

# Design Meets Data

## *A Journey Through A/B Test Findings*



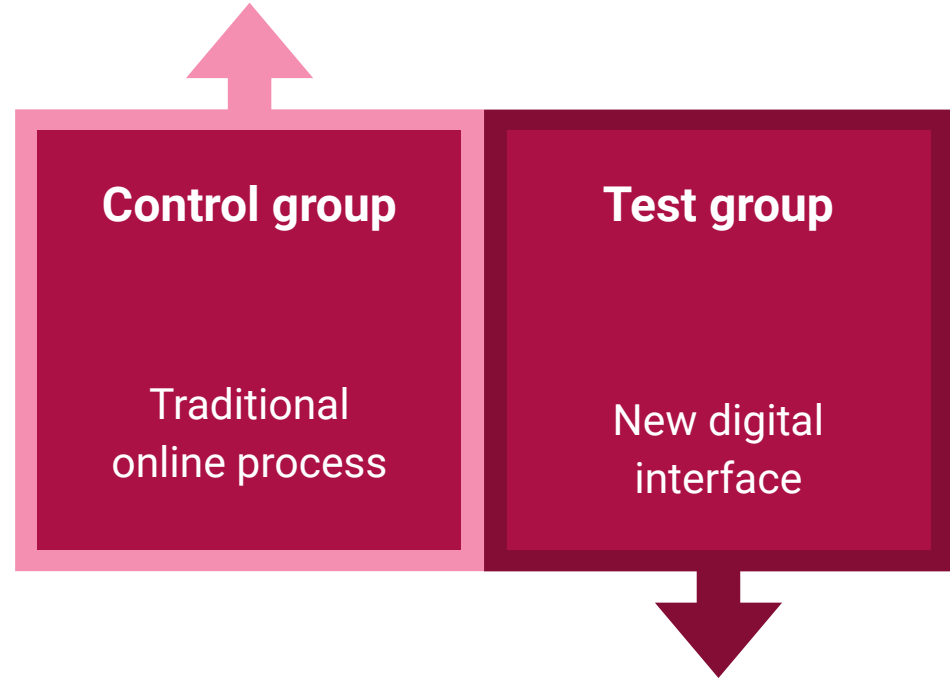
*Amirreza & Lorena*

# Vanguard

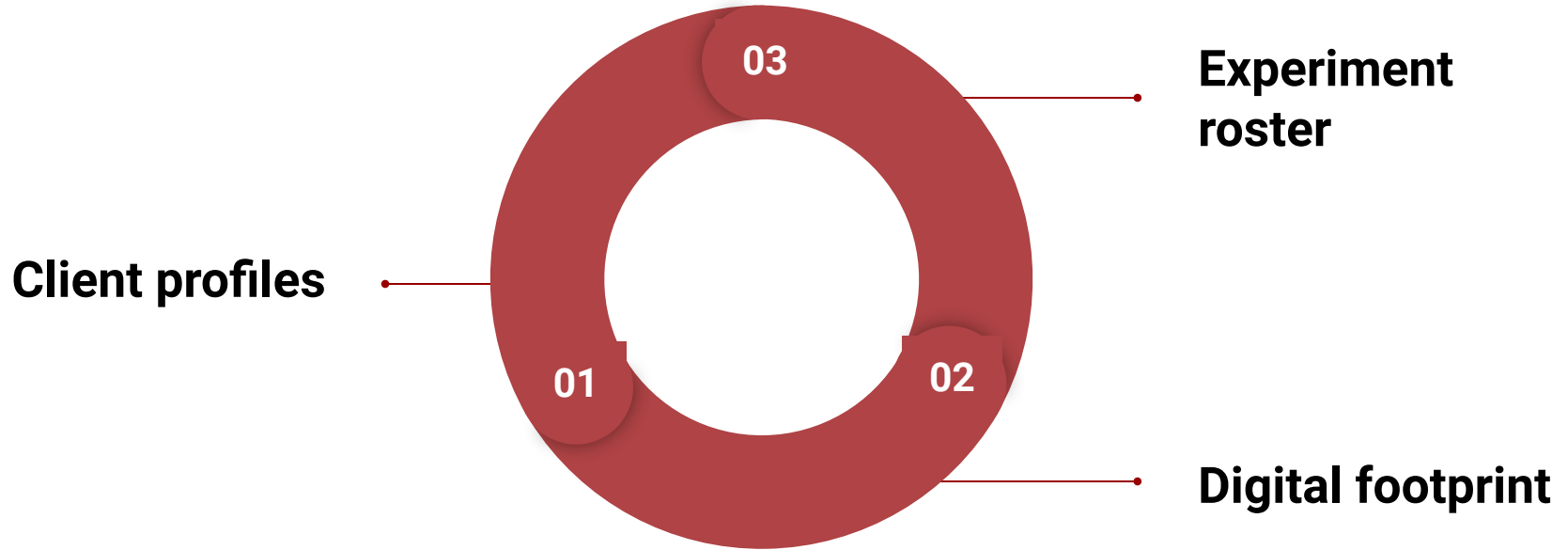
- Vanguard is one of the world's most respected investment management companies, offering a broad selection of **investments, advice, retirement services**, and **insights** to individual investors, institutions, and financial professionals.
- Purpose: To take a stand for all investors, to treat them fairly, and to give them the best chance for investment success.

