Body Performance
Exploratory Analysis

Ben Griffis, bgriffis@bellarmine.edu
Peyton Fosnough, pfosnough@bellarmine.edu

I. INTRODUCTION

This is a data set that measured different ages of different heights and weights on how well they performed on different athletic measurements, based on their performance they were placed in a class. These classes are scored using an ABCD grading scale. We chose this data set because we are both athletes and have interest in fitness/health and thought it would be interesting to see how these different variables correlated with each other.

https://www.kaggle.com/datasets/kukuroo3/body-performance-data

II. DATA SET DESCRIPTION

This data set contains 13,393 samples with 12 columns and had different data types.

Table 1: Data Types and Missing Data

Variable Name	*Data Type*	*Missing Data (%)*
V1 age	**Nominal, int64**	**0 %**
V2 gender	**Nominal, object**	**0 %**
V3 height_cm	**Interval, float64**	**0 %**
V4 weight_kg	**Interval, float64**	**0 %**
V5 body fat_%	**Interval, float64**	**0 %**
V6 diastolic	**Ratio, float64**	**0 %**
V7 systolic	**Ratio, float64**	**0 %**
V8 gripForce	**Interval, float64**	**0 %**
V9 sit and bend forward_cm	**Interval, float64**	**0 %**
V10 sit-ups counts	**Interval, float64**	**0 %**
V11 broad jump_cm	**Interval, float64**	**0 %**
V12 class	**Ordinal, object**	**0 %**

III. Data Set Summary Statistics

Narrative introduction to the section.

Table 2: Summary Statistics for XXX (name of dataset)

Variable Name	*Count*	*Mean*	*Standard Deviation*	*Min*	*25th*	*50th*	*75th*	*Max*
age	13393	36.775	13.626	21	25	32	48	64
Height_cm	13393	168.56	8.4266	125	162.4	169.2	174.8	193.8
Weight_kg	13393	67.447	11.95	26.3	58.2	67.4	75.3	138.1
Body fat_%	13393	23.240	7.2568	3	18	22.8	28	78.4
diastolic	13393	78.797	10.742	0.0000	71	79	86	156.2
systolic	13393	130.23	14.714	0.0000	120	130	141	201

gripForce	13393	36.964	10.625	0.0000	27.5	37.9	45.2	70.5
Sit-ups counts	13393	39.771	14.277	0.0000	30	41	50	80
Broad jump_cm	13393	190.13	39.868	0.0000	162	193	221	303
Sit and bend forward_cm	13393	15.209	8.4567	-25.00	10.9	16.2	20.7	213

There should be a table for **EACH** categorical variable.

Table 3: Proportions for XXX (n=yyy)

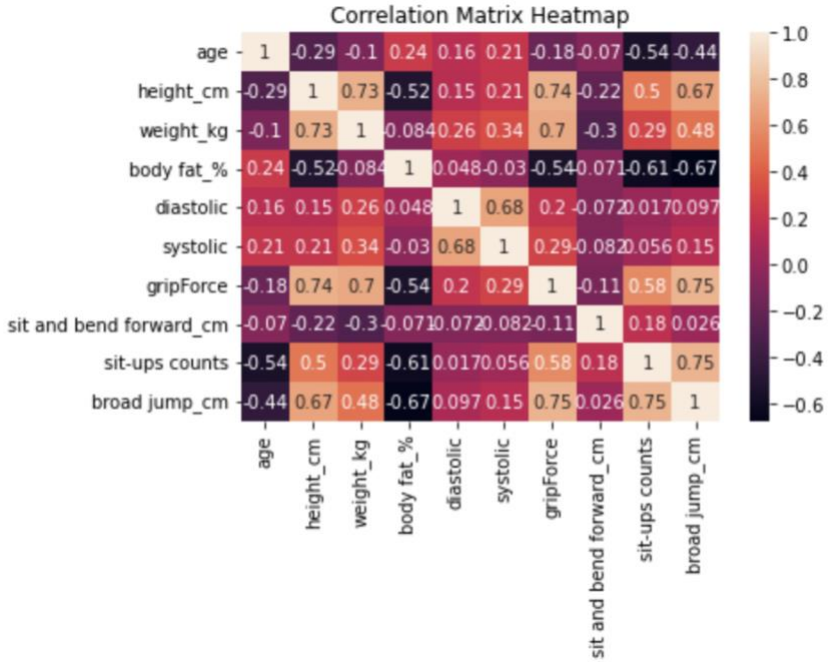
Category	*Frequency*	*Proportion (%)*
Gender	*Male - 8467* *Female - 4926*	*Male – 63.22* *Female – 36.78*
Class	*A - 3348* *B - 3347* *C - 3349* *D - 3349*	*A – 25.00* *B - 24.99* *C – 25.01* *D – 25.01*

