

Project 2, Week 5-6

Dimitra



Project Overview



Company:
Vanguard, US-based investment management company

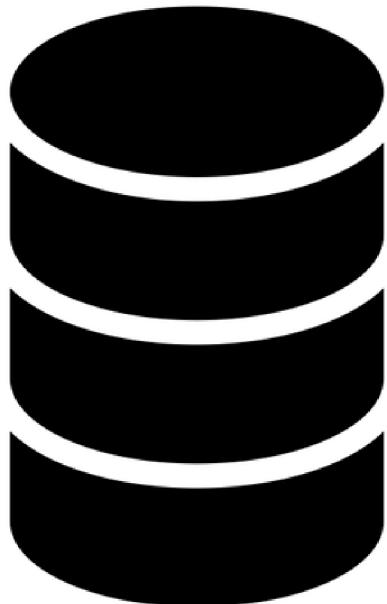


Role:
Newly employed **data analyst** in the Customer Experience (CX) team



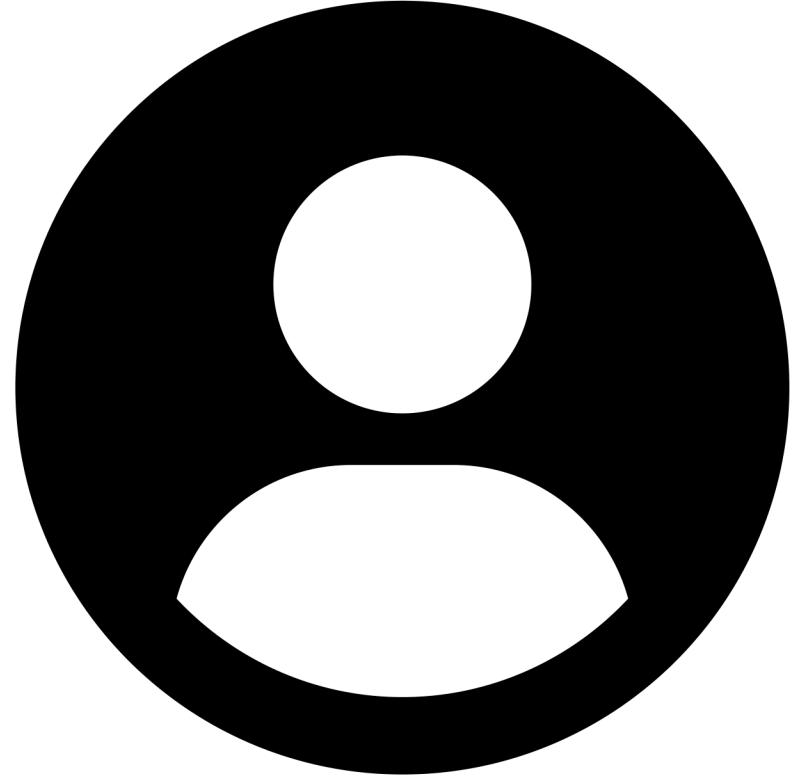
Scope:
A/B testing with more intuitive and modern UI, timely in-context prompts, with Test and Control group.

**Would the changes
encourage more clients
to complete the process?**



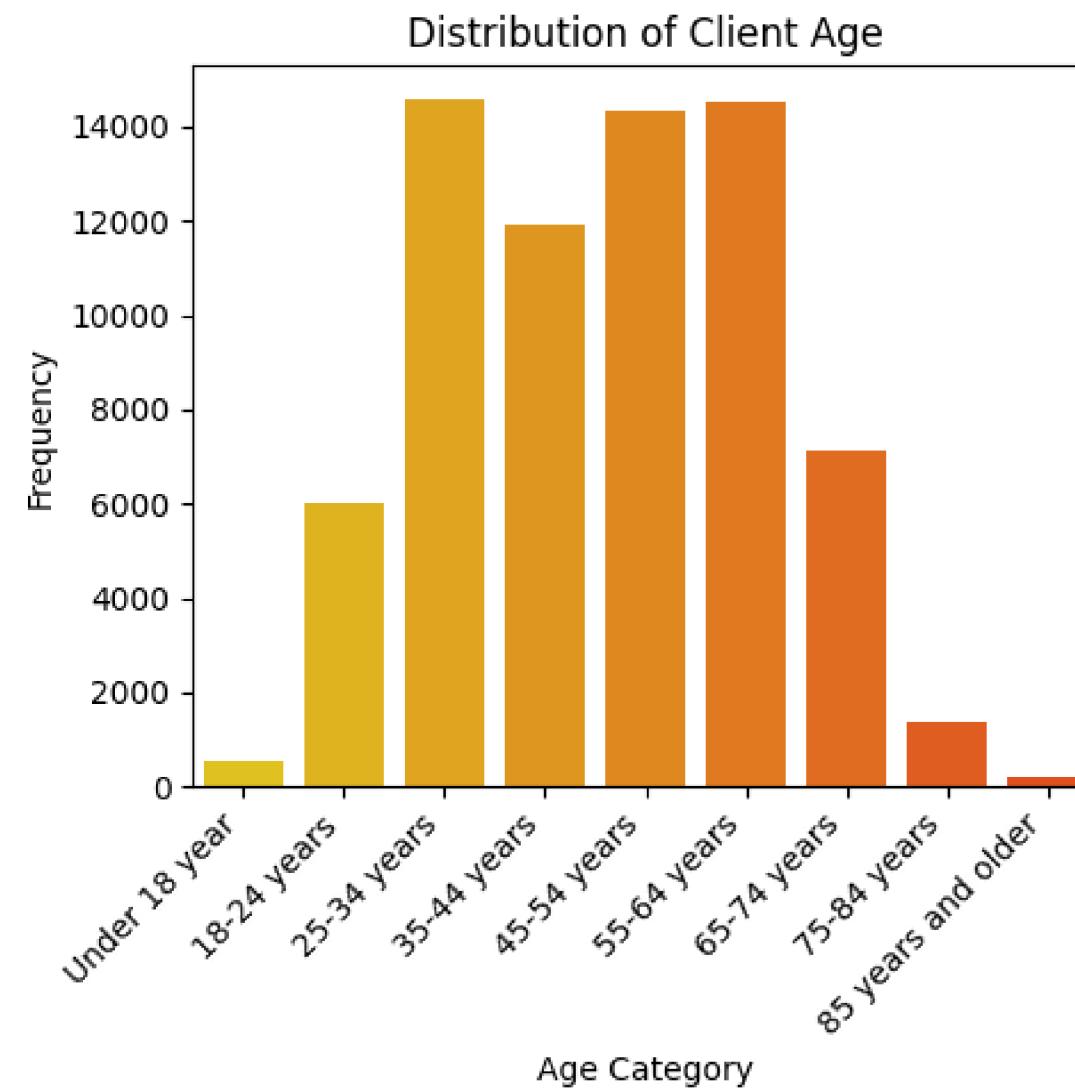
- **Clients**
- **Digital Footprint**
- **Experiment Roster**

The datasets required minimal cleaning, mostly involving null values/ duplicates/ setting datetime. Merging datasets was also necessary for the analysis. Footprint dataset presented the most significant challenge, particularly in defining valid sessions, evaluating their completion and errors.

