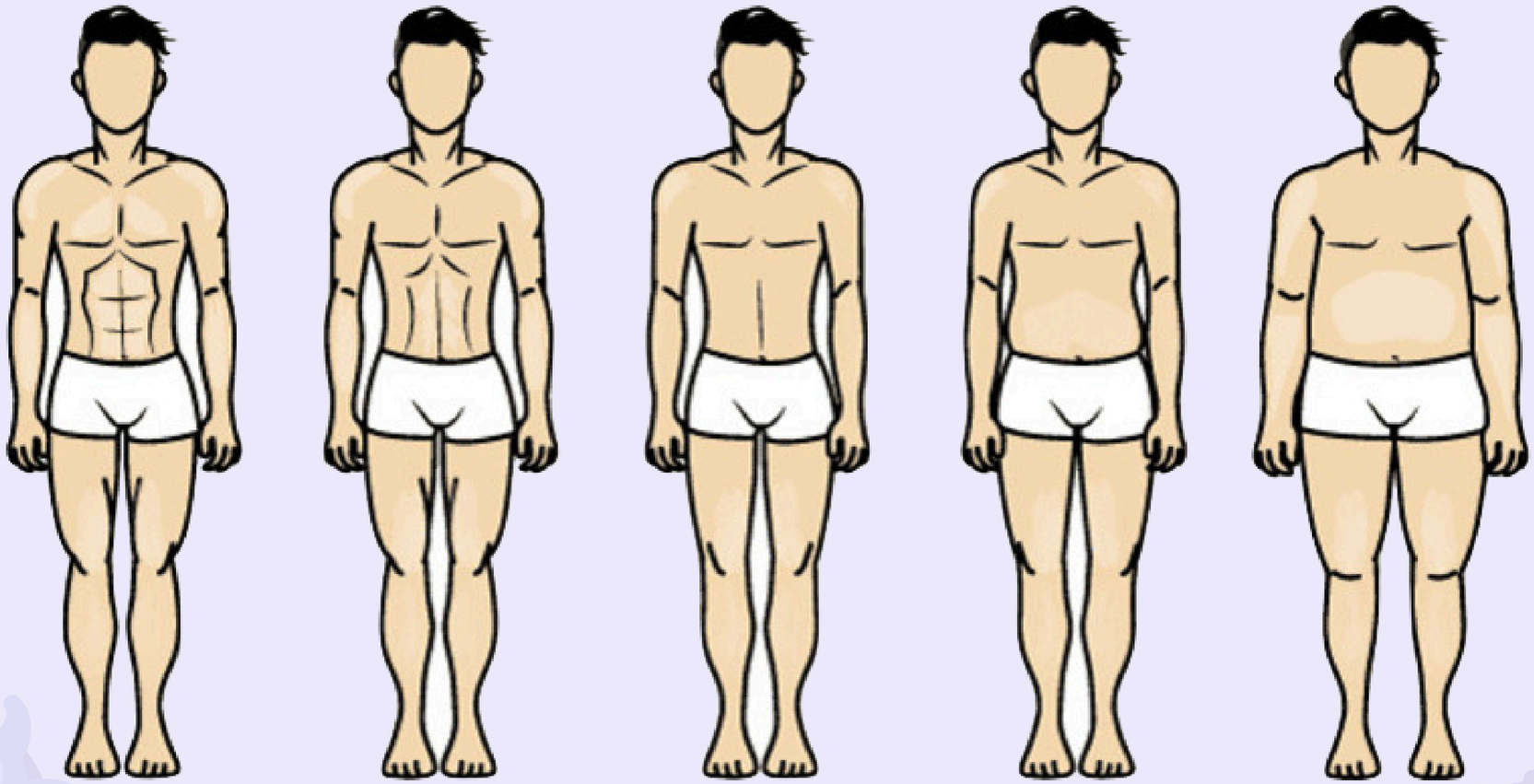
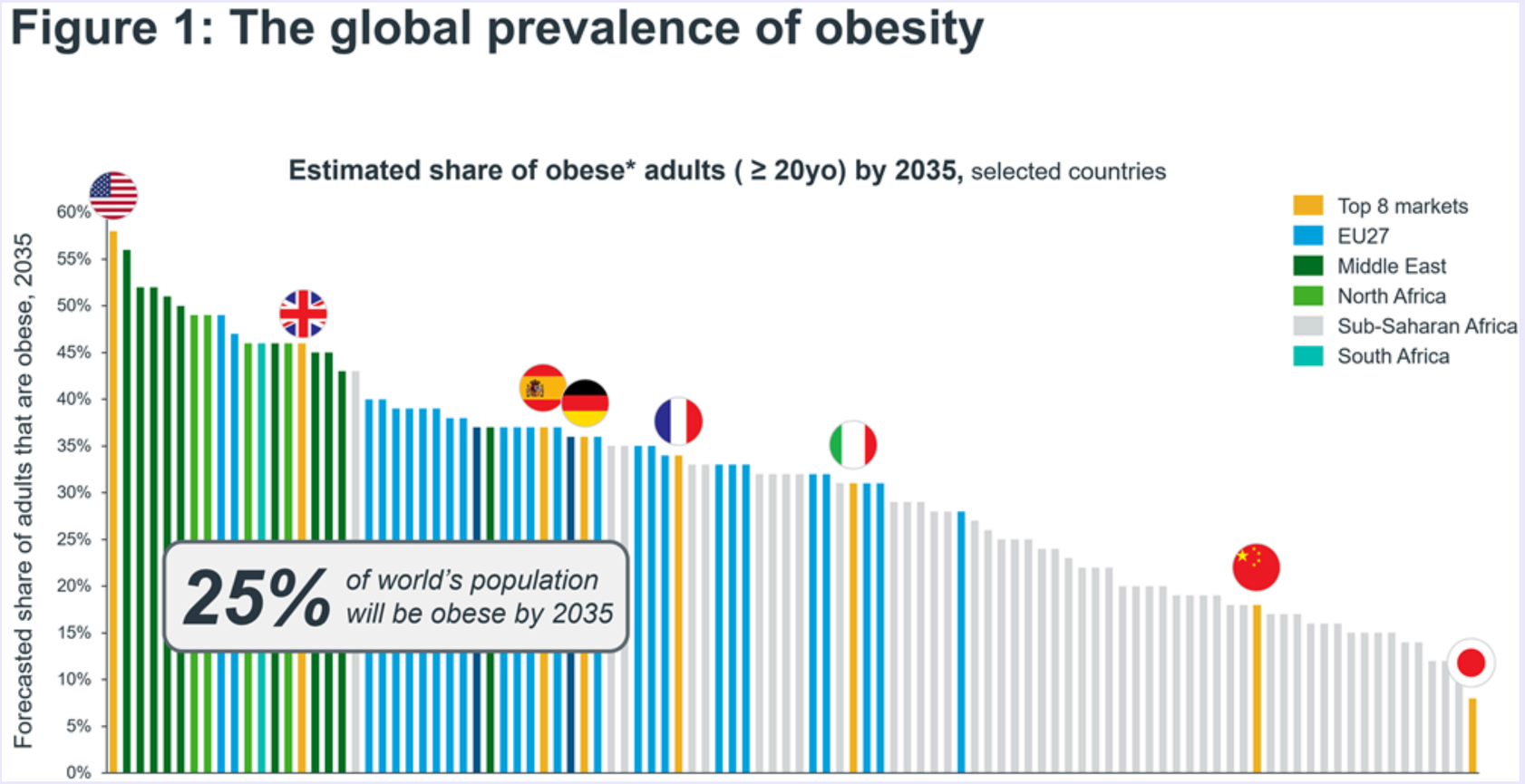
BODY FAT PREDICTION ANALYSIS



INTRODUCTION

Figure 1: The global prevalence of obesity



DATASET OVERVIEW

Density
Density determined from underwater weighing

02
Age (years)

03
Weight (lbs)

Height (inches)

