

REPORT

Sahana Kulal

NMAM INSTITUTE OF TECHNOLOGY

Data Summary:

- ✓ The dataset contains 10,000 rows, 3 columns.
- ✓ The columns are Height, Weight and Gender.

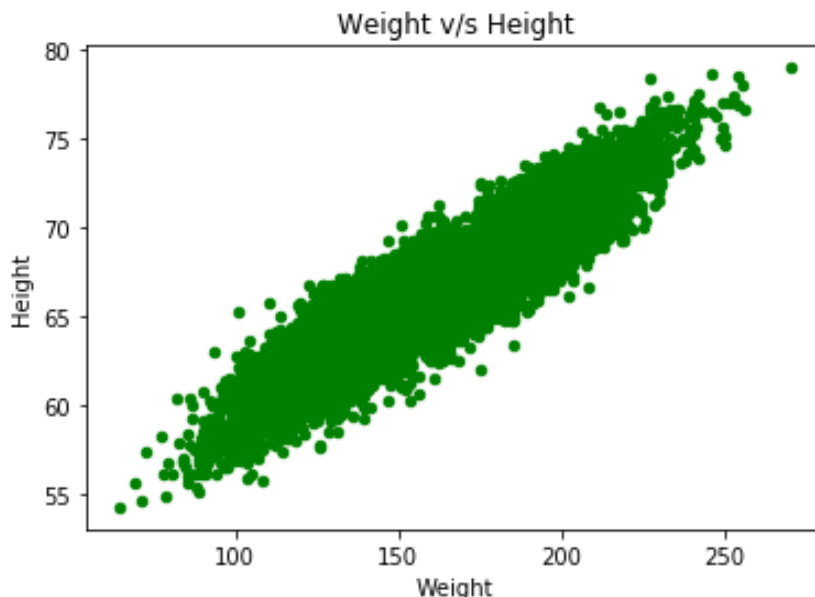
Problem Statement:

Build a machine learning model to predict the height of a person based on the weight and gender.

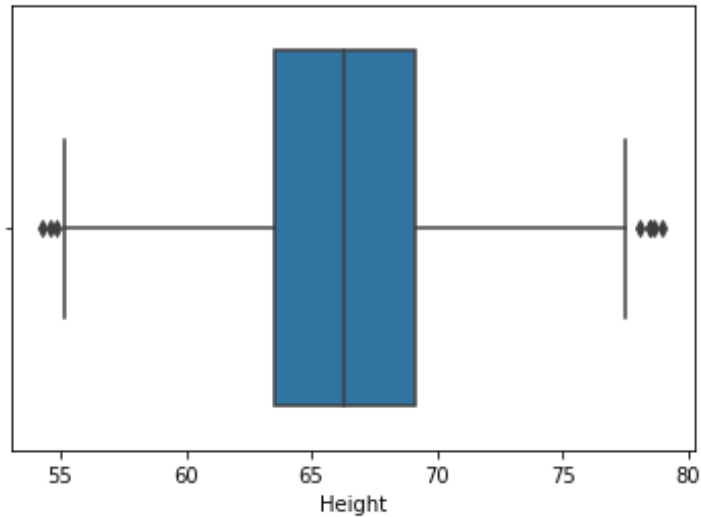
Explanation:

- Plot Weight v/s Height:

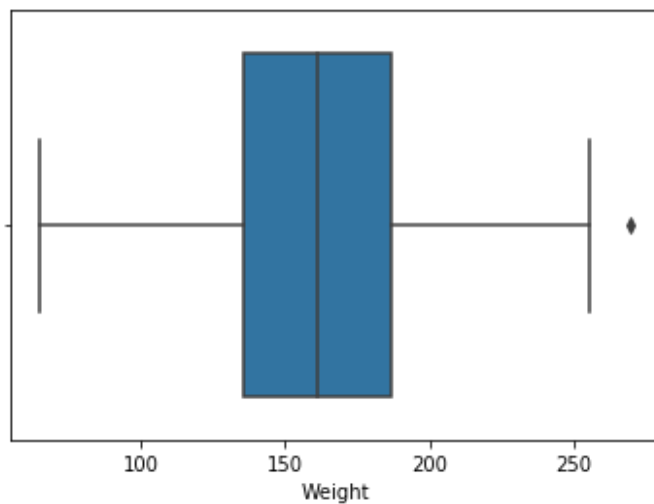
We can infer that as weight **increases** the height also **increases**.