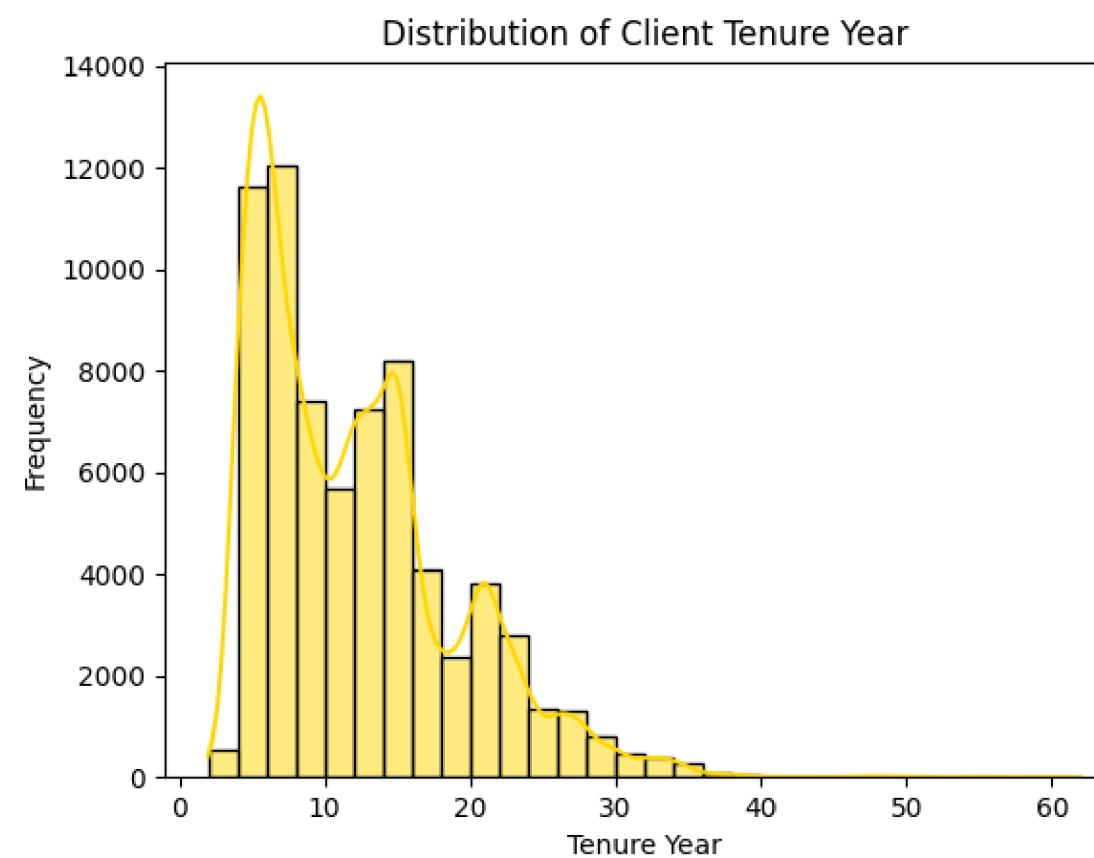


who are Vanguard's online clients?

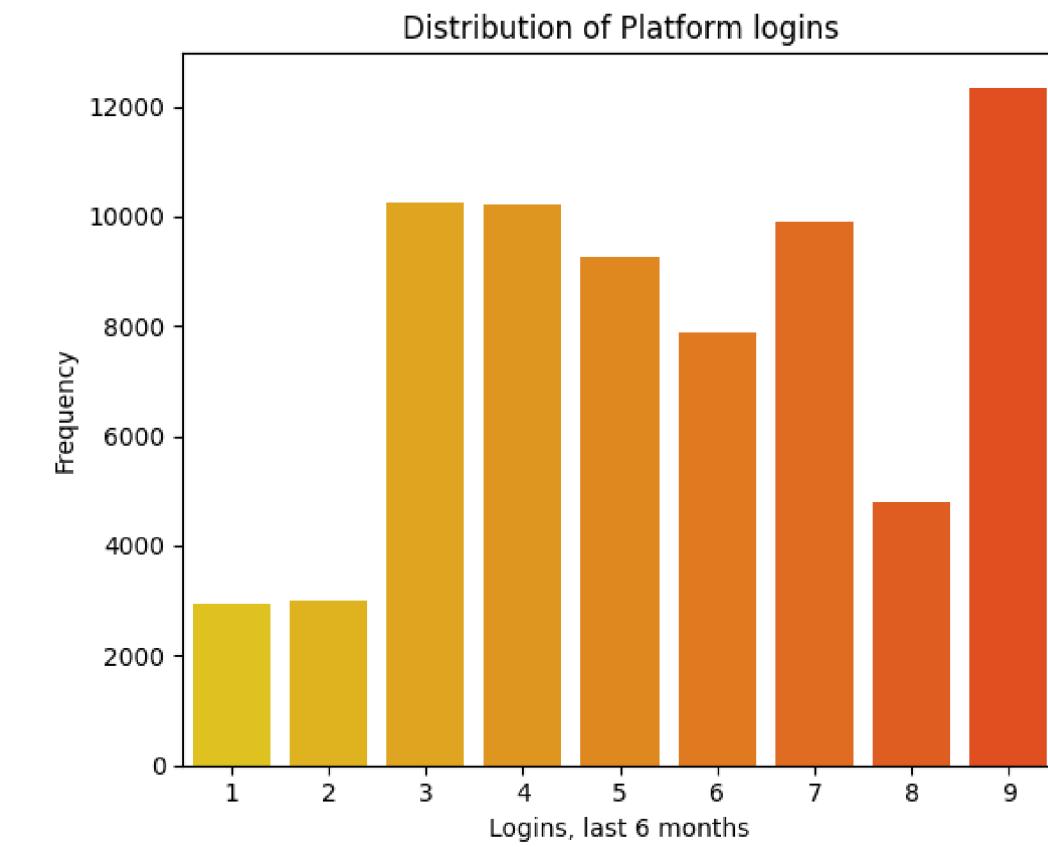
Key client characteristics



Main age groups are
**25-34 years and 45-64
years**



Most clients are with the
company **6 to 16 years**



Most clients interacted
online with the
company **at least 3
times** within the past 6
months



how is the online process looking like?

