

# Codecademy Pro Introduction to Data Analysis Capstone Project

**Option 1: MuscleHub A/B test**

Ryan T. Maloney

# Description of the MuscleHub A/B test

- MuscleHub is a fashionable inner-city fitness club. Currently, when a prospective new member is interested in joining the gym, they undertake three steps:
  - 1. Take a fitness test with a qualified fitness instructor
  - 2. Fill out a (no-obligation) application form for joining
  - 3. Accept the terms and conditions of membership and make a payment for the first month's membership.
- Janet, the MuscleHub manager, has hypothesised that the initial fitness test might be off-putting for prospective new members. She sets up an A/B test to investigate this, randomly assigning a large sample of first-time visitors to the gym to either:
  - **Group A:** who undertake the fitness test with the instructor, or
  - **Group B:** who do not undertake the fitness test and instead proceed directly to being given the membership application.

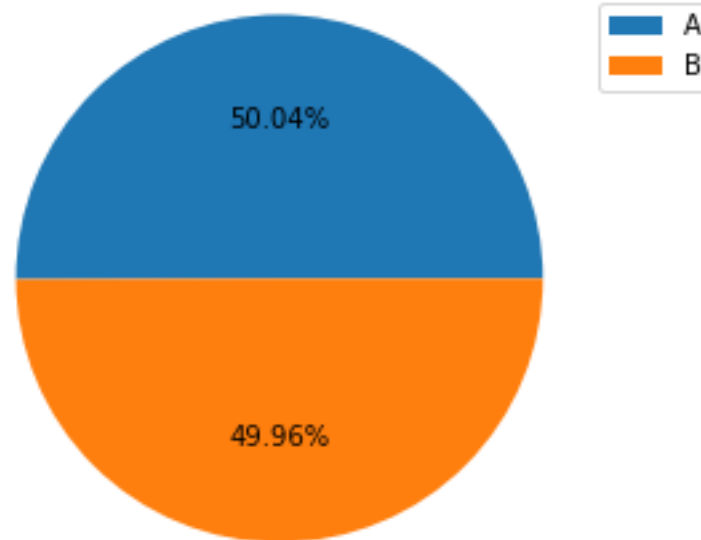
# Dataset summary

- Janet has given us access to an SQLite database containing a number of tables. Each table provides information relevant to each stage of the A/B test, and is useful in determining what proportion of visitors eventually go on to purchase memberships, which is the desired outcome.
- Each table contains information including the first and last name, email address, and sex of each visitor to MuscleHub. The date information varies across tables, according to which group visitors were assigned to, and which later stages of the process they undertook. They are:
  - **visits**: includes date of visit
  - **fitness\_tests**: includes date of fitness test of visitors assigned to Group A
  - **applications**: includes date at which visitors from either group submitted the application form, if at all.
  - **purchases**: includes date at which visitors from either group made a membership purchase, if at all.

# A/B test group assignment

- Not all visits occurred during the A/B test, so we only pull data following 1 July, 2017.
- This leaves us with 5004 visitors assigned to the A/B test: approximately half randomly assigned to each group, as illustrated in the pie chart below.

