Codeacademy – Introduction to Data Analysis Capstone Project – Musclehub A/B Test

Jussi Niemistö

Helsinki, Finland

18th of June 2018

What happened in the Musclehub A/B test

- Musclehub's manager Janet assumed that the fitness test that the Group A visitor took with a personal trainer did not add value for the business meaning that the visitor who took the fitness test is less likely to become a member compared with a visitor from Group B who skipped the fitness test
- Our objective was to find out if Janet needs the additional phase of visitors taking the fitness test with a personal trainer before making the application and becoming a member
- In the A/B test of Musclehub, we wanted to find out if the two test groups, A and B, had significantly different outcomes
- Group A took a fitness test with a personal trainer and Group B skipped the fitness test and proceeded directly to the application phase
- We found out that there was a statistically significant results in these groups in the application phase and the results are in favour of the Group B, the visitors who skip the fitness test

Summary of the data set

- In this A/B test we had a data set of Musclehub visits, fitness tests, applications for a membership and membership purchases
- We combined the customer visit data table with fitness test, membership application and membership purchase data tables by linking them by customer name and e-mail
- After combining these data tables we had a larger data table with 5004 rows, which we could use for Musclehub's A/B Test and more importantly test our different hypotheses / approaches
- The combined data table had 2504 visitors from group A (fitness test group) and 2500 from group B (no fitness test group)
- The first rows for the data set used for Musclehub's A/B test:

	first_name	last_name	gender	email	visit_date	fitness_test_date	application_date	purchase_date
0	Kim	Walter	female	KimWalter58@gmail.com	7-1-17	2017-07-03	None	None
1	Tom	Webster	male	TW3857@gmail.com	7-1-17	2017-07-02	None	None
2	Edward	Bowen	male	Edward.Bowen@gmail.com	7-1-17	None	2017-07-04	2017-07-04
3	Marcus	Bauer	male	Marcus.Bauer@gmail.com	7-1-17	2017-07-01	2017-07-03	2017-07-05
4	Roberta	Best	female	RB6305@hotmail.com	7-1-17	2017-07-02	None	None
5	Joseph	Foley	male	JosephFoley81@gmail.com	7-1-17	None	None	None
6	Carrie	Francis	female	CF1896@hotmail.com	7-1-17	2017-07-05	None	None

The result of the hypothesis test 1 – Who picks up an application?

- We wanted to know if there was a difference in the two visitor groups in the application phase
- Before the application, group A took the fitness test and group B did not. Was one group more likely to fill out an application?
- Group B was more likely to fill out an application – 325 applications of 2500 visitors (13 % picked up the application for group B versus 10 % for group A)

is_application	ab_test_group	Application	No Application	Total	Percent with Application
0	A	250	2254	2504	0.09984
1	B	325	2175	2500	0.13000

- For this hypothesis test, we used the Chi Square test, because it is useful in situations, where we want to compare two or more categorical data sets.
- The Chi Square test indicates that the results are statistically significant (p-value < 0.05)
- The group B, who did not take the fitness test, were more likely to fill out an application

The result of the hypothesis test 2 – Who purchases a membership?

- We wanted to know how many people of those who filled an application ended up purchasing a membership for Musclehub
- 80 % of the group A visitors who applied ended up purchasing a membership. For group B 77 % of applicants purchased.

is_member	ab_test_group	Member	Not Member	Total	Percent Purchase
0	A	200	50	250	0.800000
1	B	250	75	325	0.769231

- For this hypothesis test, we used the Chi Square test for the same reason as in the hypothesis test 1 (two categorical data sets)
- The Chi Square test indicates that the results are not statistically significant (p-value > 0.05)

The result of the hypothesis test 3 – Who purchases a membership?