After you summarize the categorical variables, generate a correlation matrix for all continuous variables (not categorical – this doesn't make sense)

Table 4: Correlation Table/Tables

	age	**height_cm**	**weight_kg**	**body fat_%**	**diastolic**	**systolic**	**gripForce**	**sit and bend forward_cm**	**sit-ups counts**	**broad jump_cm**
age	1.000000	-0.293980	-0.099966	0.2423	0.158508	0.211167	-0.18	-0.07	-0.545	-0.435172
height_cm	-0.294	1.000000	0.734909	-0.515	0.145933	0.210186	0.735024	-0.221970	0.500424	0.674589
weight_kg	-0.09996	0.734909	1.000000	-0.084	0.26232	0.338943	0.700119	-0.296249	0.294899	0.479564
Body fat_%	0.242302	-0.51544	-0.084065	1.000000	0.048059	-0.03038	-0.54179	-0.071225	-0.60891	-0.673273
diastolic	0.158508	0.145933	0.262317	0.048059	1.000000	0.67631	0.202062	-0.072098	0.016547	0.097243
systolic	0.211167	0.210186	0.338943	-0.03037	0.676309	1.000000	0.286012	-0.082434	0.056276	0.152894
gripForce	-0.17958	0.735024	0.700119	-0.5418	0.202062	0.286012	1.000000	-0.112577	0.576669	0.746853
sit and bend forward_cm	-0.07003	-0.22197	-0.2962	-0.07123	-0.0721	-0.08243	-0.11258	1.000000	0.177153	0.026487
sit-ups counts	-0.54458	0.500424	0.2949	-0.60891	0.01655	0.056276	0.576669	0.177153	1.000000	0.748273
broad jump-cm	-0.43517	0.674589	0.47956	-0.6733	0.097243	0.152894	0.74685	0.026487	0.748273	1.000000

After the table with the raw data, include a heatmap of the correlation matrix as a figure.



IV. DATA SET GRAPHICAL EXPLORATION

FIGURE 1 – Looking at the bar chart comparison of ages to body fat %, from this graph we can observe multiple things. First, we can see that between the age of around 27 to 34 is where there is the lowest percentage of body fat. Besides from an outlier at the age of 21 we can see that the age and body fat % is consistent.

Figure 2: (a) Function Output (b) A against B (multiple plots) (8 pt)

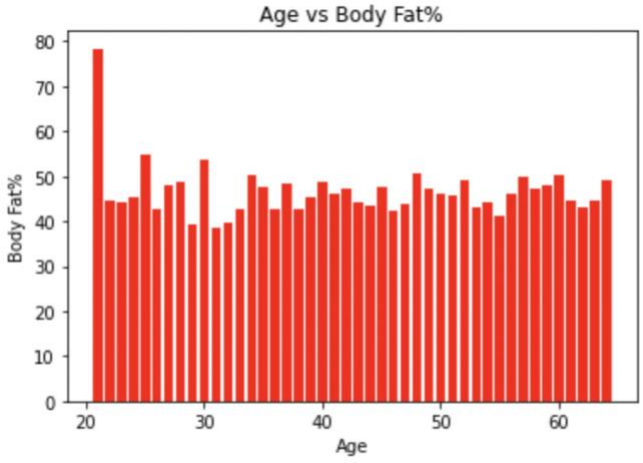


FIGURE 2/3 – Using the two joint plots below we were trying to test to see if there was a correlation between size and grip force. Using the weight and height variables as our x axis and the grip force as the y axis we can see that as the size, whether that be weight or height increases, the grip force follows that trend and increases.