Percentage of gym-goers assigned to each group

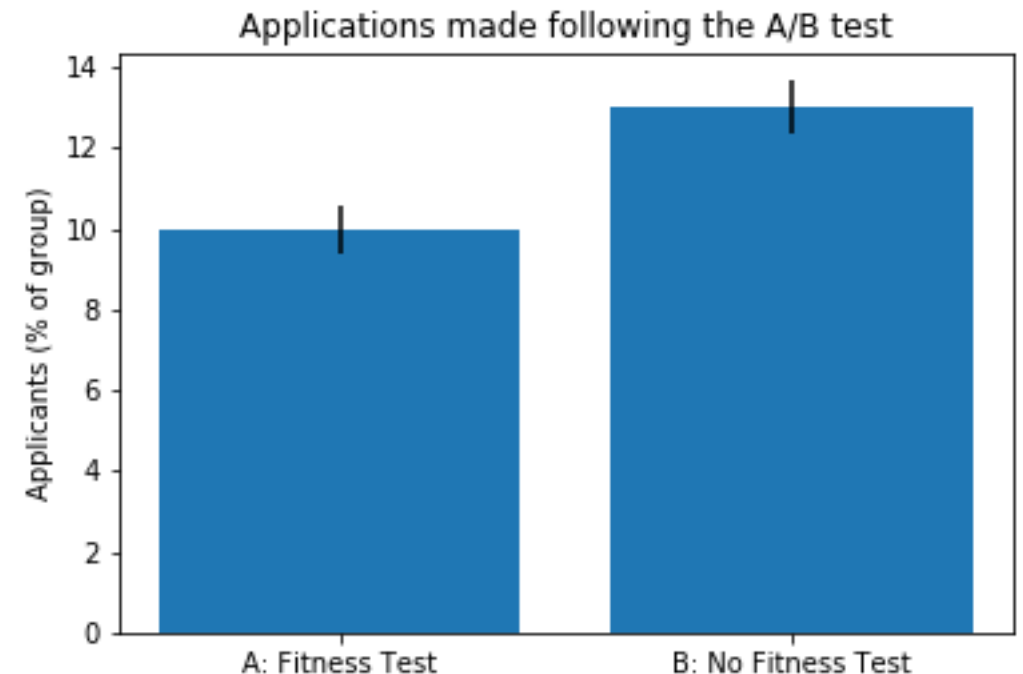


# Summary of qualitative data

- 4 brief interviews with participants in the A/B test are also provided.
- The information we can glean from the interviews alone is quite limited.
- While one visitor was clearly in favour of the fitness test (Cora, 23), another clearly was not (Sonny, 26).
- It's not clear which group the other two visitors (Jesse, 35 and Shirley, 22) were assigned, though based on their testimonials it would seem that they would prefer not to undergo a fitness test. While both were positive about their introduction to MuscleHub, only Shirley went on to purchase a membership. Jesse declined, but for reasons other than the fitness test (dirty gym equipment).

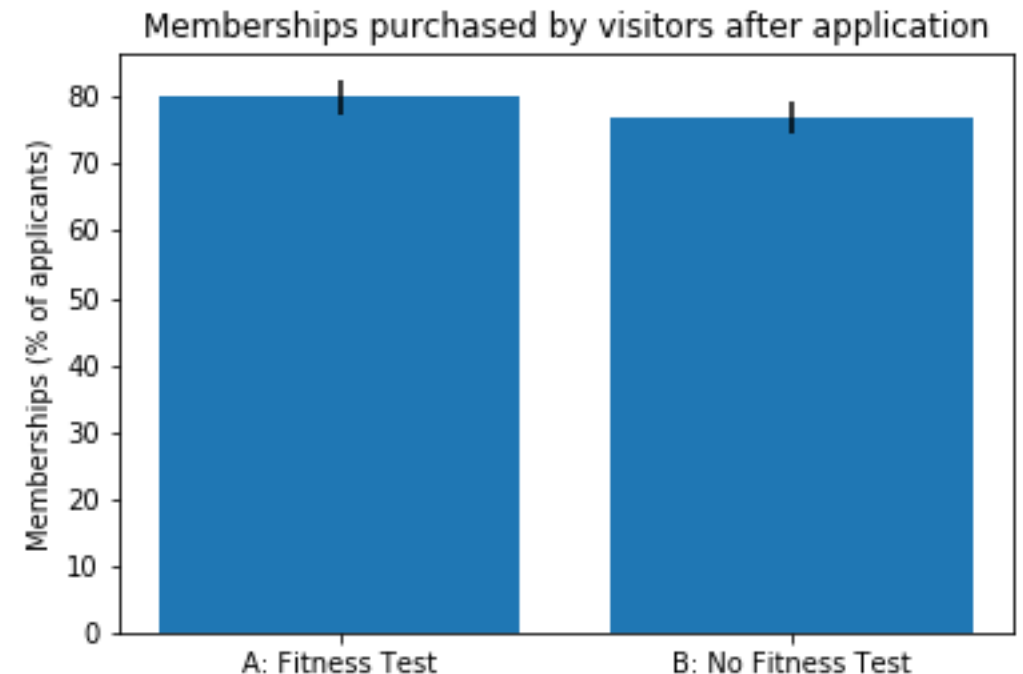
# Results 1: Proportion of each group who go on to make an application

- Firstly, we want to examine the proportions of Groups A and B that actually submit an application. We can see from the bar plot below that the proportion of visitors submitting applications is slightly larger in Group B.
- Because these data represent proportions of categorical variables, we will perform the **chi-squared test for independence**.
- $\chi^2(1) = 10.89, p < 0.001$ . We can therefore reject the null hypothesis and conclude that the proportion of visitors in Group B that make an application is significantly greater than that in Group A.