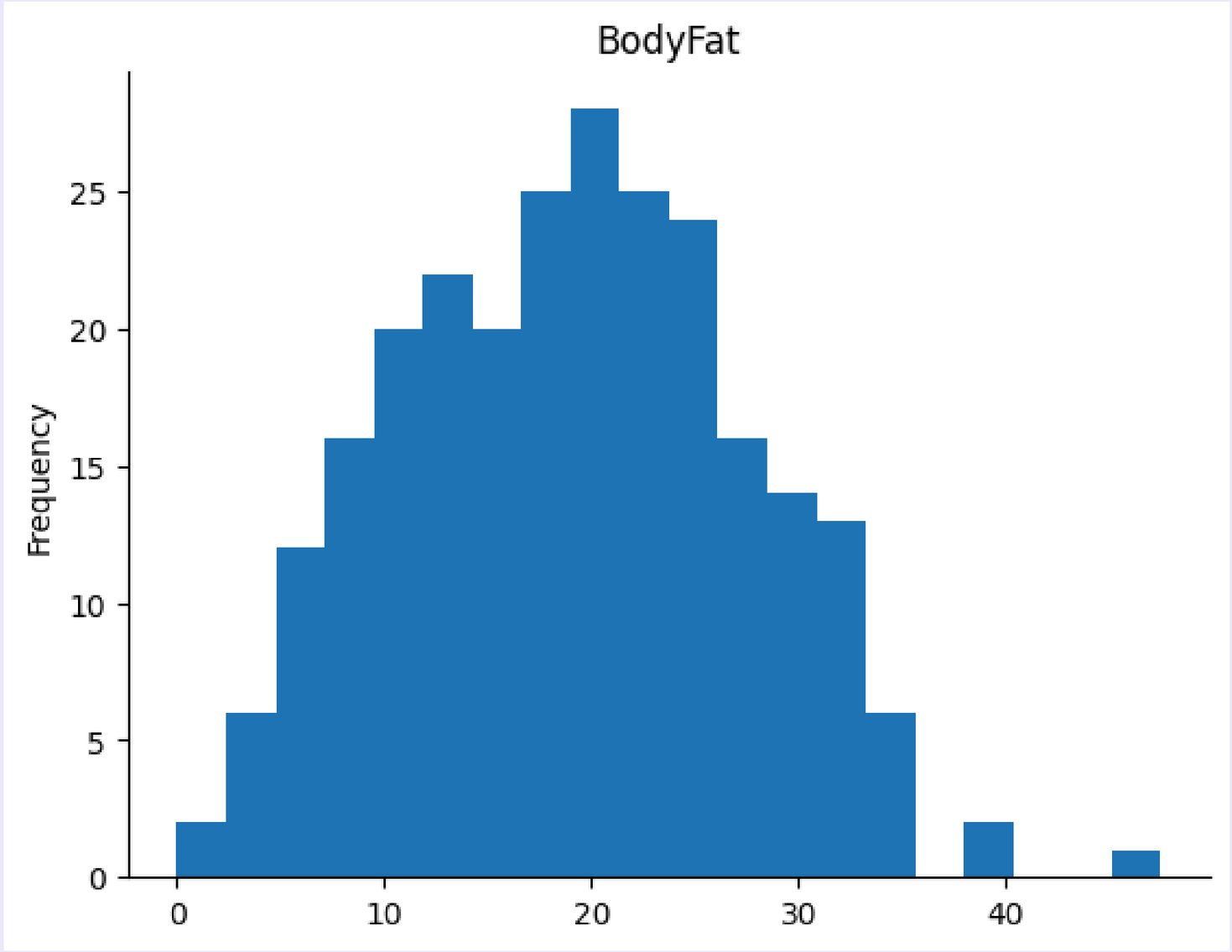
05
circumference (cm)
Neck, Chest, Biceps etc....

06
BodyFat (%)
Our Target,
Percent body fat from Siri's (1956) equation

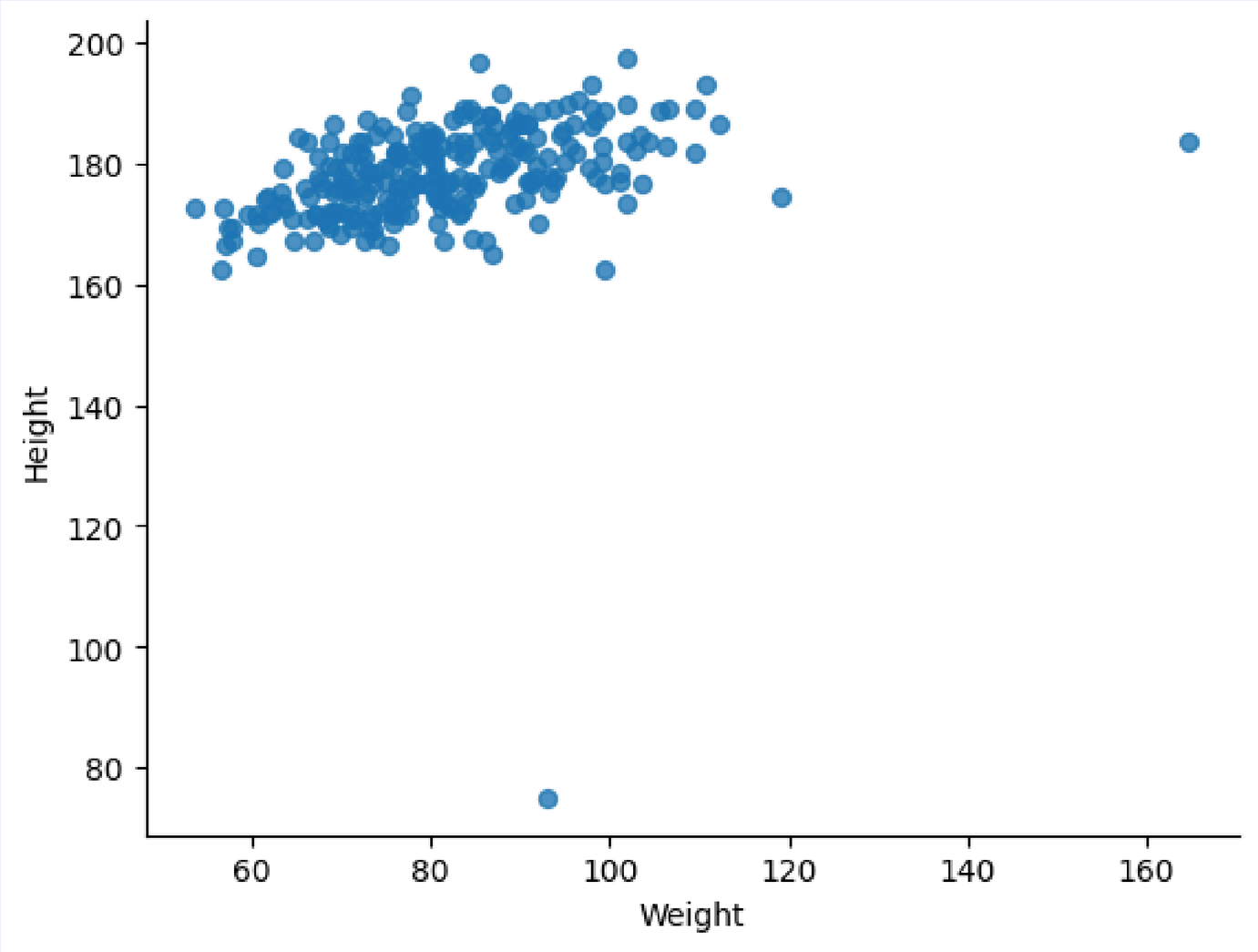
VISUALIZATIONS

- Histogram of Body Fat
- Scatter plot of Weight vs. Height
- Heatmap of feature correlations

