**Muscle Hub A/B Testing**

**Introduction**

**Background**

Muscle Hub provides gym services to customers by charging their membership fees. In order to boost revenue, or attract more customers to join the gym, the owner, Janet, offers a fitness test for any potential customers. The test is to tell the potential customers what level they are at and to let the gym plan out better fitness plans for the customers accordingly.

**Problem**

The data might contribute to determining whether the fitness test is effective might include the date customers take the test, application date/no application, purchase date/no purchase. This project aims to determine whether the fitness test is effective in attracting more customers.

**Interest**

The owner, Janet, will be the most interested in this report. The gym employees, such as instructors, should be interested in it as well.

**Data**

**Data source**

All the data is from Codecademy ([click here](https://www.codecademy.com/paths/analyze-data-with-python/tracks/ida-8-final-projects/modules/ida-8-2-capstones/informationals/ida-capstone-musclehub)).

**Data cleaning**

Data are separated in 4 different csv files and combined into one data frame. I only focus on the data after July 1st, 2017, since when the fitness test was provided.

**Feature selection**

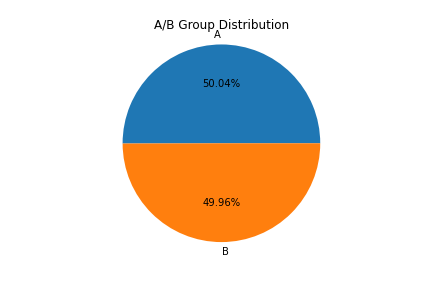
The features are visit date, fitness test date, application date, and purchase date. Customers’ gender and whether customers take the fitness tests will be also considered in the analysis. Feature like customer email will be dropped since it has no use in this case.

By joining the 4 different files by matching the customer name, we will get the information that customers who take the tests, customers who file applications, customers who make the purchases, etc.

**Exploratory Data Analysis**

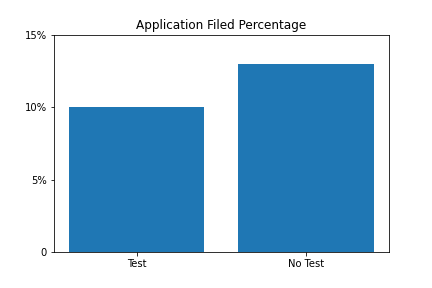
**A/B Group Distribution**

There are total of 5,004 customers, 2,504 of them in group A (with fitness test) and 2,500 in group B (without fitness test). Considered the number is distributed evenly into two groups, the analysis will provide more accuracy.

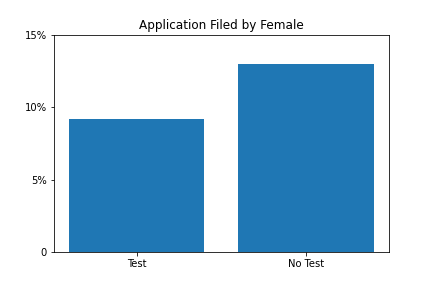
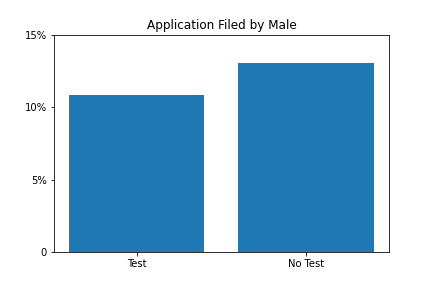


**Percentage of Filing Application**

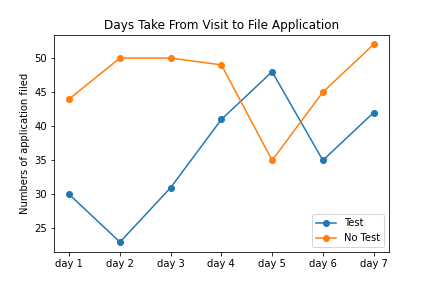
I use the number of filling application to divide the total number of the customers for each group to calculate the percentage of filing application. 13% of the customers who don’t take the fitness test file applications, while only 10% of the customers who take the fitness tests file applications. Supported by z-test, where p < 0.01, the result of the two groups filing applications are significant different. We can say customers who don’t take the fitness tests are more inclined to file applications.



By looking into the gender, the percentage of female customers who take the fitness tests and file the applications is only 9%, while for those who don’t take the fitness test, 13% file the applications (z-test, p < 0.01). For male customers, the difference is not as big as the difference of female customers. However, z-test with p < 0.11 could say something that male customers don’t feel comfortable filing the applications either.

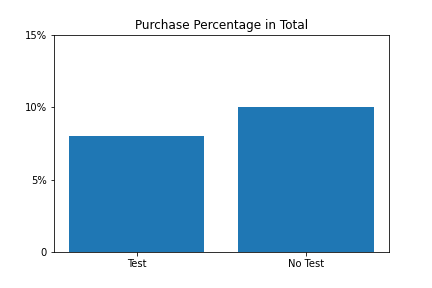
I also look into the number of applications filed after days since the customers have visited the club. Applications filed within four days from the customers who don’t take the fitness tests are way more than the applications filed from the customers who take the tests. Even though more and more test-taken customers file the applications, we can still surely conclude that the fitness test turns some customers away, especially female customers.