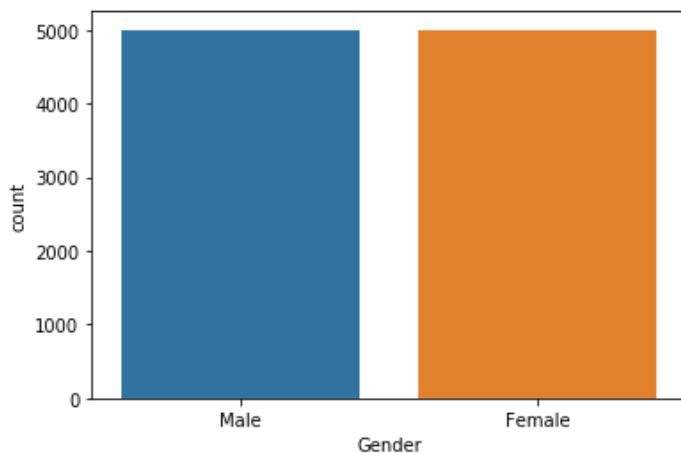
- The boxplot of the height column shows that the value of heights ranges approximately between **63 to 70 inches**.



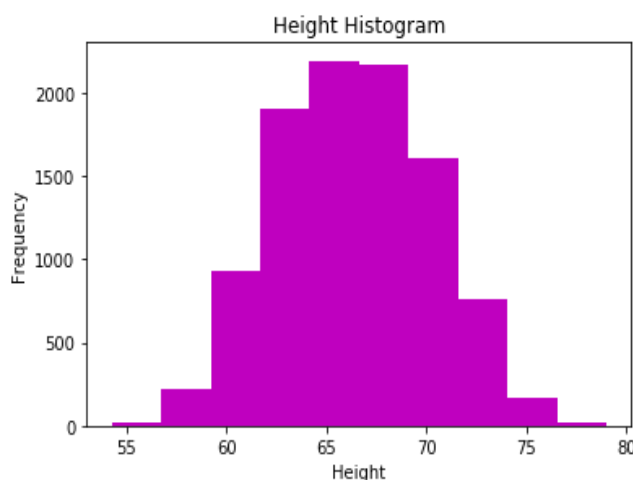
- The boxplot of the weight column shows that the value of weights ranges approximately between **140 to 190 pounds**.



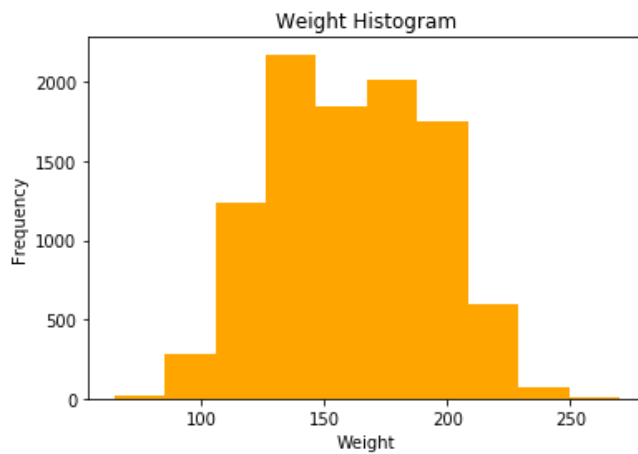
- The dataset contains **equal** number of Males and Females, i.e. 5000 each.