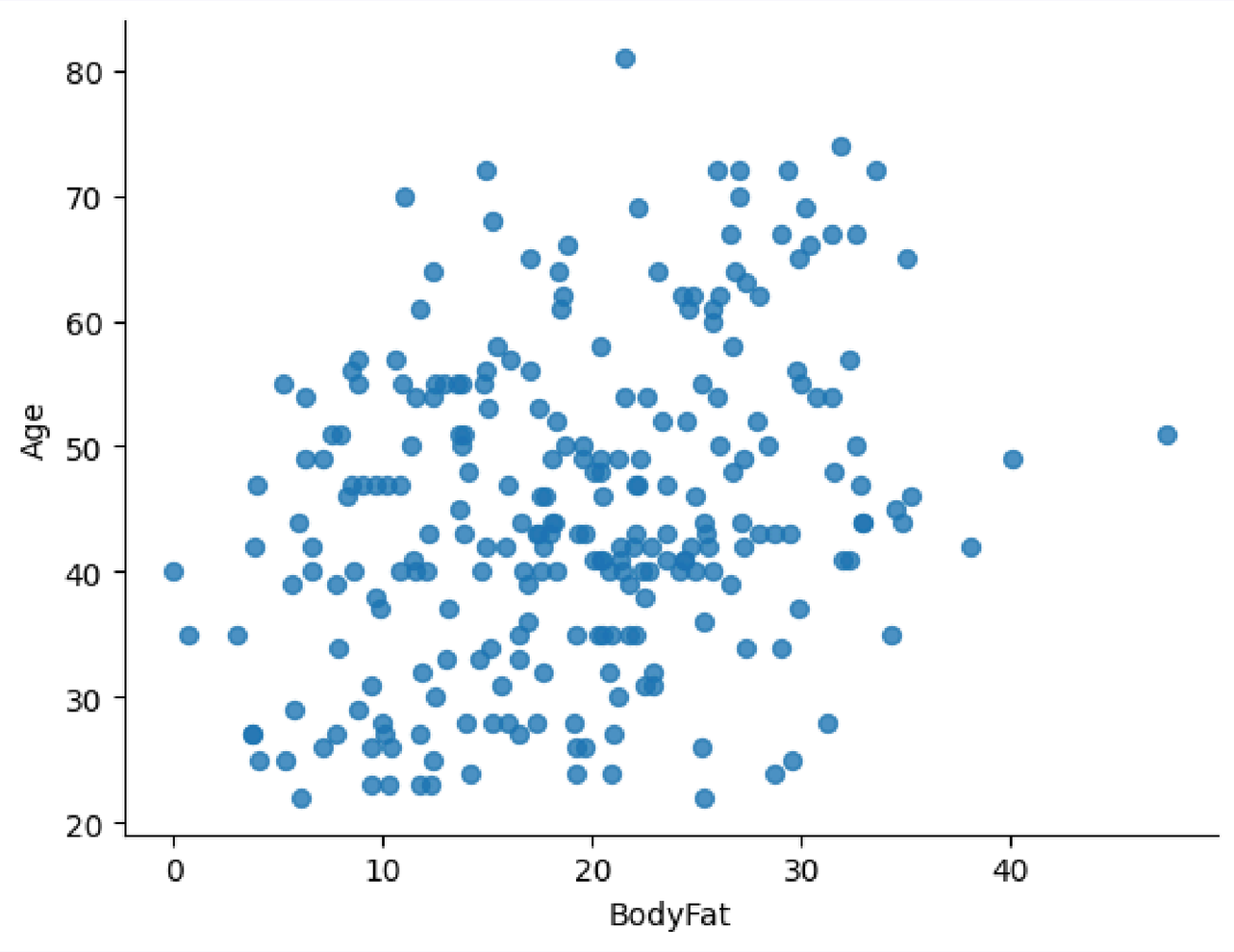
VISUALIZATION - BODYFAT DISTRIBUTION



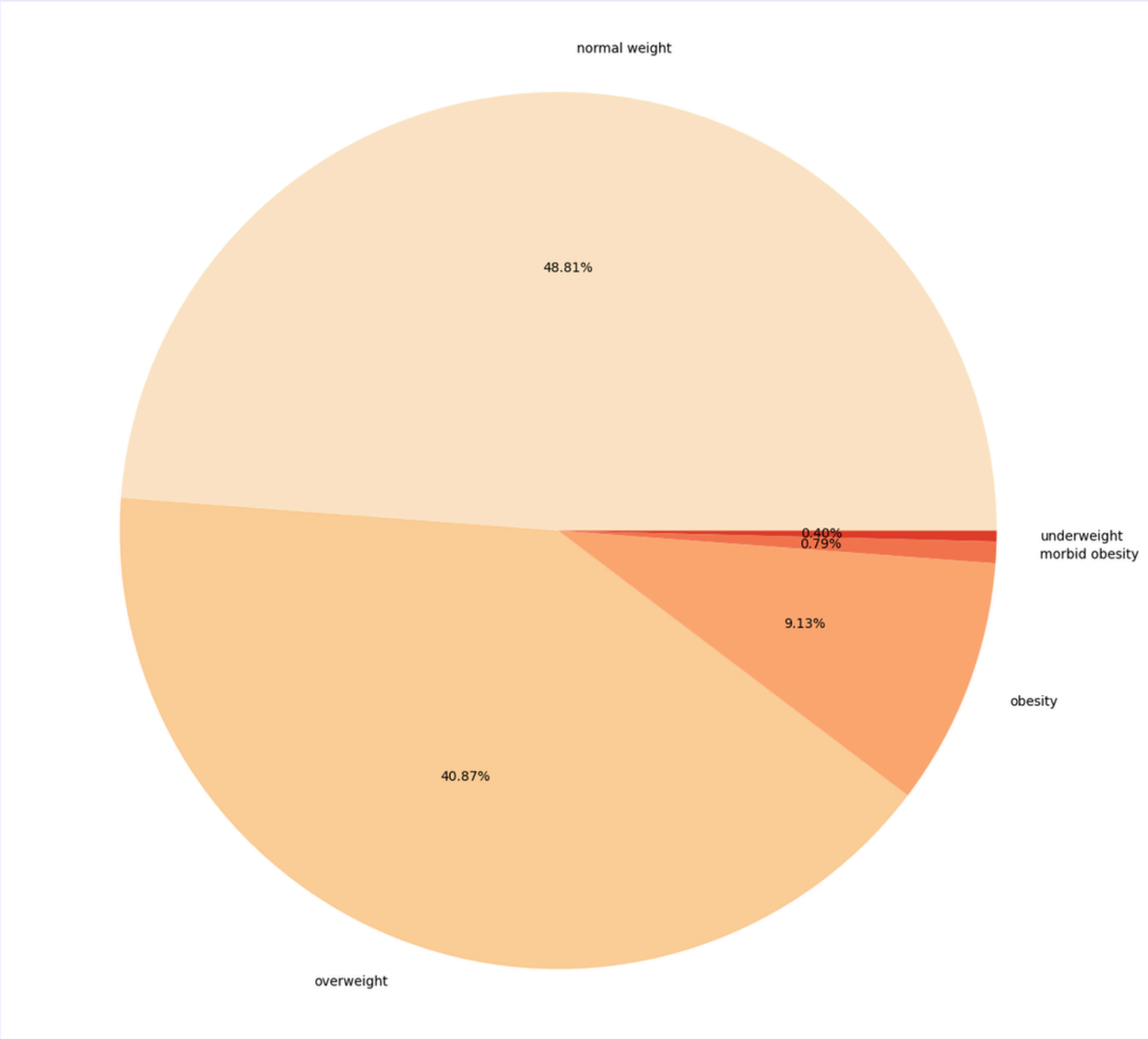
VISUALIZATION - WEIGHT IN FUNCTION OF HEIGHT



VISUALIZATION - BODYFAT IN FUNCTION OF AGE



VISUALIZATION - FREQUENCIES OF WEIGHT CLASSIFICATION BASED ON BMI



Dataset before data processing

	Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
0	1.0708	12.3	23	154.25	67.75	36.2	93.1	85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
1	1.0853	6.1	22	173.25	72.25	38.5	93.6	83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
2	1.0414	25.3	22	154.00	66.25	34.0	95.8	87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6
3	1.0751	10.4	26	184.75	72.25	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2
4	1.0340	28.7	24	184.25	71.25	34.4	97.3	100.0	101.9	63.2	42.2	24.0	32.2	27.7	17.7

