Team Machanus

DSC 423 Project: Final Report

Team Members:
Srinilay Jalagam
Hasmitha Chatla
Alyssa Robinson
Endri Rama

Introduction

Machanus, is a team of data scientists which was set up by their CEO Jonathan Gemmel, in order to investigate how video game statistics could actually be of use in real life. Machanus the name originates from the game Dungeons and Dragons and it operates on a strict schedule that is planned, measured, and controlled perfectly. In this report, the Data Scientists will reveal through the process of regression as to how one can use video game data, in order to predict how it could correlate with real life. The data which is analyzed and modeled in this report correlates with FIFA. The data scientists in this report analyze what real-life players have to maintain in order to keep themselves from retiring from the game, and what variables are relevant in maintaining a good contestant weekly wage.

Data Selection & Data Preparation

The dataset we've chosen for this project is "Fifa 2020", Fifa which is produced by EA Sport for over 20 years, it is known to be one of the largest sports video games amongst PS4 and Xbox. The main focus of this project will be to predict what are the key factors of a player's weekly salary, and then further determine what key factors would also be relevant for soccer players which would like to continue to play the beautiful game as they get older. The dataset which we (Data Scientists) have chosen already includes the player age, preferred position overall, potential, value in euro, weekly wages in euro, common positions the player plays, release clause in euro, team position, their attributes, and the potential growth of the player for all positions available on the field, with their respective formations. In order for this to be executed properly, we would have to determine what are independent variables and what would be the response variable. The FIFA 20 data set has been gathered from the website www.kaggle.com, the source is down below:
https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset?select=players_20.csv. In this dataset, there are 18,278 professional football players and 104 variables. We will pre-process the raw data and consider only a few variables which are useful to create better models. From the raw dataset, we have taken the weekly wage of the players as the response variable or dependent variable and other variables such as value_euro, overall, potential, name, age, nationality, club, height, weight, etc. as explanatory variables. One of the challenges we had is the "NA" values in the data or rather missing data, and how to deal with it. This may be due to the player not having a particular attribute or never playing a particular position, and it will most likely throw off the analysis. Another challenge we had was deciding what variables were truly useful.

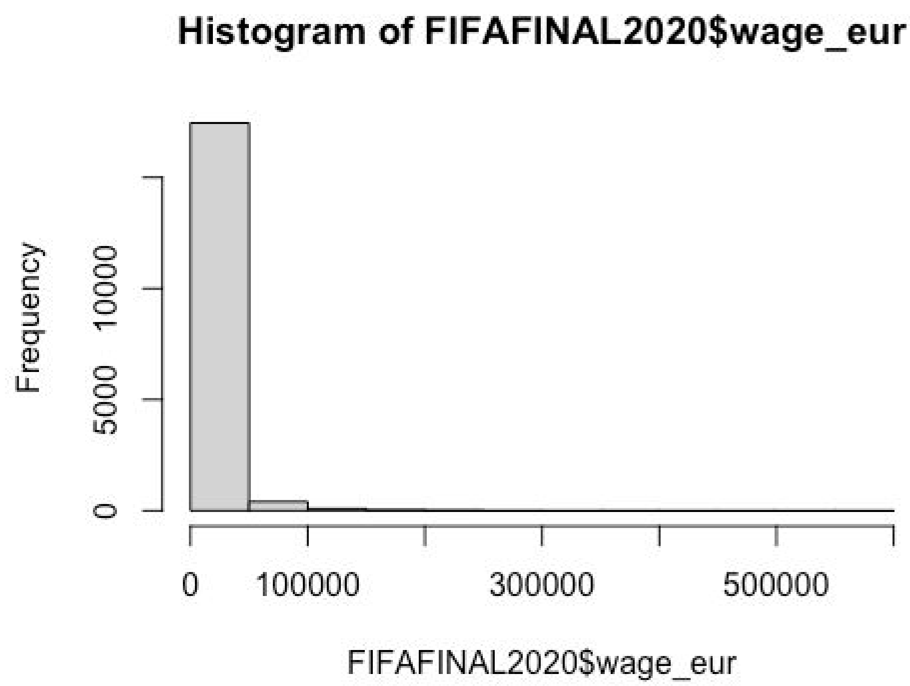
We have pre-processed the data and removed columns with “NA” values as well as rows with “NA” value_euro variables. The total number of entries removed was 250 entries. This brought our data set to a total of 18,028 entries to work with. The files are available in the following Github repository: https://github.com/ERAMA237/Team-Machanus. We have categorized the positions in four categories: Forwards, Midfielder, Defenders, and Goalkeepers. These CSV’s can also be accessed in the link above. Our final independent variables are down as listed, value_eur, age, height_cm, weight_kg, overall, potential, pace, shooting, passing, dribbling, defending, physic, attacking_crossing, attacking_finishing, attacking_heading_accuracy, attacking_short_passing, attacking_volleys, skill_dribbling, skill_curve, skill_fk_accuracy, skill_long_passing, skill_ball_control, movement_acceleration, movement_sprint_speed, movement_agility, movement_reactions, movement_balance, power_shot_power, power_jumping, power_stamina, power_strength, power_long_shots, mentality_aggression, mentality_interceptions, mentality_positioning, mentality_vision, mentality_penalties, mentality_composure, defending_marking, defending_standing_tackle, defending_sliding_tackle, gk_diving, gk_handling, gk_reflexes, gk_speed, gk_positioning, and goalkeeping_kicking. These variables will be used to predict our response variable, wage_eur, in order to see what truly causes change to the salary and to further hypothesize what variables a player may need in order to maintain a position on the field rather than end their career and retire.

