CODECADEMY

INTRODUCTION TO DATA ANALYSIS

Final Project:
MuscleHub A/B Test

Morteza Mashayekhi

Date: March 17, 2018

PROBLEM DESCRIPTION

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.

Visitors will randomly be assigned to one of two groups:

- Group A will still be asked to take a fitness test with a personal trainer
- Group B will skip the fitness test and proceed directly to the application

Janet's hypothesis is that visitors assigned to Group B will be more likely to eventually purchase a membership to MuscleHub.

PROBLEM DESCRIPTION

Null Hypothesis:

H0: There is no significant difference between visitors in group A and B in terms of purchasing a membership to MuscleHub

Alternative Hypothesis:

H1: Visitors assigned to Group B will be more likely to eventually purchase a membership to MuscleHub.

DATA UNDERSTANDING

Data sets used in this analysis:

- visits.csv including: first name, last name, email, gender, and visit date
- applications.csv including: first name, last name, email, gender, and application date
- fitness_tests.csv including: first name, last name, email, gender, and fitness test date
- purchases.csv including: first name, last name, email, gender, and purchase date
- Interviews.txt including comments from a few different gym visitors who participated in the A/B test – can be used for qualitative analysis

DATA UNDERSTANDING

- visits = pd.read_csv('visits.csv')
- application=pd.read_csv("applications.csv")
- fitness_test=pd.read_csv("fitness_tests.csv")
- purchases=pd.read_csv("purchases.csv")

- print visits.describe()
- print application.describe()
- print fitness_test.describe()
- print purchases.describe()

Variable	Frequency
Visits	6000
Application	575
Fitness_test	2500
Purchase	450

DATA PREPARATION

- Merging all data sets using left join function and select cases with visit_date after 7-01-17
- Creating new variables: ab_test_group, is_application, and is_purchase using lambda function

