

Analyze Data with Python

Capstone Project

MuscleHub A/B Test

Paulina Kossowska
April 2021

Introduction

Task Details: For this project, you'll have to access SQL in a slightly different way. You'll be using a special Codecademy library that lets you type SQL queries directly into this Jupyter notebook. You'll have pass each SQL query as an argument to a function called `sql_query`. Each query will return a Pandas DataFrame.


Get Dataset

We have several tables that will be helpful:

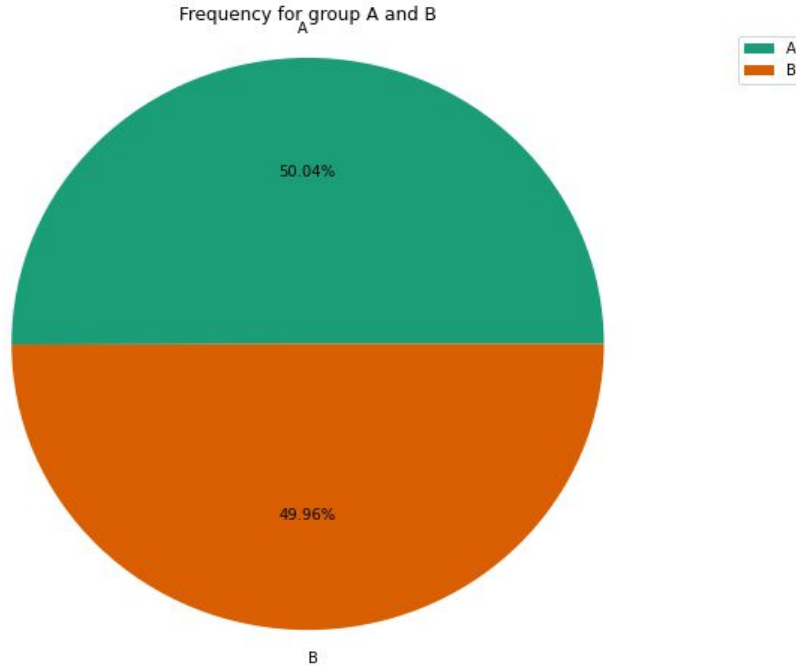
- **visits** contains information about potential gym customers who have visited MuscleHub
 - **fitness_tests** contains information about potential customers in "Group A", who were given a fitness test
 - **applications** contains information about any potential customers (both "Group A" and "Group B") who filled out an application. Not everyone in visits will have filled out an application
 - **purchases** contains information about customers who purchased a membership to MuscleHub
-

SQL query that connect all this data into one DataFrame

```
df = sql_query("""
SELECT visits.first_name,
       visits.last_name,
       visits.visit_date,
       fitness_tests.fitness_test_date,
       applications.application_date,
       purchases.purchase_date
FROM visits
LEFT JOIN fitness_tests
  ON fitness_tests.first_name = visits.first_name
  AND fitness_tests.last_name = visits.last_name
  AND fitness_tests.email = visits.email
LEFT JOIN applications
  ON applications.first_name = visits.first_name
  AND applications.last_name = visits.last_name
  AND applications.email = visits.email
LEFT JOIN purchases
  ON purchases.first_name = visits.first_name
  AND purchases.last_name = visits.last_name
  AND purchases.email = visits.email
WHERE visits.visit_date >= '7-1-17'
""")
```