Data Analysis

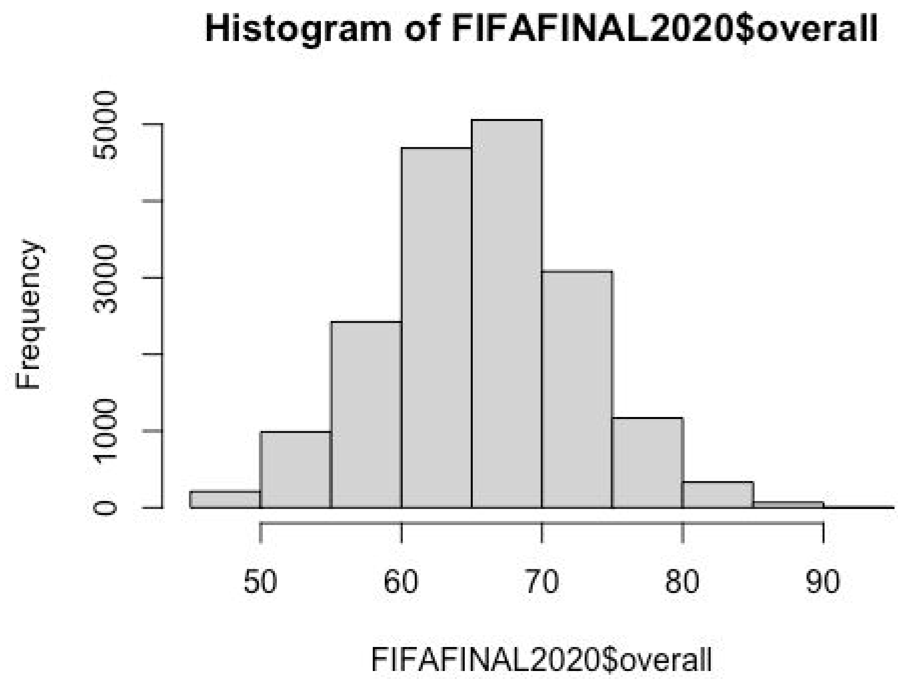
Now we will be going over some of our analysis of the final data we have accumulated. We are doing this in order to determine the best response variable, as well as analyze some of our independent variables.

First, we will analyze the variables we deem relevant down below.

1. Histogram of Wage_Eur (weekly salary aka response variable).
    a. As we can see this would make perfect sense to make this variable our response variable, this is because of the direptency in the distribution of this variable.

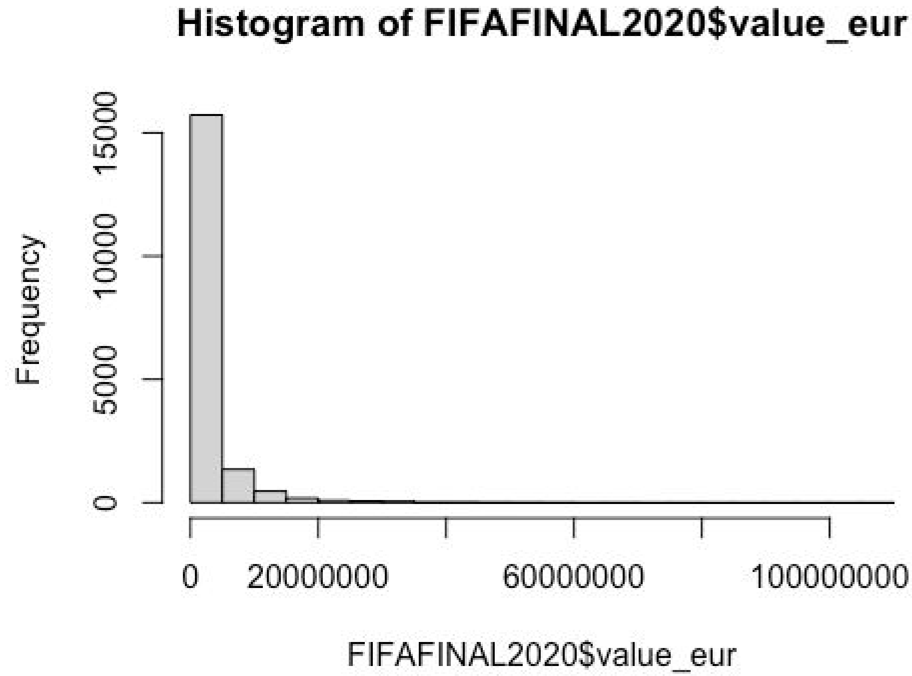


    b.
2. Frequency of Overall of the Players
    a. As we can see, the overalls of the players for the most part are normally distributed throughout the data. This means that even though the overalls for these players are completely random due to how well they play in their games, we are still dealing with a bell-shaped curve which is good



b.

3. Frequency of the Value_euro (market value) of the Players
    a. This is also very disproportionate as due to the market value of the players, now, at first sight, one might say that this is not a great variable to have within our model, but it is as per our mission being to predict the weekly salary and the determination of whether or not the players should retire or not. How is this a good predictor, well we expect this to play a key role within the project?



b.