# The digital experiment

A/B Test for 98 days  
(3/15/2017 - 6/20/2017)



# Data overview



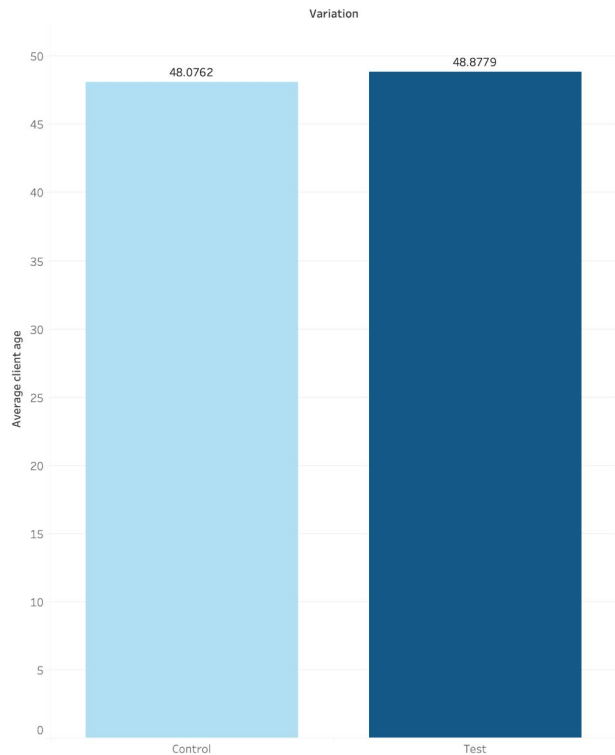
# Data cleaning and merging

- Merge the datasets
- Drop missing values
- Create new columns for time calculations
- Separate data sets (Test and Control)