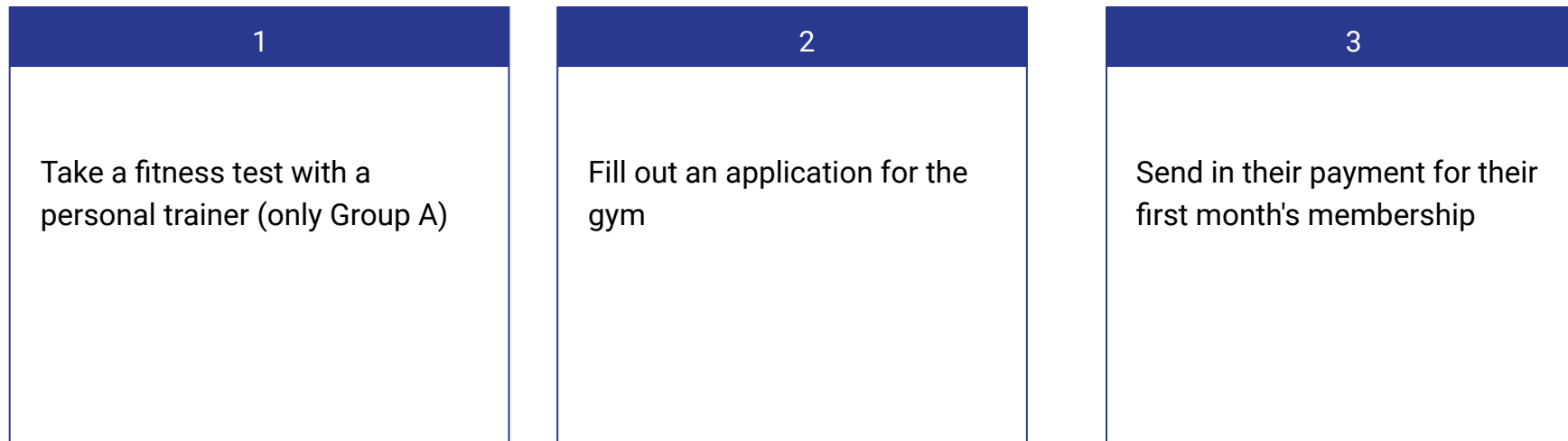
Investigate the A and B groups



As we can see **50.04%** of visitors belong to A group and **49,96%** of visitors belong to B groups.

Who picks up an application?

Sign up process for MuscleHub has several steps



Let's examine how many people make it to Step 2, filling out an application.

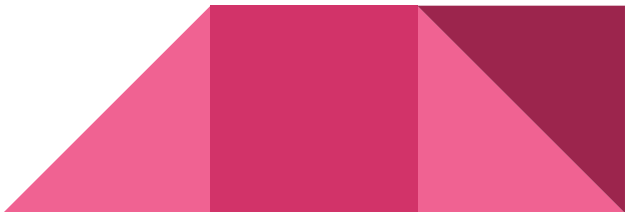
After some calculations I got a pivot table with following results

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

It looks like more people from Group B turned in an application.

We need to know if this difference is statistically significant.