4. Five Number Summary of Wage_Euro
   a. We have analyzed the five-number summary of the response variable Wage_euro (weekly salary) and we can see that the minimum a soccer player is making is $1,000 per week, we can also see that the median of the weekly salary is $3,000 dollars per week, we can see that the mean is $9,585 per week and that the high-value players are making $565,000 per week, with the first quartile being $1,000 and the third quartile being $8,000. This further proves that what we observed from the histogram of the wage_eur, this is because this data set does not rely on only one specific league but rather all of the leagues within the Fifa 20 dataset.
   b.
```
> summary(FIFAFINAL2020$wage_eur)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   1000    1000    3000    9585    8000  565000 
```

5. Five Number Summary of Overall
   a. We have analyzed the independent variable Overall, to evaluate the spread of the overalls of the soccer players. We can see that the minimum overall in all of the Fifa 20 soccer data is 48, the first quartile being 62, the median being 66, the mean being between 66 and 67 aka 66.2 which is not something you will see on an overall card. The third quartile is 71 and the maximum overall which is held by Lionel Messi being 94. This is

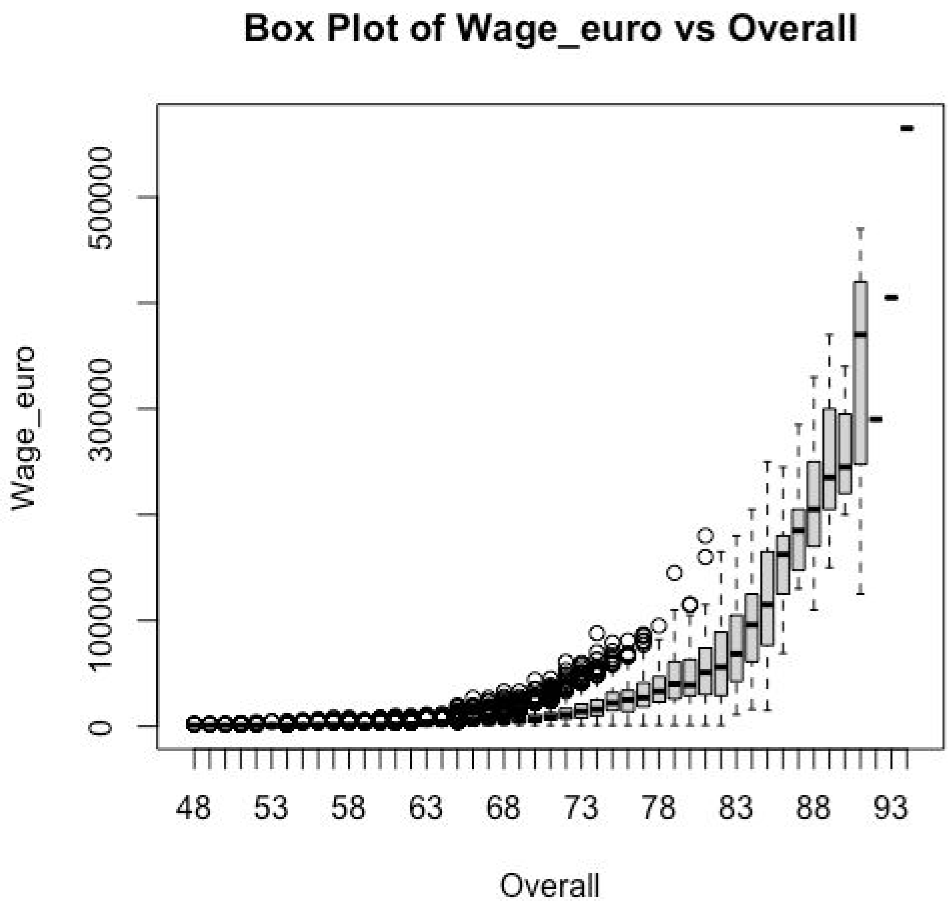
important as it explains the Value_Euro (market value) and Wage_eur (weekly salary) of the soccer players.

b.
```
> summary(FIFAFINAL2020$overall)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  48.00   62.00   66.00   66.21   71.00   94.00 
```

6. Five Number Summary of Value_Euro
   a. We have analyzed the market value of the soccer player five-number summary. We can see that the minimum market value of a soccer player is $10,000, the first quartile being $325,000, the median being $725,000, the mean being $2,518,485 for soccer players, the third quartile being $2,100,000 and the maximum market value which once again is held by Lionel Messi being 105,500,000. This tells us what the current market values of players are and how they are distributed, which once again explains the histogram for Value_Euro above.

