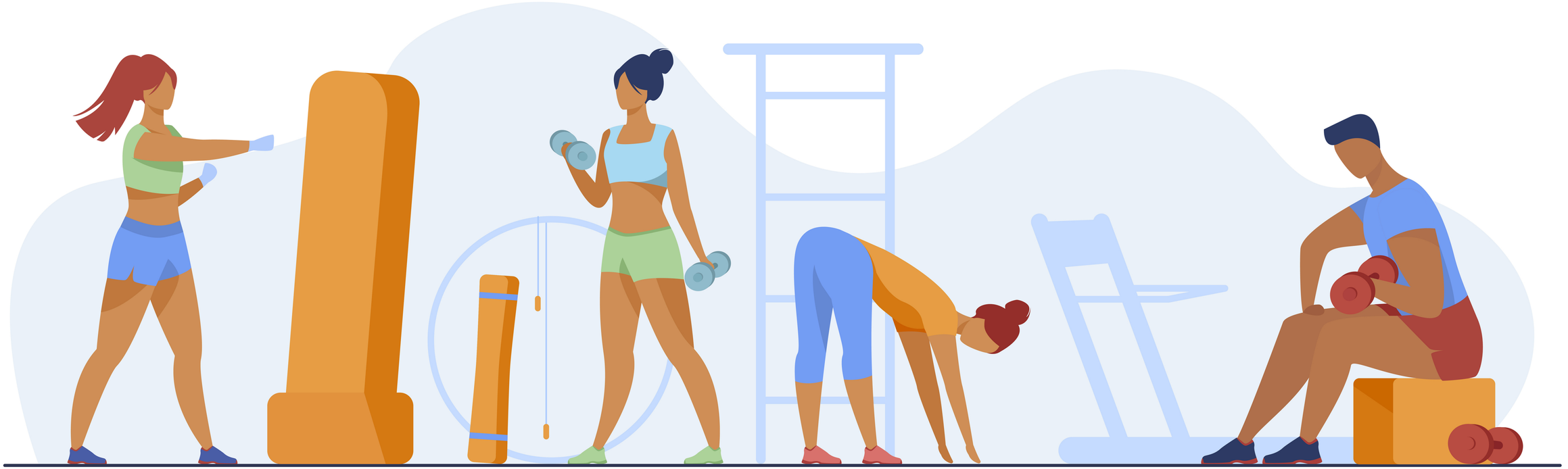
LIUDMILA LEVITINA, APRIL 2020



A/B TEST FOR ‘MUSCLEHUB’

DO WE NEED A FITNESS TEST WITH A PERSONAL TRAINER?

STUDY PROJECT OBJECTIVES

- Import data and practice table transformation with SQL
- Manipulate data with Pandas
- Test statistical significance with SciPy
- Visualise data with Matplotlib and Plotly
- Analyse results of A/B test
- Present the findings

A/B TEST

ARE VISITORS MORE LIKELY TO PURCHASE A MEMBERSHIP IF THEY SKIP THE FITNESS TEST?

Currently, when a visitor to MuscleHub is considering buying a membership, he or she follows the following steps:

- Take a fitness test with a personal trainer
- Fill out an application for the gym
- Send in their payment for their first month’s membership

Janet, the manager of MuscleHub, thinks that the fitness test intimidates some prospective members, so she has set up an A/B test.

LIUDMILA LEVITINA, APRIL 202

Visitors were randomly assigned to one of two groups:

GROUP A

Took a fitness test with a personal trainer

2504

VISITORS

GROUP B

Skipped the fitness test and proceed directly to the application

2500

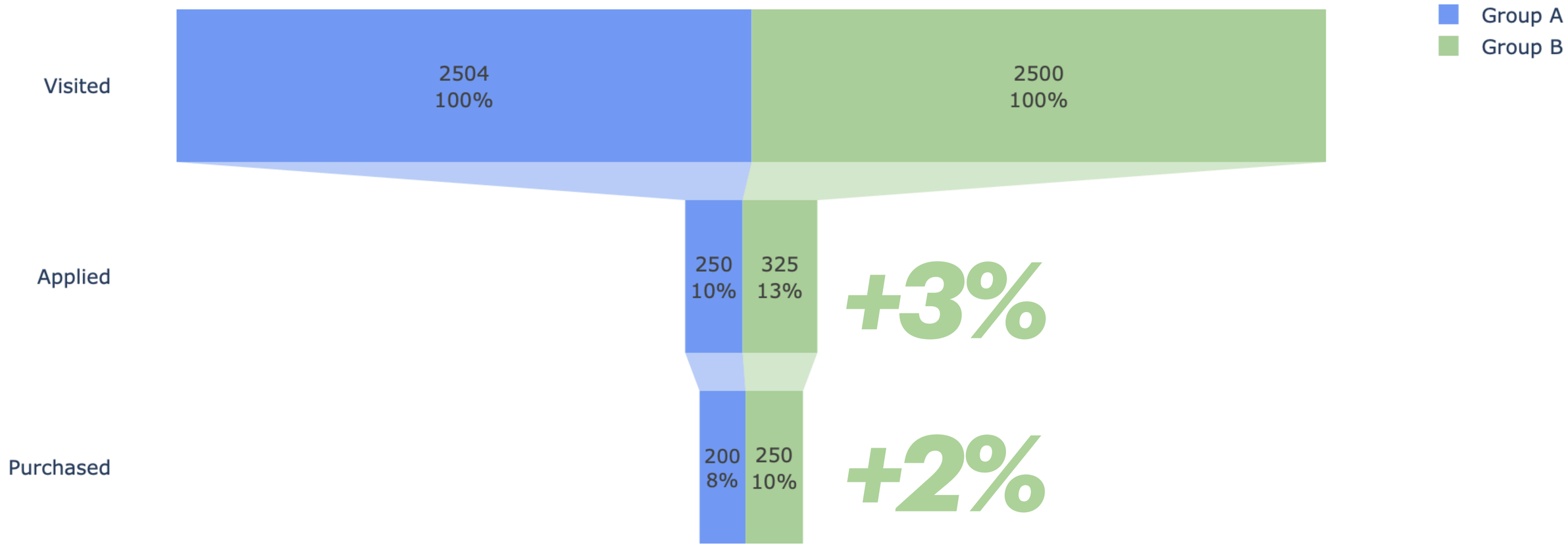
VISITORS

Hypothesis: visitors assigned to Group B are more likely to eventually purchase a membership to MuscleHub.

Calculations using Pandas

CONVERSION RATE

It seems, that visitors assigned to Group B are more likely to purchase a membership, but is this result significant?



Funnel chart using Plotly

SIGNIFICANCE LEVEL

Null Hypothesis (Ho): There is no difference between groups.

	Ha: Group B has higher purchases		Ha: Group B has lower purchases	
	Accept Null	Reject Null	Accept Null	Reject Null
Null true	Skipping test doesn't increase purchases. Keep test	Skipping test does not increase purchases. Cancel test	Skipping test doesn't decrease purchases. Cancel test	Skipping test doesn't decrease purchases. Keep test
Null false	Skipping test increase purchases. Keep test	Skipping test increase purchases. Cancel test	Skipping test decrease purchases. Cancel test	Skipping test decrease purchases. Keep test

Chi-square test tells only if difference is significant, not if it is significantly higher or lower.

Red sells represent possible errors in our analysis.

Based on the all possible actions it is more important not to accept null hypothesis if there is difference between groups.