So, the next step will be performing hypothesis testing - **Chi Square Test**



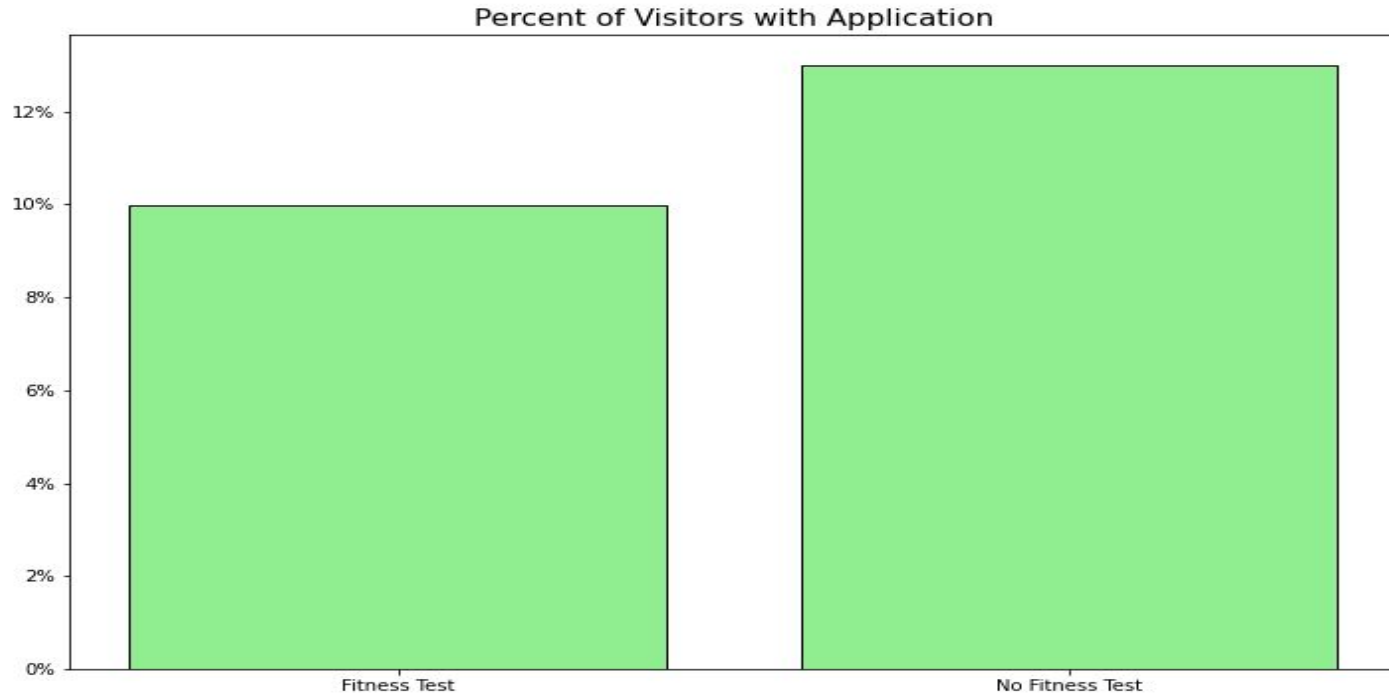
Chi2 Square Test

pvalue
0.001018198175755847

threshold level
0.05

result
statistically significant
there was significant difference in applications between Group A and Group B

Graph showing Percent of Visitors with Application



Who purchases a membership?

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

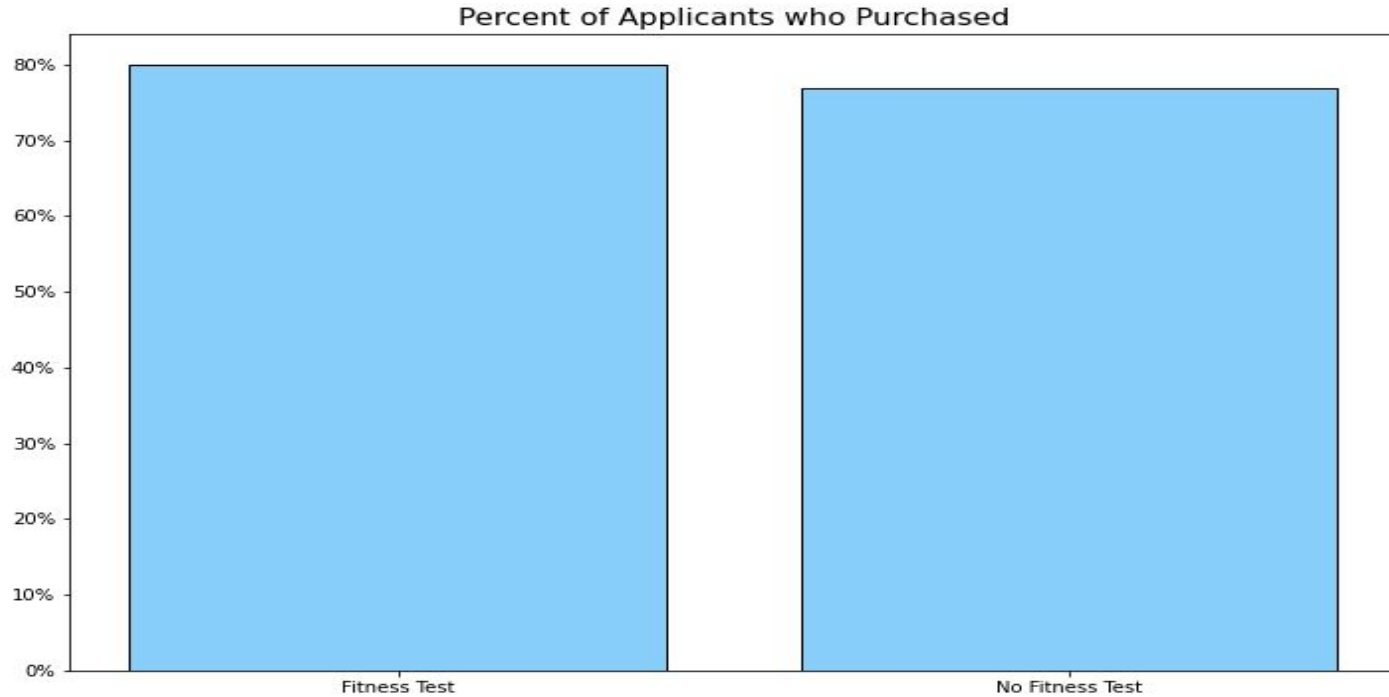
This pivot table shows us a total number of member and not member in groups A and B but and it contains only people who picked up an application.