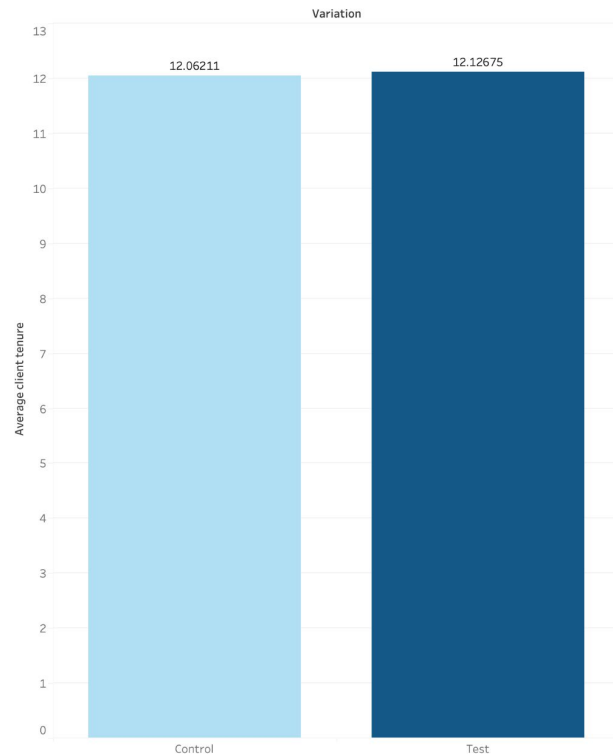
# Test and Control Groups Comparison

# Distribution of customers by age and tenure

Average clients age in test and control groups

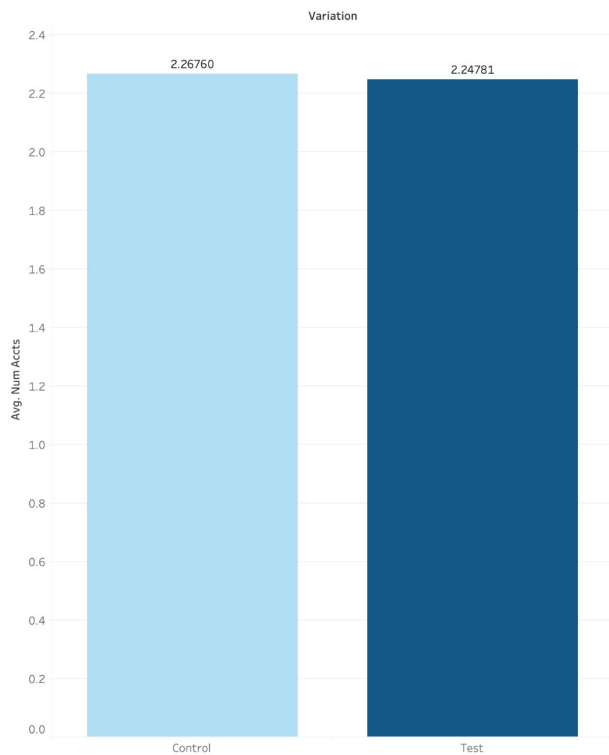


Average clients tenure in test and control groups

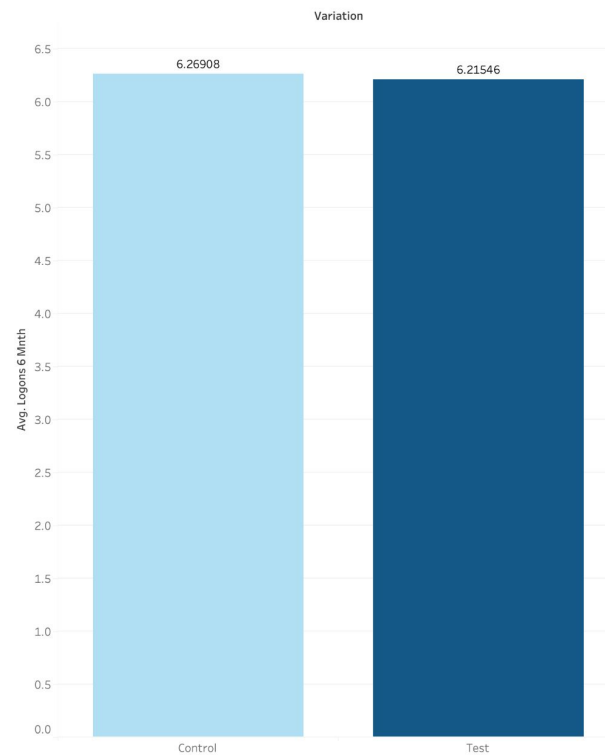


# Distribution of customers by number of accounts and logins

Average number of accounts in test and control groups



Average logins in 6 months in test and control groups



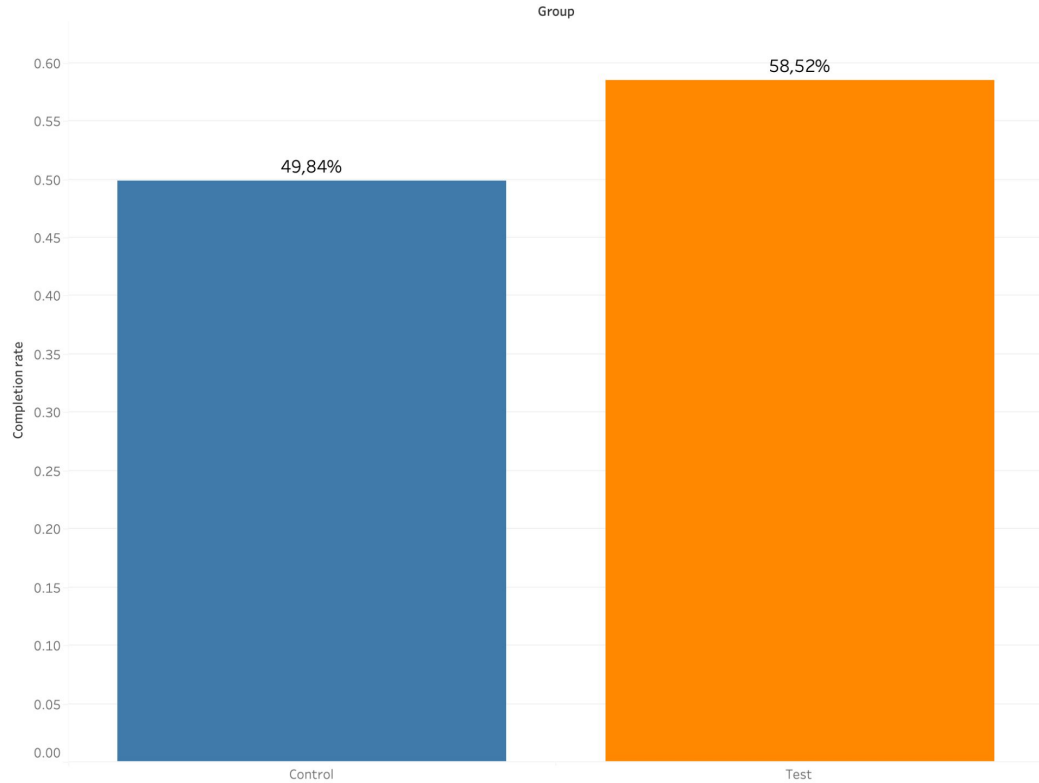


# Performance metrics

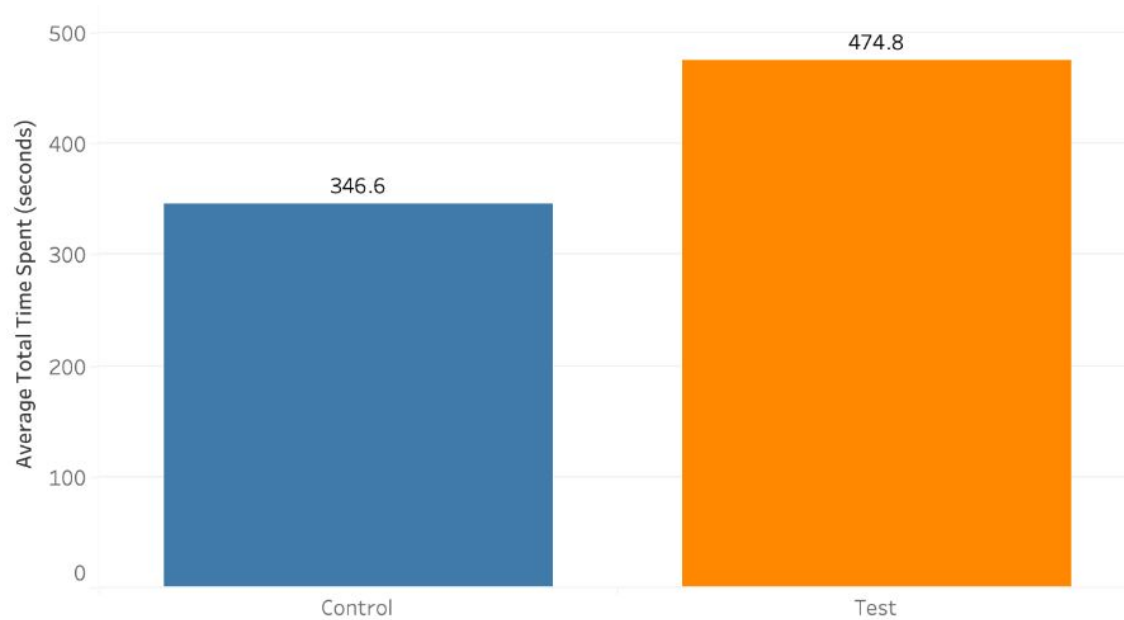
# Performance Metrics

- Completion rate
- Time Spent in the process
- Error rate

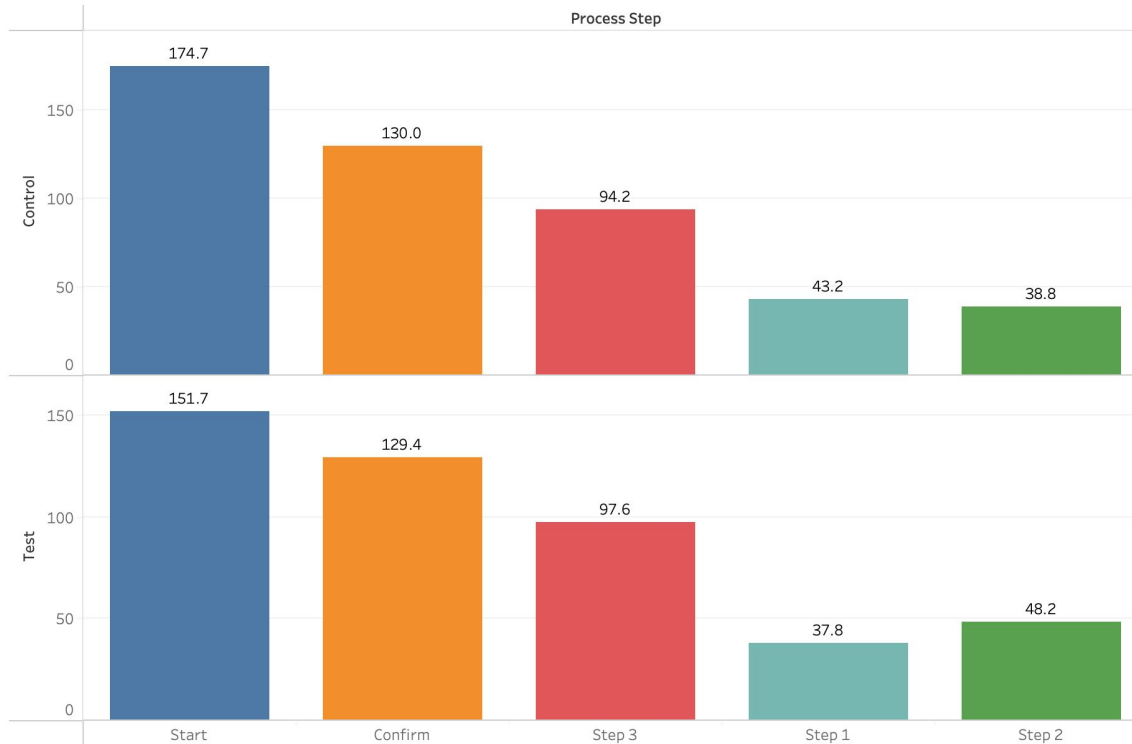
# Completion rate



# Total time spent in the process

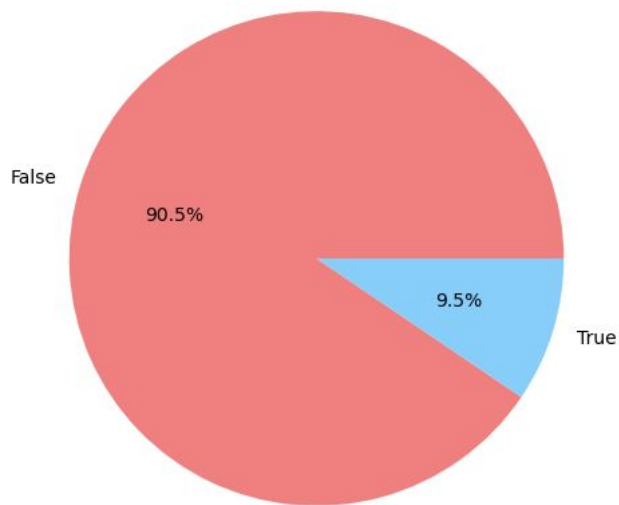