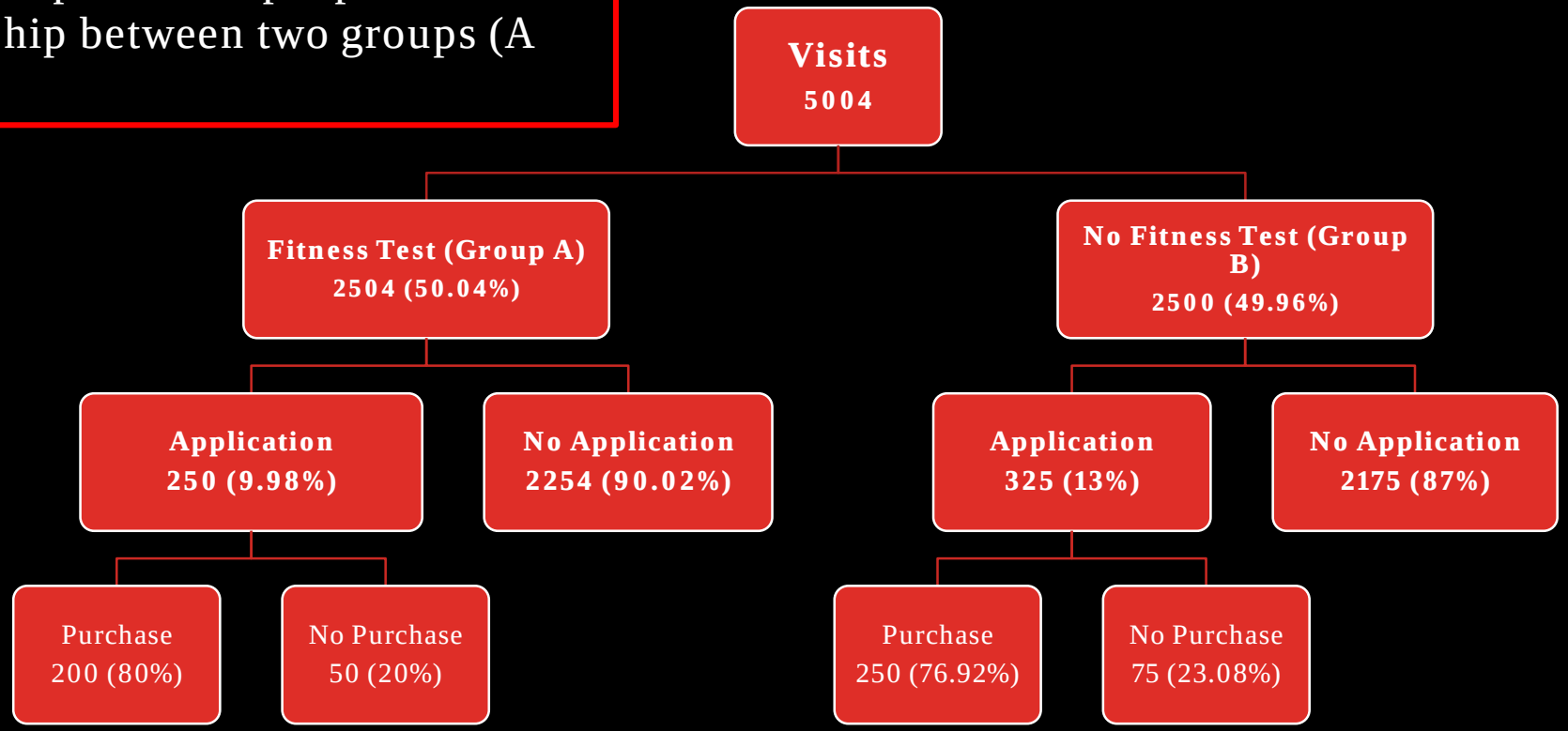
DATA ANALYSIS

- Using groupby function to count the number of cases for following variables:
  - Number of application in each of group A and B
  - Number of purchases in each of group A and B

DATA ANALYSIS

- Why Chi-square test for testing hypotheses?

In this analysis we use chi-square test to test the hypotheses because we aim to compare the proportion of people who applied or purchased a membership between two groups (A and B).



DATA ANALYSIS

- Using chi square test we can test the following hypothesis:
  - There is no significant difference between group A and B visitors in terms of number of applications

Ab_test_group	Application	No Application	Total	Percentage	Reject null hypothesis
A	250	2254	2504	9.98%	Yes Chi-square=0.001 Less than 0.05
B	325	2175	2500	13%	

- Therefore, we found support for alternative hypothesis and can conclude that there is indeed a significant difference between group A and B visitors in terms of number of applications

DATA ANALYSIS

- Using chi square test we can test the following hypothesis:
  - There is no significant difference between group A and B members in terms of number of purchases

Ab_test_group	Member	Not Member	Total	Percentage	Reject null hypothesis
A	200	50	250	80%	No Chi-square=0.432 more than 0.05
B	250	75	325	76.92%	

- Therefore, we found no support for alternative hypothesis and can conclude that there is indeed not a significant difference between group A and B members in terms of number of purchases