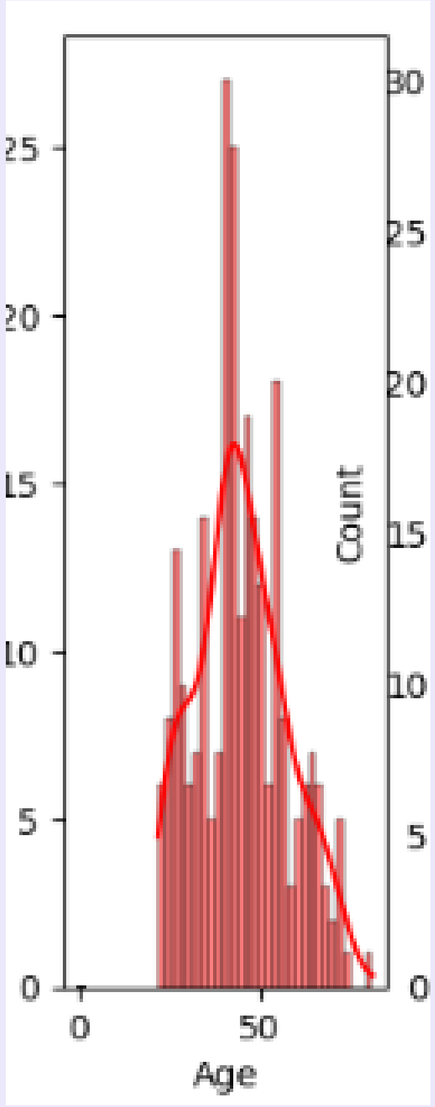
Data Cleaning and Preprocessing

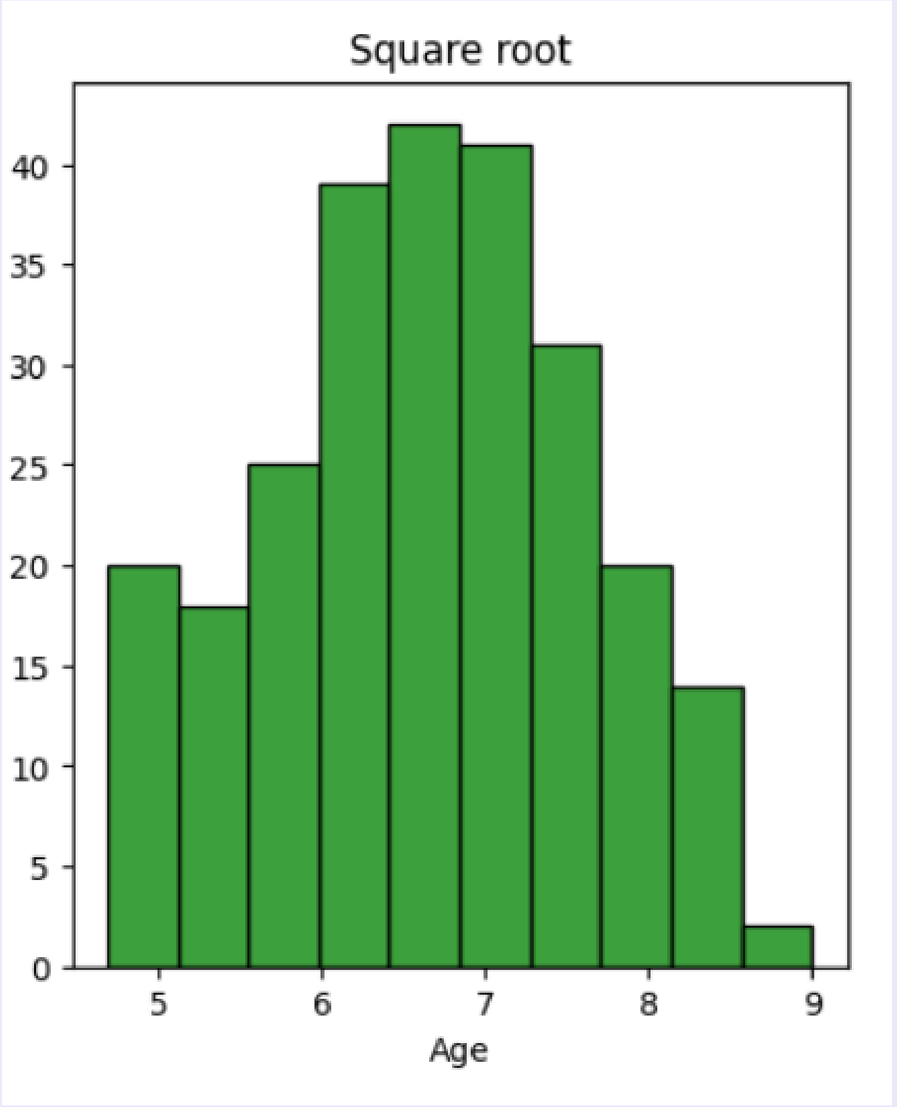
- Handling missing values and outliers
- Weight (lbs) to (kg) & Height (inches) to (cm)

- New column: BMI (Body mass index) = Weight/((Height/100)**2)

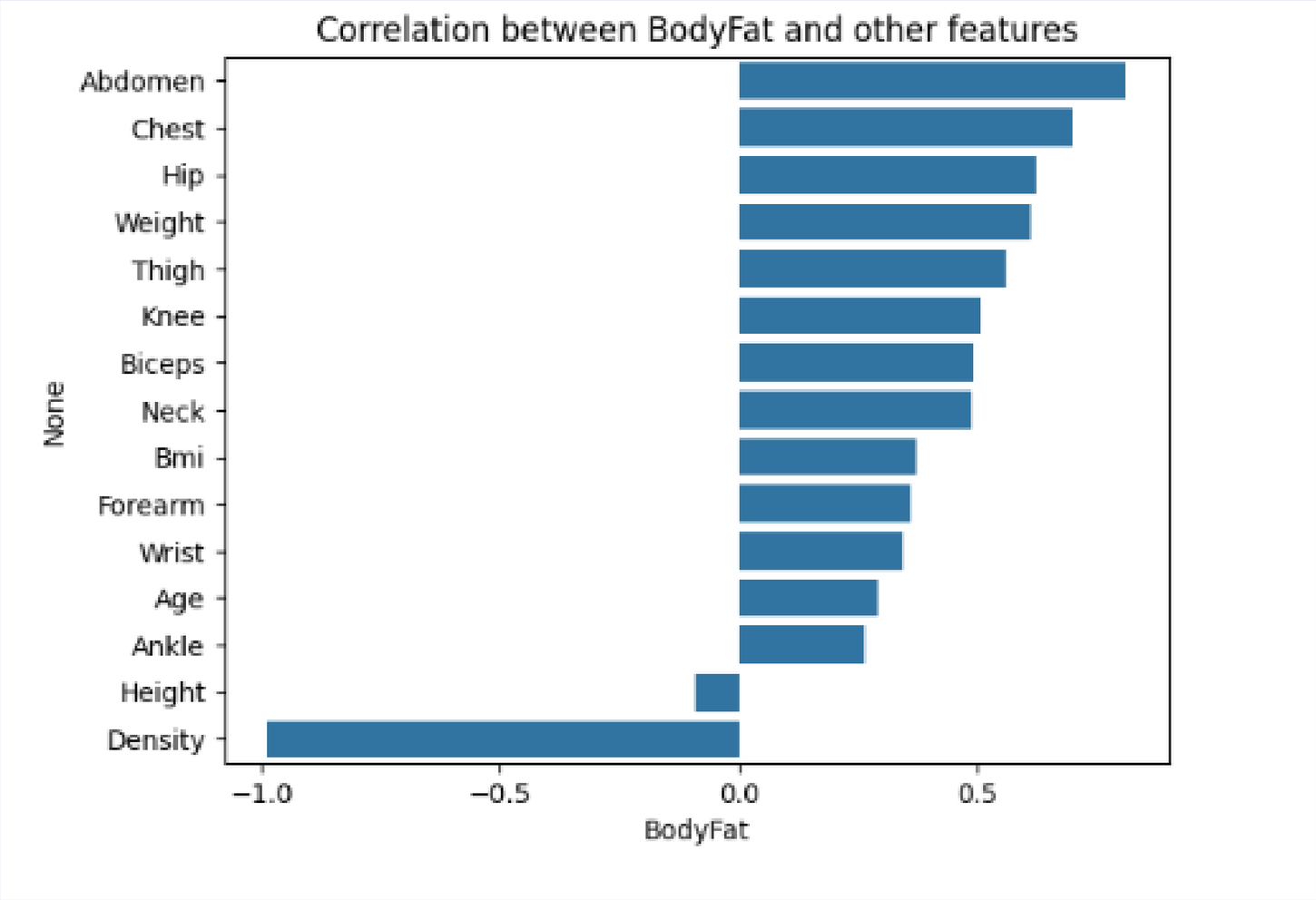
  BMI is an inexpensive and easy screening method for weight category – underweight, healthy weight, overweight, and obesity.

