# Capstone Project

Machine Learning Fundamentals
Ted Stetzel
February 2019

Table of Contents

- Exploration of the Dataset
- Question(s) to Answer
- Augmenting the Dataset
- Classification Approaches
- Regression Approaches
- Conclusions/Next steps

Exploration of the Dataset

I was interested in finding out more about the association between how individuals describe their diet in relation to income and body type.
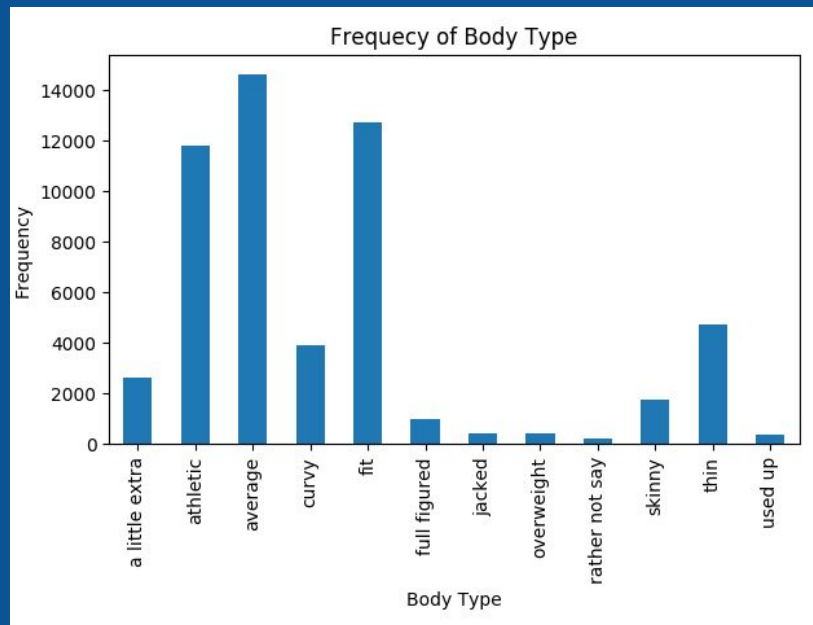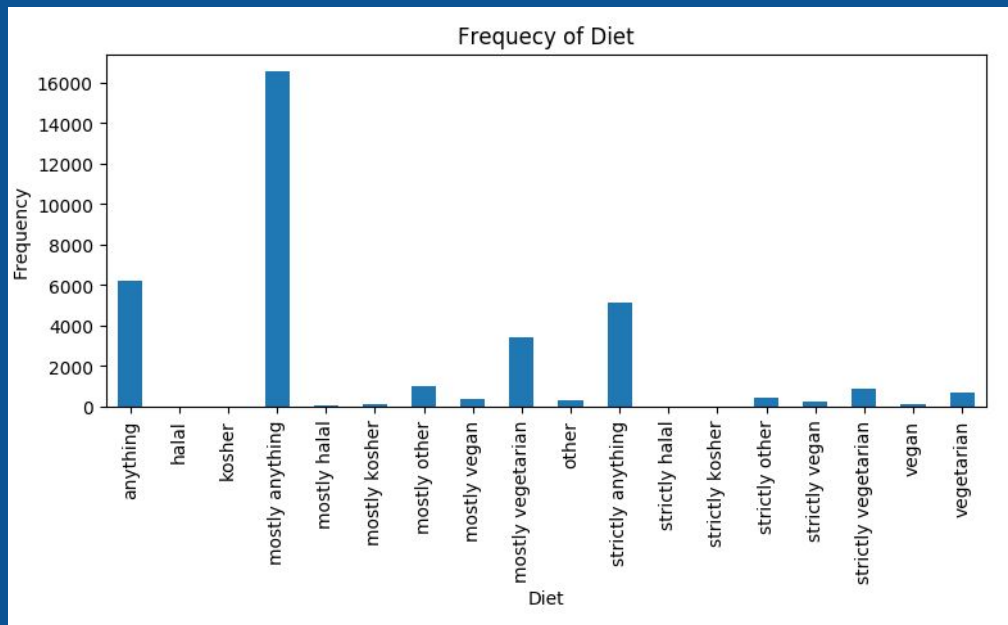
Question(s) to Answer

- Does the data in this data set follow some common norms?

- Question 1: Does income increase with age?

- Question 2: Can factors such as age and diet influence a person's perceived body type?

Augmenting the Dataset

- Diet gives the option of "mostly" and "strictly" for each diet.

- I combined the "mostly" and "strictly" options with the base option to simplify the model

- Each group was assigned a number

```
| diet               | diet_code   |
|--------------------|-------------|
| anything           | 0           |
| mostly anything    | 0           |
| strictly anything  | 0           |
| halal              | 1           |
| mostly halal       | 1           |
| strictly halal     | 1           |
| kosher             | 2           |
| mostly kosher      | 2           |
| strictly kosher    | 2           |
| vegan              | 3           |
| mostly vegan       | 3           |
| strictly vegan     | 3           |
| vegetarian         | 4           |
| mostly vegetarian  | 4           |
| strictly vegetarian | 4          |
| other              | 5           |
| mostly other       | 5           |
| strictly other     | 5           |
```

Augmenting the Dataset

- There are also a wide variety of body type choices

- I assigned each to one of three "perception" groups

- Each group was assigned a number

```
| body_type      | bt_code      |
|----------------|--------------|
| thin           | 0            |
| skinny         | 0            |
| athletic       | 1            |
| fit            | 1            |
| jacked         | 1            |
| average        | 2            |
| curvy          | 3            |
| a little extra | 3            |
| full figured   | 3            |
| overweight     | 3            |
| used up        | 4            |
| rather not say | 4            |
```

Regression Approaches

- **Question 1:** Does income go up with age?