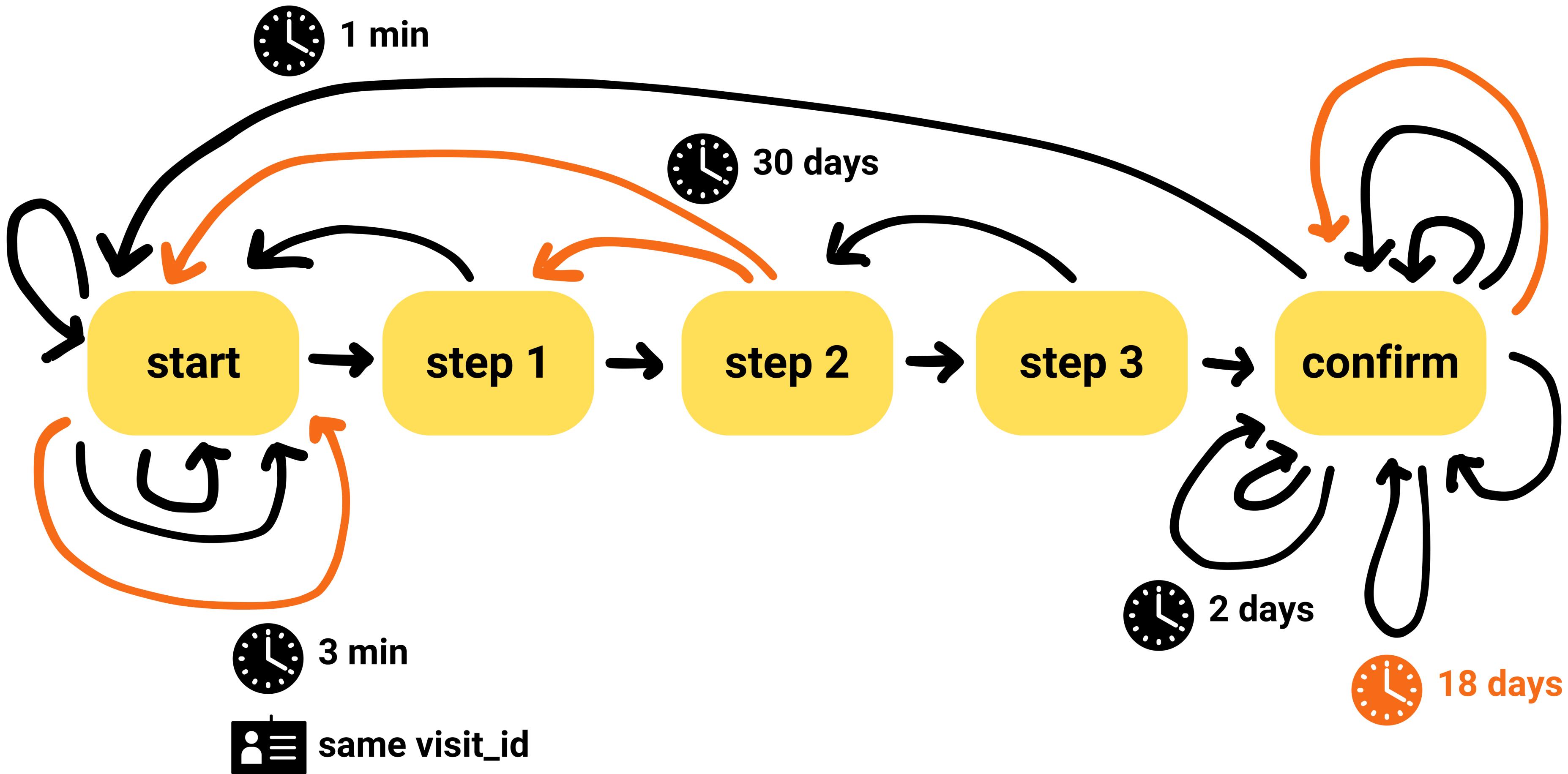
Expectations...