Normalization of Age column

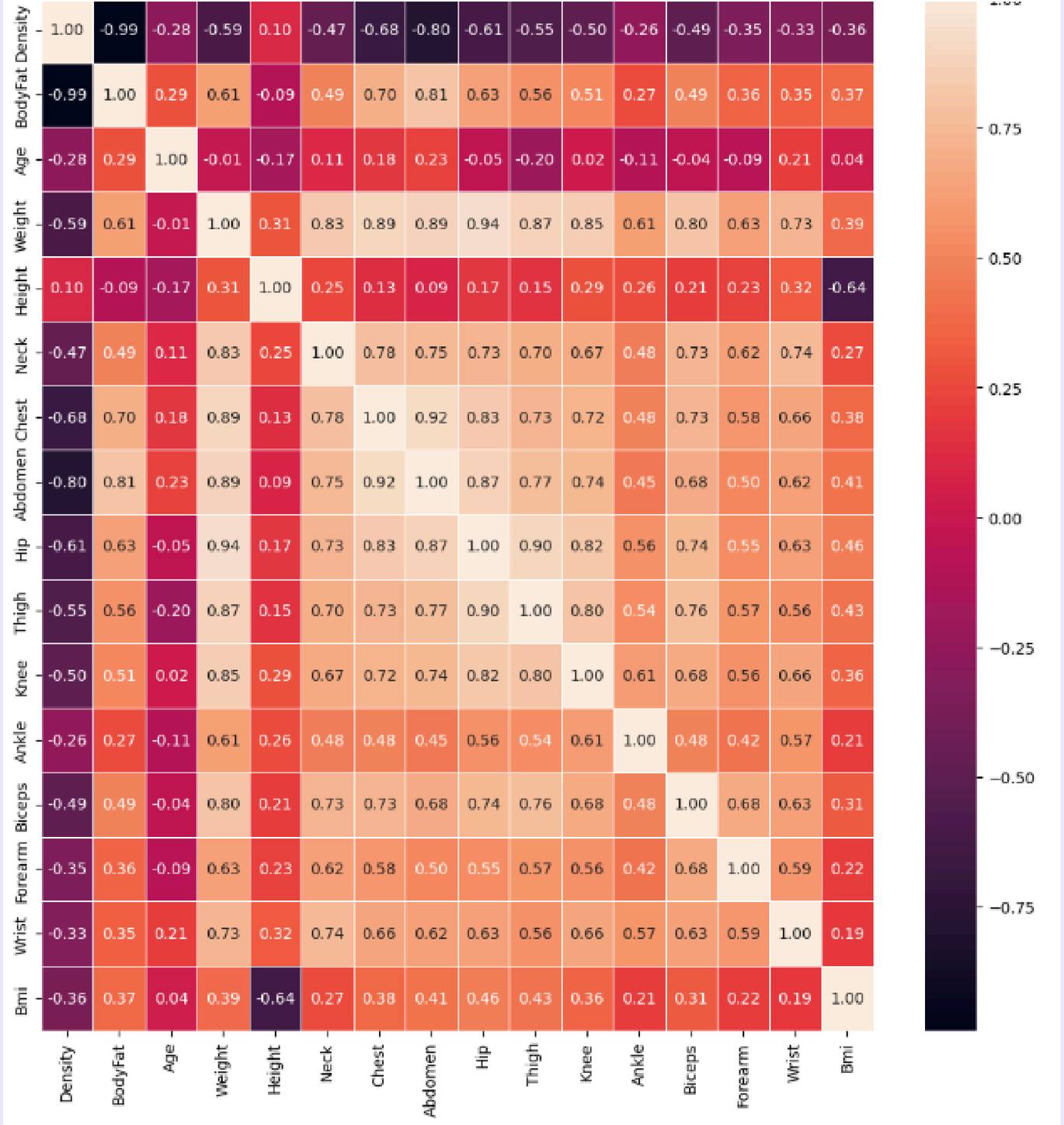




TACKLING MULTICOLLINEARITY BY REMOVING HIGHLY CORRELATED FEATURES



Body Fat = (495/Density) – 450.



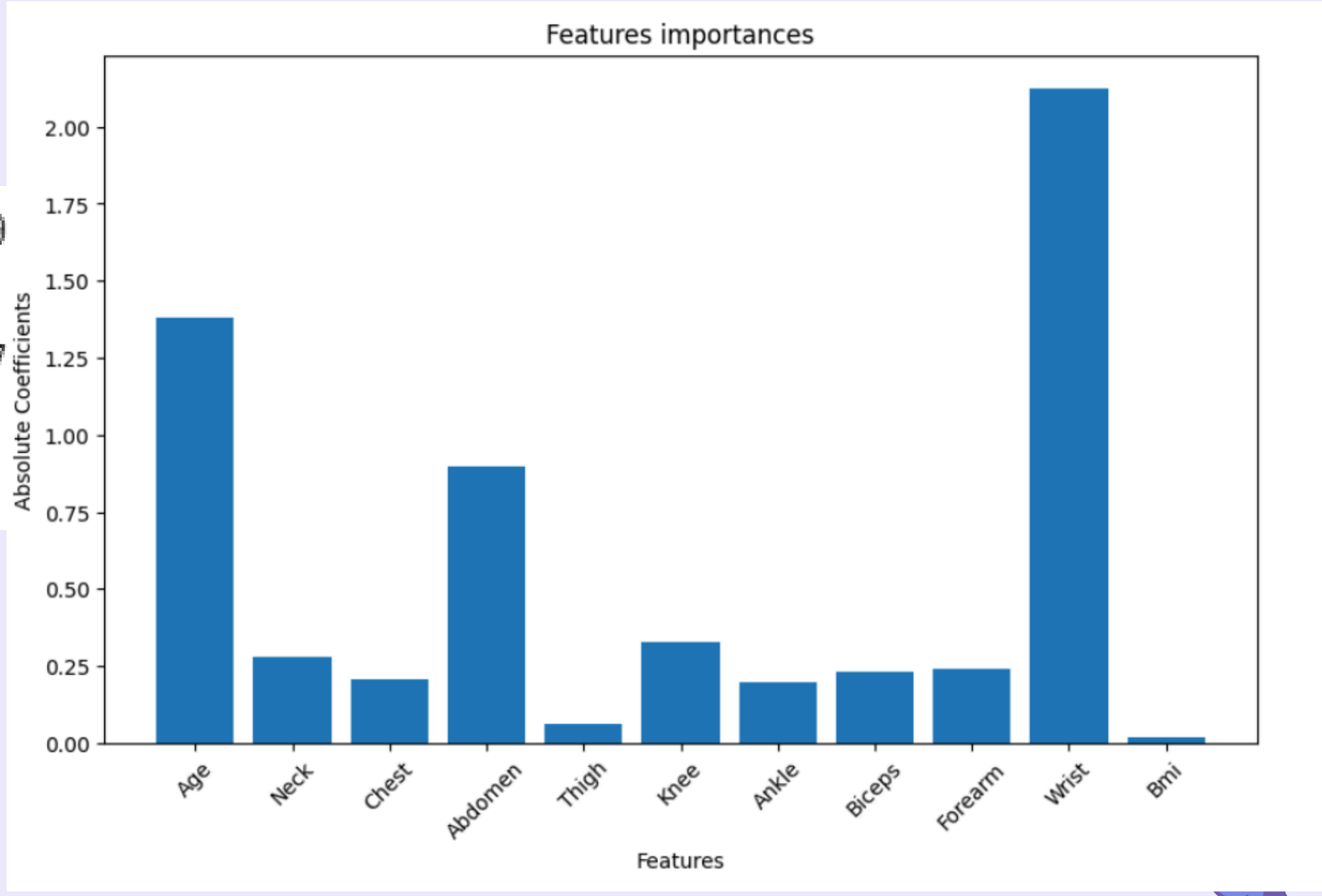
Dropping columns that are too highly correlated such as ‘Weight’ and ‘Hip’

Final dataset for prediction

	BodyFat	Age	Neck	Chest	Abdomen	Thigh	Knee	Ankle	Biceps	Forearm	Wrist	Bmi
0	12.3	4.795832	36.2	93.1	85.2	59.0	37.3	21.9	32.0	27.4	17.1	23.626779
1	6.1	4.690416	38.5	93.6	83.0	58.7	37.3	23.4	30.5	28.9	18.2	23.334338
2	25.3	4.690416	34.0	95.8	87.9	59.6	38.9	24.0	28.8	25.2	16.6	24.668737
3	10.4	5.099020	37.4	101.8	86.4	60.1	37.3	22.8	32.4	29.4	18.2	24.883227
4	28.7	4.898979	34.4	97.3	100.0	63.2	42.2	24.0	32.2	27.7	17.7	25.517358

