Codecademy Data Analysis Intensive Course CAPSTONE project

MuscleHub A/B Test

By

Francisco R. Silverio

Description of A/B Test

- The A/B Test for MuscleHub resulted in a process which showed the following:

- Grouping of test subjects into Groups A and B.

- 5004 people were brought into the study, beginning in July 1st, 2017, were brought into the A/B test.

- 2504 into Group A which required a physical test prior to offering an application and

- 2500 into Group B which did not require a physical test prior to offering an application.

Description of A/B Test Continued

- After importing all of the data sets; visits, fitness_tests, applications, and purchases, I left joined them using SQL into one dataframe called 'df'.

- Once I had 'df' established, I manipulated the data in order to answer the following question:

    - 1. How many people were assigned to Group A and B? This was visualized into a pie chart using matplotlib.
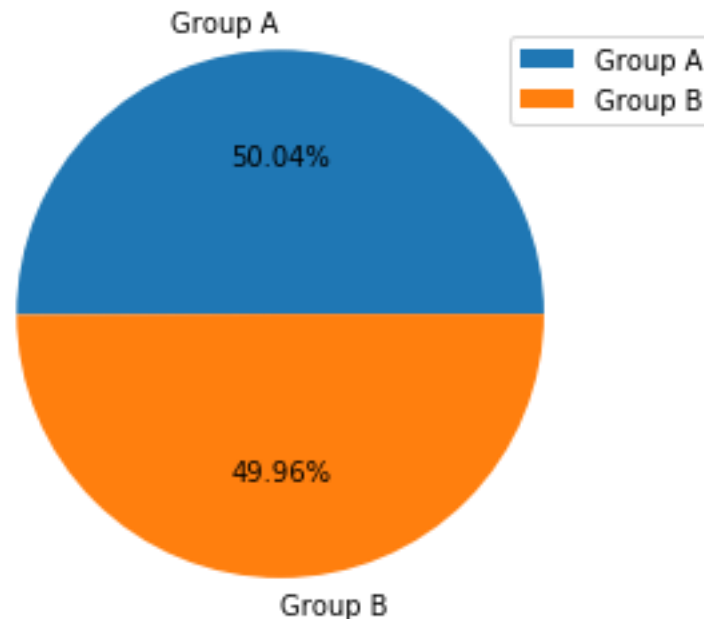
Description of A/B Test Continued

- The next question that was answered: How many people picked up applications?

- I created a value called app_counts and then had to pivot the data in order to obtain the result: Group A: 250 (9.984%) took applications and 2250 did not. Group B: 325 (13.00%) took applications and 75 did not.

- Therefore, more people from Group B took applications.

- Here is a bar chart visualizing this result:
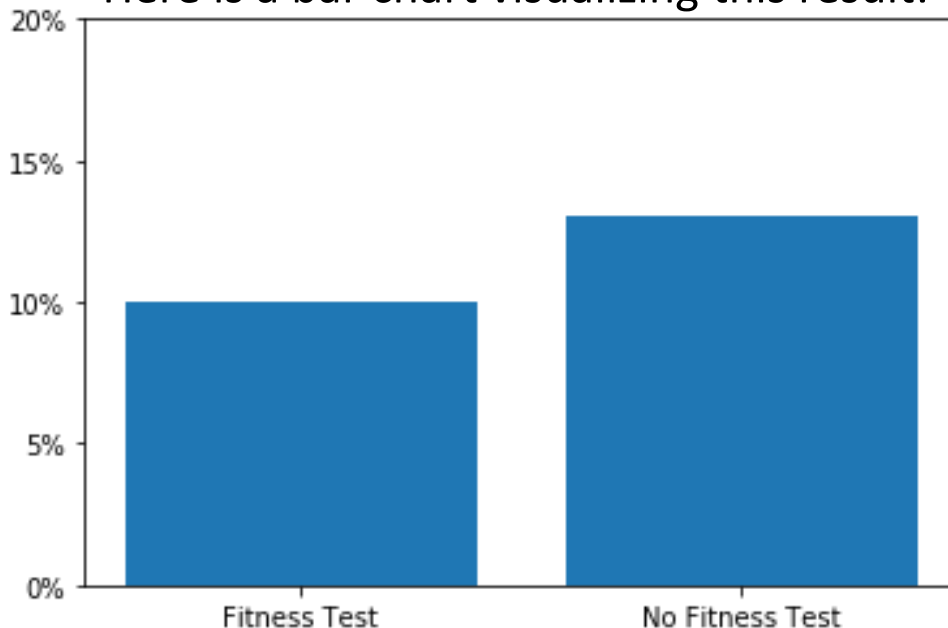
I believed that by not making the physical test mandatory, more people took applications. Therefore, in order to ensure that the statistical differences were significant or not, and because there are more than two categorical datasets to compare, I ran a Chi Square test to look for the p-value. The P-value result was less than 0.05%, so I reject the Null hypothesis that there no significance between the two Groups. The difference was significant and not random.

Description of A/B Test Continued

- Next I used a lambda function to determine who purchased a membership by adding two new columns called 'is_member' and 'Not Member.