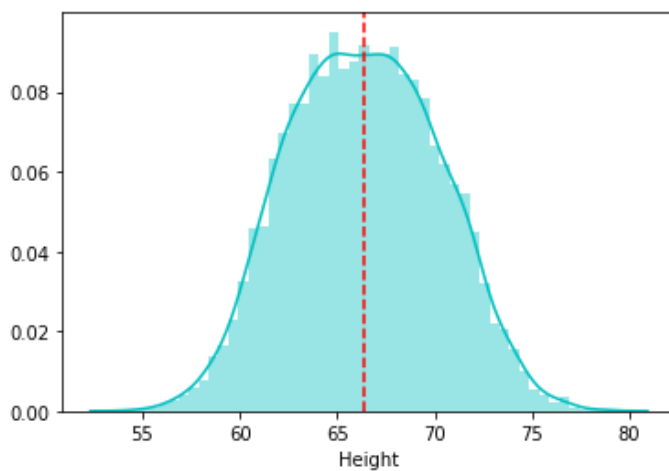
- The Height Histogram shows us that about 2000 people have the height approximately between **65 to 70 inches**.



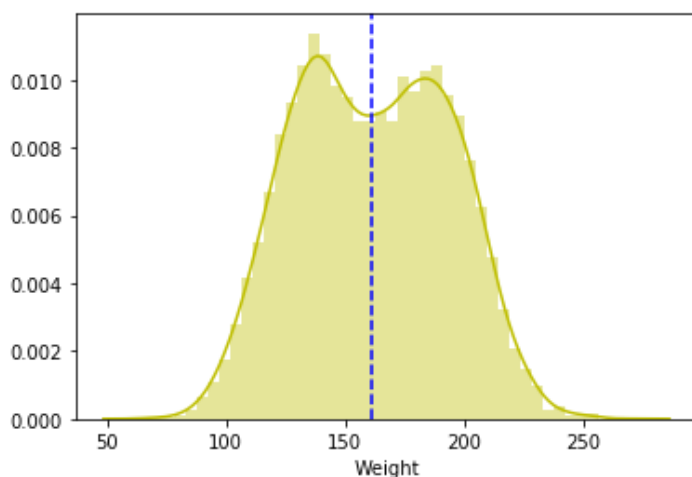
- The Weight Histogram shows us that about **2000** people have the height approximately between **130 to 150 pounds**. And about **1750** people have the weight above **150 and below 200 pounds**.



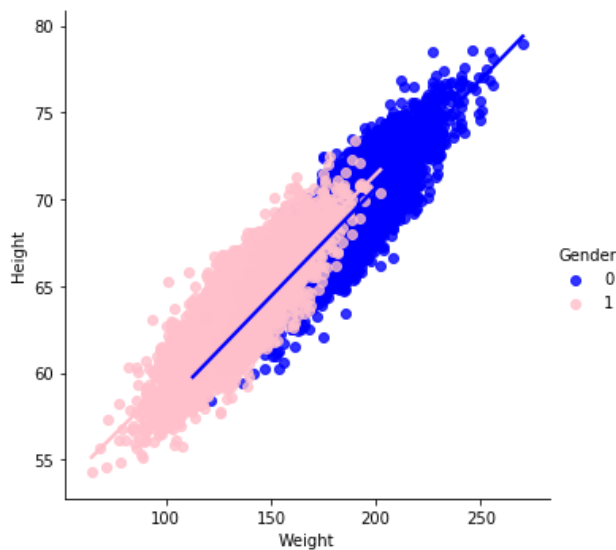
- The distribution plot of the Height values shows that it follows a **bell curve**. The **mean** is roughly around **66 inches**.



- The distribution plot of the Weight values shows us that the **mean** is around **160 pounds**.



- Through the Implot (Linear Model plot) we can see the best fit line for weight v/s height plot for males and females.