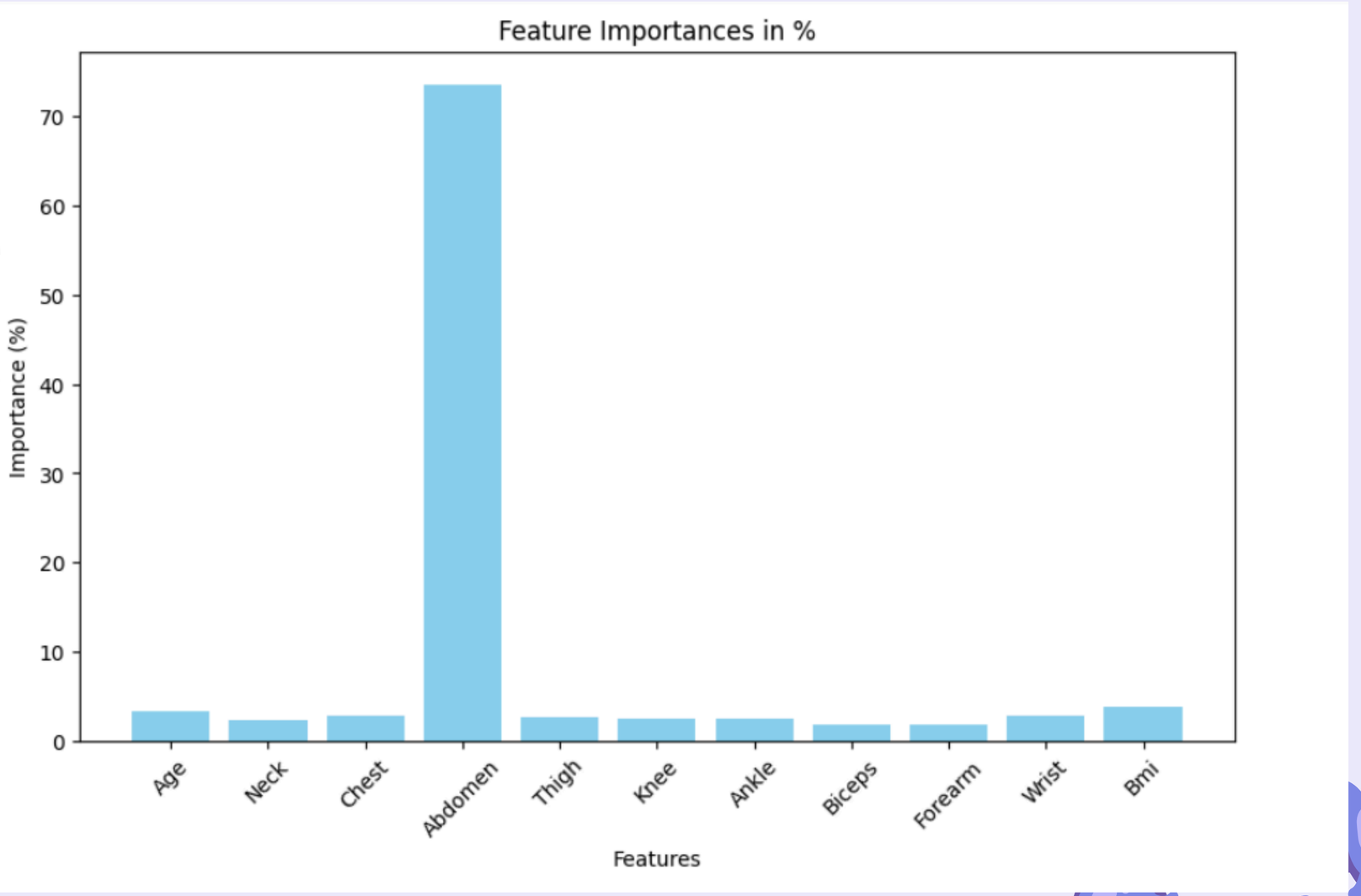
Linear Regression Model

```
Mean Squared Error: 21.435076609
r2 score: 0.539209674395323
Mean Absolute Error: 3.431108117
rmse: 4.62980308538063 4
```



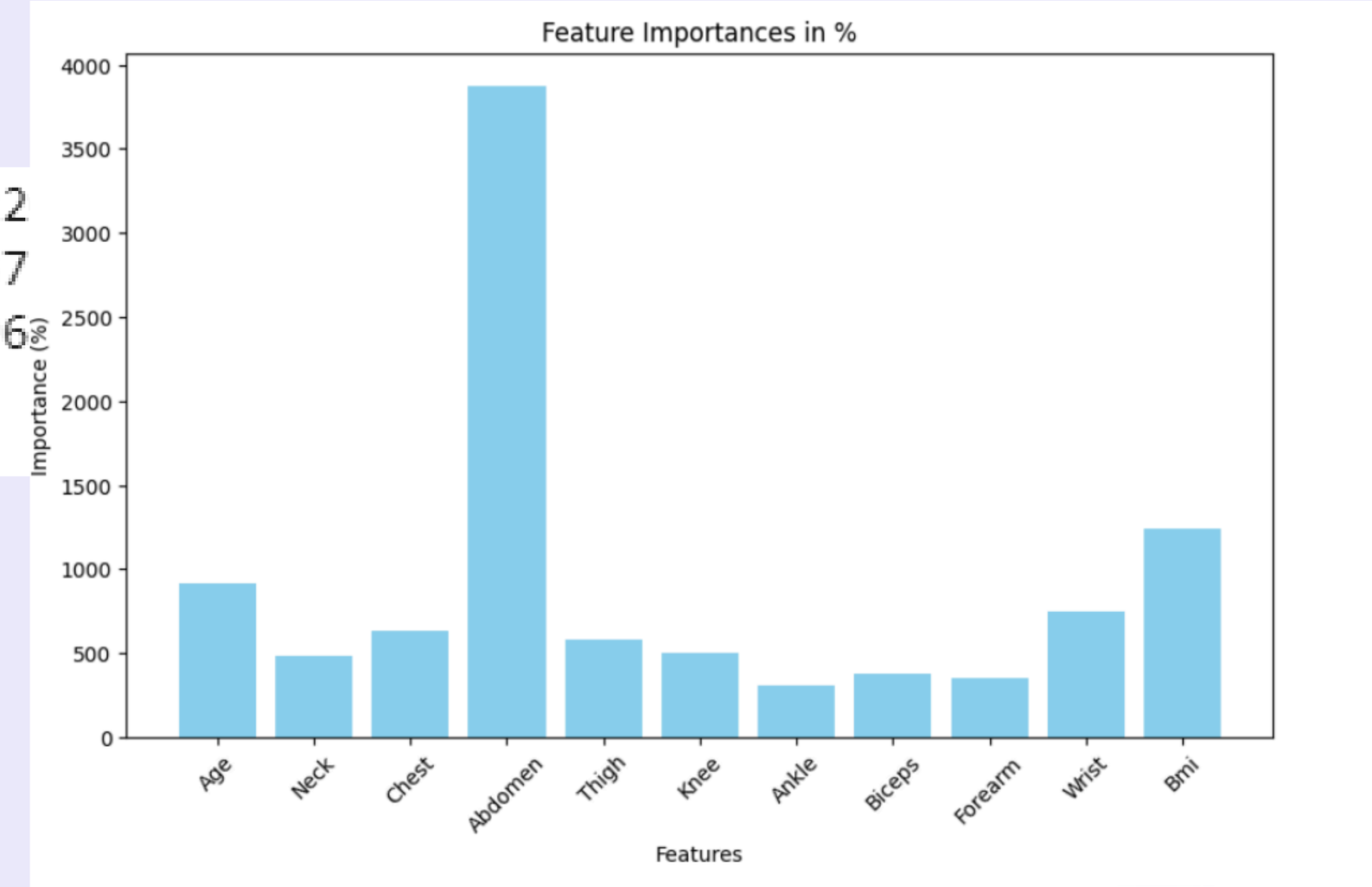
RandomForestRegressor Model

```
Mean Squared Error: 16.52770
r2 score: 0.6447035544180055
Mean Absolute Error: 3.35861
rmse: 4.065428011510278
```