different visit_id

...VS reality



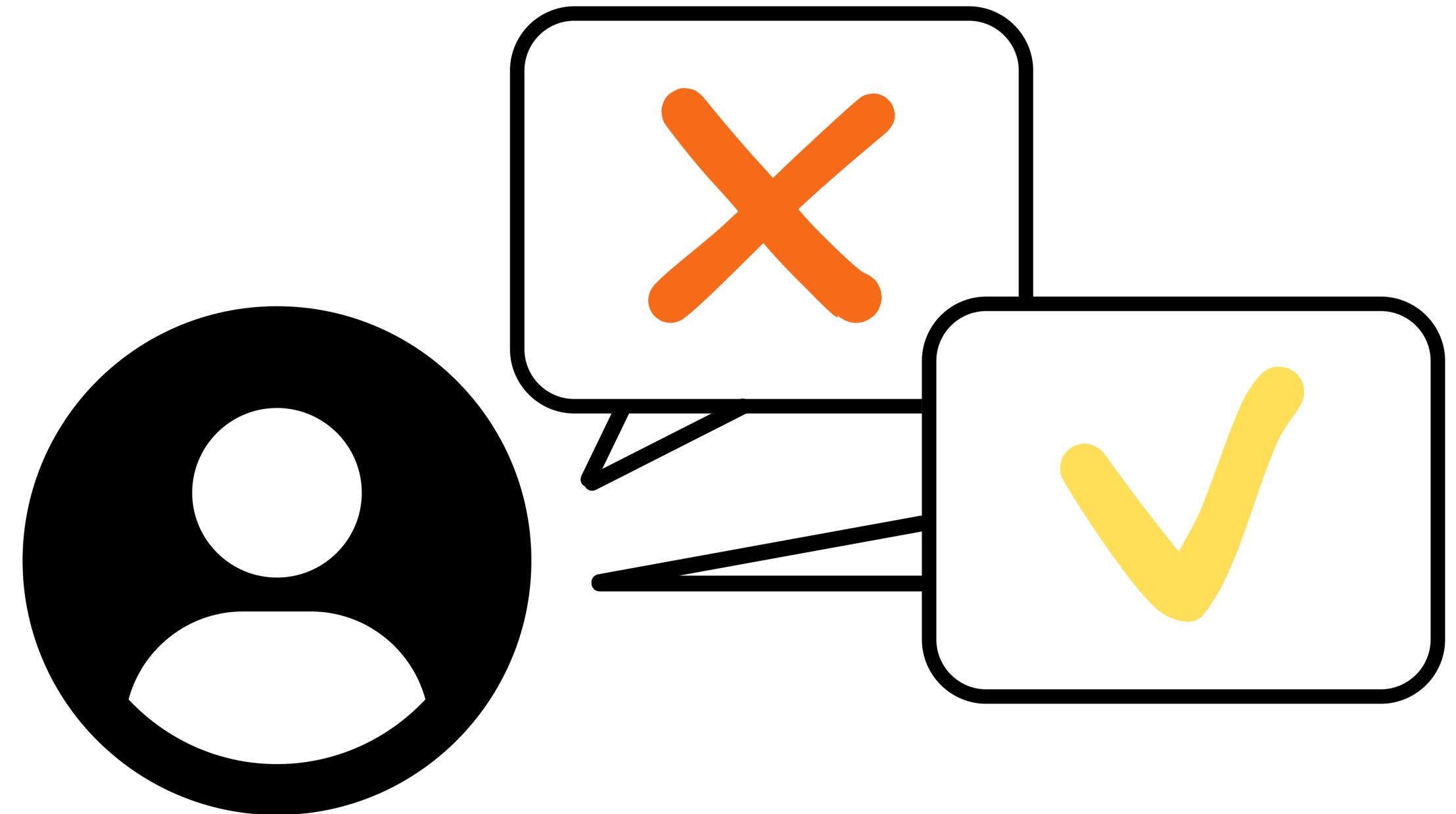
Valid session criteria

For the scope of this exercise, as there was no direct communication with the digital experiment's design team, I established my own criteria to determine a **valid session**:



After initial analysis on the time between steps, I identified outliers using the Standard Deviation method (multiplied by 3). Steps exceeding **6 minutes and 2 seconds were considered outliers**.

A session begins with the '**start**' step, triggered by either a new client or if the previous activity occurred more than 6 minutes and 2 seconds earlier.



have the changes been successful amongst clients?