DATA ANALYSIS

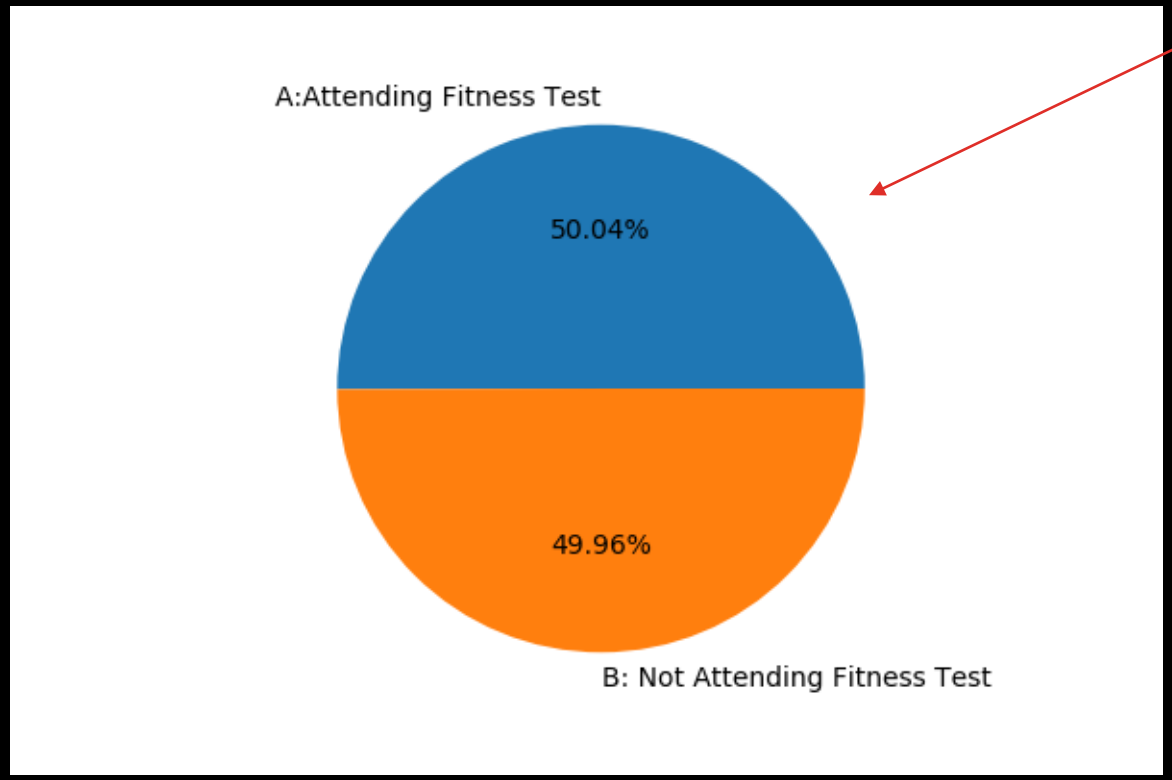
- Using chi square test we can test the following hypotheses:
  - There is no significant difference between group A and B visitors in terms of number of purchases

Ab_test_group	Member	Not Member	Total	Percentage	Reject null hypothesis
A	200	2304	2504	7.98%	Yes Chi-square=0.014 Less than 0.05
B	250	2250	2500	10%	

- Therefore, we found support for alternative hypothesis and can conclude that there is indeed a significant difference between group A and B visitors in terms of number of purchases