SIGNIFICANCE LEVEL

Usually significance level $\alpha = 0.05$

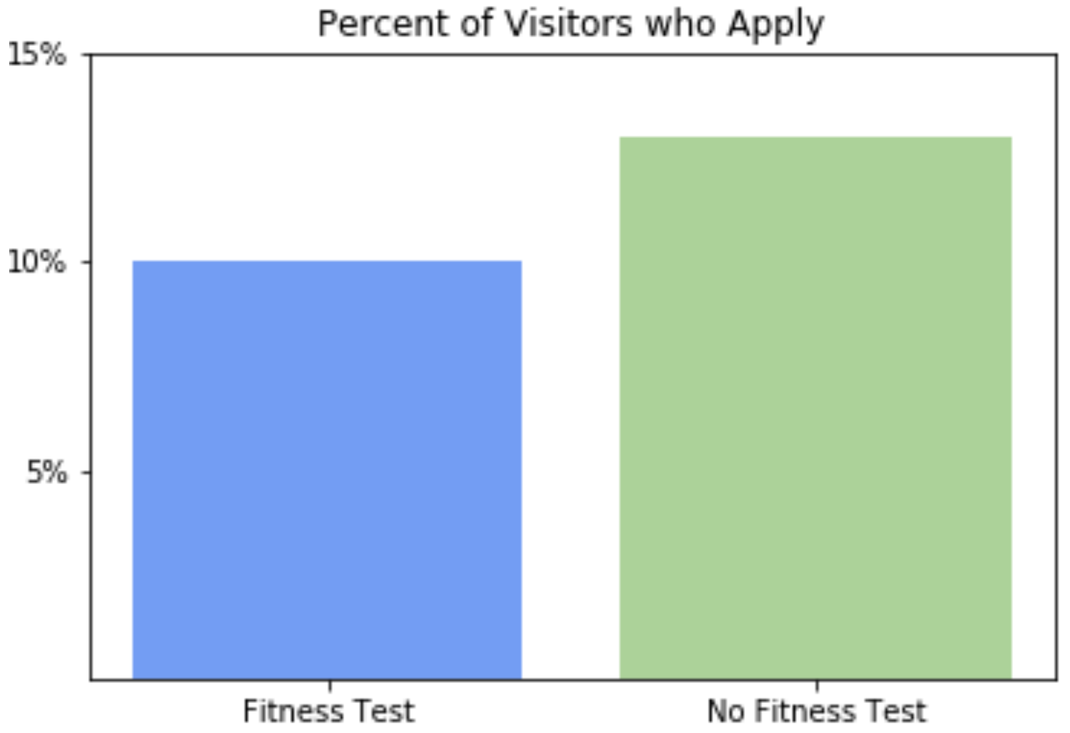
We want to **decrease our chances to accept false null hypothesis** (to keep the test that decreases purchases or to cancel the test that increases them).

Since significance level is the **probability of rejecting true null hypothesis** we want to **increase** it.

For this analysis we will set **$\alpha = 0.1$**

DO MORE VISITORS APPLY IF THEY SKIP FITNESS TEST?