# Total time spent on each step

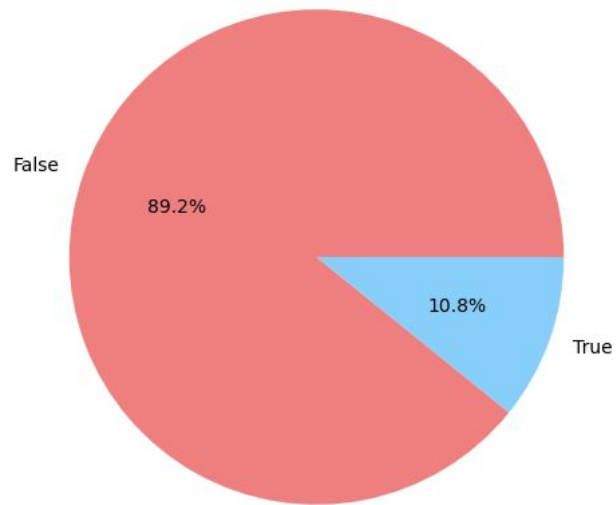


# Error rate

Control Group Error Rate



Test Group Error Rate



# Hypothesis Testing

# Hypothesis testing

KPI		H0	H1	Results
1	Completion rate		✓	<ul style="list-style-type: none"><li>• Z-Statistic= 22.89</li><li>• <math>p &lt; .001</math></li></ul>
2	Total time spent in the process		✓	<ul style="list-style-type: none"><li>• t-Statistic= 11.37; <math>p &lt; .001</math></li><li>• U-Statistic= 4,426,175.0; <math>p &lt; .001</math></li></ul>
3	Mean time spent on each step		✓	<ul style="list-style-type: none"><li>• t-Statistic -6.27; <math>p &lt; .001</math> (start)</li><li>• t-Statistic -2.53; <math>p = .006</math> (step 1)</li></ul>



# Hypothesis testing (y=total time spent)

Independent variables		H0	H1	Results
1	Tenure years	✓		<ul style="list-style-type: none"><li>coef= -.7574</li><li>p=.538</li></ul>