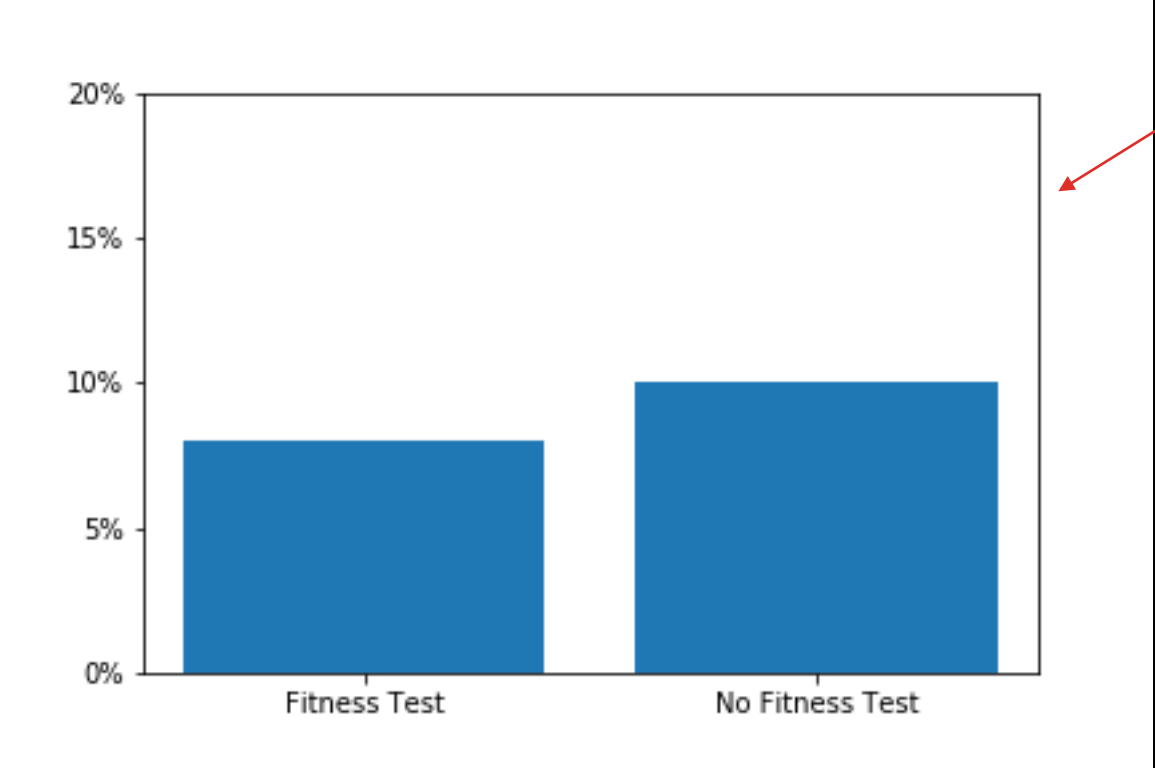
DATA ANALYSIS

There is no significance difference between group A and B



DATA ANALYSIS

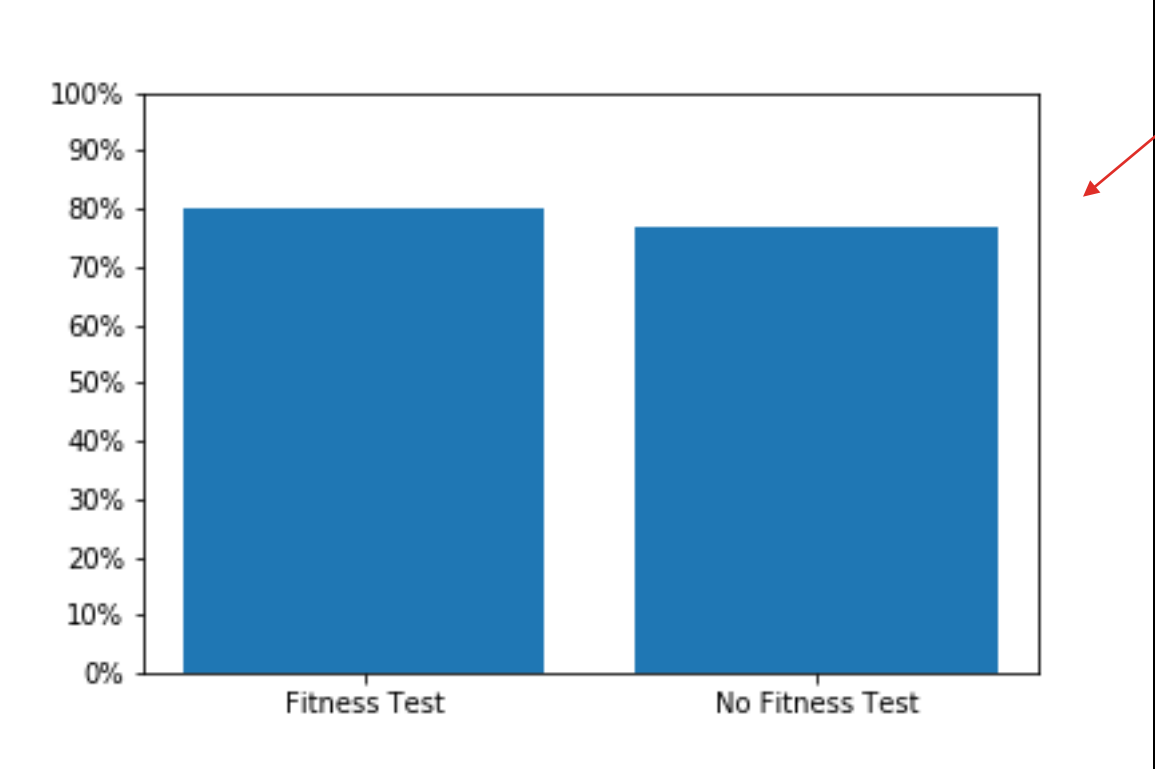
Percent of Visitors Who Applied



Results of data analysis (chi-square test) shows that There is a significance difference between group A and B visitors in terms of number of applications.

DATA ANALYSIS

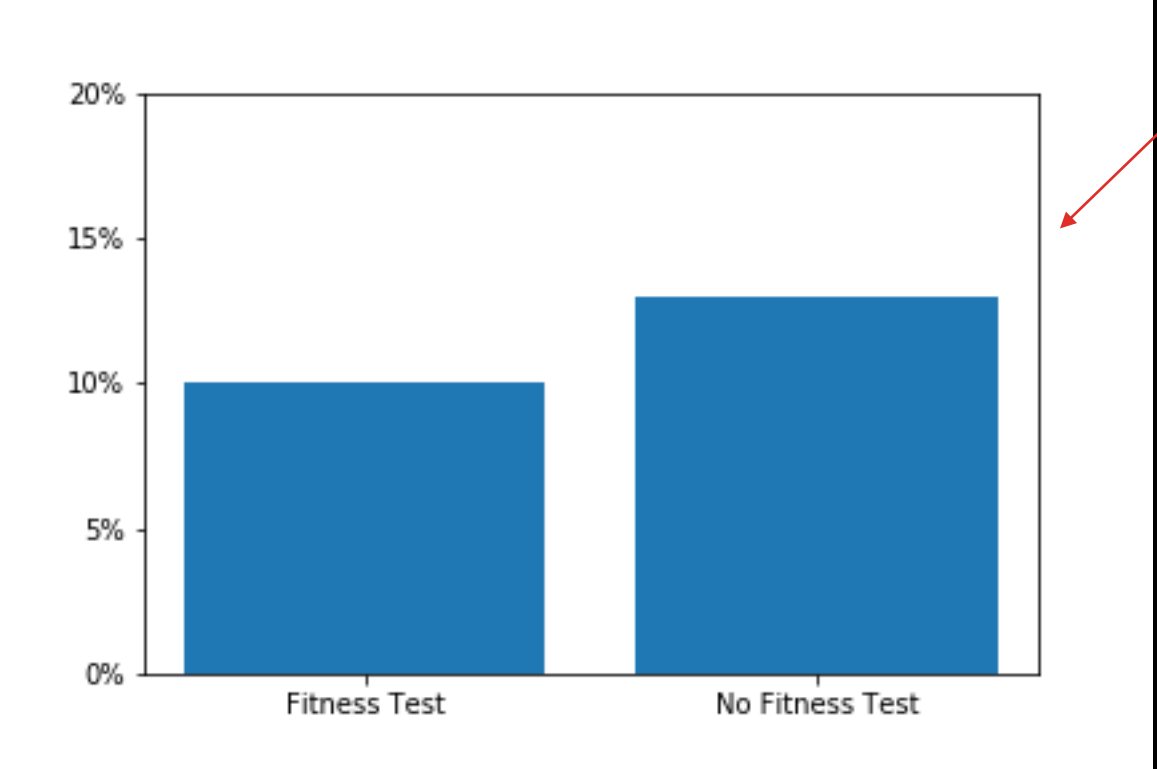
Percent of members Who purchased



Results of data analysis (chi-square test) shows that There is not a significance difference between group A and B members in terms of number of purchases.

DATA ANALYSIS

Percent of Visitors Who Purchase



Results of data analysis (chi-square test) shows that There is a significance difference between group A and B visitors in terms of number of purchases.

QUALITATIVE ANALYSIS

- The following table shows the results of qualitative analysis based on interviews.txt file

Comments	Positive regarding taking fitness test	Negative regarding taking fitness test
MuscleHub's introductory fitness test was super helpful for me! After taking the fitness test, I had to sign up	×	
I still ended up not signing up for a membership because the weight machines had all those sweat stains on them and you know, no thanks.		×
I took the MuscleHub fitness test because my coworker Laura recommended it. Regretted it		×
The people there were suuuuuper friendly and the whole sign-up process took a matter of minutes. I tried to sign up for LiftCity last year, but the fitness test was way too intense.		×

RECOMMENDATION

- Based on results of data analysis, we can conclude that visitors assigned to Group B (the group who skips the fitness test) will be more likely to eventually purchase a membership to MuscleHub. Therefore, we recommend that MuscleHub Company eliminate the step that requires an applicant takes a fitness test before filling out an application from its sign up process.