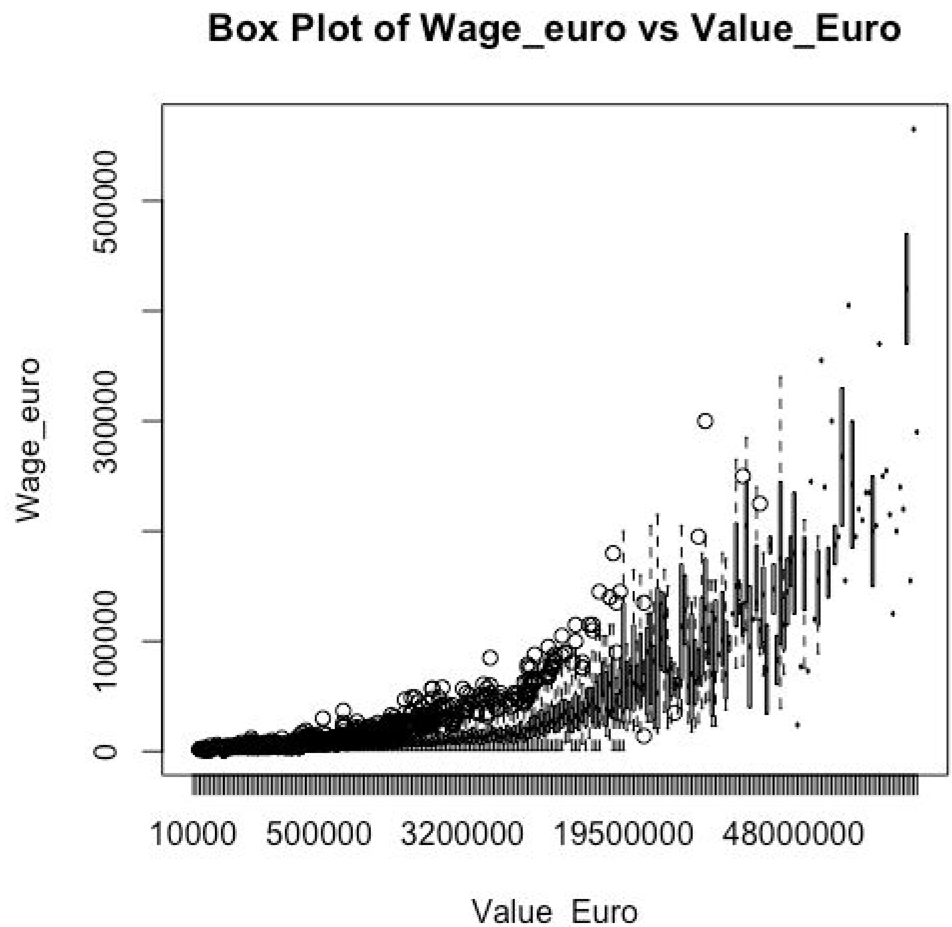
   b.
```
> summary(FIFAFINAL2020$value_eur)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
    10000    325000    725000   2518485   2100000 105500000 
```

7. Box Plot of Wage_Euro vs Overall
   a. We have analyzed how the Wage_Euro (Weekly Salary) stands with the Overalls, what we have noticed is that even though a player may have an average Overall such as 77 or 82 they also have some outliers to the weekly salaries, we believe this may be due to the Value_Euro (market value) of the soccer players at hand.



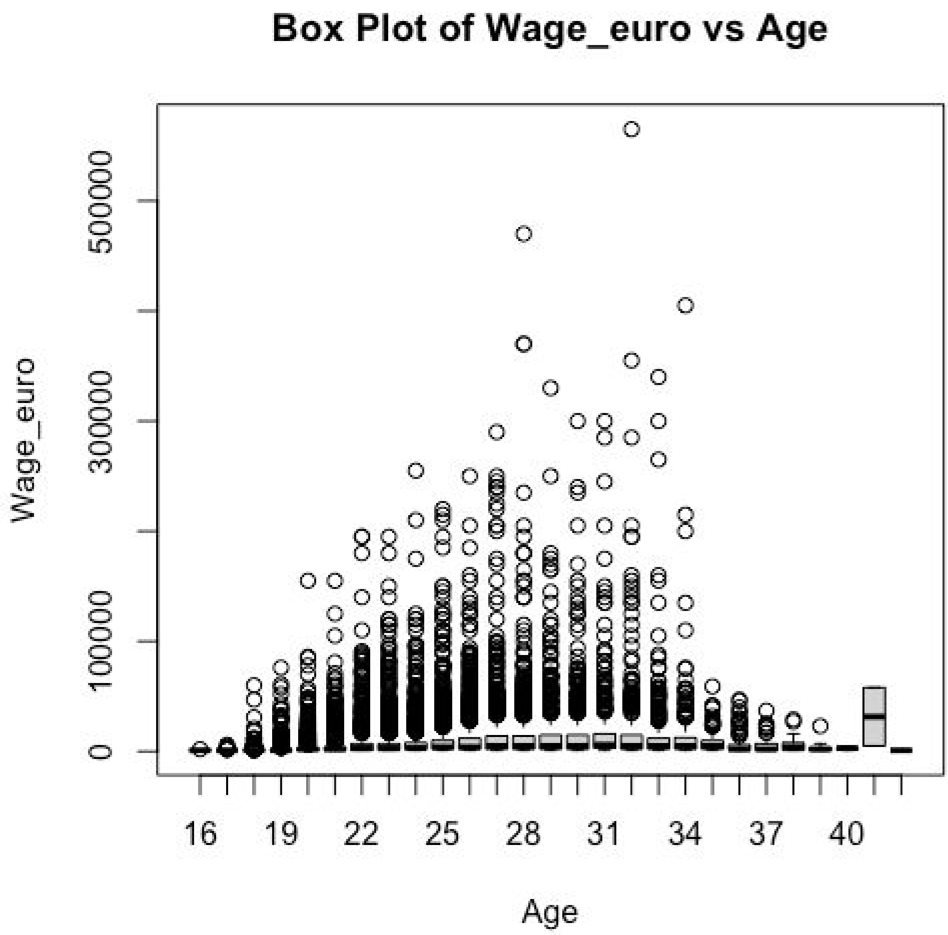
b.

8. BoxPlot of Wage_Euro vs Value_Euro
    a. We have analyzed how Wage_Euro (Weekly Salary) stands with Value_Euro (Market Value), our assumptions are correct from the above, that the market value of a soccer player (Value_Eur) does play a key role in the project when predicting the weekly salary (Wage_Euro). We can see that even players with high market values have outliers in the weekly salaries, and this is due to the Value_Euro (market value) of the soccer players.



b.

9. BoxPlot of Wage_Euro vs Age
    a. We have analyzed how Wage_Euro (Weekly Salary) stands with Age of the player, we have noticed that age also plays a key role with respect to the other independent variables when predicting the Wage_Euro (Weekly Salary) of a soccer player. This proves our second assumption correct that even with age, the players have to maintain a high market value in order to maintain or improve their weekly salary rather than retire.


