

Capstone Project: Muscle Hub A/B Test Analysis

Prepared by:
Bob Okony

Codecademy Capstone Project – Analyze Data with Python

Overview of Project

- The Muscle Hub A/B Test project combines data from four SQL data sources into one Pandas dataframe. Initial grouping of users from the dataframe is performed based on a valid fitness test date. Additional analysis is performed to determine how many users pick up an application and how many users purchase a membership. A Pandas chi2_contingency test is performed on the initial groups of user to determine if there is statistical significance between the results for the two test groups.

Preliminary Data Analysis

```
In [10]: print(df.head())  
df.describe()
```

```
first_name last_name visit_date fitness_test_date application_date \  
0      Kim    Walter    7-1-17      2017-07-03              None  
1      Tom   Webster    7-1-17      2017-07-02              None  
2   Edward   Bowen    7-1-17           None      2017-07-04  
3   Marcus   Bauer    7-1-17      2017-07-01      2017-07-03  
4  Roberta    Best    7-1-17      2017-07-02              None  
  
purchase_date  
0      None  
1      None  
2  2017-07-04  
3  2017-07-05  
4      None
```

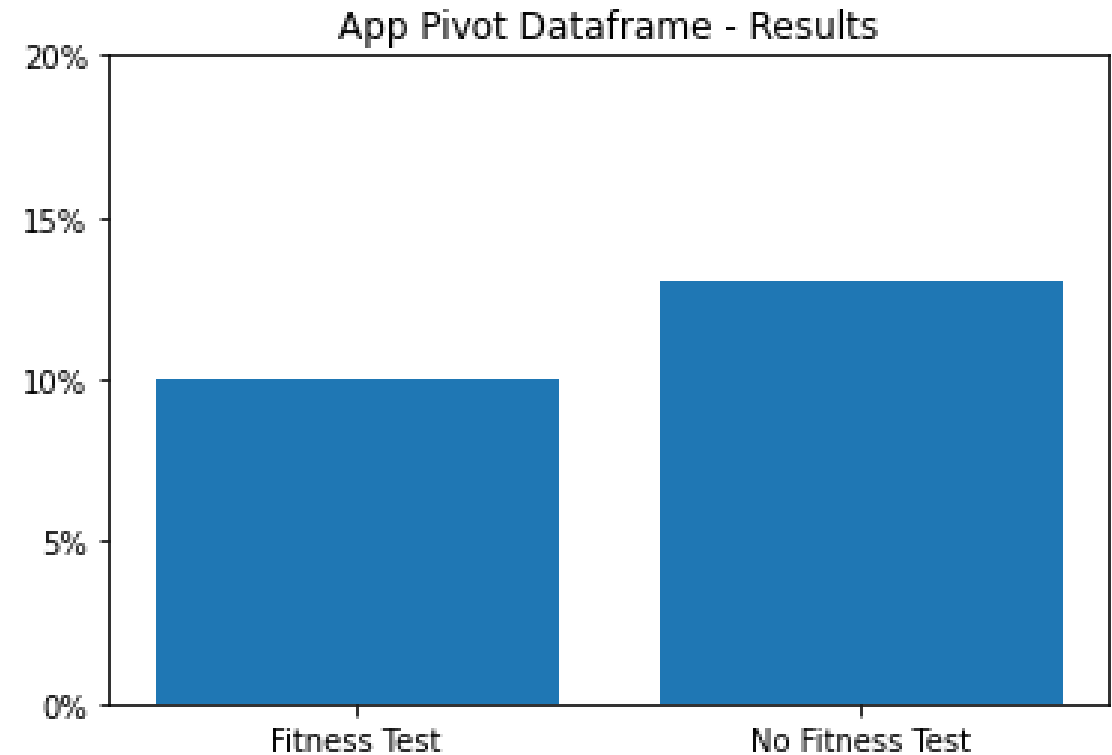
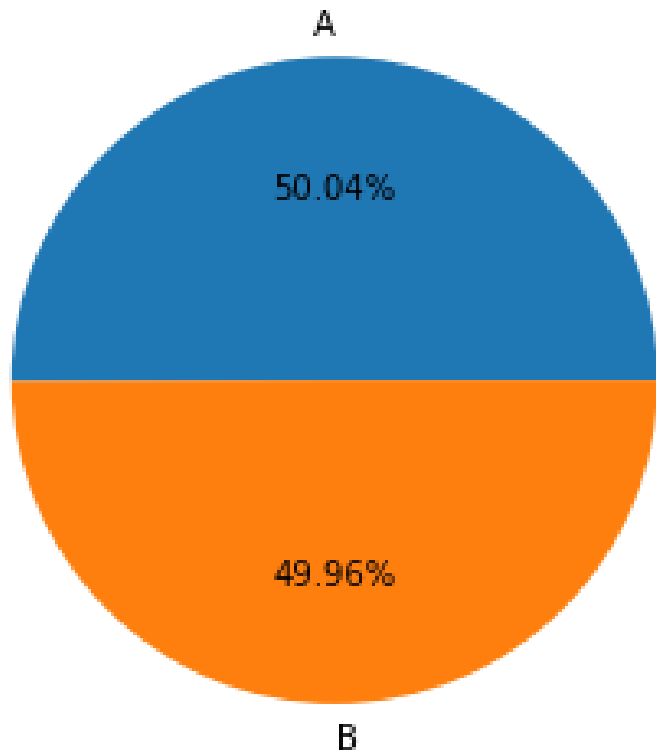
Out[10]:

	first_name	last_name	visit_date	fitness_test_date	application_date	purchase_date
count	5004	5004	5004	2504	575	450
unique	592	684	87	97	95	99
top	Leslie	Owen	9-26-17	2017-08-18	2017-07-21	2017-08-18
freq	29	16	76	40	12	10

Initial dataframe from SQL query input. Operation performed on the dataframe to group individuals by valid test date. Created two groups – A Group – valid test date, B Group – no valid test date.

Graphical Analysis –

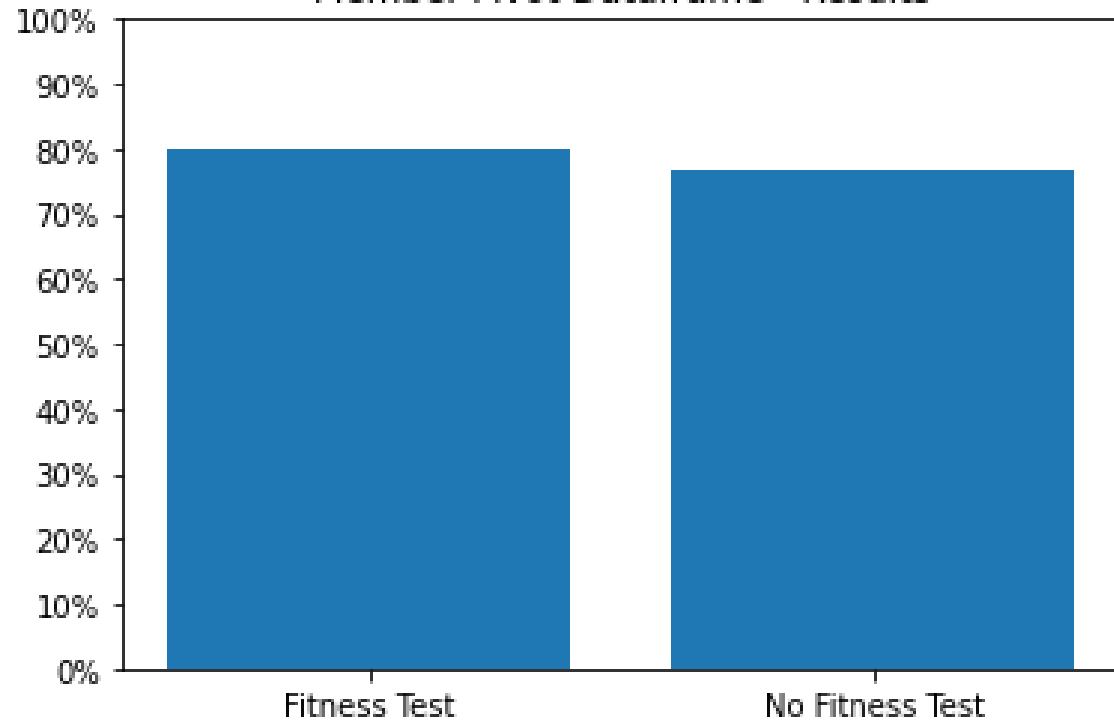
- Grouping of users by valid test date (left graph)
- Grouping of users by valid fitness test (right graph)



