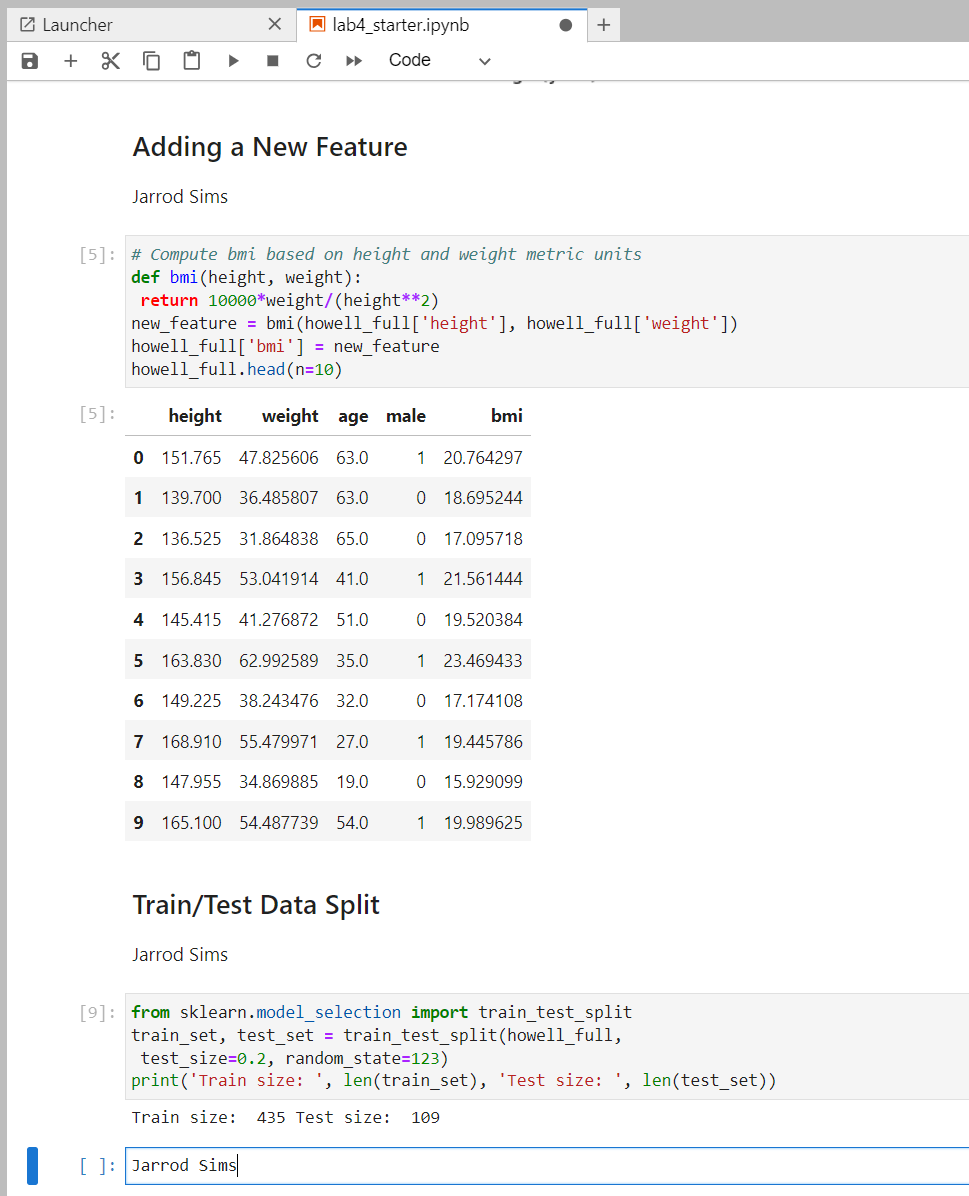
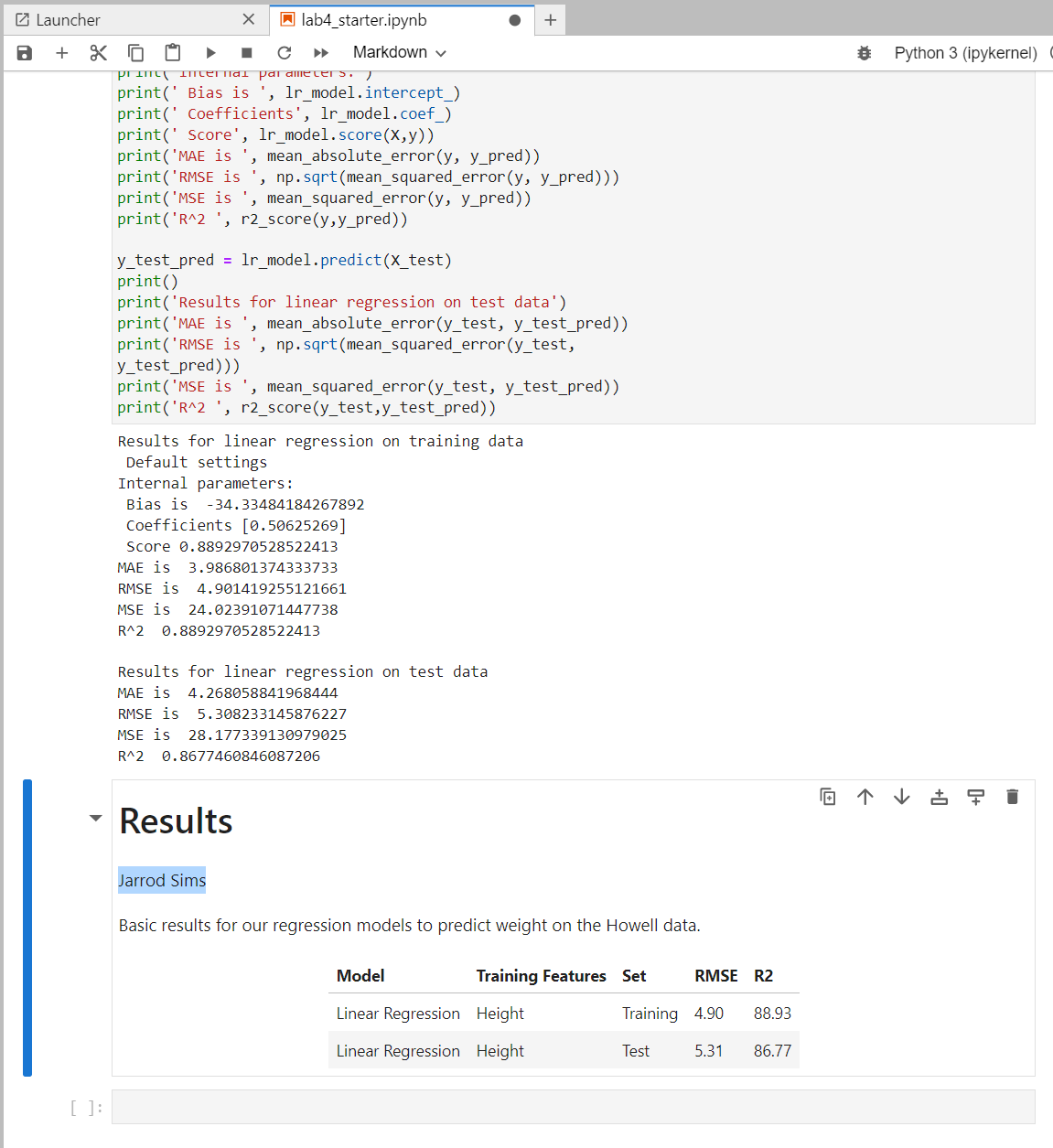
Module Four Lab: Building a Regression Model