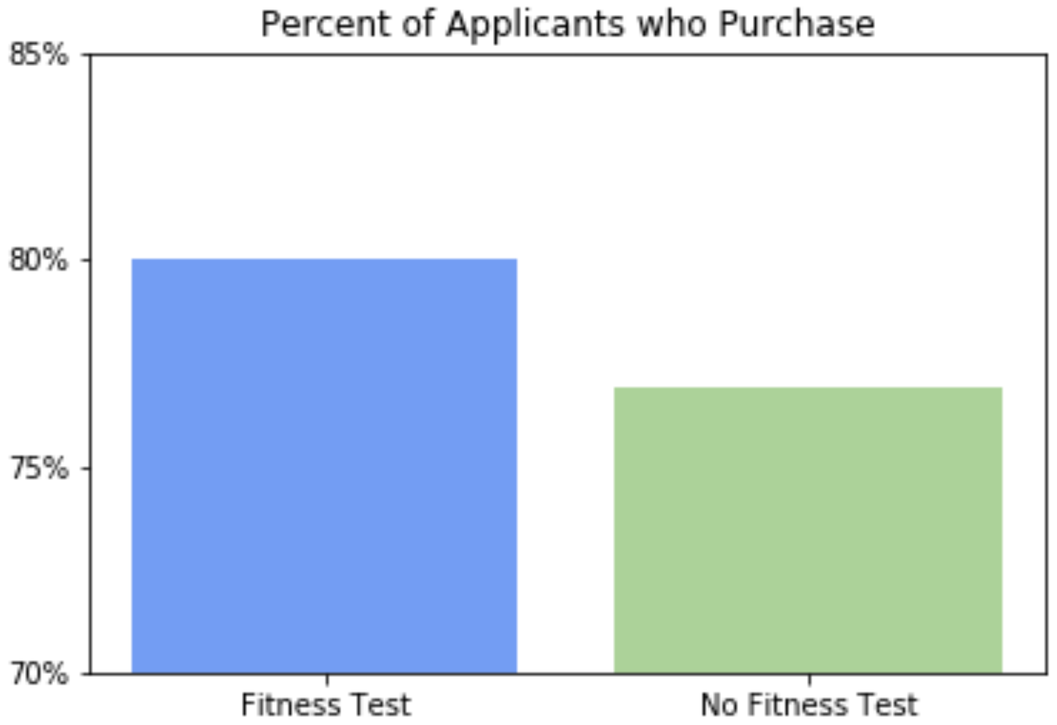
- 3% more visitors applied in Group B (skipped test)
- P-value ≈ 0.002 (probability that there is no difference between groups is 0.2%)
- α = 0.1 (probability of rejecting true null hypothesis is 10%)
- P-value < α
- Conclusion: Group that skipped fitness test indeed had more application, we do not attribute this this difference to the random chance



Bar chart using Matplotlib, calculations using Pandas and SciPy

DO MORE APPLICANTS PURCHASE IF THEY SKIP FITNESS TEST?

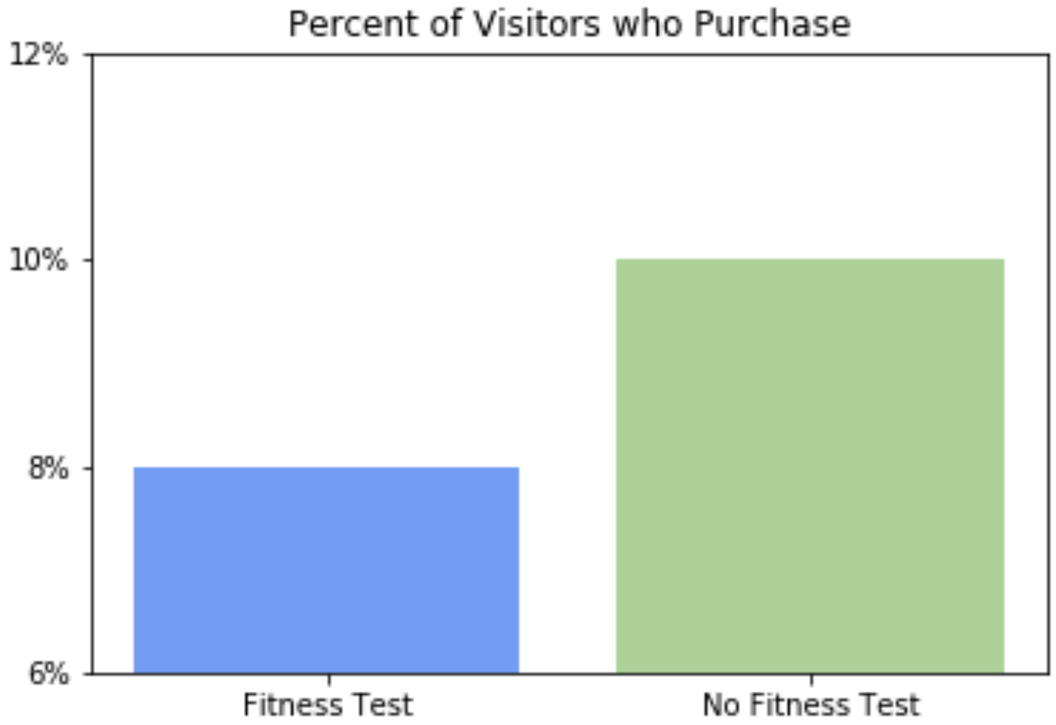
- 3% more applicants applied in Group A (had test)
- P-value ≈ 0.68 (probability that there is no difference between groups is 68%)
- α = 0.1 (probability of rejecting true null hypothesis is 10%)
- P-value > α
- Conclusion: Group that had fitness test are highly likely had more applicants due to random chance