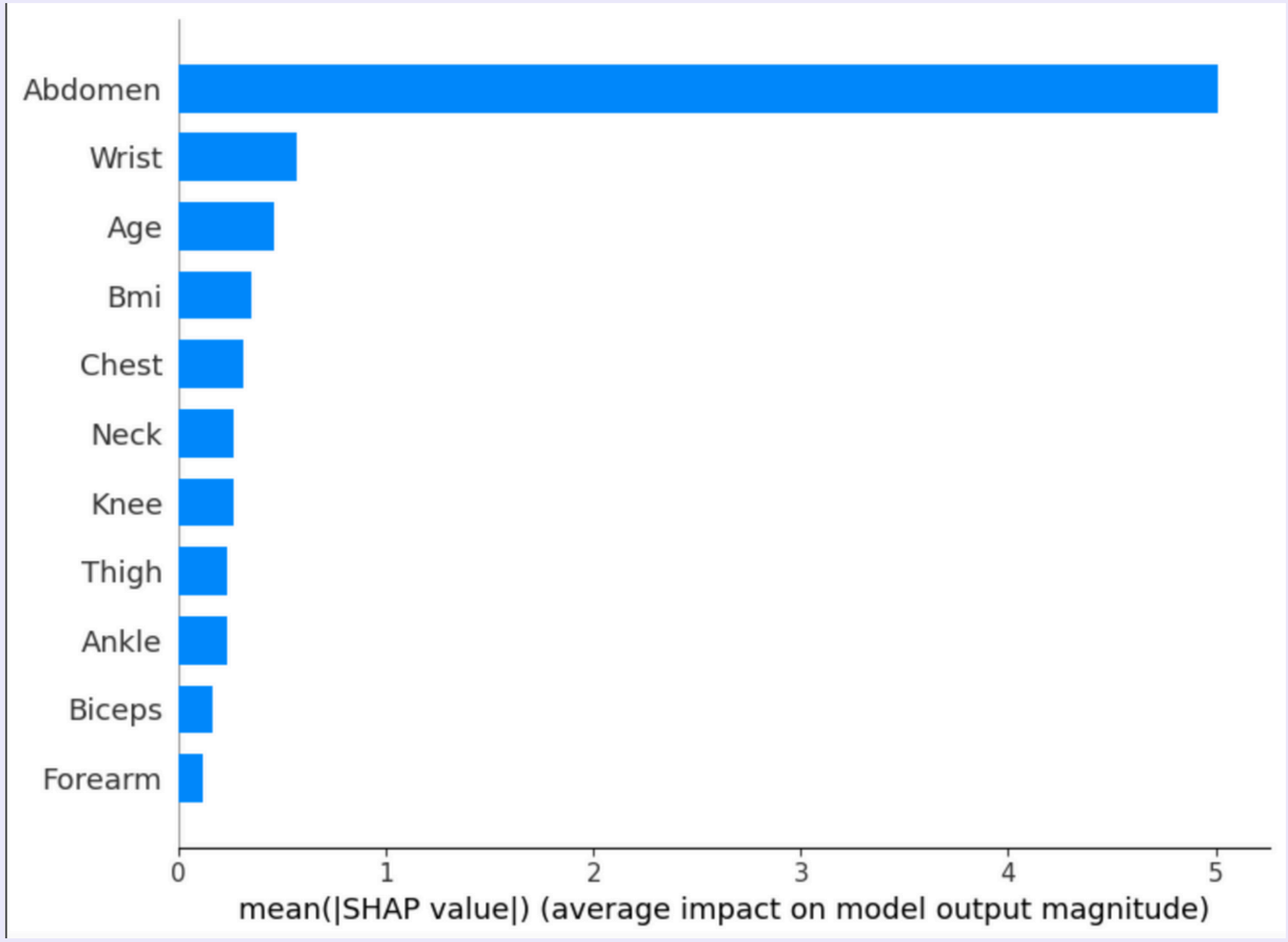


CatBoostRegressor Model

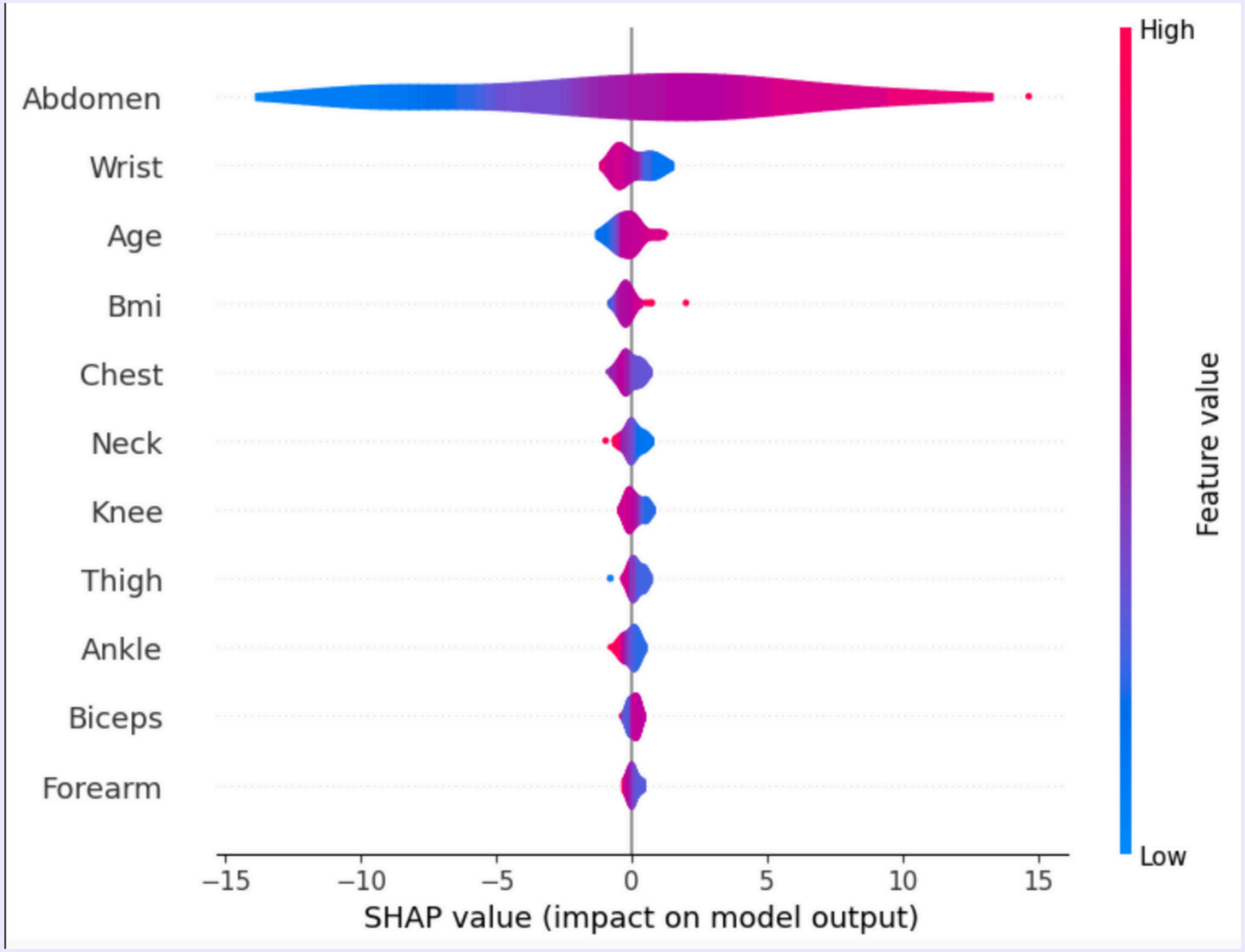
```
Mean Squared Error: 15.37122
r2 score: 0.669564526468253 7
Mean Absolute Error: 3.25996
rmse: 3.9206148088456745
```