# Author: Jarrod Sims

## Progress 1 - Screen shot pipeline ready to replace classifier with regression model



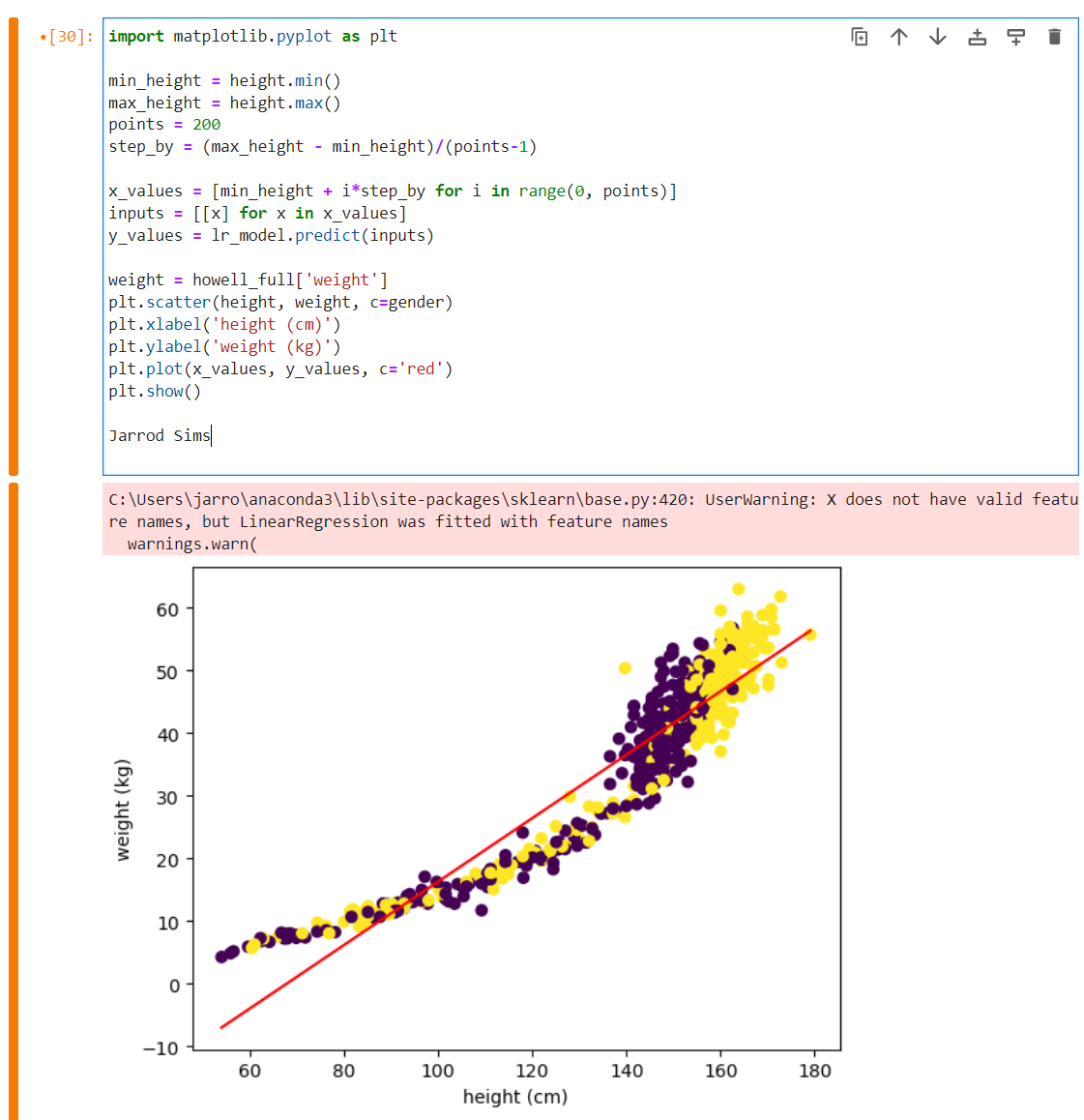
## Progress 2 - Screen shot performance of model on training and test sets.



## Progress 3 - Analysis

* Compare performance and argue if model is underfitting the training set.
  + The r2 value for both the training and the test set are high and very close indicating that the model is not significantly underfitting or overfitting. The RMSE values are relatively low for both the training and the test set as well which indicates that the model is performing well.

## Progress 4 - Screen shot linear fit



## Submission 5 - Analysis

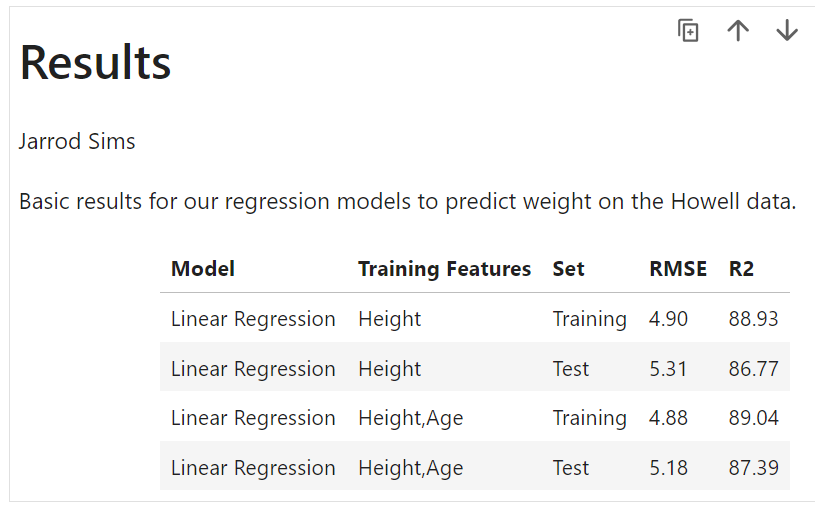
1. Will more training data improve the fit?