Box Plot of Wage_euro vs Age
Wage_euro
500000
300000
100000
0
16
19
22
25
28
31
34
37
40
Age

b.

Proposal

Team Machanus proposed that in order for the models to be nearly perfect for this project we will split the data based of the players positions down below, we have also based our models off predicting the wage_euro and finalize our assumptions about what a soccer player will have to maintain as they get older, with the years to come, in order to maintain their current weekly salary or improve it, and continue to play the beautiful game rather than end their careers and retire. Data Scientist Srinilay Jalagam has investigated the Forwards portion of the data, Data Scientist Endri Rama has investigated the Midfielders portion of the data, Data Scientist Hasmitha Chatla has investigated the Defenders portion of the data, and Data Scientist Alyssa Robinson has investigated the Goalkeepers portion of the data. Their findings and models are in the following pages, in their own words.

a. Srinilay Jalagam - Forwards

Owing to the success of football, in recent decades there has been a significant rise in demand for professional footballers, and football players salary demands. Determining income for football is an economic complex task in a broad variety of facets. In the pre-information period, the job was performed mainly with statistical research, as the numbers and records of football players have not been gathered regularly and the results and abilities of football players have to be measured in quantitative terms. Football data were gathered and released, however, and the quality of these data constantly increased. In this report, we are using player ability knowledge to quantitatively evaluate the salaries of football players according to their talents and abilities as comparison with other players.

I will be focusing on the forward position of the footballers, by using wage as a dependent variable I will be focusing on to predict the independent variables which affect the wage of the players.

The model below used has the list variables used for the forward positions and we can see that the r square value is 77% which is good value but for many of the variables their p value is more than 0.005 which is not good for the module. after these modules we will be removing variables based on collinearly p value f value

<u>*First Iteration of Model (Full Model)*</u>:

- *w2 <- lm (wage_eur ~., data=. ForwardPositionCombos.)*
- *summary(w2)*

```
Residuals:
    Min      1Q  Median      3Q     Max 
-159704   -2522    -184    2105  235621 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -3.564e+04  1.182e+04  -3.015  0.00259 ** 
age                        8.173e+02  1.037e+02   7.880 4.11e-15 ***
height_cm                  8.980e+01  5.866e+01   1.531  0.12589    
weight_kg                  1.389e+02  5.079e+01   2.734  0.00628 ** 
overall                   -3.806e+02  1.726e+02  -2.205  0.02752 *  
potential                  2.343e+02  7.878e+01   2.974  0.00296 ** 
value_eur                  3.450e-03  3.926e-05  87.889  < 2e-16 ***
pace                      -1.556e+02  6.780e+02  -0.229  0.81853    
shooting                  -1.052e+02  6.849e+02  -0.154  0.87794    
passing                    1.217e+03  6.792e+02   1.792  0.07322 .  
dribbling                  4.598e+02  6.803e+02   0.676  0.49919    
defending                  6.798e+02  6.824e+02   0.996  0.31918    
physic                     6.728e+02  6.765e+02   0.995  0.32001    
attacking_crossing        -2.021e+02  1.387e+02  -1.457  0.14524    
attacking_finishing        2.761e+01  3.142e+02   0.088  0.92999    
attacking_heading_accuracy -3.748e+01  7.416e+01  -0.505  0.61334    
attacking_short_passing   -4.975e+02  2.456e+02  -2.025  0.04291 *  
attacking_volleys         -2.634e+01  4.759e+01  -0.554  0.57988    
skill_dribbling           -2.008e+02  3.466e+02  -0.579  0.56237    
skill_curve               -5.301e+01  4.522e+01  -1.172  0.24121    
skill_fk_accuracy         -9.105e+01  4.084e+01  -2.229  0.02586 *  
skill_long_passing        -1.599e+02  1.071e+02  -1.493  0.13550    
skill_ball_control        -7.924e+01  2.172e+02  -0.365  0.71526    
movement_acceleration      6.340e+01  3.067e+02   0.207  0.83627    
movement_sprint_speed      8.984e+01  3.762e+02   0.239  0.81128    
movement_agility          -2.942e+01  7.657e+01  -0.384  0.70086    
movement_reactions        -1.896e+01  5.863e+01  -0.323  0.74649    
movement_balance           2.847e+01  4.636e+01   0.614  0.53921    
power_shot_power           5.963e+01  1.437e+02   0.415  0.67813    
power_jumping             -2.767e+00  3.987e+01  -0.069  0.94468    
power_stamina             -2.466e+02  1.704e+02  -1.447  0.14805    
power_strength            -3.977e+02  3.395e+02  -1.171  0.24150    
power_long_shots           9.811e+00  1.421e+02   0.069  0.94497    
mentality_aggression      -1.304e+02  1.365e+02  -0.955  0.33949    
mentality_interceptions   -1.344e+02  1.388e+02  -0.968  0.33299    
mentality_positioning     -1.470e+01  6.660e+01  -0.221  0.82527    
mentality_vision          -2.678e+02  1.410e+02  -1.898  0.05772 .  
mentality_penalties        2.464e+01  4.570e+01   0.539  0.58978    
mentality_composure       -4.648e+01  4.006e+01  -1.160  0.24600    
defending_marking         -2.415e+02  2.057e+02  -1.174  0.24051    
defending_standing_tackle -1.448e+02  2.070e+02  -0.699  0.48429    
defending_sliding_tackle  -9.478e+01  7.812e+01  -1.213  0.22507    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13030 on 4390 degrees of freedom
Multiple R-squared:  0.7767,	Adjusted R-squared:  0.7746 
F-statistic: 372.3 on 41 and 4390 DF,  p-value: < 2.2e-16
```

