**CS 240 Final Project Report**

**Amer Nour Eddin | 213171245**

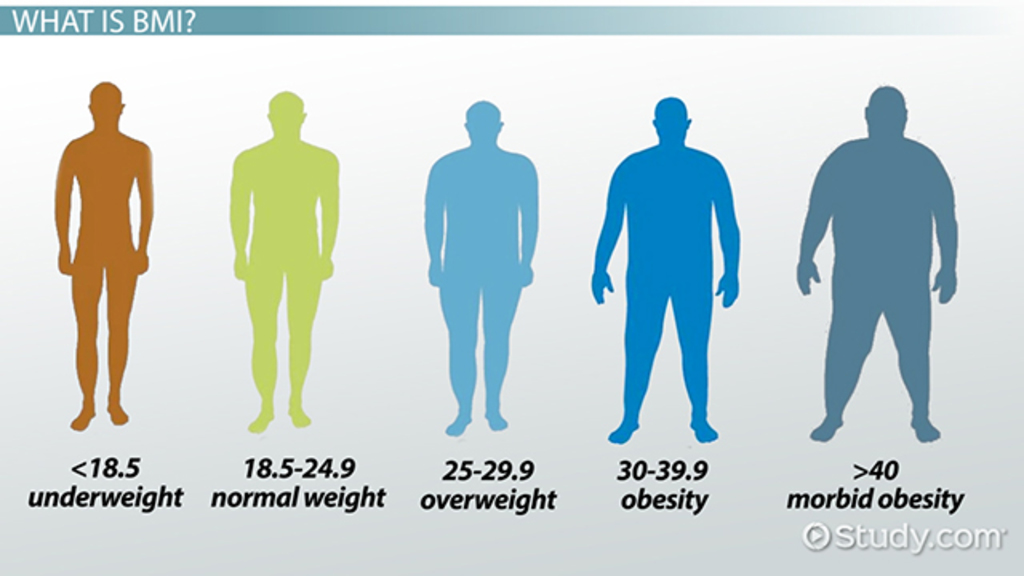
PART 1

I choose the data of the baseball players (Master.csv) which contains a general information about all the players like the (name, Id, birthdate place, weight, height etc.).

Also I choose the (Salaries.csv) data which contains the players Ids and each player salary throughout the years.

My intention is the following:

I will calculate the BMI (Body Mass Index) for each player, which is calculated using the player’s weight and height 🡪 *BMI = weight (lb.) / (height (in)) 2 x 703*

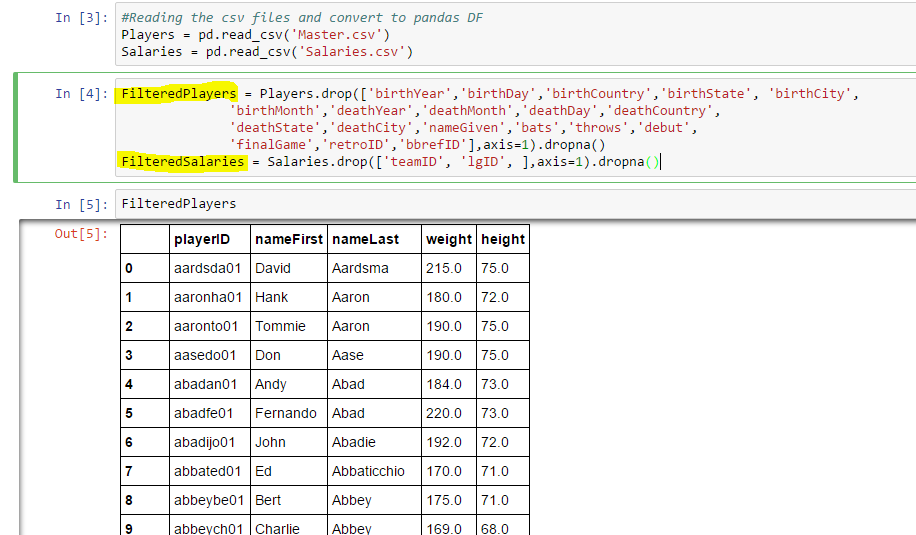


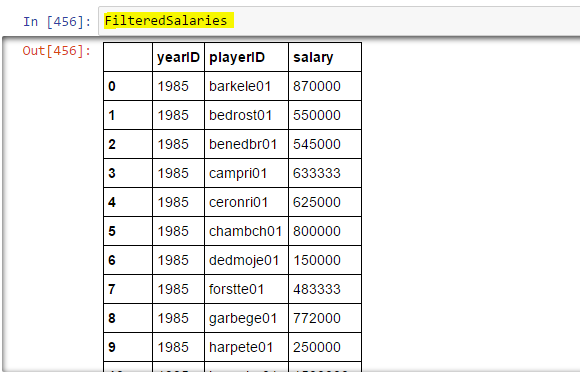
My question is there any relation between the player’s BMI and his average salary throughout the years?

