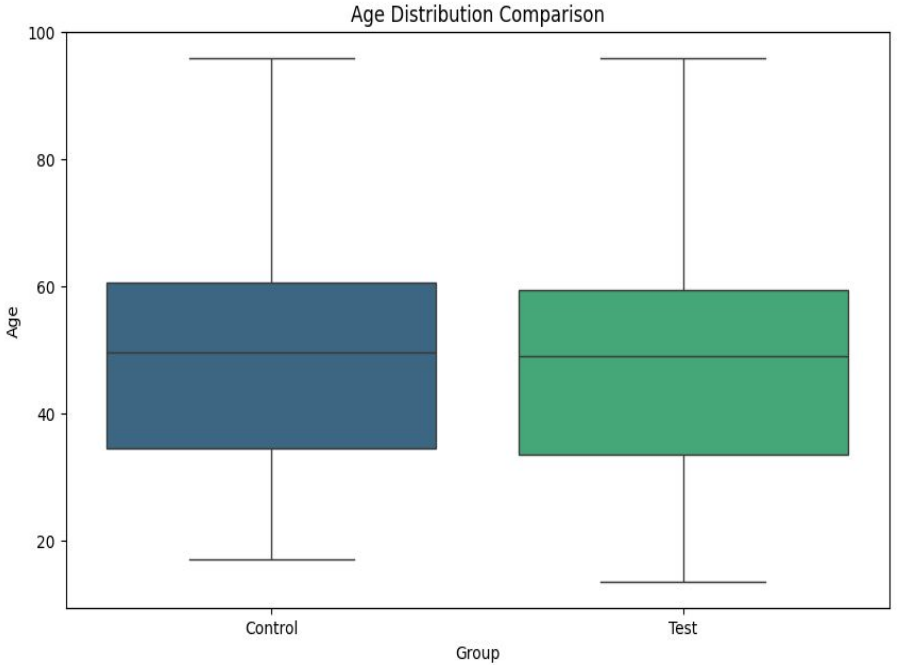
  Reject the null hypothesis: The average age of clients in the Test group is significantly different from the average age of clients in the Control group.

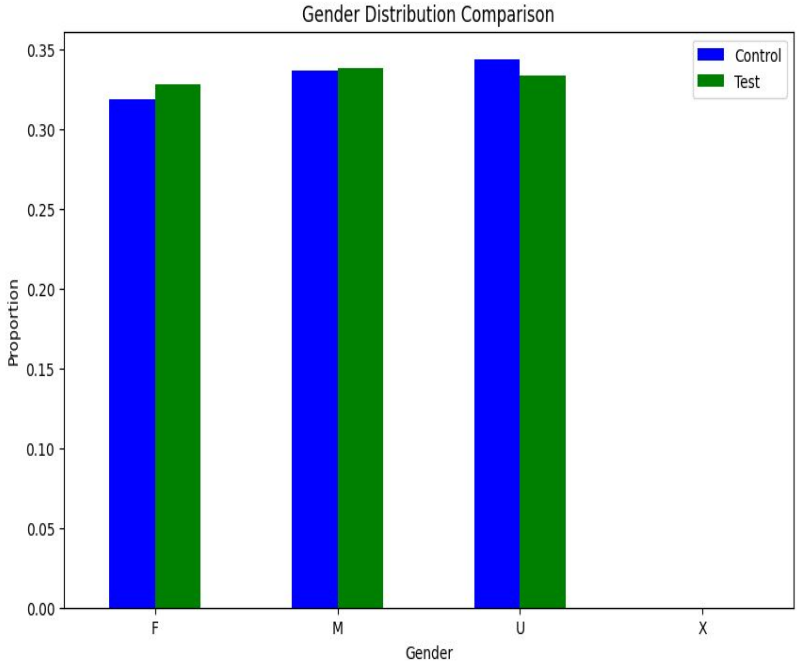
Experiment Evaluation

Were clients randomly and equally divided between the old and new designs?

