

Muscle Hub A/B Testing

Ly

Valentine's Day, 2018

Background

Visitors are assigned to either Group A or Group B to see how many become members of the gym in each group

- Group A
 - Take a fitness test with a personal trainer
 - Fill out an application
 - Purchase the membership
- Group B
 - Fill out an application
 - Purchase the membership

Hypothesis: visitors assigned to Group B will be more likely to purchase a membership to MuscleHub

Concatenate the data

Combine four datasets together

- **Visits:** potential gym customers who visited MuscleHub
- **Fitness tests:** customers in "Group A", who were given a fitness test, Group A is marked with a date, Group B without a date
- **Applications:** potential customers (Group A & Group B) who filled out an application, filled out the application is marked with a date, did not fill the application without a date
- **Purchases:** customers (Group A & Group B) who purchased a membership to MuscleHub, purchased is marked with a date, did not purchase without a date

Not every customer moves on to the next stage. We combine four tables using left join the Visitor table all the way to the Purchase table to get a clear view of the journey of the customers, primary key and foreign key are email, first name and last name.

Python Pandas manipulate visits data

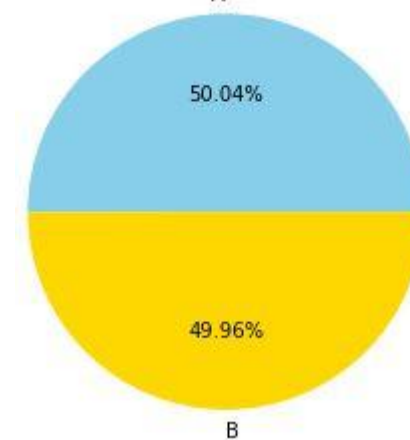
Differentiate two groups

	first_name	last_name	gender	email	visit_date	fitness_test_date	application_date	purchase_date	ab_test_group
0	Kim	Walter	female	KimWalter58@gmail.com	7-1-17	2017-07-03	None	None	A
1	Tom	Webster	male	TW3857@gmail.com	7-1-17	2017-07-02	None	None	A
2	Edward	Bowen	male	Edward.Bowen@gmail.com	7-1-17	None	2017-07-04	2017-07-04	B
3	Marcus	Bauer	male	Marcus.Bauer@gmail.com	7-1-17	2017-07-01	2017-07-03	2017-07-05	A
4	Roberta	Best	female	RB6305@hotmail.com	7-1-17	2017-07-02	None	None	A
5	Joseph	Foley	male	JosephFoley81@gmail.com	7-1-17	None	None	None	B
6	Carrie	Francis	female	CF1896@hotmail.com	7-1-17	2017-07-05	None	None	A
7	Sharon	William	female	Sharon.William@outlook.com	7-1-17	None	None	None	B
8	Teresa	Yates	female	TYates1988@gmail.com	7-1-17	2017-07-02	None	None	A
9	Salvador	Cardenas	male	SCardenas1980@gmail.com	7-1-17	2017-07-07	2017-07-06	None	A

Group by the visits by Group A and Group B. As we can see the sample size is almost equal, there is no huge bias at sample size

	ab_test_group	first_name
0	A	2504
1	B	2500

Visits by Group A and Group B



Python Pandas manipulate application data

Find those who filled out the application and those who did not using a lambda function

	first_name	last_name	gender	email	visit_date	fitness_test_date	application_date	purchase_date	ab_test_group	is_application
0	Kim	Walter	female	KimWalter58@gmail.com	7-1-17	2017-07-03	None	None	A	No Application
1	Tom	Webster	male	TW3857@gmail.com	7-1-17	2017-07-02	None	None	A	No Application
2	Edward	Bowen	male	Edward.Bowen@gmail.com	7-1-17	None	2017-07-04	2017-07-04	B	Application
3	Marcus	Bauer	male	Marcus.Bauer@gmail.com	7-1-17	2017-07-01	2017-07-03	2017-07-05	A	Application
4	Roberta	Best	female	RB6305@hotmail.com	7-1-17	2017-07-02	None	None	A	No Application

Group by two groups and application to create a pivot table to get an insightful view of the data, furthermore, get the total number of visits from Group A and Group B and calculate the percentage of the customers who filled out the application