BACKWARD ELIMINATION

```
Step:  AIC=83992.53
wage_eur ~ age + height_cm + weight_kg + overall + potential +
    value_eur + passing + defending + physic + attacking_short_passing +
    skill_fk_accuracy + movement_balance + power_stamina + power_strength +
    mentality_aggression + mentality_vision + defending_marking

                          Df  Sum of Sq        RSS   AIC
<none>                                  7.4740e+11 83993
- height_cm                1 4.0972e+08 7.4781e+11 83993
- defending                1 4.1692e+08 7.4782e+11 83993
- mentality_vision         1 5.3214e+08 7.4793e+11 83994
- movement_balance         1 6.0658e+08 7.4801e+11 83994
- mentality_aggression     1 6.2308e+08 7.4802e+11 83994
- defending_marking        1 7.1610e+08 7.4812e+11 83995
- physic                   1 7.2445e+08 7.4812e+11 83995
- skill_fk_accuracy        1 7.3178e+08 7.4813e+11 83995
- attacking_short_passing  1 8.0095e+08 7.4820e+11 83995
- power_strength           1 9.3687e+08 7.4834e+11 83996
- passing                  1 1.0396e+09 7.4844e+11 83997
- weight_kg                1 1.3474e+09 7.4875e+11 83999
- power_stamina            1 1.4165e+09 7.4882e+11 83999
- potential                1 1.4756e+09 7.4888e+11 83999
- overall                  1 1.8694e+09 7.4927e+11 84002
- age                      1 1.1133e+10 7.5853e+11 84056
- value_eur                1 1.3444e+12 2.0918e+12 88552
```

First Backward Elimination:

- *backw <- stepAIC(w2,direction = "backward")*

After performing the backward elimination, we can see that few variables are removed based on these models. even though few variables are removed we need perform backward elimination as long as we can get a model which has good p value collinearity

MULTICOLLINEARITY

First Multicollinearity Check:

- *vif(backw)*

```
> vif(backw)
                age             height_cm             weight_kg               overall             potential
           4.797011              3.768720              3.188974             12.137884              5.879212
          value_eur               passing             defending                physic attacking_short_passing
           2.019077             21.566399              4.417215            277.582672             10.942582
  skill_fk_accuracy      movement_balance         power_stamina        power_strength   mentality_aggression
           2.715056              3.194881             22.778797            132.972169             28.047333
   mentality_vision     defending_marking
           5.432465              3.342551
```

VIF is used to perform multicollinearity on the model and we can see that the collinearity of the variables is not good. we can determine the multicollinearity value if it is good enough by checking the variables value, if it is more than 10 then the variables should be removed. After Multicollinearity the variables are removed with high multicollinearity and the second model is created after multicollinearity variables are removed. after these for the next step, backward elimination is applied to the second model

BACKWARD ELIMINATION FOR SECOND MODEL

```
Step:  AIC=84036.97
wage_eur ~ age + value_eur + height_cm + defending_marking +
    mentality_vision + movement_balance

                    Df  Sum of Sq        RSS   AIC
<none>                             7.5869e+11 84037
- height_cm          1 3.8150e+08 7.5907e+11 84037
- movement_balance   1 4.4211e+08 7.5913e+11 84038
- defending_marking  1 6.4112e+08 7.5933e+11 84039
- mentality_vision   1 1.3405e+09 7.6003e+11 84043
- age                1 1.5457e+10 7.7415e+11 84124
- value_eur          1 1.8200e+12 2.5787e+12 89457
```

These is backward elimination of the second model, here we can see the more variables are removed but further variables should be eliminated based on their p value and multi collinearity

MULTICOLLINEARITY FOR THE SECOND MODEL

```
> vif(m5)
            age       value_eur       height_cm defending_marking mentality_vision movement_balance
       1.242829        1.424560        2.661649          1.205630         1.883897         2.851928
```

In these second models the multicollinearity of the variables is good as the values are below 10.

FINAL MODEL

```
Residuals:
    Min      1Q  Median      3Q     Max 
-160334   -3199     884    4375  234833 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -1.270e+04  4.531e+03  -2.804  0.00513 ** 
age               8.862e+02  1.331e+02   6.658 4.19e-11 ***
value_eur         3.489e-03  6.210e-05  56.175  < 2e-16 ***
mentality_vision -1.985e+02  6.612e+01  -3.002  0.00274 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18780 on 1219 degrees of freedom
Multiple R-squared:  0.7797,	Adjusted R-squared:  0.7791 
F-statistic:  1438 on 3 and 1219 DF,  p-value: < 2.2e-16
```

This is the final model of forward position; the p value is good as every variable is below 0.05. adjusted r square is good which has 77%. The final model has no multicollinearity and has good p value.While doing these models I found out that the value of the player in the market places an important role in the wage as it shows the demand of the player. Mentality vision is important as a player with a better mentality can maintain calmness and pass the ball to the forwards to create chances.

Interaction terms:

1. *.ForwardPositionCombos.$skills=((.ForwardPositionCombos.$skill_ball_control + .ForwardPositionCombos.$power_shot_power)/2)*

For the forwards, I used the combination of skill ball control and power shot into skills variable. From the two variables we can check the skills which are effective to determine the wage of the players. Ball control is a skill which shows the player control over the ball and for power shots says how much power is used to shoot a ball towards the goal to score.