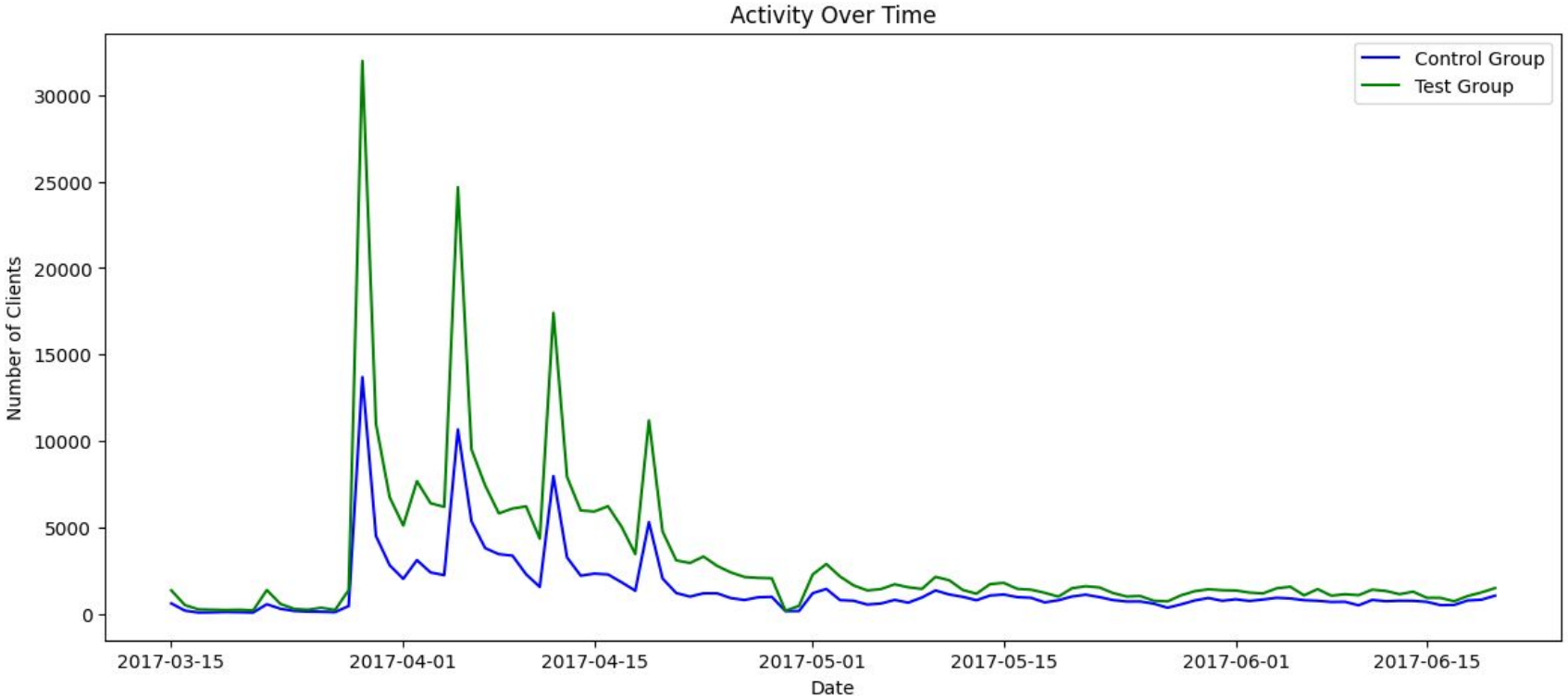


Were there any biases?



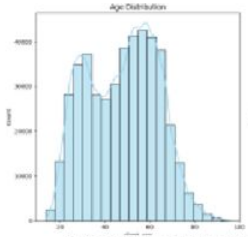


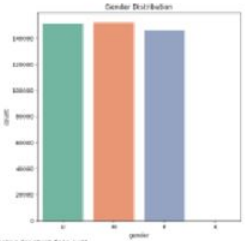
Activity over time

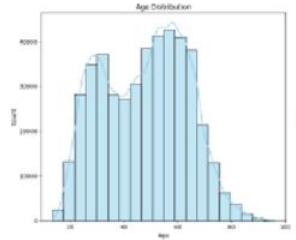


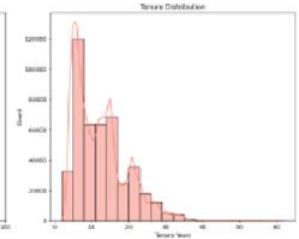
project dashboard

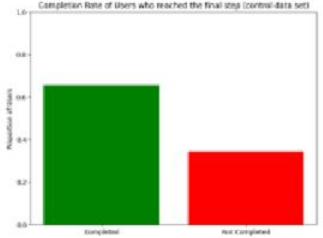
File created on: 6/12/2024 3:21:27 PM

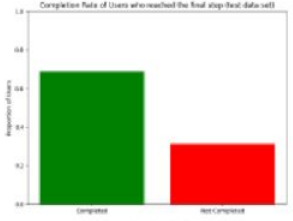

Age Distribution
Count
Age

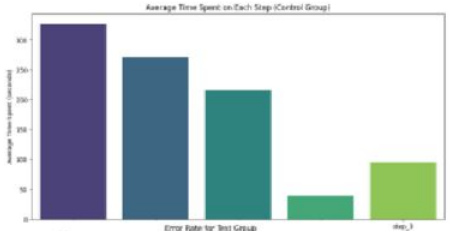

Gender Distribution
Count
gender

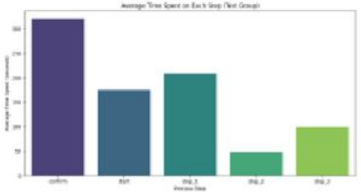

Age Distribution
Count
Age

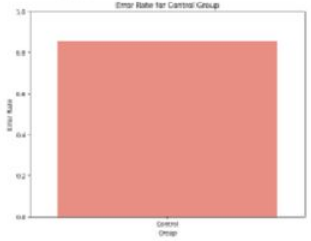

Tenure Distribution
Count
Tenure Years

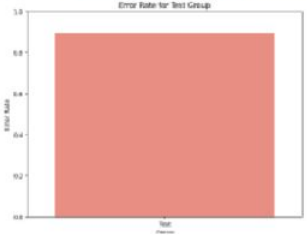

Completion Rate of Users who reached the final step (control data set)
Proportion of Users
Completed
Not Completed


Completion Rate of Users who reached the final step (test data set)
Proportion of Users
Completed
Not Completed


Average Time Spent on Each Step (Control Group)
Average Time Spent (seconds)
step_1


Average Time Spent on Each Step (Test Group)
Average Time Spent (seconds)
Process Step
confirm
start
step_1
step_2
step_3


Error Rate for Control Group
Error Rate
Control Group


Error Rate for Test Group
Error Rate
Test

Challenges & Learnings

- Data Quality and Cleaning
- Data Integration
- Multiple sessions
- Back tracking
- Library Deprecations

- Importance of Data cleaning
- Derived metrics
- Statistical testing
- Updated libraries

Recommendation & Conclusions

- Combining or reducing the number of process steps which could reduce error rate
- They can also consider giving detailed information for each step or question
- Getting feedback from all the participants
- User support & Education
- Referral bonus

Thank you for your time guys!!

-Govarthini G S