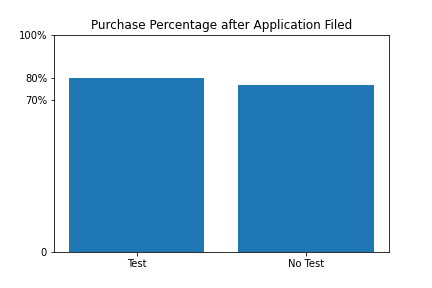
**Purchase Rate**

I also calculate the purchase rates in two different ways. First is that total number of purchase is divided by total number of visits. Second is that total number of purchase is divided by total number of applications. The two different ways of calculation give us two different result.

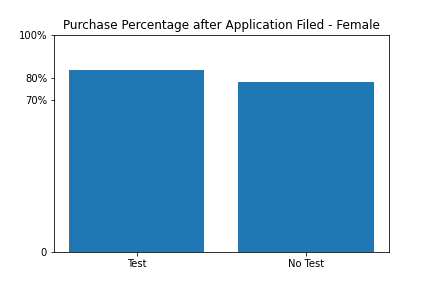
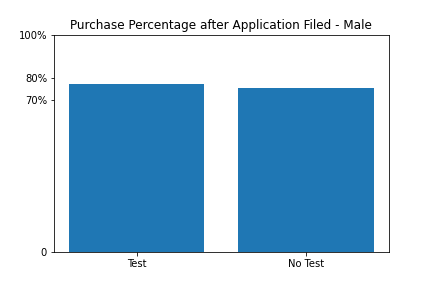
Not a surprise, for the first calculation method, higher purchase rate shows in the group without fitness test, the rate that is 10% compared with 8% purchase rate for the group with the test. The z-test with p < 0.02 strongly supports this calulation.



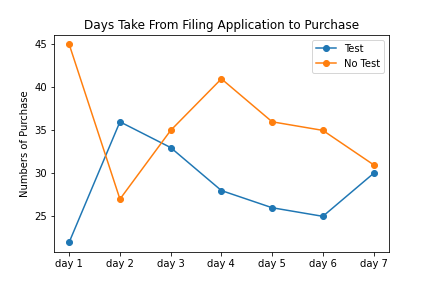
On the other hand, the conversion rate is much higher. 80% of the customers who take the tests successfully make the purchase and 77% of the customer who don’t take the tests make the purchase. This result clearly shows that for the both group the customers are commitment enrolling the membership if they file applications. Moreover, we are 57% confident that customers who take the fitness test has more commitment making the purchase.



I also look into gender. 83.5% of test-taken female customers enroll the membership and 78% for those no test taken(z-test, p < 0.347). For male customers, the purchase rate of two groups is pretty much the same, 77% and 75.5% (z-test, p = ~0.864). And we are 73% confident that for test-taken customers female customers show more commitment in making purchase than male customers do.

I also analyze the number of purchase for each day after customers file their applications. The biggest gap is showed on day one. The number of purchase by no-test group doubles the number of purchase by test-taken group. Even though customers who take the tests seem to show more commitment in enrolling the the club, more customers who don’t take the tests enroll the club eventually.



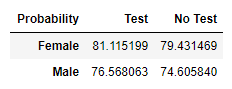
**Predictive Modeling**

**Model selection**

I use logistic regression model for this analysis, reasons as follow. First, the predicting result is binary, either purchase or no purchase (1 or 0 for the labels). Second, we more focus on the probability of the enrollment.

**Score, coefficient and probability**

The model score is 0.97. For the features, the model coefficients of gender, A/B testing, and filing application or not are 0.27, 0.10 and 7.4 respectively. The prediction probability for customers who have filed the applications shows as follow table.



**Issues**

Features are very limited. There are only three features available for use in this model. And the feature of filing application is decisive, which could make the other two features negligible. If we use filing application as our label, there will be only two features we can use to predict. Besides, we still can’t say filling application means making the purchase.

**Recommendations**

**Target the right customers**

Not every customer is suitable for the fitness test. Therefore, I recommend MuscleHub offer surveys and only offer fitness test to selective customers. For example, some customers only interested in light exercises or certain equipments, such as treadmill, exersise bike, etc. For those customers the fitness test can be too hardcore and scary.

I also recommend MuscleHub offer different tests to different group of customers accordingly. For example, some customers aim at gaining muscles and some customers aim at cutting weights. The purposes of going to gym are different from customers to customers. Therefore, MuscleHub should offer different fitness tests.

**Increase the conversion rate**

As we can see, MuscleHub lost about 20% of the customers who have already filed their application but choose not to enroll. Therefore, I recommend that MuscleHub offer promotion plans for those who have filed applications to motivate them to enroll as soon as possible. For example, MuscleHub can offer some discount on subcription fee for those who have filed applications if they can enroll within certain days.

I also recommend MuscleHub offer fitness plans to the test-taken customers. The action not only shows a better customer service, customers will be more likely to enroll the club as the right fitness plans are already showing to them.

**Conclusion**

In conclusion, the fitness test offered by MuscleHub is counter-productive. It turns away customers as a result. Therefore, the fitness test should be removed or should be adjusted. Secondly, MuscleHub shouldn’t lose any customers who have already filed the application. Thridly, MuscleHub should provide more customer details as features used in the predictive model, details that can be obtained by surveys.

Once MuscleHub increase their conversion rate as closely as 100%, we can use filing application as our labels and use more features, such as gender, age, smoking, diet types, test levels, purposes, workout time per week, and so on, to make a better prediction. We will focus more on how to let customers file their applications at that time.