Adding another feature to the training data would likely improve the fit.

1. Is the model under or over fitting?

The model is underfitting the data. The model appears to have high bias and the model’s assumption that the data is linear is incorrect.

## Submission 6 – Explain



1. Did adding age improve the model?

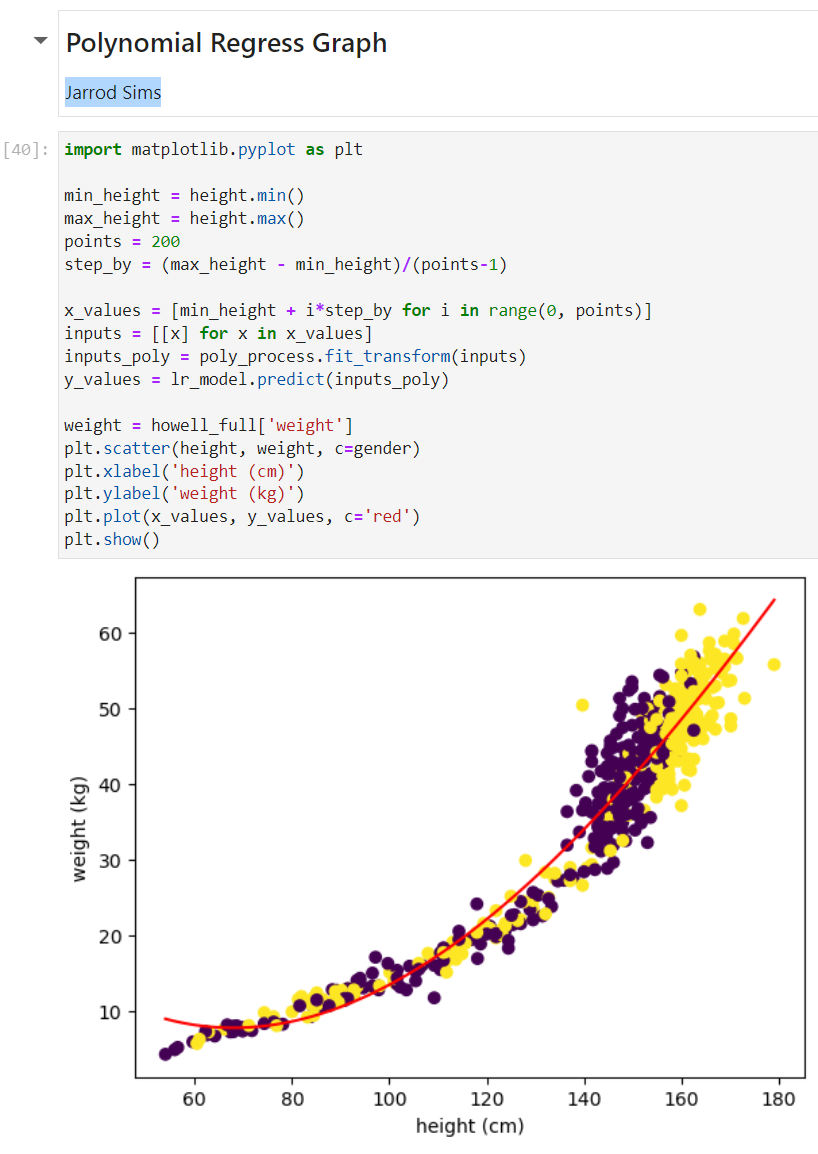
Adding age did not have significantly improve the performance of the model.

1. Propose an explanation for the results.

The most likely reason that adding age did not improve the performance of the model is that the model is still assuming a linear relationship between the input and output variables and the relationship is non-linear.