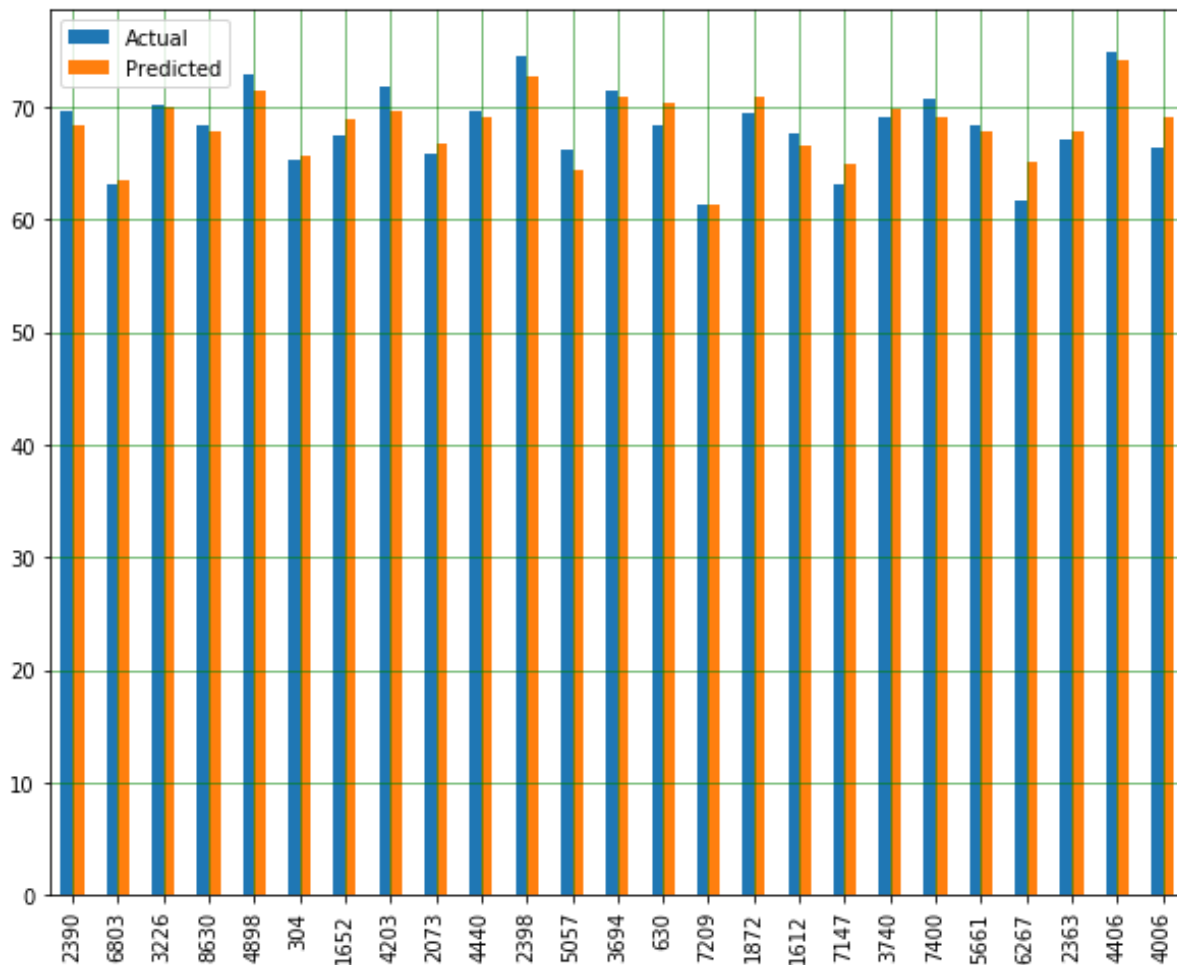


Algorithm:

A **regression problem** is when the output variable is a real or continuous value. The algorithm taken here for the analysis is linear regression, as we can clearly see it's a regression problem.

Accuracy:

| | |
|-------------------------|-------|
| r2_score | 86.19 |
| mean squared error | 2.086 |
| mean absolute error | 1.156 |
| root mean squared error | 1.444 |



In the graph, we can see that the predicted values are nearly equal to the actual values and the error is quite less.

Jupyter notebook link: [GitHub](#)

THANK YOU!