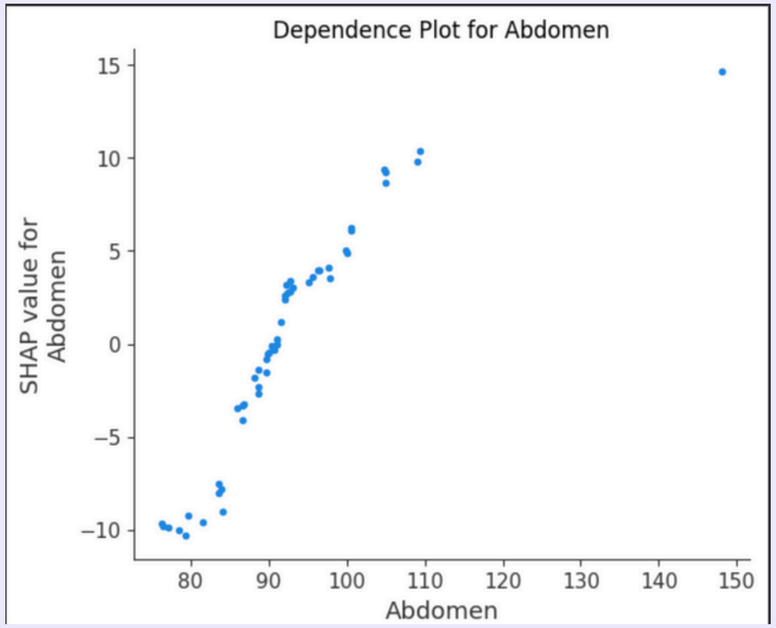
Shapley values

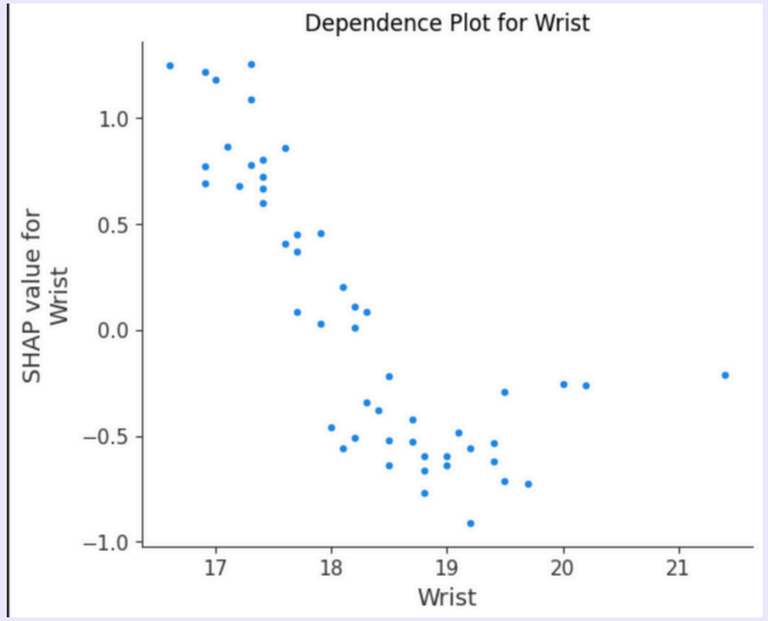


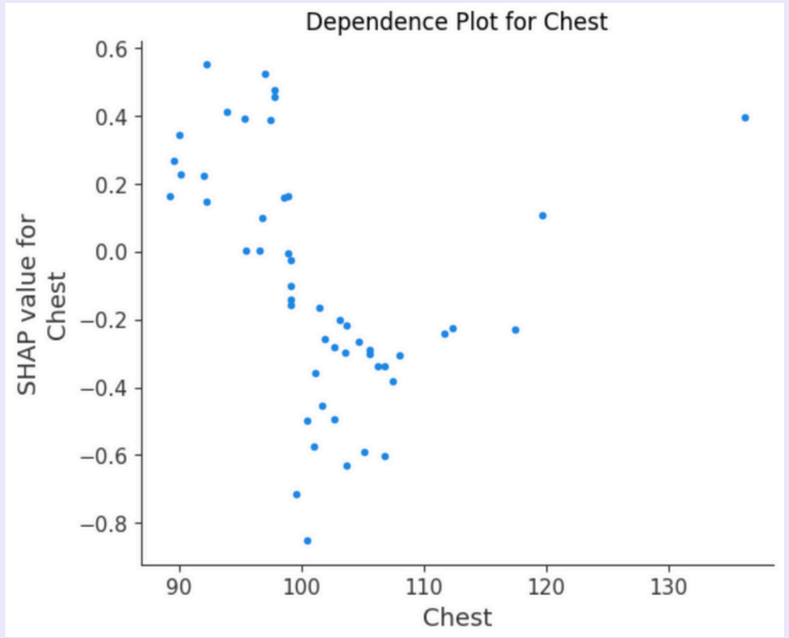
Beeswarm plot

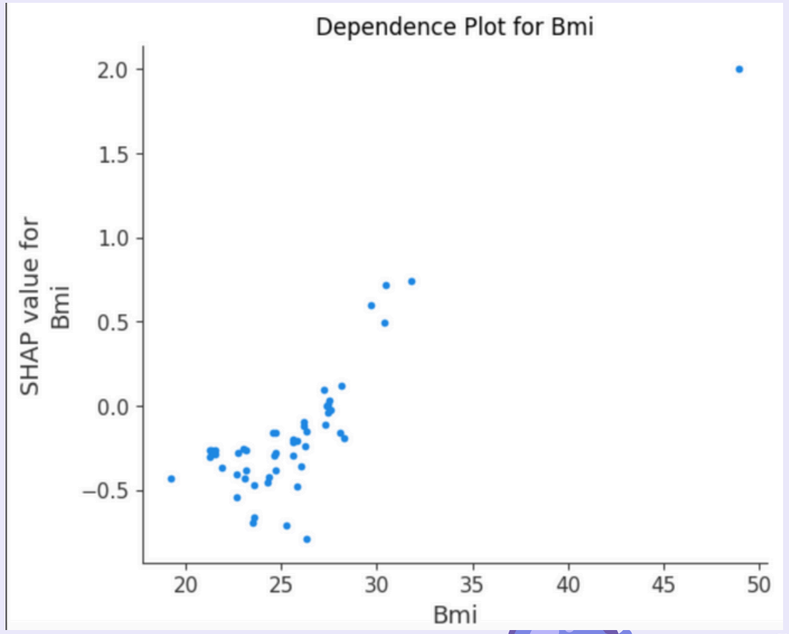


Dependance plots

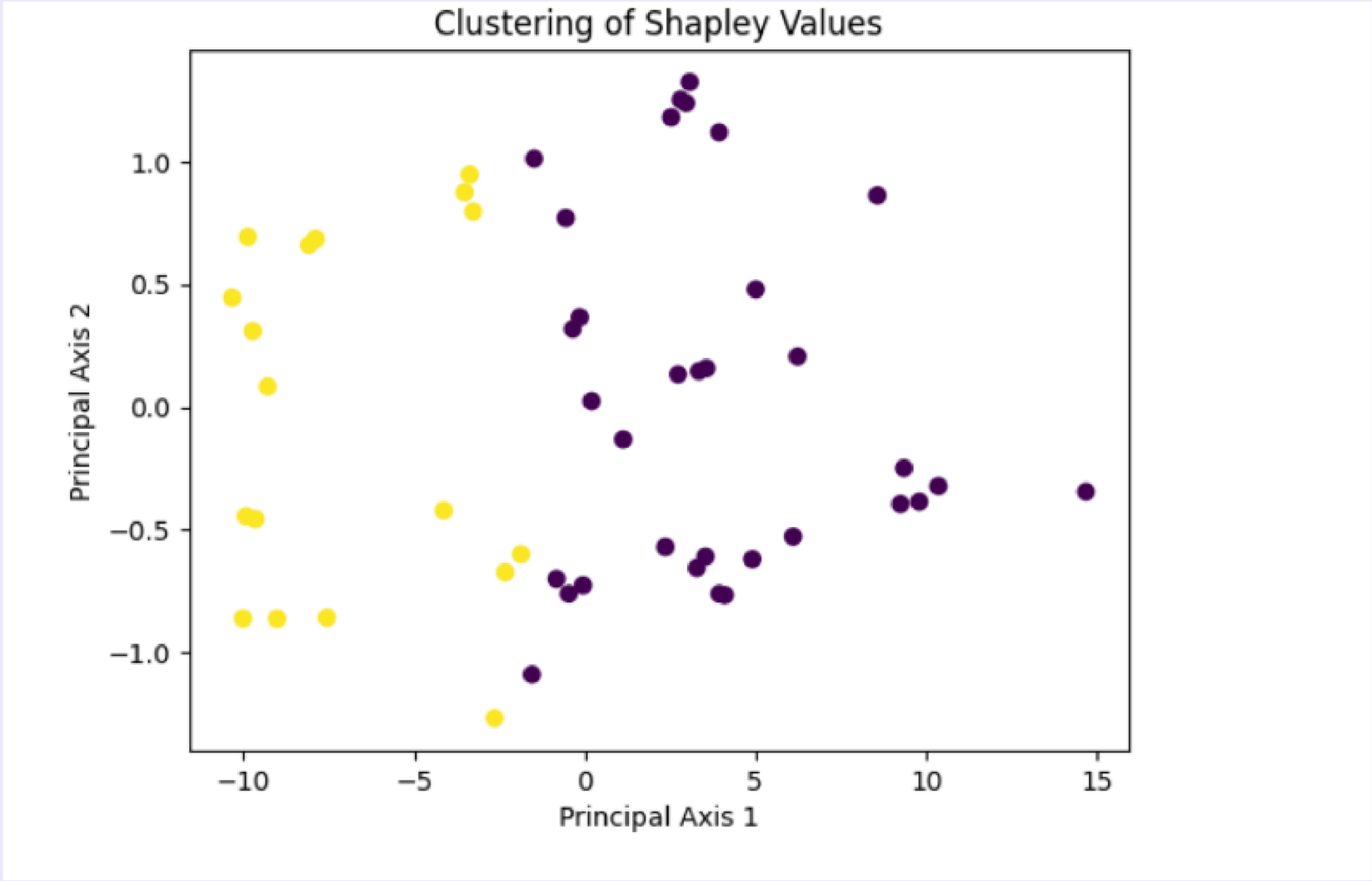








Clustering


Clustering of Shapley Values
Principal Axis 2
1.0
0.5
0.0
−0.5
−1.0
−10
−5
0
5
10
15
Principal Axis 1

CONCLUSIONS

01 ABDOMEN PLAYS PIVOTAL ROLE

02 ALIGNMENT WITH MEDICAL UNDERSTANDING

03 PERSONALIZED HEALTH INSIGHTS

THANK YOU

FOR WATCHING THIS PRESENTATION

Q&A