Performance Metrics



Completion rate

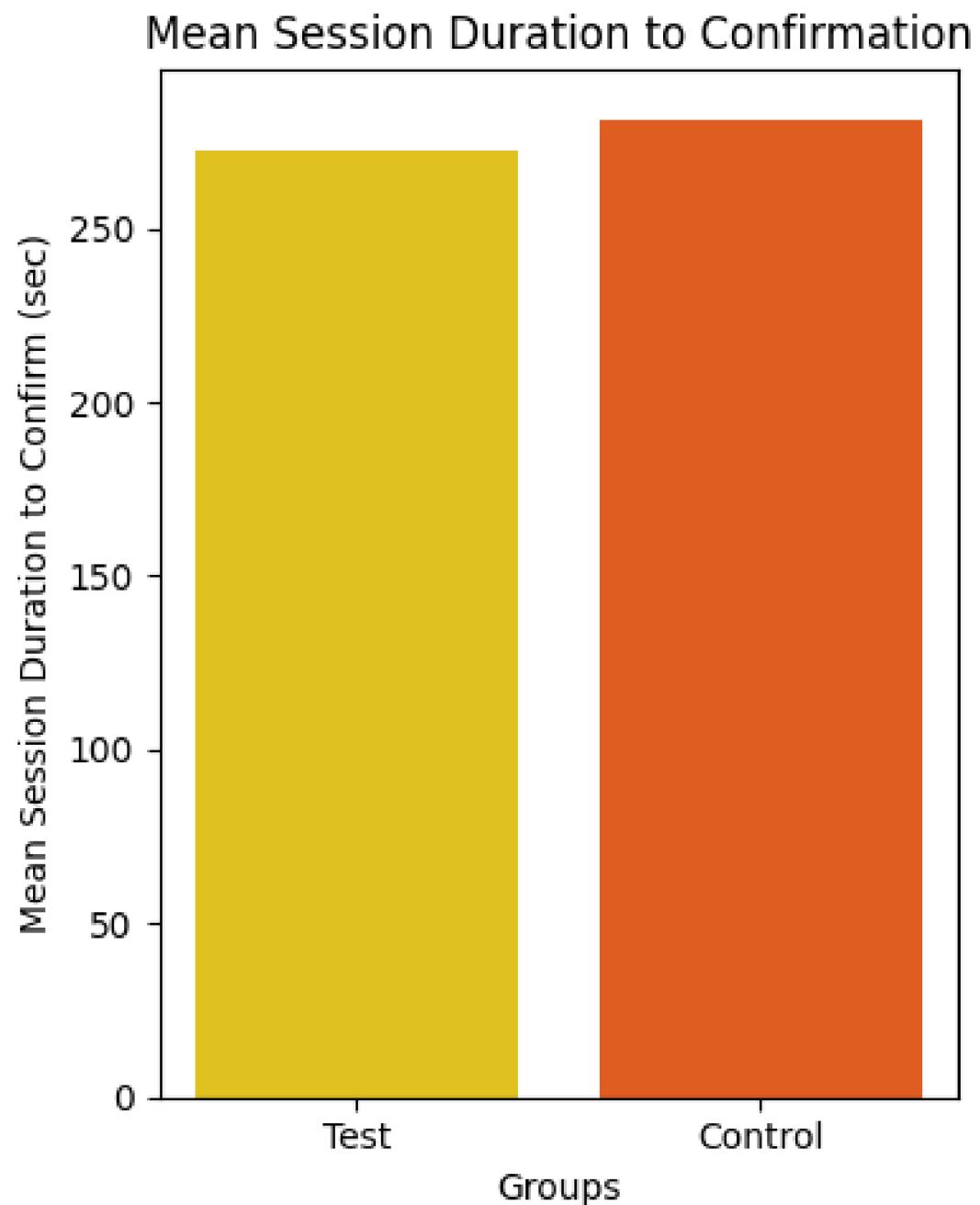


Time spent on
each step
&
Total session
duration



Error rates (steps
back and multiple
'confirm')
&
Total session
steps

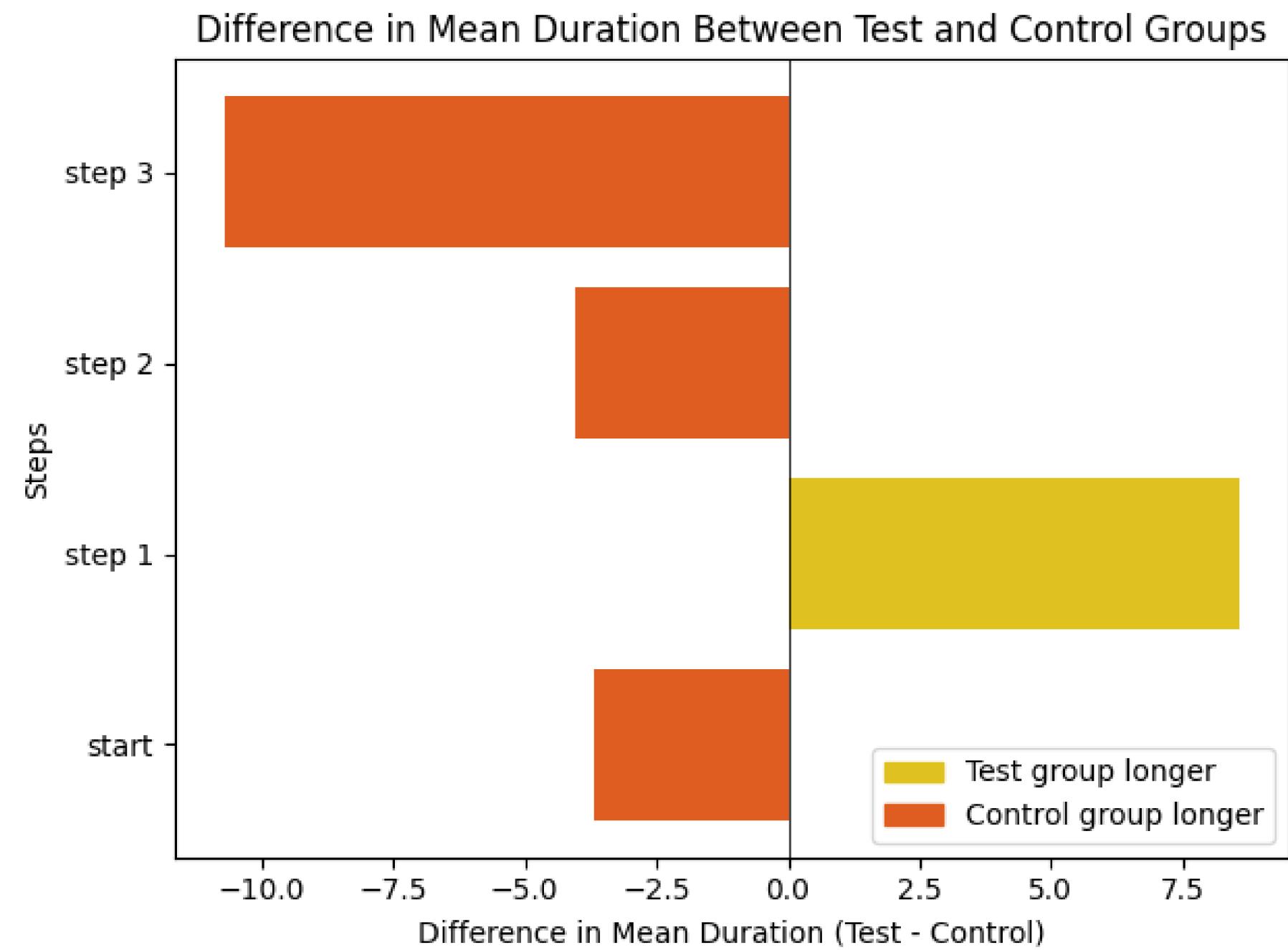
Time spent on each step



