Vincent Zimmerman

Dr. Butler

DSC 140S M.W.F 10-11:05

1 December 2024

NFL Statistics Report

For this project I analyzed the Basic Stats data set. This data set was comprised of data and information of NFL players. This data is comprised of columns such as Age, current team, birthplace, college, Height, Weight, birthday, etc. To give you some context on ti why I chose this data set is because I play football at Mount Here and consider myself an undersized athlete. So, I wanted to see just how undersized I was. In this data set I analyzed each bit pf data comparing the accuracy of the data set, the correlation between certain variables, and answered questions. The two-research question I had that I wanted to answer were 1. Where is majority NFL players born? 2. What is the average height (inches) and weight (lbs) of NFL athletes.

First, I started off by simply printing the data head which is just shows us 5 rows in each column, so 5 weights, 5 ages etc. I did this just to be able to visualize and make sure the data is okay and looks fine off a quick glance. But, I also did this to get all the data columns on my google colab. Next, I condensed my search by using stats.desscribe() and print (descriptive\_stats) to get the mean avg count min and max of each column of data. I then jumped into finding the relevance between the Birthplace of NFL players and their current team.

A close-up of a chart

Description automatically generated

This is the scatterplot comparing the Birth Place vs Current Team numbers.

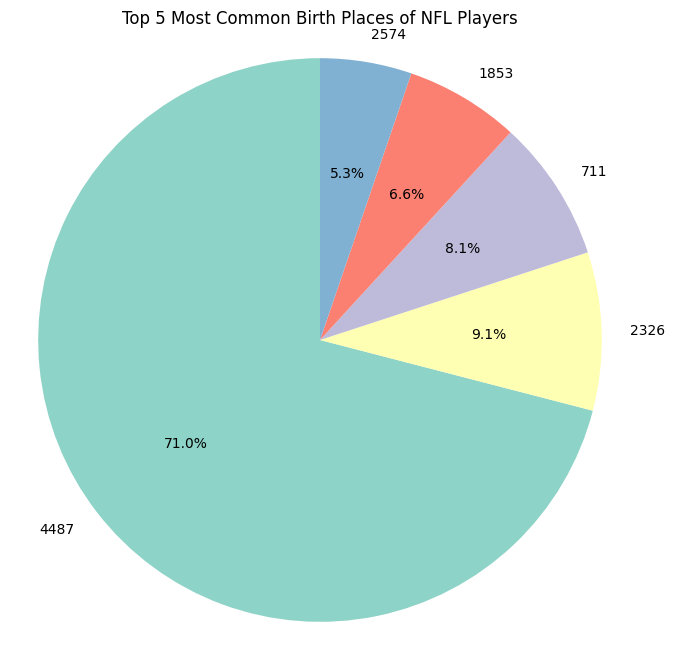
Secondly, I moved on to using the pearson\_corr function to determine the correlation between the current team column and the birthplace column. This determines the correlation these two data sets and how much they directly affect one another. I then preformed another test to determine the accuracy of these two columns this is he KS Test. This KS value was .99668 which is close to one but the p-value is 0 which shows there is no correlation between these two values. I also got a p-value when calculating my chi squared test which again proved zero correlation.

Lastly, I used x\_train y train and x and y test to find the accuracy through a linear regression. I predicted the accuracy of these two columns used label\_encoder and import logistic regression from sklearn to find that these two columns has an accuracy of .81.

Research Question 1: Where is the majority of NFL players born?

I printed the birthplace of the NFL players just to get familiar with the places from these players. I think used the print(f) tool to print the most common birthplace based on the highest number of people born there. And found the most players were born in 4487 with 2387 players.

The Chart below shows the top 5 most common birth places among nfl players.

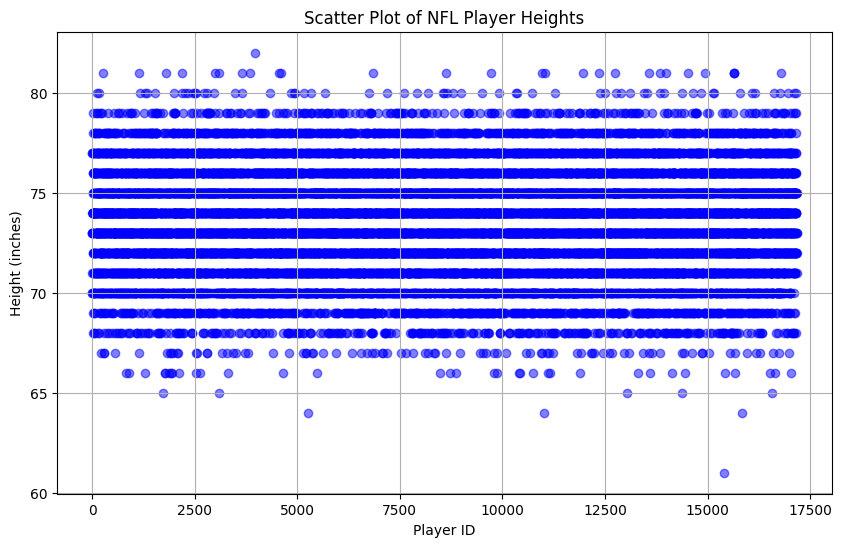


This chart directly shows that the most nfl players were born in 4487.

Research Question 2: What is the Average Weight and Height of NFL Players?

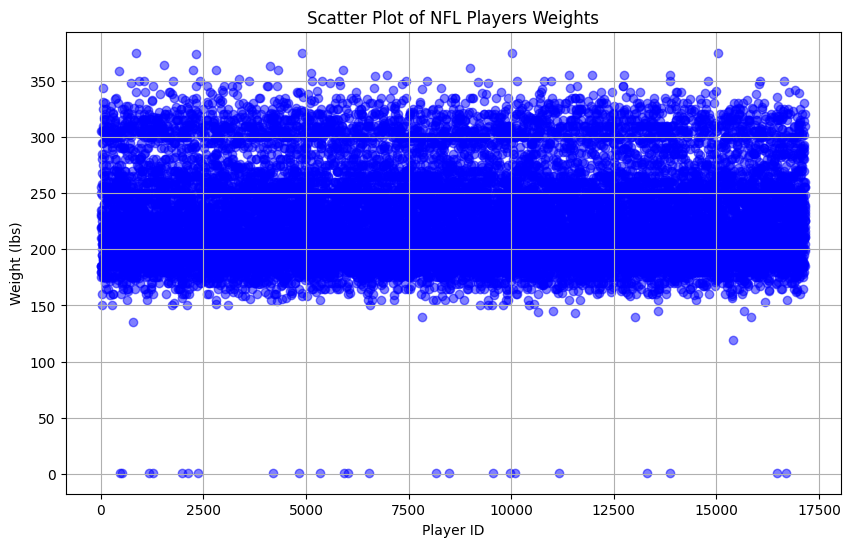
To start off my second research question, I used avg\_ height and weight mean to find the avg height and weight of NFL players. I found the average Height of NFL players was 73.51 inches and the average weight was 229.22 pounds.

Secondly, I wanted to visualize show this so I created a scatter plots for the NFL players heights and weights.



From this data you can see that everyone of these points is really 68 inches and above. While the average and most is around 73-74 inches.

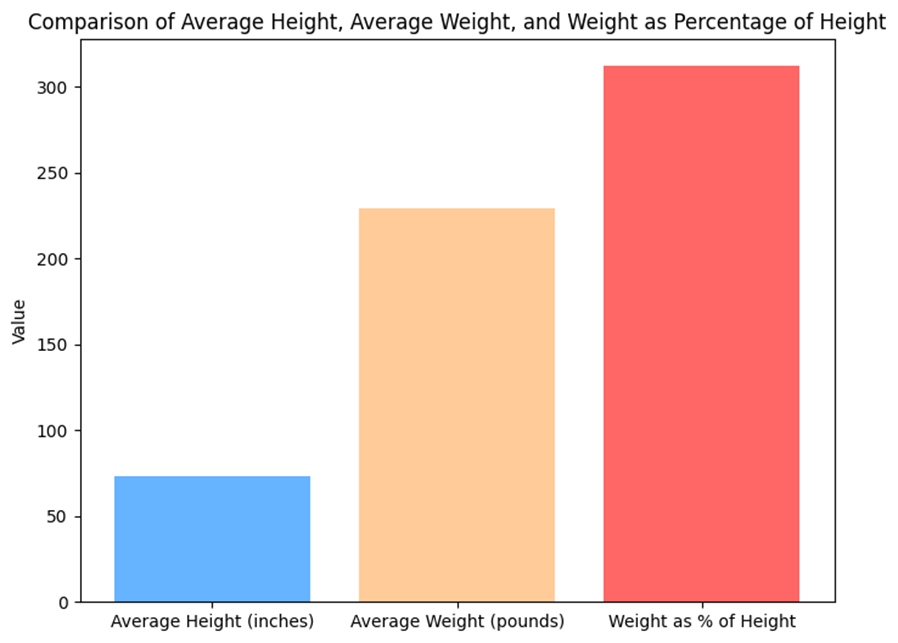
Next, the average weight of NFL players is also shown through a similar scatterplot:



The majority of the points are above 170LBS with the average being around 230LBS. There are a few outliers, but these are NAN or people who have been below the average significantly.

Lastly, I wanted to compare the average weight/ the average height and multiply it by 100 to find the the weight as a percentage of Height. I did this by doing percentage\_weight\_vs\_height = (average\_weight / average\_height) \* 100.

Then I plotted this using a bar graph by using plt.figure to get this



As you can see the weight as percentage of height is far more signification sitting around 311.83%

Ultimately, I really enjoyed analyzing and researching this data set. From my findings the average height of an NFL player is around 74 inches while the average weight is around 230LBS. Also, most NFL players are from 4887.