2	Age		✓	<ul style="list-style-type: none"><li>coef= 5.7533</li><li>p&lt;.001</li></ul>
3	Balance	✓		<ul style="list-style-type: none"><li>coef= 6.622</li><li>p=.770</li></ul>
4	Calls		✓	<ul style="list-style-type: none"><li>coef= -7.8541</li><li>p=.028</li></ul>
5	Group (dichotomic)		✓	<ul style="list-style-type: none"><li>coef= 48.8970</li><li>p=.002</li></ul>

# Key findings

# Key findings

- The test group has a greater completion rate:
  - The changes encourage more clients to complete the process →  $8.68\% > 5\%$
- The test group spent more time in the process:
  - We need to improve the step 2 and step 3 of the new website

# Prototype customer (2.28 minutes)

- 10 tenure years
- 46 years old
- \$111,208 in balance
- 2 accounts
- 3 calls in 6 months
- 6 logins in 6 months



# Evaluation and Recommendations

# Experiment evaluation

1	Design		✓	<ul style="list-style-type: none"><li>• Short-term focus</li></ul>
2	Potential biases	✗		<ul style="list-style-type: none"><li>• New potential customers</li></ul>
3	Randomization		✓	<ul style="list-style-type: none"><li>• Both groups are very similar</li></ul>
4	Duration		✓	<ul style="list-style-type: none"><li>• 3 months</li></ul>

# Recommendations

- Apply a customer satisfaction survey to collect more data and measure more KPIs → engagement and customer satisfaction score
- Collect the correct data for the variable Gender
- Having more demographic data for a more accurate evaluation

# Challenges and learnings



# Challenges

- The size of the dataset
- Missing information
- Derive insights and draw conclusions

# Learnings:

- Hypothesis testing
- Figures to communicate the results

# Thank you

Amirreza and Lorena