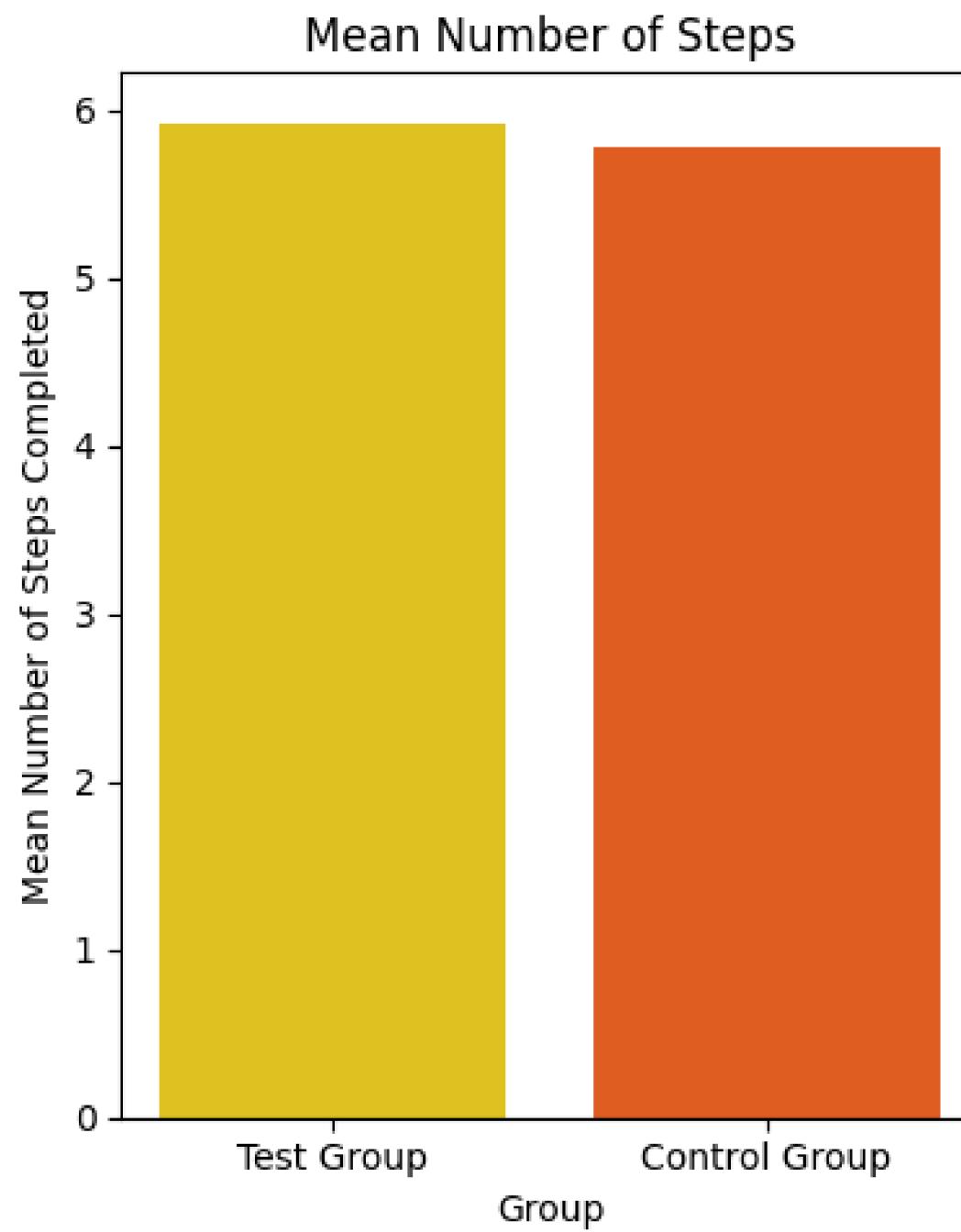
'Test' mean session duration: 4 min and 32 sec

'Control' mean session duration: 4 min and 41 sec

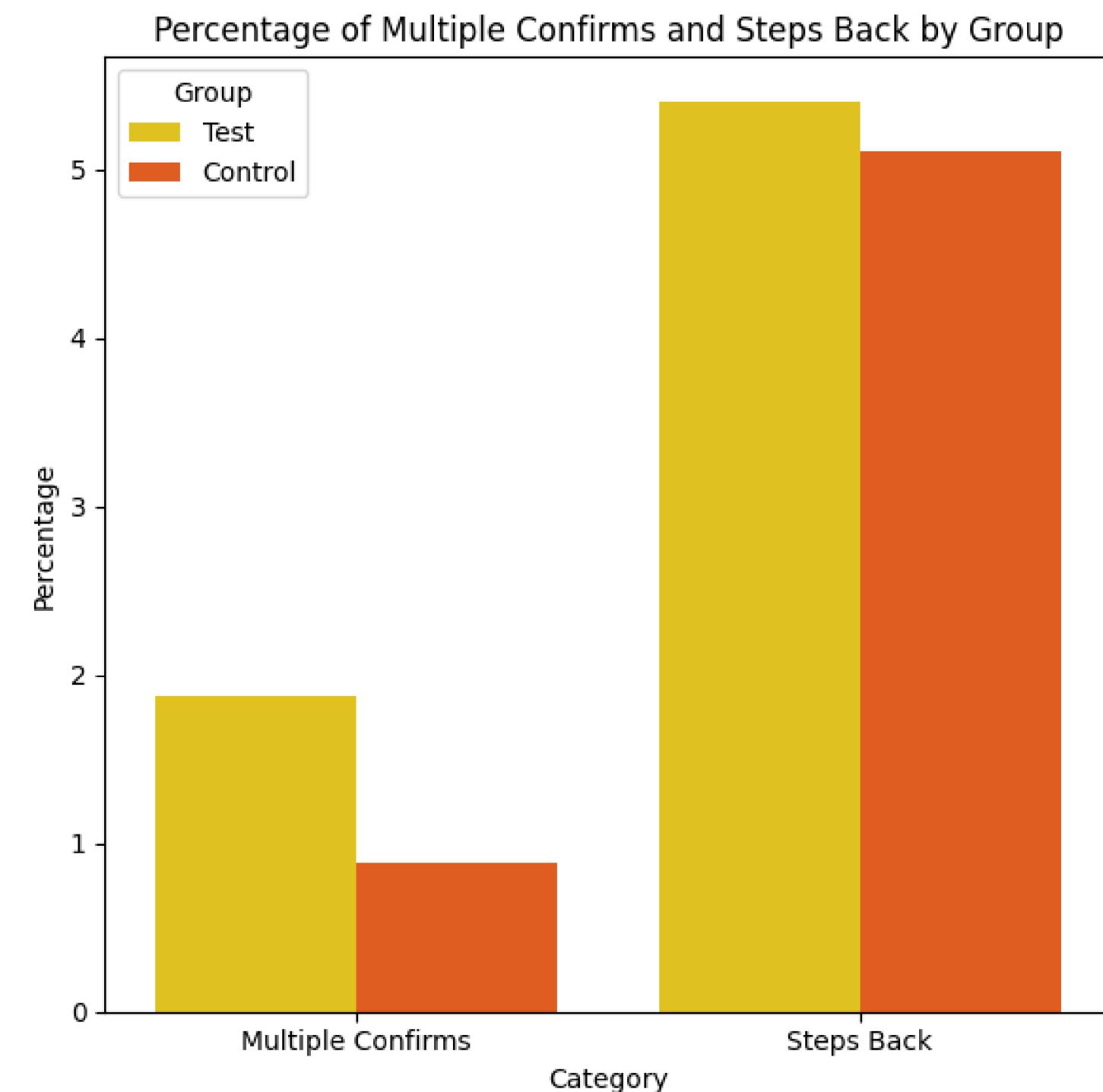
Interestingly, the only step where the 'Test' group required more time than the 'Control' group was during **step 1**.