2. *.ForwardPositionCombos.$power2=((.ForwardPositionCombos.$power_stamina +.ForwardPositionCombos.$power_strength)/2)*

I combined power stamina and power strength to show the stamina and strength of the players within the forward position.

Second Order Term:

3. *ForwardPositionCombos.$poten =((.ForwardPositionCombos.$potential)^2)*

The second order term used in the forwards is potential. Potential is the expected growth of skills in the coming years.

First Iteration of Model:

- *a3 <- lm(wage_eur ~ value_eur + skills + attack + interception + shooting + passing + dribbling + defending + physic + attacking_heading_accuracy + +attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + skill_long_passing + movement_agility + movement_balance + power2 + mentality_aggression + mentality_vision + skills + poten + mentality_penalties + mentality_composure +*

defending_marking + defending_standing_tackle + defending_sliding_tackle , data = .ForwardPositionCombos.)

- *summary(a3)*

This model includes both the second order terms and interaction terms. The variables used in the interaction and second order models are not used

```
Residuals:
    Min      1Q  Median      3Q     Max
-159935   -2304    -328    1947  240472

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 1.839e+04  9.174e+03   2.004 0.045110 *
value_eur                   3.397e-03  4.602e-05  73.807  < 2e-16 ***
skills                      8.557e+01  7.852e+01   1.090 0.275870
attack                      1.751e+02  1.697e+02   1.032 0.302188
interception                1.892e-01  9.198e-01   0.206 0.837047
shooting                   -5.077e+01  7.925e+01  -0.641 0.521811
passing                    -2.620e+02  3.443e+02  -0.761 0.446616
dribbling                   3.564e+00  2.225e+00   1.602 0.109335
defending                   4.811e+01  3.347e+02   0.144 0.885692
physic                      4.426e+02  1.338e+02   3.307 0.000949 ***
attacking_heading_accuracy  4.859e+01  4.318e+01   1.125 0.260512
attacking_volleys           1.709e-01  3.426e+01   0.005 0.996019
skill_dribbling            -4.390e+02  2.860e+02  -1.535 0.124861
skill_curve                 2.710e+01  3.288e+01   0.824 0.409951
skill_fk_accuracy          -5.340e+00  2.814e+01  -0.190 0.849498
skill_long_passing          4.532e+01  6.680e+01   0.678 0.497527
movement_agility           -3.380e+00  2.920e+01  -0.116 0.907839
movement_balance            1.828e+01  2.635e+01   0.694 0.487933
power2                     -4.358e+02  1.038e+02  -4.198 2.74e-05 ***
mentality_aggression       -7.317e+01  3.388e+01  -2.160 0.030844 *
mentality_vision            3.870e+01  8.279e+01   0.467 0.640190
poten                      -1.957e+00  3.201e-01  -6.114 1.06e-09 ***
mentality_penalties         5.126e+01  3.152e+01   1.626 0.103976
mentality_composure        -7.439e+00  3.785e+01  -0.197 0.844188
defending_marking          -4.967e+01  1.023e+02  -0.485 0.627353
defending_standing_tackle   4.034e+01  1.083e+02   0.373 0.709532
defending_sliding_tackle   -3.993e+01  4.991e+01  -0.800 0.423744
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13120 on 4405 degrees of freedom
Multiple R-squared:  0.7727,    Adjusted R-squared:  0.7714
F-statistic: 575.9 on 26 and 4405 DF,  p-value: < 2.2e-16
```

Most of the variables from the models have high p values which is not good for the model so in order to eliminate the variables i used backward elimination

Second Model

In these models few variables are removed based on the multi collinearity and p values by backward elimination and for the next process these variables are again checked with multicollinearity to remove variables and perform backward elimination.

<u>*Second Iteration of Model:*</u>

- *a4 <- lm(wage_eur ~ value_eur + attack +attacking_heading_accuracy + power2 + mentality_aggression + poten + mentality_penalties + defending_marking + defending_standing_tackle,data = .ForwardPositionCombos.)*
- *summary(a4)*

```
Residuals:
    Min      1Q  Median      3Q     Max 
-160821   -2239    -392    1812  241089 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                6.359e+03  2.290e+03   2.777 0.005516 ** 
value_eur                  3.432e-03  3.734e-05  91.903  < 2e-16 ***
attack                     4.837e+01  2.657e+01   1.821 0.068735 .  
attacking_heading_accuracy 7.319e+01  2.341e+01   3.127 0.001777 ** 
power2                    -1.018e+02  2.737e+01  -3.719 0.000202 ***
mentality_aggression       2.581e+01  1.719e+01   1.501 0.133324    
poten                     -1.951e+00  2.859e-01  -6.825 9.97e-12 ***
mentality_penalties        4.735e+01  2.729e+01   1.735 0.082808 .  
defending_marking         -3.765e+01  2.003e+01  -1.880 0.060139 .  
defending_standing_tackle  3.617e+01  2.138e+01   1.692 0.090716 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13130 on 4422 degrees of freedom
Multiple R-squared:  0.7717,    Adjusted R-squared:  0.7712 
F-statistic:  1661 on 9 and 4422 DF,  p-value: < 2.2e-16
```

```
> vif(a4)
      value_eur                 attack attacking_heading_accuracy                     power2       mentality_aggression
       1.856290               2.077429                   2.091817                   1.835770                   1.738800
          poten    mentality_penalties          defending_marking  defending_standing_tackle
       1.723446               1.669062                   1.685972                   1.946133
```

```
                              Df  Sum of Sq        RSS   AIC
<none>                                      7.6218e+11 84063
- mentality_aggression         1 3.8853e+08 7.6257e+11 84064
- defending_standing_tackle    1 4.9345e+08 7.6267e+11 84064
- mentality_penalties          1 5.1885e+08 7.6270e+11 84064
- attack                       1 5.7130e+08 7.6275e+11 84065
- defending_marking            1 6.0936e+08 7.6279e+11 84065
- attacking_heading_accuracy   1 1.6854e+09 7.6386e+11 84071
- power2                       1 2.3840e+09 7.6456e+11 84075
- poten                        1 8.0293e+09 7.7021e+11 84108
- value_eur                    1 1.4558e+12 2.2180e+12 88795
```

```
> vif (a5)
                 value_eur                     attack attacking_heading_accuracy                     power2                      poten
                  1.819210                   1.480741                   1.581622                   1.581794                   1.706062
```