## Progress 7 - Scatter graph

1. Scatter graph with cubic model displayed



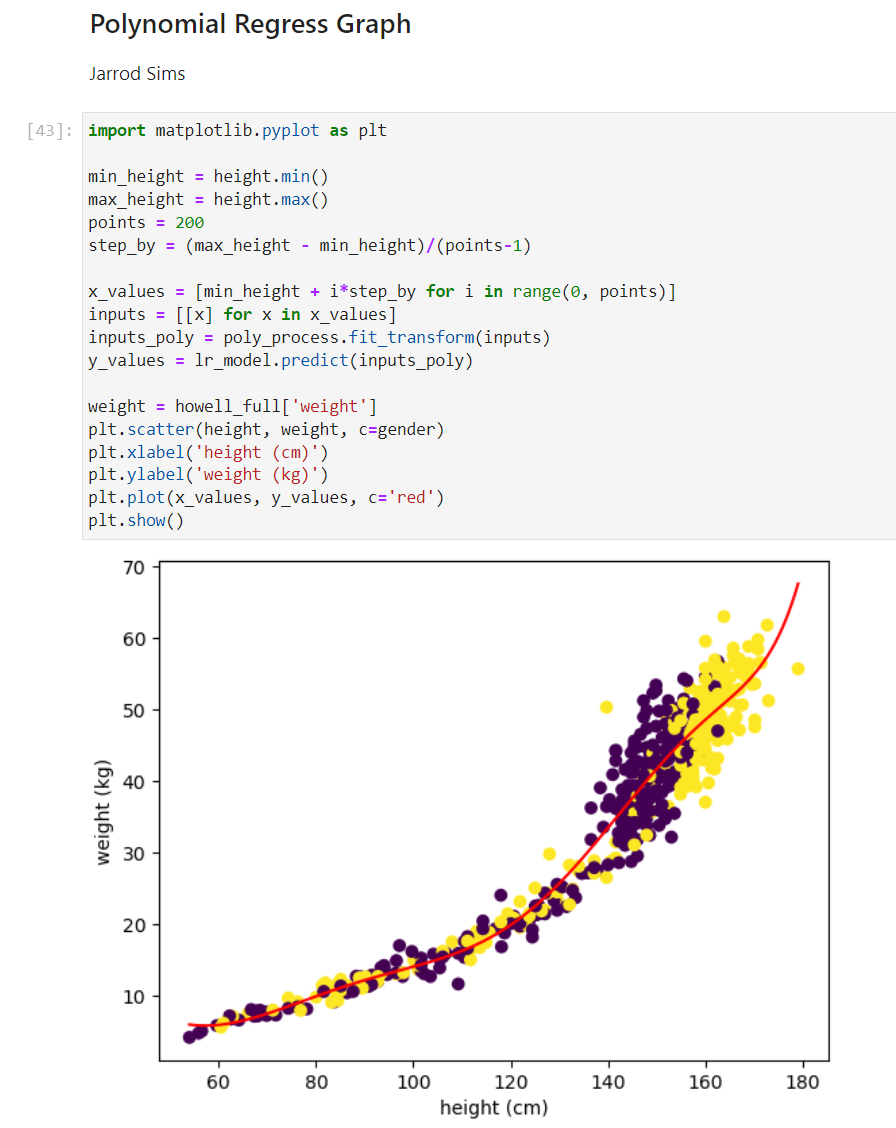
1. Propose an explanation

The polynomial model fits the data much better than the linear model. The polynomial does look like it fits the data better for individuals with a height above around 140cm.