Error rates and total number of steps



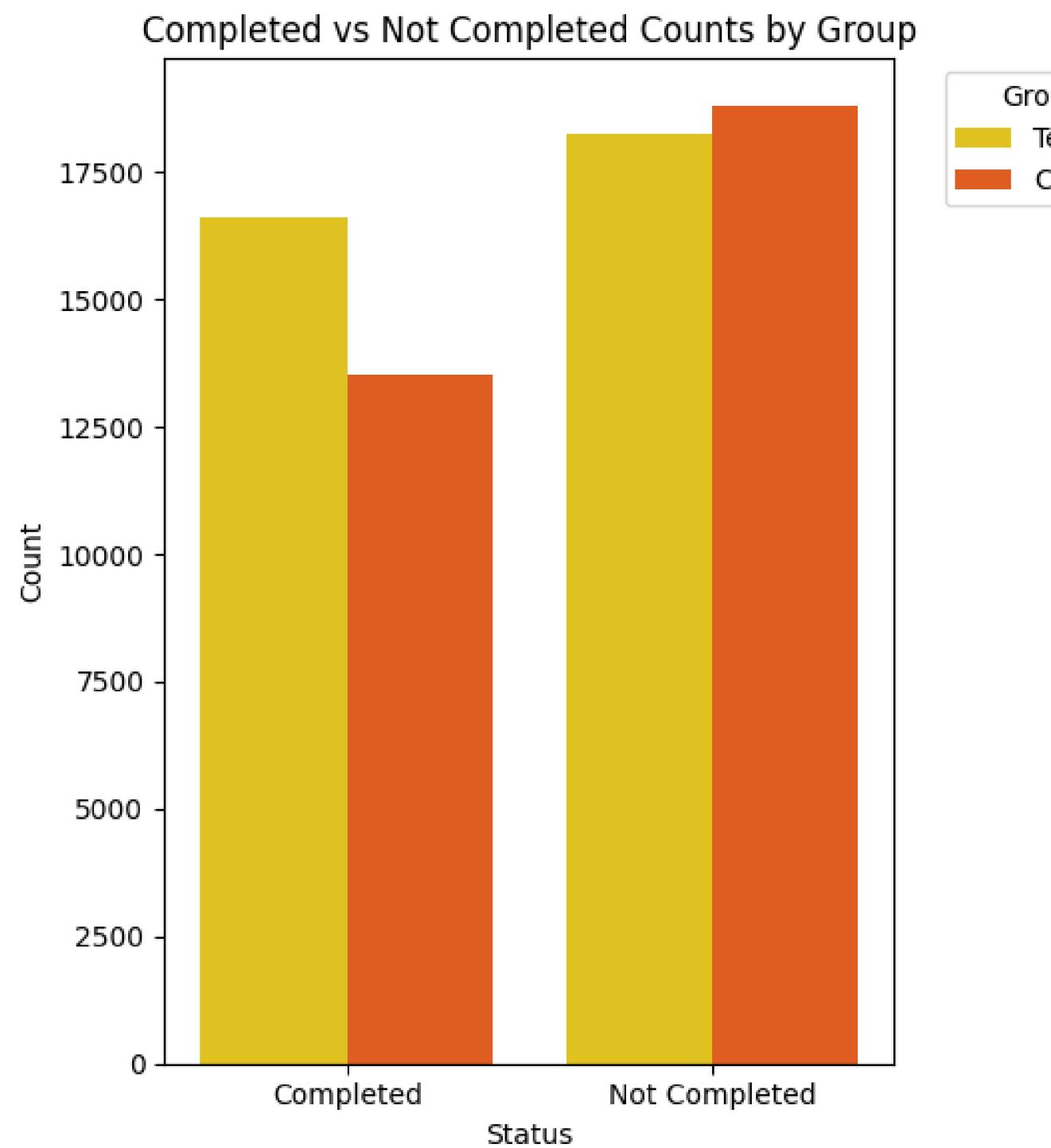
The 'test' group shows **higher error rates, more step backs, and a greater number of 'confirms'**.