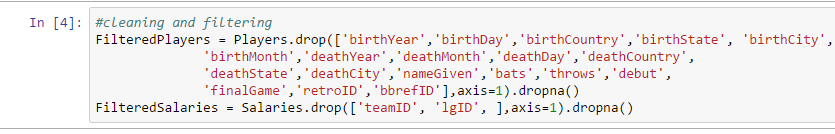
*“If the player’s BMI value is HIGHER than 26 then his salary value will be larger than other players”*

PART 2

These are the two main columns that I used:



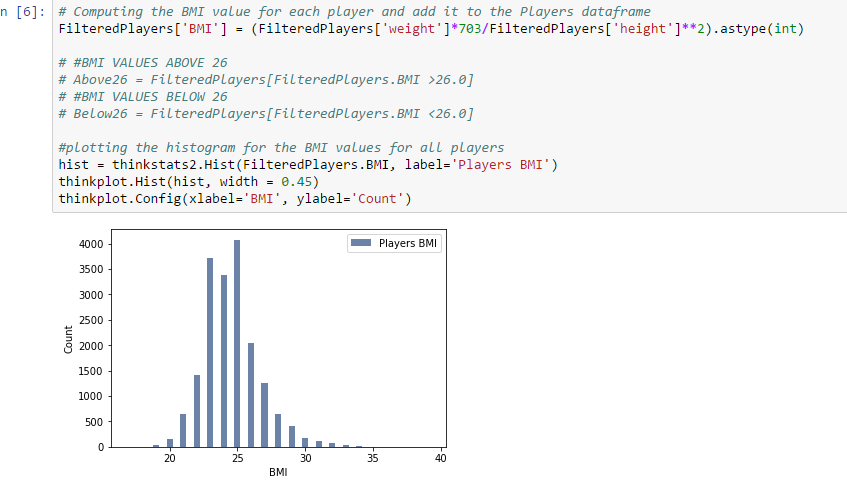




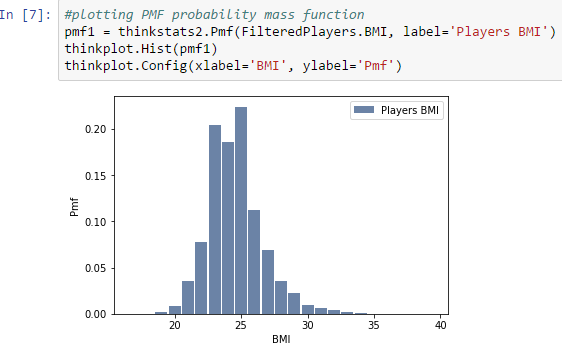
PART 3

I computed the BMI values and add it to the players data frame as the following…

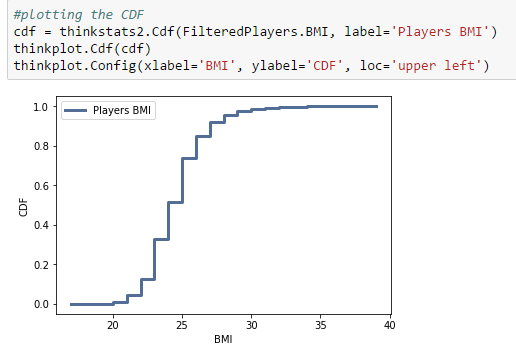
Also I made the histogram chart for the BMI values.



Probability Mass Function (PMF)



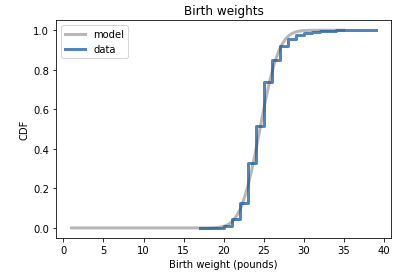
Cumulative Distribution Function (CDF)

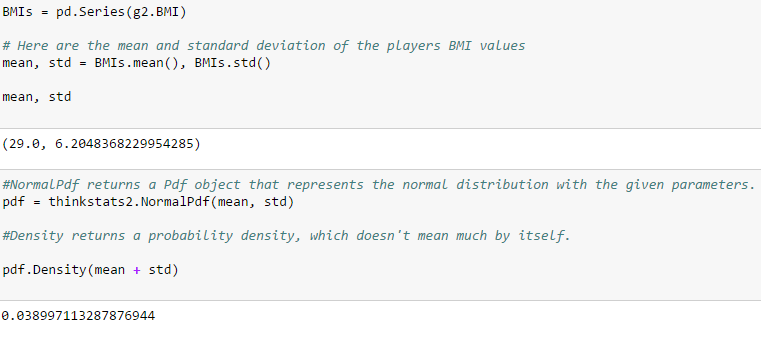


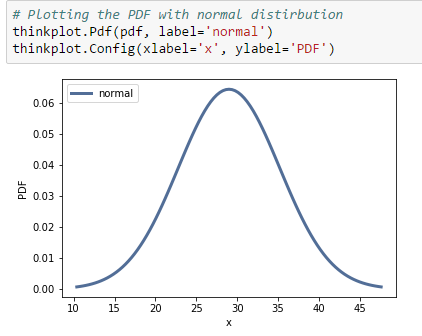
PART 4

I used the Normal distribution for modeling as I saw that it is the most that fits the data shape