## Submission 8 - Explain

1. Does the polynomial fit do better?
2. Where does it fit the best?

## Progress 9 - Screen shot of scatter plot with degree 8 polynomial fit



## Submission 10 - Explain

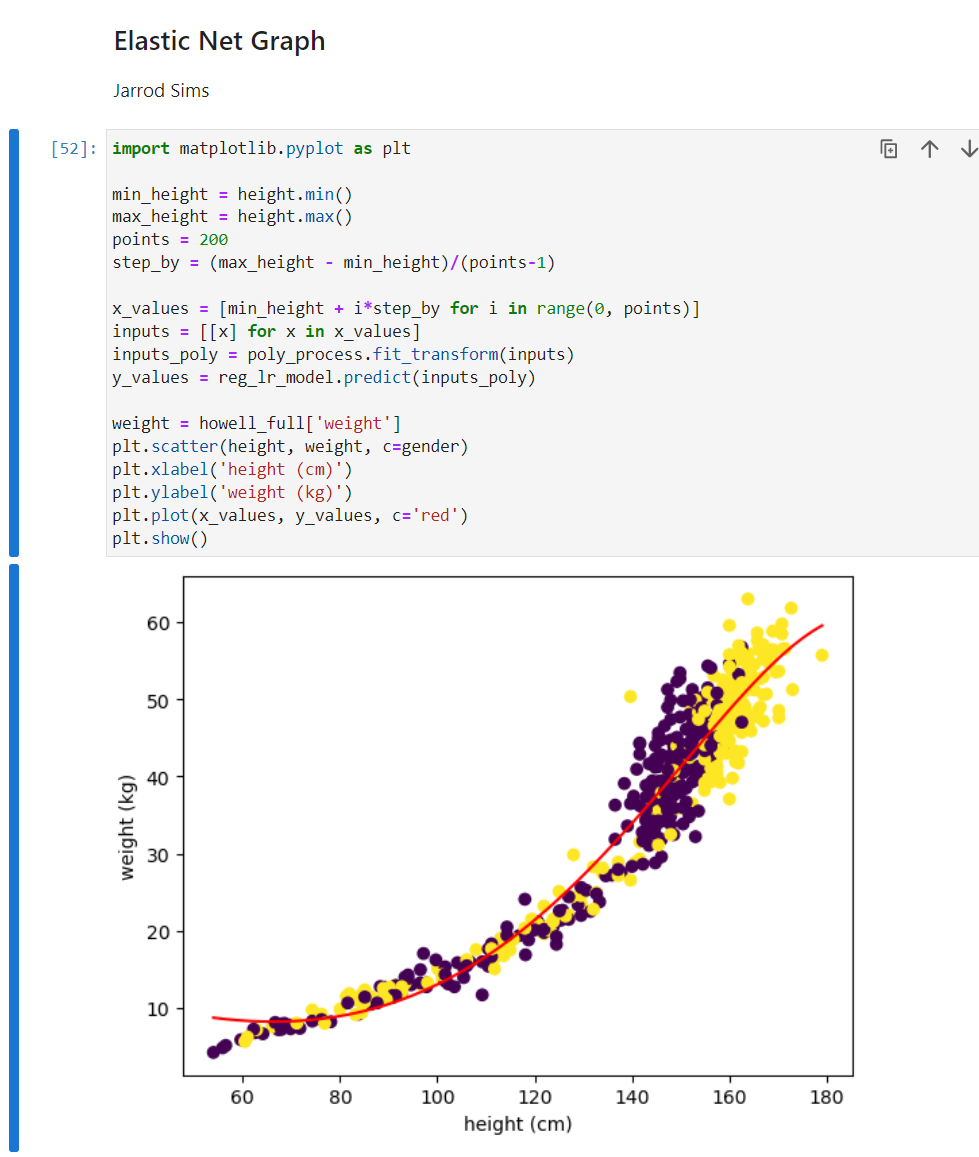
1. Compare degree 3 and degree 8 fits

The 8-degree polynomial model fits the data much better than the 3-degree model. The r2 value is better and the model has less error for both the training and the test set.

1. Is the increase in performance big enough to justify the degree 8 fit?

While the 8-degree model fits the data significantly better than the 3-degree model is possible that this model is overfitting the data.

## Progress 11 - Screen shot of scatter plot with degree 8 Elastic net.



## Submission 12- Explain

1. Compare coefficients for regular and elastic net for degree 8.

|  |  |  |
| --- | --- | --- |
|  | 8-Degree Polynomial | 8-Degree Elastic Net |
| Bias | 30.623665706183964 | 18.846113403922274 |
| Linear coefficient | -1.05821512e-06 | -2.55756771e-01 |
| Quadratic coefficient | -5.84608126e-05 | 6.87204298e-04 |
| Cubic coefficient | -1.70146004e-03 | 8.52528900e-06 |
| Quartic coefficient | 6.23656892e-05 | 3.67563933e-08 |
| Degree 5 coefficient | -9.32819301e-07 | 1.17760939e-10 |
| Degree 6 coefficient | 7.00941575e-09 | 1.01431642e-13 |
| Degree 7 coefficient | -2.62048570e-11 | -2.55533600e-15 |
| Degree 8 coefficient | 3.88233282e-14 | -3.04757085e-17 |

1. Which ones have been reduced in elastic net?

Bias, cubic coefficient, quartic coefficient, degree 5 coefficient, degree 6 coefficient, degree 7 coefficient, degree 8 coefficient were all reduced in the elastic net model. These coefficients help explain why the regularized elastic net model is less susceptible to overfitting than the polynomial model.