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

Group A filled out the application is 9.98%, whereas Group B is 13%. Based on the number, the hypothesis is correct. However we need to prove it statistically

Python Scipy proves the statistical significance

We have two samples (Group A and Group B) , and they are categorical data. Hence we use Chi Square Test to prove there is no significant difference between the datasets

As we can see, the p-value is way less than .05, meaning it is significant, we reject the null hypothesis, there is a significant difference between two groups. Statistically we are confident that Group B shows a higher percentage of customers filled out application

P-Value= .000964782760072

Python Pandas manipulate purchase data of those who fill out application

Let's find out how many people actually become members after filling out the application. First find those who purchased the membership and those who did not using a lambda function, and then filter out who filled out the application.

first_name	last_name	gender	email	visit_date	fitness_test_date	application_date	purchase_date	ab_test_group	is_application	is_member
Edward	Bowen	male	Edward.Bowen@gmail.com	7-1-17	None	2017-07-04	2017-07-04	B	Application	Member
Marcus	Bauer	male	Marcus.Bauer@gmail.com	7-1-17	2017-07-01	2017-07-03	2017-07-05	A	Application	Member
Salvador	Cardenas	male	SCardenas1980@gmail.com	7-1-17	2017-07-07	2017-07-06	None	A	Application	Not Member
Valerie	Munoz	female	VMunoz1998@gmail.com	7-1-17	2017-07-03	2017-07-05	2017-07-06	A	Application	Member
Michael	Burks	male	MB9820@gmail.com	7-1-17	None	2017-07-07	2017-07-13	B	Application	Member

Group by two groups and member to create a pivot table to get an insightful view of the data, furthermore, get the total number of visits from Group A and Group B and calculate the percentage of the customers who purchased the membership. As we can see, the percentage is less than .3% apart, seems Group A is doing slightly better than Group B, we must prove this statistically

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

Python Scipy proves the statistical significance

We have two samples (Group A and Group B) , and they are categorical data. Hence we use Chi Square Test to prove there is no significant difference between the datasets

As we can see, the p-value is greater than .05, meaning it is not significant, we accept the null hypothesis, there is NOT a significant difference between two groups. Statistically we are NOT confident that Group A shows a higher percentage of customers purchased the membership

P-value= .432586460511

Python Pandas manipulate purchase data

Find those who purchased the membership and those who did not using a lambda function

first_name	last_name	gender	email	visit_date	fitness_test_date	application_date	purchase_date	ab_test_group	is_application	is_member
Kim	Walter	female	KimWalter58@gmail.com	7-1-17	2017-07-03	None	None	A	No Application	Not Member
Tom	Webster	male	TW3857@gmail.com	7-1-17	2017-07-02	None	None	A	No Application	Not Member
Edward	Bowen	male	Edward.Bowen@gmail.com	7-1-17	None	2017-07-04	2017-07-04	B	Application	Member
Marcus	Bauer	male	Marcus.Bauer@gmail.com	7-1-17	2017-07-01	2017-07-03	2017-07-05	A	Application	Member
Roberta	Best	female	RB6305@hotmail.com	7-1-17	2017-07-02	None	None	A	No Application	Not Member

Group by two groups and membership to create a pivot table to get a holistic view of the entire data, moreover, get the total number of visits from Group A and Group B and calculate the percentage of the customers who purchase the membership

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

Overall, Group B has 10% of purchase rate, Group A has about 8%. The hypothesis is correct according to the percentage. We must prove it statistically