After performing Chi Square Test for that results I got:

Chi2 Square Test

pvalue	threshold level	result
0.43258646051083327	0.05	statistically not significant there was no significant difference in membership between Group A and Group B

Graph showing Percent of Applicants who Purchased



Who purchases a membership?

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

Previously, we looked at what percent of people **who picked up applications** purchased memberships.

What we really care about is what percentage of **all visitors** purchased memberships

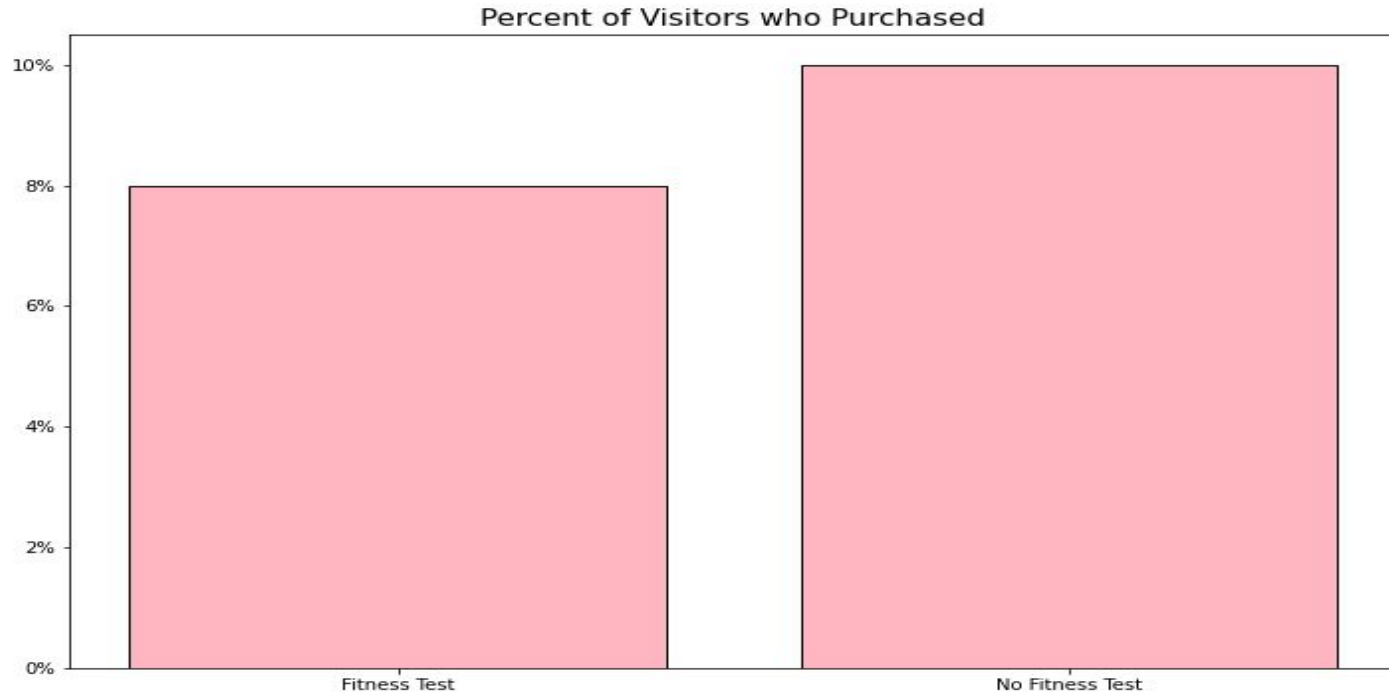
This pivot table shows us a total number of visitors with member and not member in groups A and B.

After performing Chi Square Test for that results I got:

Chi2 Square Test

pvalue	threshold level	result
0.001018198175755847	0.05	statistically significant there was significant difference in membership between Group A and Group B

Graph showing Percent of all Visitors who Purchased





Thanks for attention