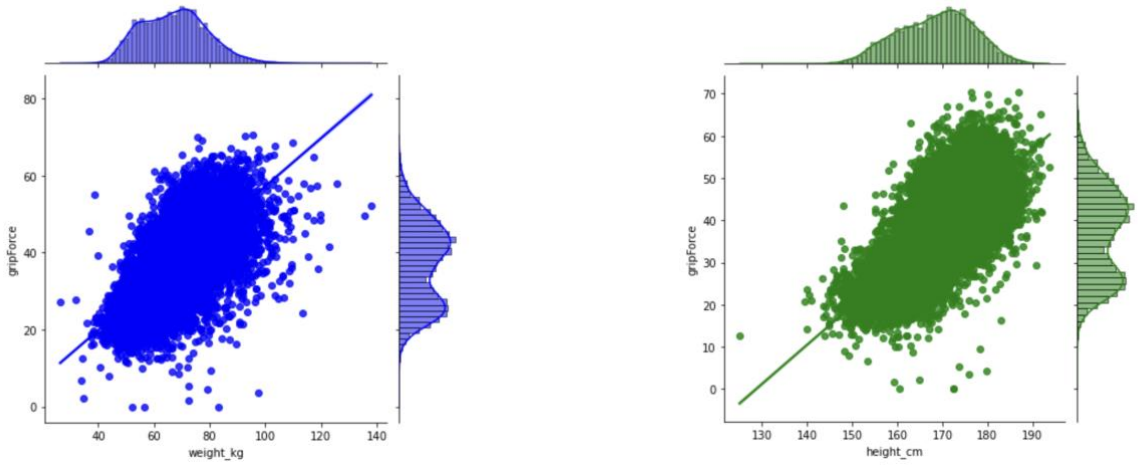


FIGURE 4 - By the looks of this graph we can see that the higher the weight the higher the systolic, except for a few outliers which could be because of various reasons that trigger high heart rate. The measurement of systolic is your heart rate so we can see that the heavier someone is the more their heart beats which could potentially cause health problems.

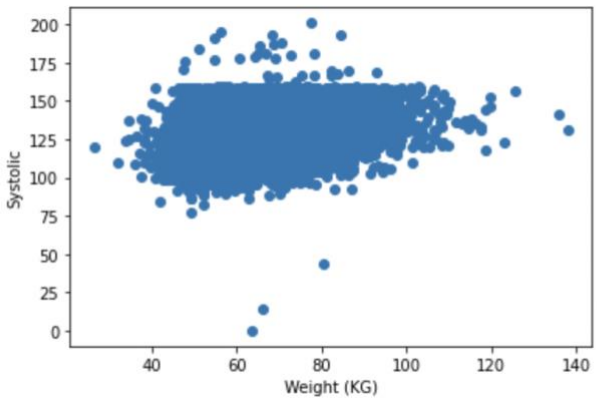
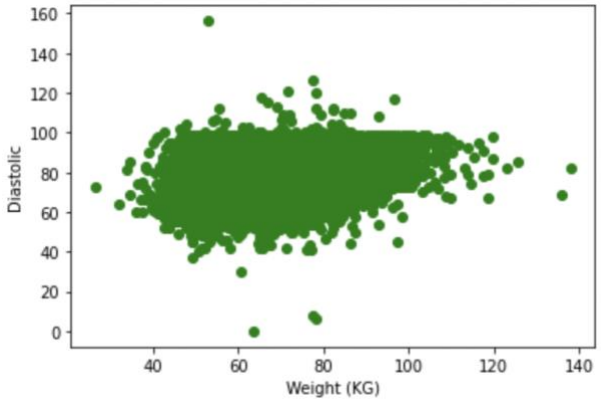


FIGURE 5 - Similarly to the figure above, by the looks of this graph we can see that the higher the weight the higher the diastolic, except for a few outliers which could be influenced to outside related factors rather than pure health. The measurement of diastolic is your blood pressure so we can see that the heavier someone is the higher their blood pressure which could potentially cause health problems like heart failure and heart attacks or stroke.



V. SUMMARY OF FINDINGS

Finish up with a paragraph or two of summarizing your findings about this data set.

In conclusion there are many relationships and ways we can compare this data to itself. We as humans must focus on our health and take care of ourselves to put us in the best shape as possible in order to survive for the greatest amount of time possible. When looking and using this data set and the different measurements in each specific column we can see how our health may change due to bodily factors. When looking at weight and how it compares to different variables such as systolic (heartrate) we can see that the more you weigh the higher your heart rate and that can mean your heart may not pump enough blood to the rest of the body, and as a result the organs and tissues may not get enough oxygen. A similar conclusion can be found when looking at weight and how it relates to diastolic, which is your blood pressure. With a high diastolic/blood pressure health problems such as heart attack, stroke, and heart failure is more common. By looking at simple relationships like the ones given as examples we can see that taking responsibility with certain aspects of our health can be detrimental to our bodies.