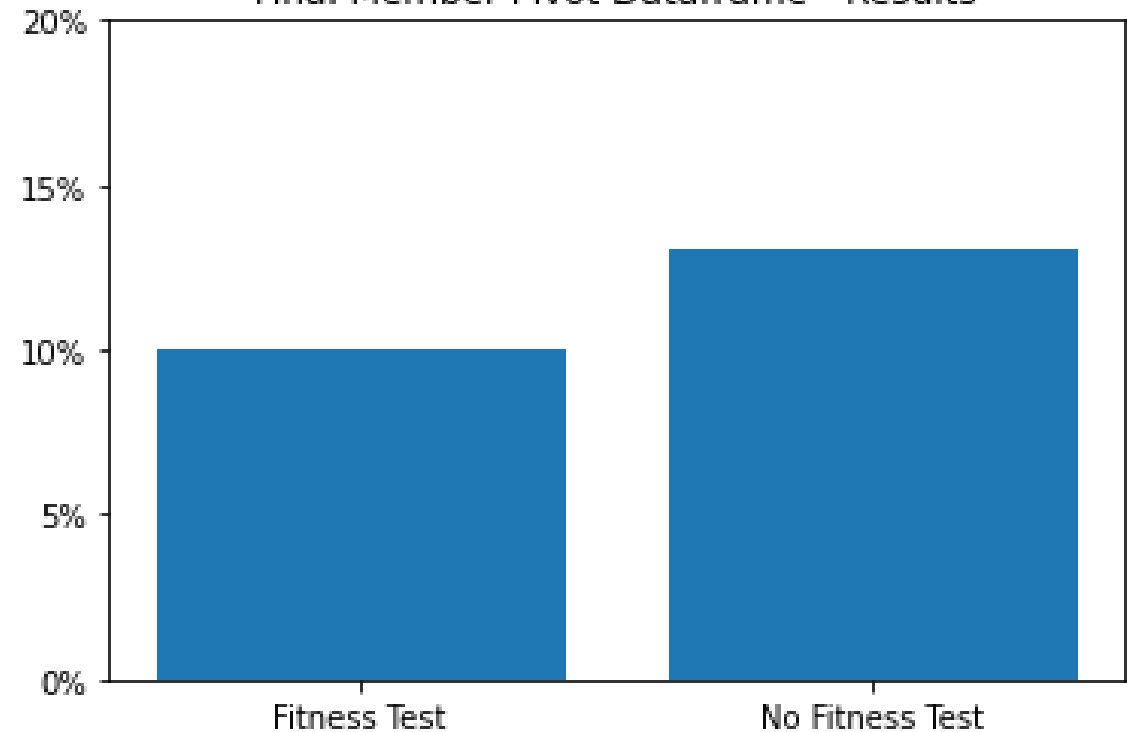
Graphical Analysis –

- Percent of applicants who purchase a membership (left graph)
- Percent of visitors who purchase a membership (right graph)

Member Pivot Dataframe - Results



Final Member Pivot Dataframe - Results



Recommendations/Findings

- No statistical significance indicated by chi2_contingency analysis of the two groups – A (fitness test) and B (no fitness test)
- P-value of .014724
- Continue to market and promote to both current members and non-members.
- Visitors are less likely to complete the fitness test than current members
- It appears that people who took the fitness test were more likely to purchase a membership **if** they picked up an application