

INTRODUCTION

The data set is provided in two csv files containing

- breed
- minimum and maximum height and weight
- classification of obedience
- probability of obeying a command
- minimum and maximum repetitions to learn command

How I cleaned the data set

- combined the two csv files
- removed index columns
- converted obey, height, and weight columns to floats
- saved one csv file with cleaned data

In this project, I explore the relationship between the minimum and maximum height and weight of various breeds of dogs by classification of obedience to using linear regression.

The six classifications of obedience are

- brightest dogs
- excellent working dogs
- above average working dogs
- average working/obedience intelligence
- fair working/obedience intelligence
- lowest degree of working/obedience intelligence

RESULTS

For each of the classifications of obedience, I compared the minimum and maximum heights and weights of each breed to see if there was any relationship. To do this, I used linear regression.

Linear regression uses labeled data to create a model which is then used to predict new values. The line of best fit is the equation of the model. I then used the F-statistic in the OLS Regression Results from Python to determine if linear regression is a good fit for this data.

Figure 1 below is an example of linear regression for the Average Working/Obedience Intelligence classification of dogs. We see there is a decent relationship between the minimum height and weight of dogs. We also see that there is not one specific size dog that falls in this classification. The F-statistic for the regression is $4.17e - 14$ tells me that there is a strong relationship between the minimum height and weight for this classification of dogs.



Figure 1: Correlation between Minimum Height and Weight for Average Working/Obedience Intelligence Classification

The scatter plot in Figure 2 below is another example of linear regression for the Fair Working/Obedience Intelligence classification. In this plot, we see there is less of a relationship between the maximum height and weight. This can be seen with the F-statistic for this regression model: $5.14e - 05$. There is still a decent relationship since this number is small, but it is not as good as in Figure 1. It also appears that more dogs that are smaller in size will fall into this classification.

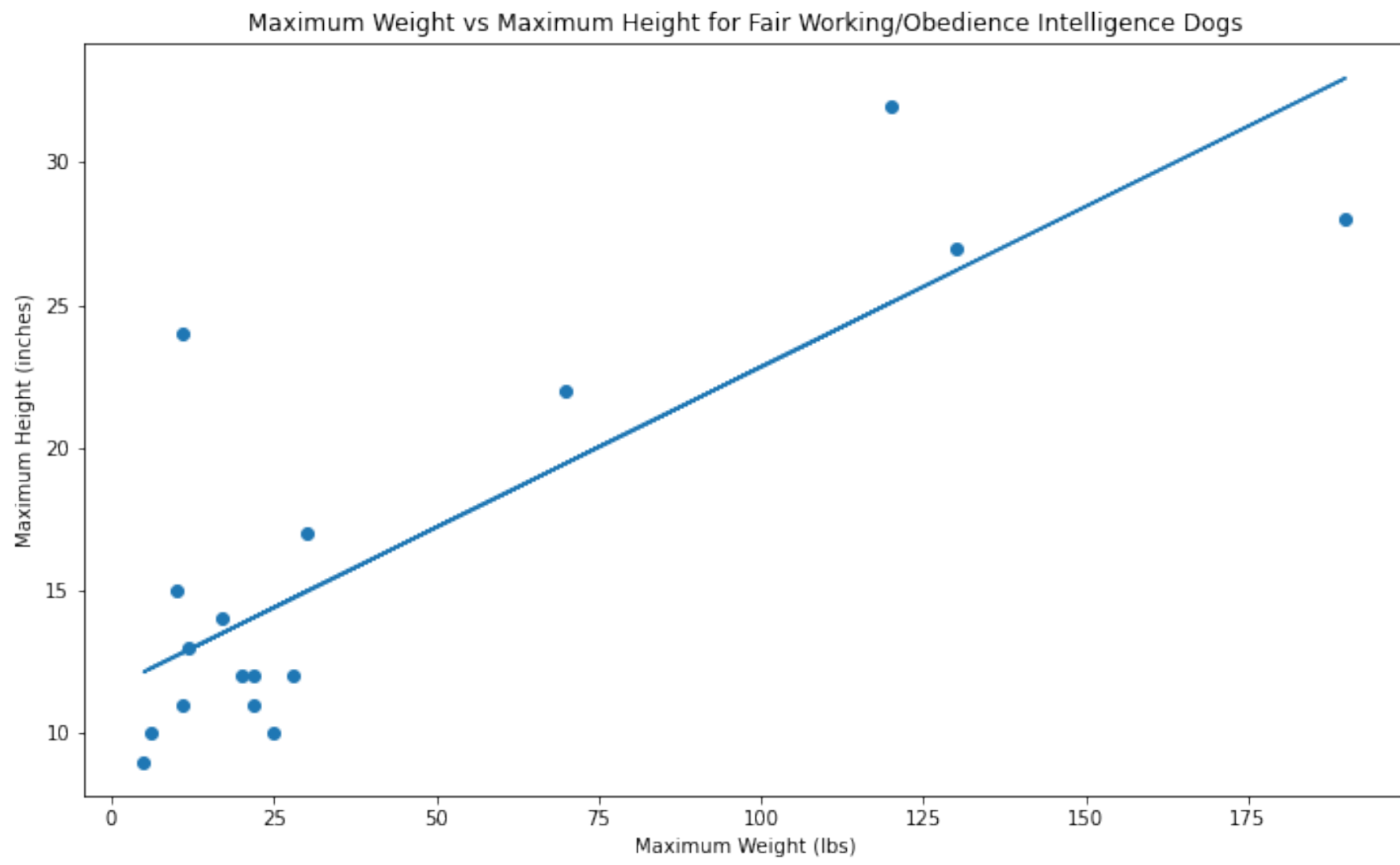


Figure 2: Correlation between Maximum Height and Weight for Fair Working/Obedience Intelligence Classification

CONCLUSION

Discuss results from the linear regression on the other classifications of dogs. (Will update draft when completed)