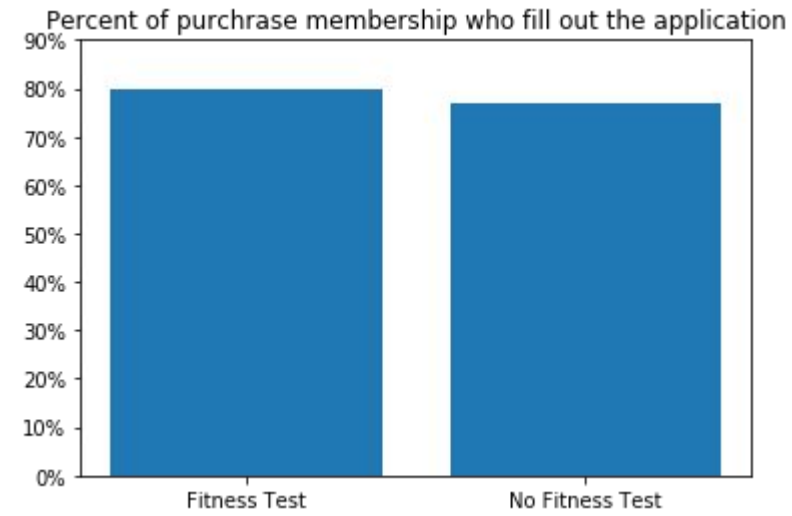
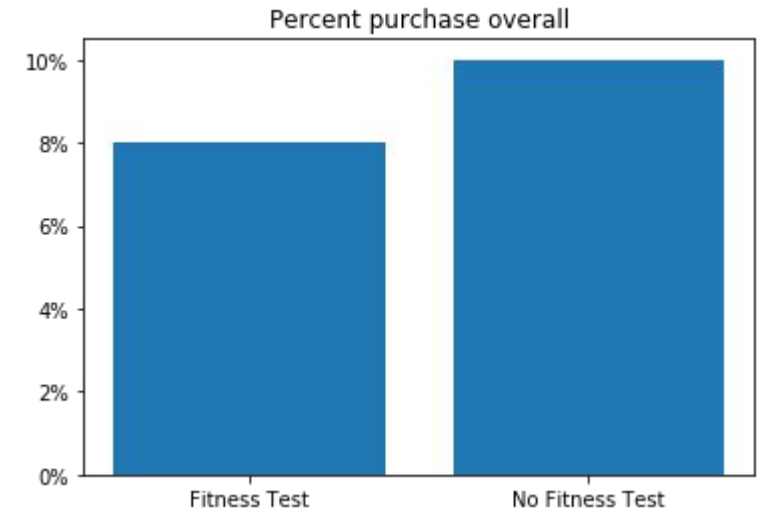
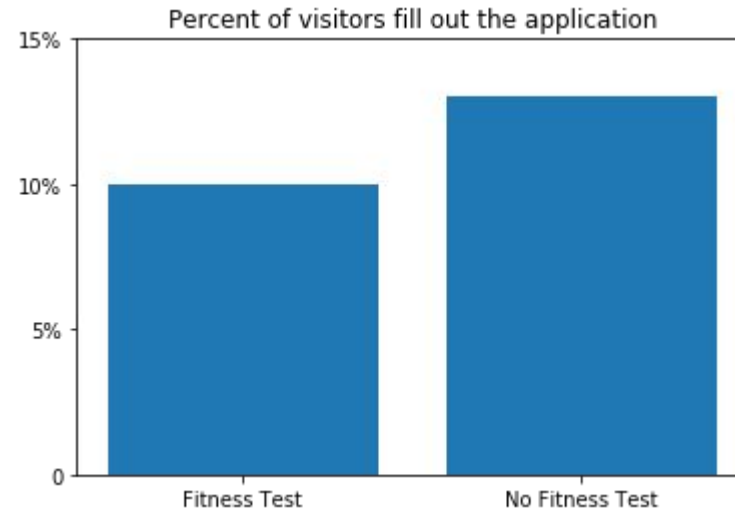
Python Scipy proves the statistical significance

We have two samples (Group A and Group B) , and they are categorical data. Hence we use Chi Square Test to prove there is no significant difference between the datasets

As we can see, the p-value is less than .05, meaning it is significant, we reject the null hypothesis, there is a significant difference between two groups. Statistically we are confident that Group B shows a higher percentage of customers purchased the membership

P-value= .0147241146458

Summary



Overall, we are confident that no fitness test does bring more people to sign up and become members.

Suggestions

- Provide safe, clean and nice gym equipment in the gym. Put up a Clorox Disinfecting Wipes tub and encourage gym-goers to wipe after use(Jesse complained about the sweat stained equipment) If we have enough members, we can convince Clorox to sponsor us for a place to exposure their products to people
- Make the Fitness Test a voluntary one and convey to the customer we have the Fitness Test to accurately decode your body before become a member, however it is all up to you to take the test or not(Cora appreciated the fitness test, Sonny did not)
- Keep the sign-up process fast and easy(Shirley love Musclehub for this)

Thank you

Q&A?