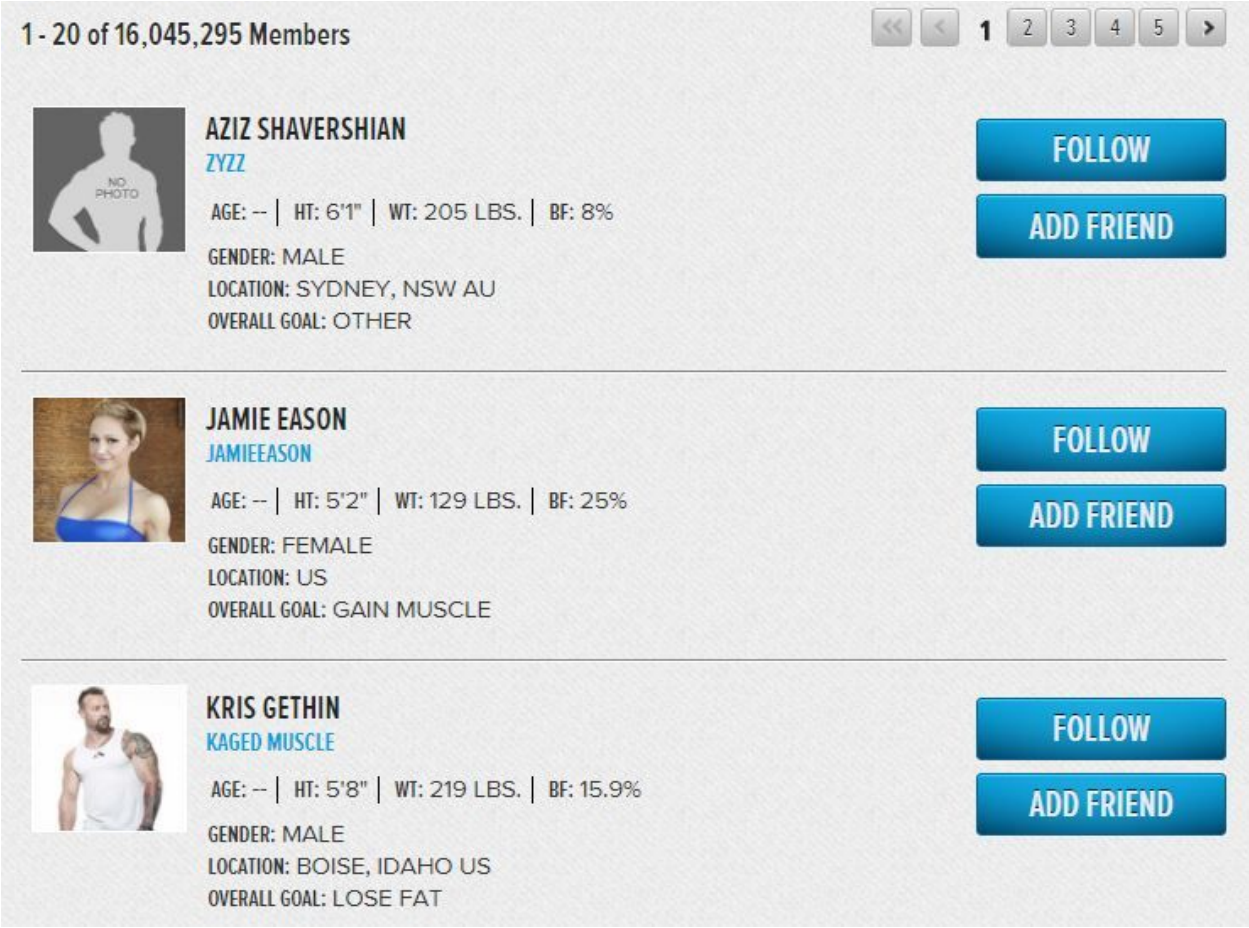
Intermediate Report

By Elbert Kim

Objective: Train a model using a supervised machine learning technique to predict a customer's fitness goal based on four attributes (age, weight, height, and gender). I thought this was an interesting solution to beginners wanting to start a fitness journey.