- When all ages are included, the relationship is actually negative, which I found to be a surprise.

- When I excluded those older than 50, the relationship was positive

- **Conclusion**: Linear Regression doesn't account for income decreases at retirement

age vs income

Age vs Income - 50 or Younger
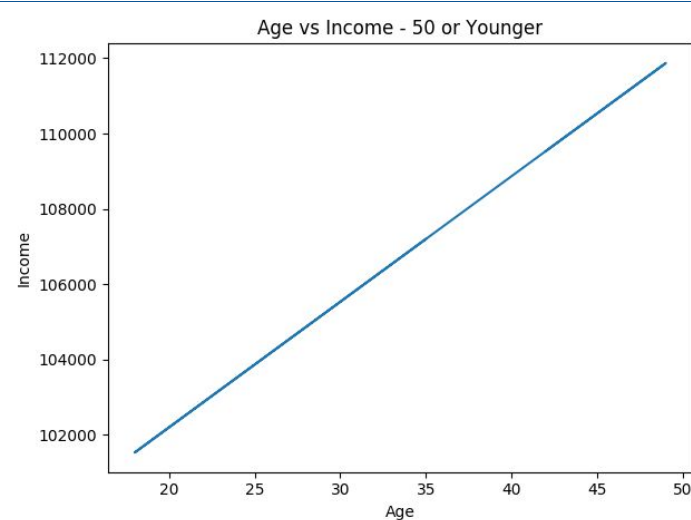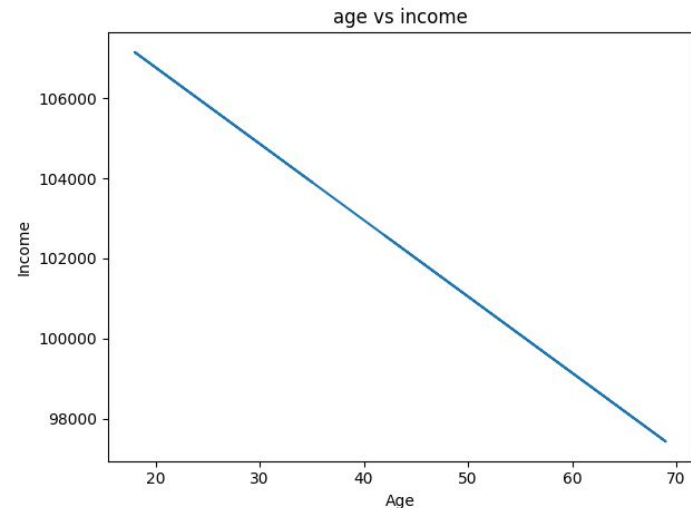
Classification Approaches

- I wanted to determine if income could determine diet?
- I supplied income as datapoints and my diet group code labels and performed an 80/20 split
- I then looped through k 1:100 to see that the best possible k was set right away
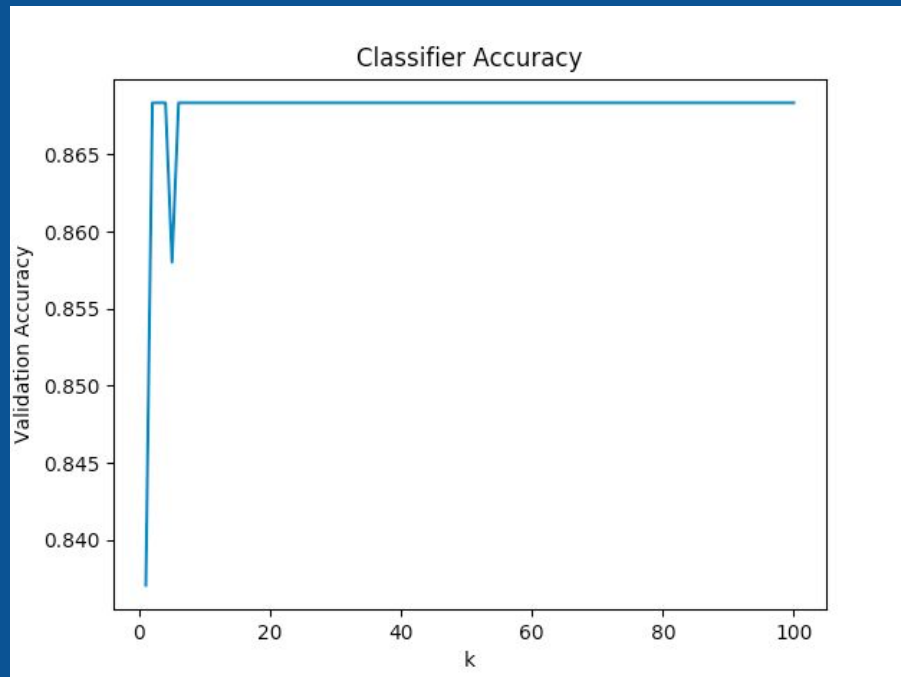
Conclusions/Next steps

- Do people who fall outside of "anything" perceive their body type differently than other diet types?
- Is there a difference between men and women?
- Could other variables be used to predict body type?