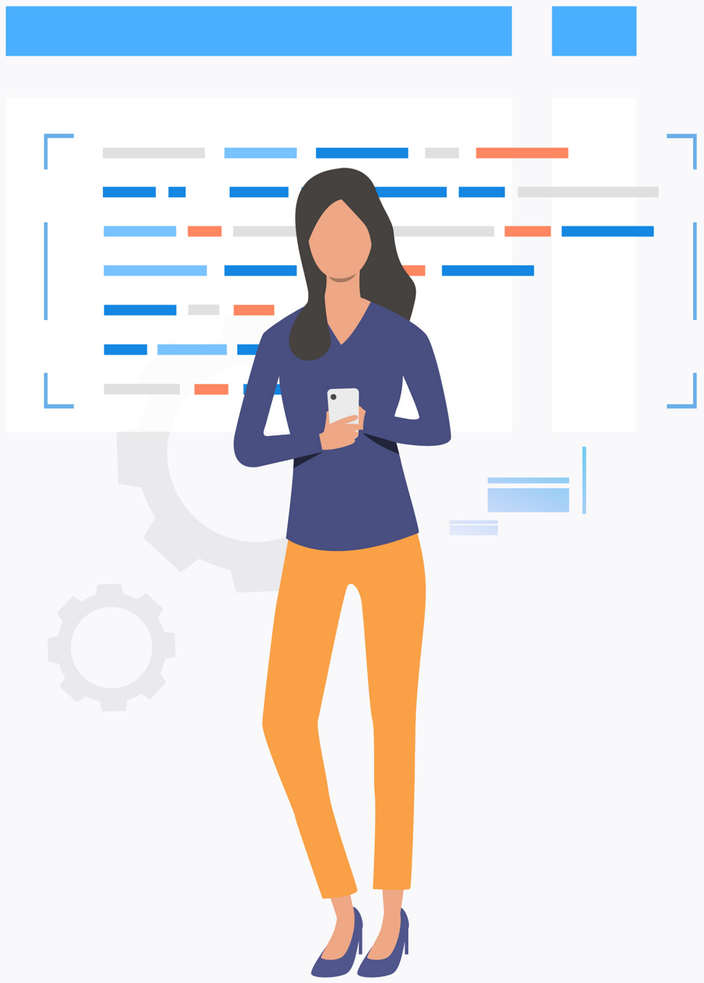
Bar chart using Matplotlib, calculations using Pandas and SciPy

DO MORE APPLICANTS PURCHASE IF THEY SKIP FITNESS TEST?

- 2% more visitors purchased membership in Group B (skipped test)
- P-value ≈ 0.018 (probability that there is no difference between groups is 1.8%)
- α = 0.1 (probability of rejecting true null hypothesis is 10%)
- P-value < α
- Conclusion: Group that skipped fitness test indeed purchased more memberships, we do not attribute this this difference to the random chance



Bar chart using Matplotlib, calculations using Pandas and SciPy



RECOMMENDATIONS

It seems that Janet is right and **fitness test intimidates some visitors.**

But we recommend to conduct more split tests and determine:

1. What is the difference in revenue from one member? It is possible that those members who had fitness test with trainer could purchase more personal training sessions, have more results and give more recommendations to friends.

2. Why so few visitors apply? Numbers are strangely low in both groups, while people usually visit fitness club if they are already interested in membership. May be another split test can be done in groups with different tours.