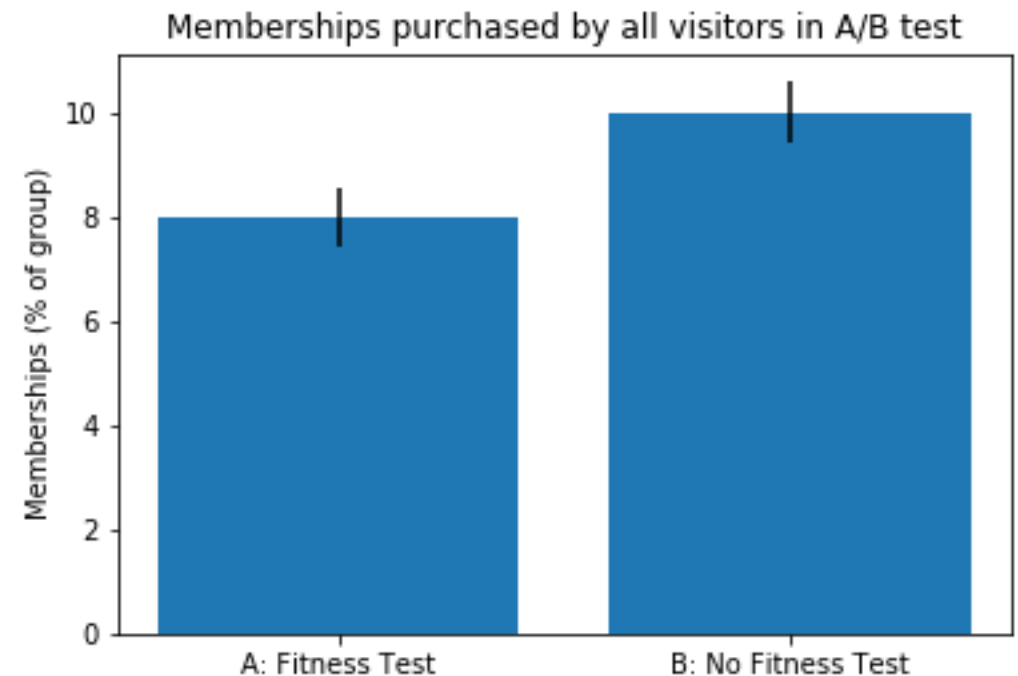
## Results 2: Proportion of applicants who purchase memberships

- Secondly, we'll examine the subgroup of visitors who made applications, and the proportions of those who went on to actually purchase memberships. From the figure below, it seems that the proportion is slightly larger in Group A.
- Why? It could be that those undergoing the fitness test feel more pressured by the instructor to make an application, or are more motivated to do so after failing the fitness test.
- We're again dealing with proportions of categorical variables, so we'll use the **chi-squared test** to establish whether this difference is significant.
- $\chi^2(1) = 0.62$ ,  $p = 0.43$ . We will therefore retain the null hypothesis and conclude that there is no significant difference in the proportion of applicants from Groups A and B who then go on to purchase memberships.



# Results 3: Proportion of all visitors in the A/B Test who purchase memberships

- The final comparison is of the most interest in the experiment: we examine what proportion of all visitors taking part actually go on to purchase memberships. From the bar plot below, it seems that Janet's original hypothesis may be correct: visitors may be more likely to purchase a membership if they do not undertake the initial fitness test.
- We will test this assumption using the **chi-squared test**, because we are again dealing with simple proportions of categorical variables.
- $\chi^2(1) = 5.95$ ,  $p = 0.015$ . We can therefore reject the null hypothesis, and conclude that there is evidence to support Janet's original prediction.





# Conclusions and recommendations to MuscleHub

- Based on the series of chi-squared tests performed, we can conclude that a significantly larger proportion of visitors who do NOT undertake the initial fitness test go on to make applications for membership, and furthermore, go on to actually purchase memberships.
- This supports Janet's original hypothesis that the fitness test might be dissuasive for prospective new members.
- Qualitative interviews with a handful of visitors suggested some were happy with the fitness test, others not, but these interviews alone were not particularly informative and could not be used to make an informed decision about the value of the fitness test to the business.
- My recommendation to Janet and the MuscleHub management would be to remove the fitness test requirement from the steps undertaken by prospective new members; or better yet, make it optional, so that new visitors can take the test if they so choose.