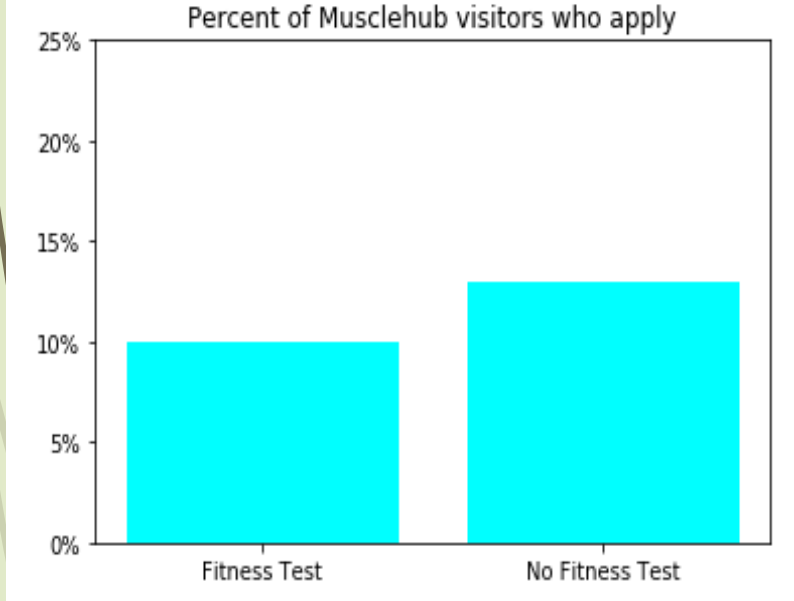
- We wanted to know how many people of all visitors in the two groups ended up purchasing a membership for Musclehub
- 8% of the group A visitors ended up purchasing a membership. For group B 10 % of visitors purchased.

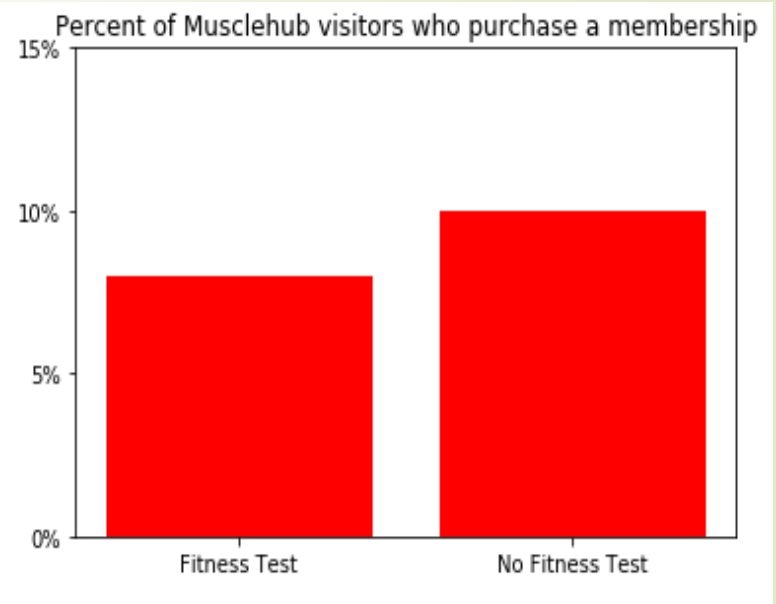
is_member	ab_test_group	Member	Not Member	Total	Percent Purchase
0	A	200	2304	2504	0.079872
1	B	250	2250	2500	0.100000

- For this hypothesis test, we used the Chi Square test for the same reason as in the hypothesis test 1 (two categorical data sets)
- There is no statistical significance (p-value > 0.05), when we consider all people who visit MuscleHub

C00BE532.py

Results in 3 Graphs





Summary of qualitative data

- The interviews gave us more information on the customer experiences of visitors:

1. Female visitors were impressed by the fitness test
2. Female visitors had seen ads in social media
3. Female visitors recommended the fitness tests to their friends and colleagues
4. Male visitors did not like the fitness test and regarded it as useless

- Conclusion is that Musclehub's Janet could start a social media campaign to target female visitors under the age of 30 to come for a fitness test visit

Recommendation for Musclehub

- Janet, the Musclehub manager, should consider removing the fitness test from the pre-application phase and maybe offer the fitness test as a separate instructor service.
- Musclehub could also consider offering the fitness test for people, who have already filled the application. This might give better application-to-membership ratio for Musclehub's business.
- Janet might want to take a look at the quality of the personal trainer's fitness test and see if there would be ways to improve the visitor experience and get more members this way.
- After reviewing the qualitative data, it would be effective for Musclehub to start targeting female visitors with a social media campaign and get this target group to visit for a fitness test, because female visitors seem to like the fitness test more than other visitors.