'Test' mean number of steps: 5.93

'Control' mean number of steps: 5.78

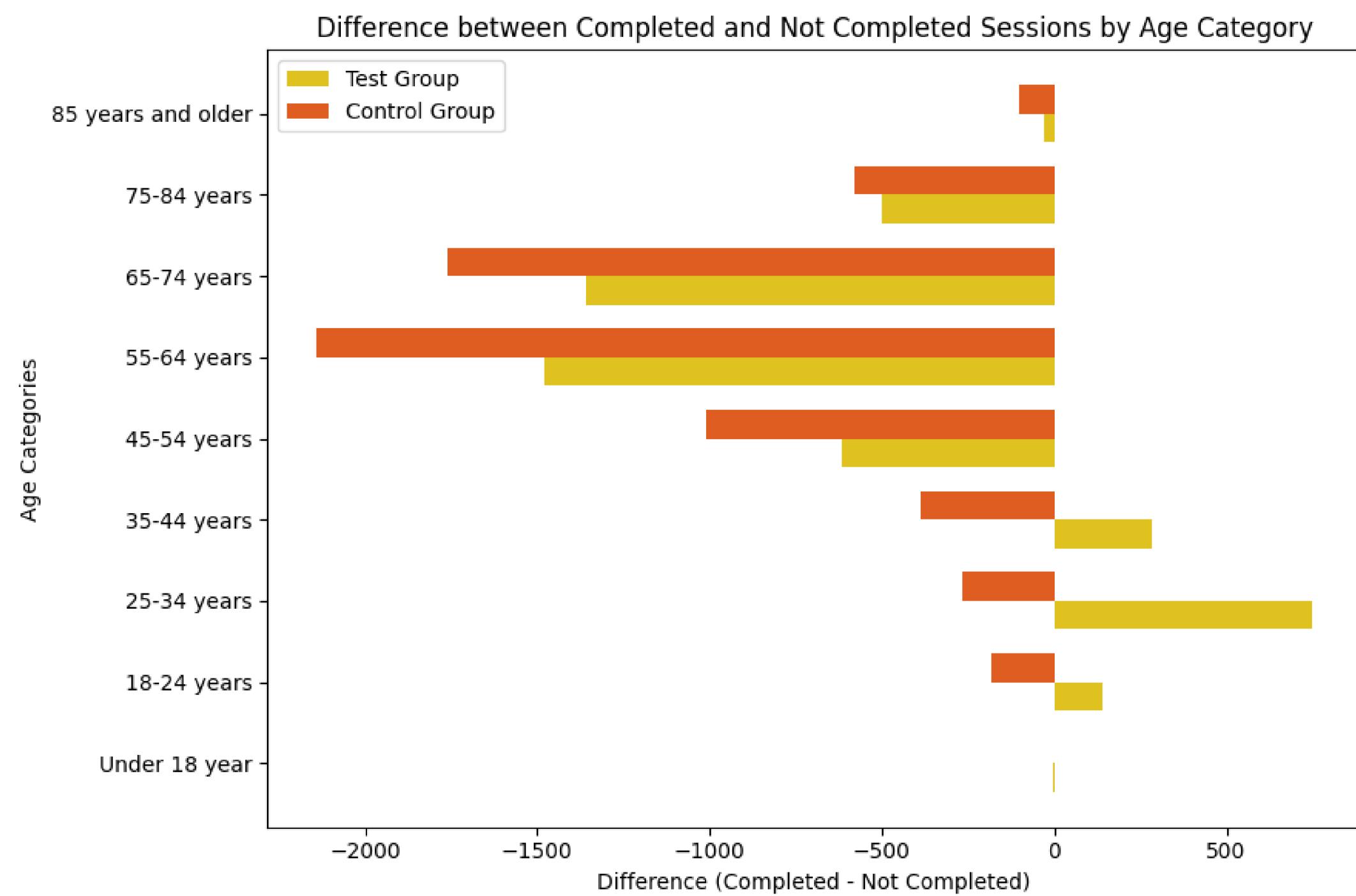
Completion Rate



'Test' completion rate: 47.74%

'Control' completion rate: 41.90%

Only the '**Test**' group within the 18-44 age range shows more completed sessions than not completed ones.



**Is the observed increase in
completion rate from the A/B test
meeting the 5% threshold?**

Hypothesis testing

A **Test of Proportions** was conducted, to compare the rates from the two groups.

Null hypothesis: the ‘Control’ group completion rates increased by 5%, is smaller than the completion rate of the ‘Test’ group.

$$\begin{aligned} H_0 &: p_1 - p_2 \leq 0 \\ H_{1/a} &: p_1 - p_2 > 0 \\ z &= \frac{(p_1 - p_2)}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \\ p_1 &= \frac{x_1}{n_1}; p_2 = \frac{x_2}{n_2} \\ p &= \frac{(x_1 + x_2)}{(n_1 + n_2)} \end{aligned}$$

```
p1 = control_completion_rate + 5
p2 = test_completion_rate
p = ( test_confirm + control_confirm ) / ( test_sessions + control_sessions )

rd = ( p1 - p2 )
p_1_p = p * (1-p)
inverse = (1 / control_sessions) + ( 1 / test_sessions )

z = rd / np.sqrt(p_1_p * inverse)
```

The value of rd is -0.81
The value of p_1_p is 0.25
The value of inverse is 0.0001
The value of the root is 0.0038
The value of the z statistic is -210.92

The calculated z-statistic of -210.92 indicates a significant difference between the completion rates of the two groups, and the negative value fails to reject the hypothesis. That means:

‘Test’ group completion rate > ‘Control’ group completion rate + 5%

Conclusion and suggestions



'Test' group has a higher completion rate than the 'Control' group!

→ In the 18-44 age group, completion rates were notably higher. With a significant number of clients aged 45-64, we can explore the effectiveness of hints and strategies tailored specifically to this demographic.

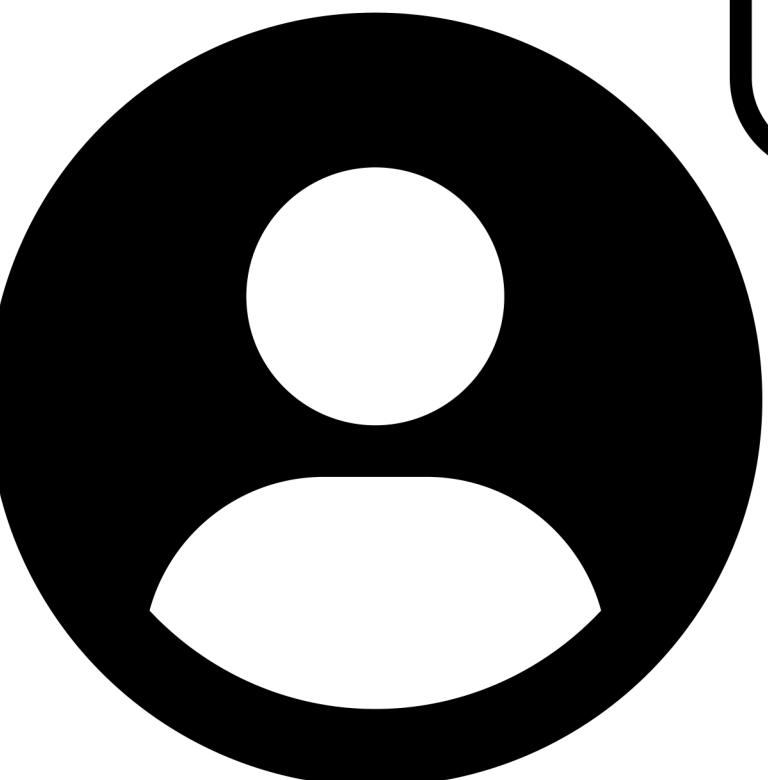
'Test' group completed the online process faster!

→ The test group needs to improve from step 1 to step 2 to increase its speed even more.

'Test' group has higher percentage of step backs and confirms compared to 'Control'.

→ To investigate reasons for the steps back, were they f.e. double checking steps?

Thank you!



Icon attributes:

Person by Tamzid Hasan from Noun Project (CC BY 3.0)

Online transaction by popcornarts from Noun Project (CC BY 3.0)

Clock by Wartini from Noun Project (CC BY 3.0)

Identification by Creative Stall from Noun Project (CC BY 3.0)

Company by Alice Design from Noun Project (CC BY 3.0)

Analyst by Alice Design from Noun Project (CC BY 3.0)

Ab testing by Adrien Coquet from Noun Project (CC BY 3.0)

database by YANDI RS from Noun Project (CC BY 3.0)

confirm by Kevin White from Noun Project (CC BY 3.0)

Error by Deemak Daksina from Noun Project (CC BY 3.0)