Data: Originates from Bodybuilding.com (Has over 1 million publicly shared user data)



(Example of customer information)

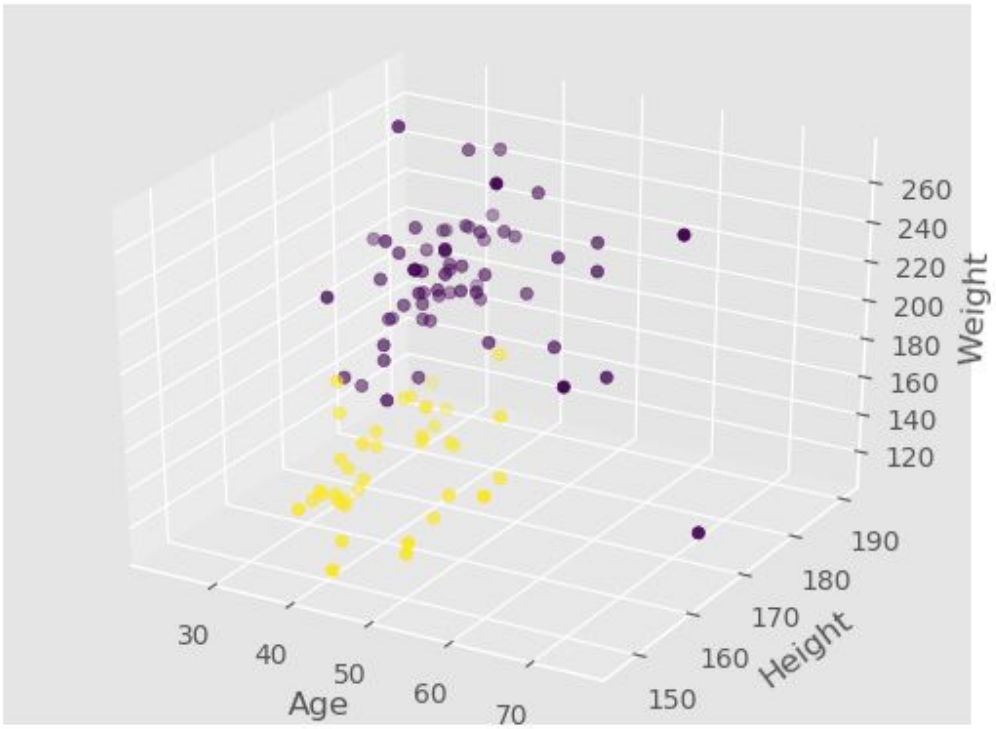
Challenges:

- Web crawling
    - Due to the fact I was facing many technical difficulties with extensive web crawling, I decided to forego the original objective and focus on recommending fitness goals. I believed this pivot would make the project more manageable and executable.
- Web scraping
    - Bodybuilding.com shows and stores customer information via pagination, but each page when clicked stays on the same URL. I was able to scrape 20 to 100 customer information, but have not yet figured out how to go beyond that because the URL stays the same. Consequently, I had to manually perform data entry to gather data.
- Preprocessing data
    - Around ⅓ of customers had an empty value for an attribute. To fill in empty values, I had to guesstimate. Eventually I would like to fill empty values with an average of values (from people with similar attributes).

Analysis:

```
PS C:\Users\Elbert Kim\Desktop\svm> python index.py
Training set size: 100
Testing set size: 20
Analysis: 13/20 correct.
```

Training data set: 100
Test data set: 20
Correct: 13/20



(Purple: Male, Yellow: Female)

Solutions to increasing success ratio:

- Collect more data to increase the training set (Ex. 100 -> 5000)
- Add an attribute or implement feature engineering (Ex. Body fat %)

Next steps:

- Figure out how to scrape more data
- Analyze data and identify any correlations