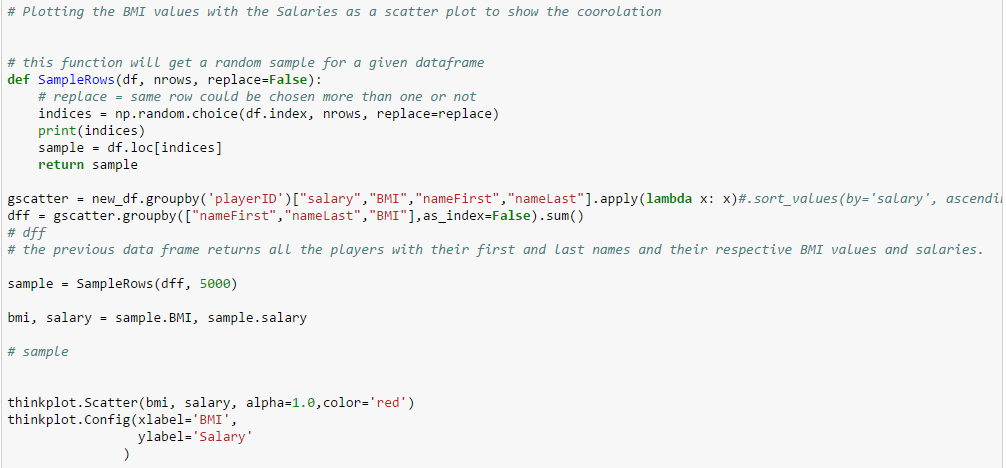


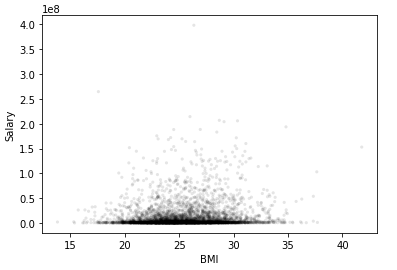
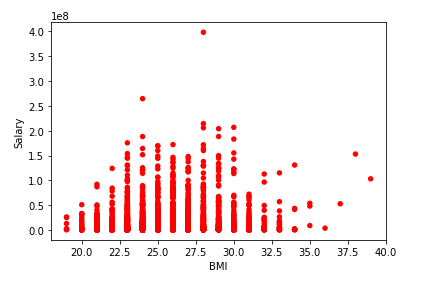




PART 5

I planned to show the correlation between 2 variables which are the BMI values and the salaries of the players and I used the scatter plot in order to visualize this correlation.



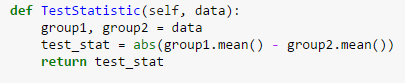


PART 6

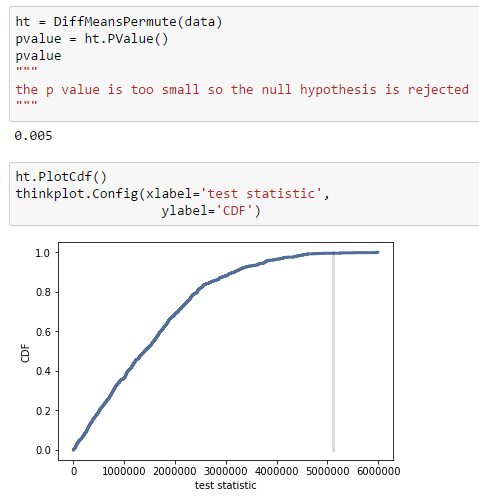
Here I tested my hypothesis that “the players with BMI above 26 have higher salary than others”

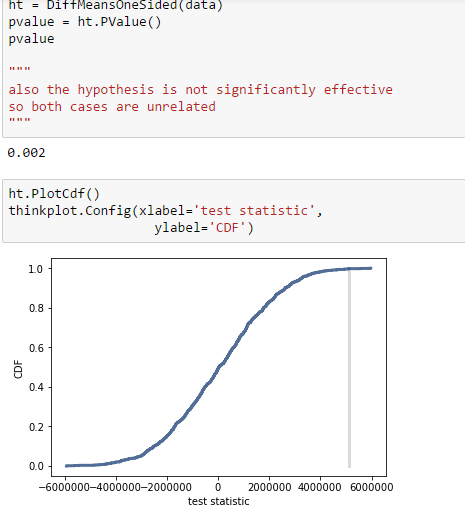
Using the hypothesis testing methods provided by thinkstats2.

My test statistic was taking the average of the 2 groups (salaries with BMI above 26 and below 26) and taking the absolute difference



The results was as the following:





PART 7

My final conclusion was:

There is no real relation or cause and effect in my proposed hypothesis, so I couldn’t prove that players with BMI higher than 26 will have higher salary.

And I’ve done the following to show that:

