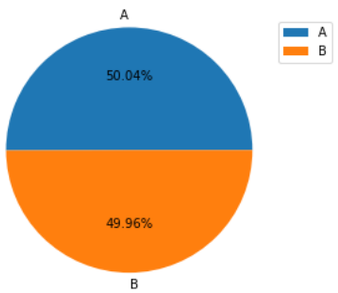
Capstone Project No.1

MuscleHub A/B Test

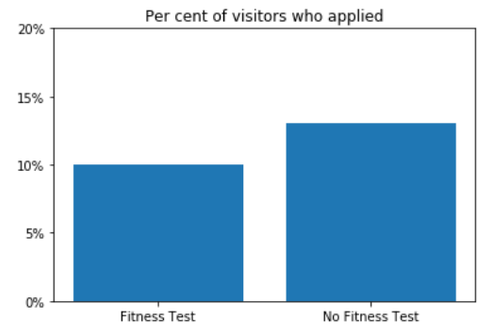
The purpose of the project was to test an hypothesis regarding the efficiency of a method to attract clients of a gym. The gym normally offers a fitness test before the application to the gym course. The hypothesis of the gym's manager is that this method would intimidate perspective clients instead of attracting them.
So we were given datasets about two groups of people:
Group A - people which were asked to sit a fitness test
Group B - people which were asked to skip it and proceed to the application.
During the project the relative percentages of applicants and purchasers were compared between the 2 groups and meaningful statistical metrics were extracted to reject or confirm the starting hypothesis.
Lets start by dividing almost equally the population between the 2 groups and check with a pie plot if indeed we have 50% in group A and 50% in group B.



Then I subsequently compared, as it can be seen from the following 3 bar-plots:

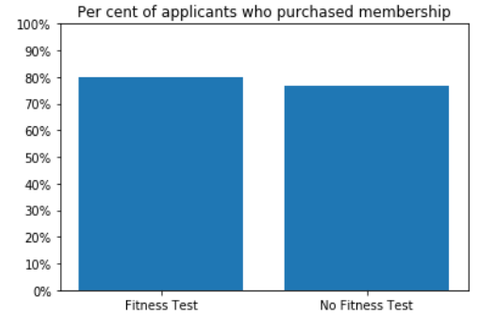
1. **the percentage of visitors who applied to a gym course belonging to group A and group B:**
2. **the percentage of applicants belonging to the 2 groups which finally bought a membership**
3. **the percentage of global visitors which finally became members**

1)



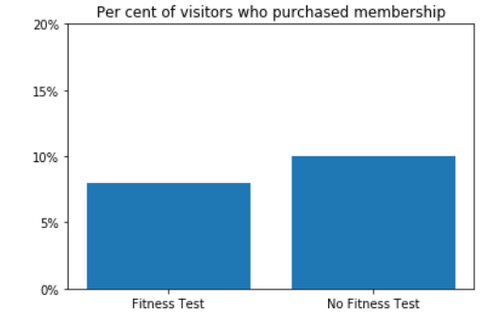
To test whether those fluctuations are statistically significant or are random I selected the chi_square method.
Regarding the plot 1) we see that there are more people from Group B applying for the gym with respect to Group A but we have to test if this is a statistically significative result or not, so basically we have to check if p-value is larger than the target value 0.05. If it is then the observed fluctuation is the result of chance, otherwise it is dictated by a different statistics.

2)



Applying the chi_square method for data in plot 1) gives out p_val = 0.000964782760072 so the difference in data is not random fluctuation = the fitness test has negative impact on applicants (there are less applicants coming from Group A and this difference is)

3)



Applying the chi_square method for data in plot 2) gives out p_val = 0.432586460511 so the difference in data is a random fluctuation = the fitness test has no impact on applicants which are buying a membership at the end.

Applying the chi_square method for data in plot 3) gives out p_val = 0.0147241146458 so the difference in data is not random fluctuation = the fitness test has impact on visitors buying a membership at the end (and this last result confirms the negative effect of the fitness test already discovered in the first chi_square test).

Final recommendation for the gym's manager: quit the fitness test as a method to attract members and invest the spare resources in different ads - try out brochures to distributes in sport centres or to disseminate into bars/restaurants

C.Cantini 17.04.18