- Then, I pivoted this dataframe to show how many of the applications taken became members. Group A took 250 application and Group B took 325 applications. The result was that Group A had a better pull through percentage (80%) from application to membership than the larger application pool from Group B (76.9%).
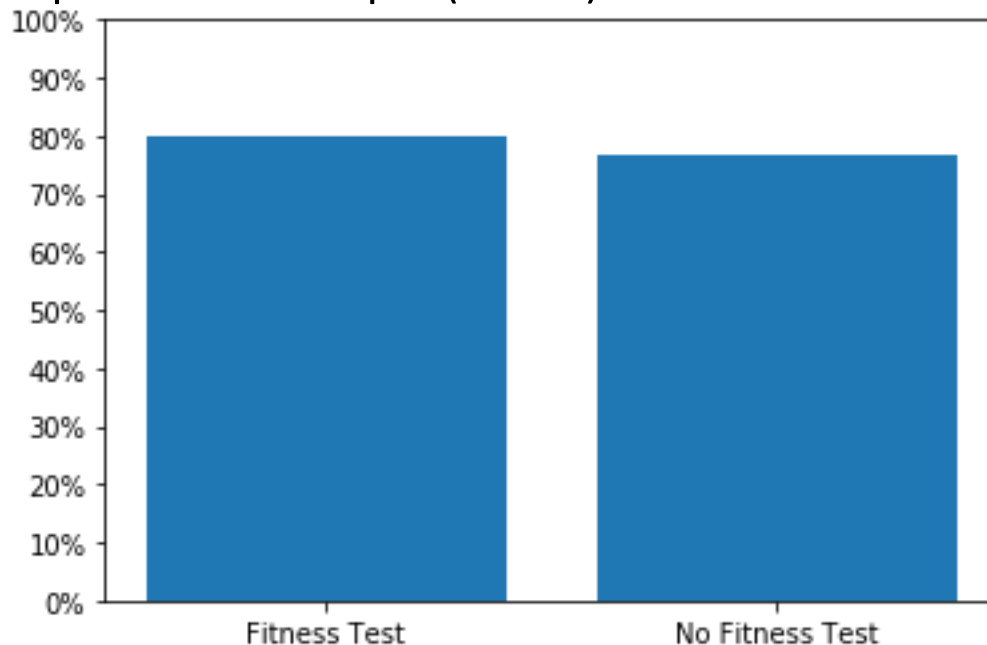
After identifying the percentage of applicants that became members from each group, I decided to run another Chi Square Test in order to obtain the P-value and find any significant difference.
This time the P-value was 0.43. Since the result was well over 0.05, this means that the result is not significant.

Description of A/B Test Continued

- After completing the 2nd Chi Square Test, I worked on my final dataframe: final_member_pivot. This dataframe was created by running a groupby of the original 'df' for 'is_member' and then pivoting it by 'ab_test_group and 'first_name', then adding a two new columns; 'Total' = 'members' + 'Not Members' and 'Percent Purchase' = 'member' divided by the 'Total' number of members from each Group.

- The result, based on overall visitors, was that more people from Group B (10%) became members than from Group A(7.9%)
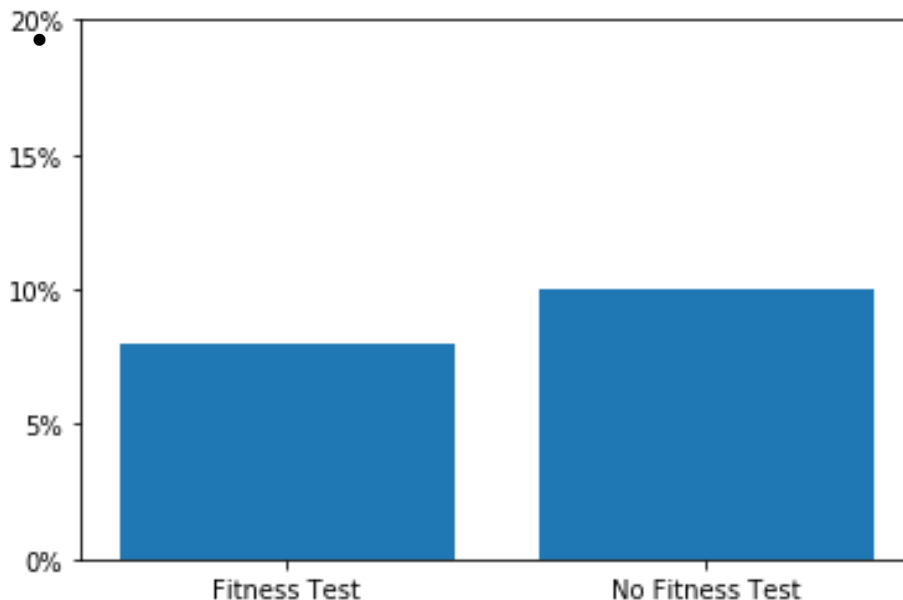
I ran my last Chi Square Test in order to test for the P-value and verify if there was any significance between Groups A and B. The result indicates strong evidence against the null hypothesis, so I reject the null hypothesis that there is no relationship between two Groups The conclusion is that there was significant difference because the P-value was lower than 0.05. (0.0147)

Recommendation

- Based on the aforementioned data results which showed a greater interest in Group B to take Application, but a greater percentage of Group A actually becoming Members after taking the Application, I would recommend offering a physical test to those who are interested, but not making it mandatory. Essentially, you could still receive a bigger lift in the number of applications taken from people who fit the Group B category and who don't want to take a physical test, but you could still benefit from offering the test to people who fit the Group A personality and are the type of person who appreciate the test. This means that they would make them more likely to sign up for a membership. Take the best aspects of both data results and blend them together to maximize your membership enrollment.