Final model

Some variables are removed from the previous model because of the p values which are greater than 0.005.

- *a5 <- lm(wage_eur ~ value_eur + attack + attacking_heading_accuracy + power2 + poten , data = .ForwardPositionCombos.)*
- *summary(a5)*

```
Residuals:
    Min      1Q  Median      3Q     Max 
-161745   -2280    -398    1731  240259 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                7.166e+03  2.146e+03   3.339 0.000849 ***
value_eur                  3.439e-03  3.699e-05  92.968  < 2e-16 ***
attack                     6.745e+01  2.245e+01   3.005 0.002672 ** 
attacking_heading_accuracy  9.765e+01  2.037e+01   4.794 1.69e-06 ***
power2                    -8.880e+01  2.542e+01  -3.493 0.000482 ***
poten                     -1.969e+00  2.847e-01  -6.916 5.33e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13140 on 4426 degrees of freedom
Multiple R-squared:  0.7712,    Adjusted R-squared:  0.7709 
F-statistic:  2983 on 5 and 4426 DF,  p-value: < 2.2e-16
```

Most of the variables from the first model variables are removed due to more p values and multicollinearity. The remaining variables still have second order terms and interaction terms which shows the effect on the model. This model is created for the forwards as the second order model and interaction model created are based on the attacking power which is helpful to recognize the players ability which depends on the wages.

Conclusion

To conclude, the forwards are highly rated and in demand because of their ability to score goals. The process of score depends mostly on the position placed on the field, receive passes, scoring goals with calmness and battling for the ball in the air these are the few instances where the goal scoring ability is determined as well as there demand with wage. The adjusted r square shows the 77% probability of the modeland p value can show that the build model is good. The poten is an variable which is to estimate the ability increase of a player in the coming years and attacking head accuracy shows the dominance of scoring a goal in the air.

b. Endri Rama - Midfielders

For our project, the main issue with today's Fifa is that when a player gets old, they tend to retire, rather than attempting to play a different position which they are comfortable with. We will be attempting to predict what significant variables cause the change of the weekly salary (wage_eur) for our data, this means that wage_eur is a dependent variable.

When building this model, what I found out is that the first model may not be the final model. With this being stated, the process which I partook in was trial and error repetition. The first model which I created was a full model, with wage_eur being the dependent variable and the other numerical variables being explanatory variables. The model code is down below, with a screen capture of the results.

First Iteration of Model (Full Model):

- *wageprediction <- lm(wage_eur ~ value_eur + age + height_cm + weight_kg + overall + potential + pace + shooting + passing + dribbling + defending + physic + attacking_crossing + attacking_finishing + attacking_heading_accuracy + attacking_short_passing + attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_control + movement_acceleration + movement_sprint_speed + movement_agility + movement_reactions + movement_balance + power_shot_power + power_jumping + power_stamina + power_strength + power_long_shots + mentality_aggression + mentality_interceptions + mentality_positioning + mentality_vision + mentality_penalties + mentality_composure + defending_marking + defending_standing_tackle + defending_sliding_tackle, data = Midfielders)*
- *summary(wageprediction)*

```
> summary(wageprediction)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + weight_kg +
    overall + potential + pace + shooting + passing + dribbling +
    defending + physic + attacking_crossing + attacking_finishing +
    attacking_heading_accuracy + attacking_short_passing + attacking_volleys +
    skill_dribbling + skill_curve + skill_fk_accuracy + skill_long_passing +
    skill_ball_control + movement_acceleration + movement_sprint_speed +
    movement_agility + movement_reactions + movement_balance +
    power_shot_power + power_jumping + power_stamina + power_strength +
    power_long_shots + mentality_aggression + mentality_interceptions +
    mentality_positioning + mentality_vision + mentality_penalties +
    mentality_composure + defending_marking + defending_standing_tackle +
    defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128883   -2348    -254    1683  194565

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -2.565e+04  6.896e+03  -3.719 0.000201 ***
value_eur                   3.116e-03  2.887e-05 107.928  < 2e-16 ***
age                         4.869e+02  6.068e+01   8.024 1.16e-15 ***
height_cm                   7.794e+01  3.469e+01   2.247 0.024666 *
weight_kg                   4.486e+01  3.052e+01   1.470 0.141657
overall                    -7.821e+01  7.179e+01  -1.089 0.275986
potential                   7.236e+01  4.723e+01   1.532 0.125527
pace                        1.215e+02  3.913e+02   0.311 0.756150
shooting                   -3.303e+02  3.967e+02  -0.833 0.405101
passing                     1.067e+03  3.927e+02   2.716 0.006621 **
dribbling                  -1.104e+02  3.957e+02  -0.279 0.780333
defending                   1.547e+02  3.990e+02   0.388 0.698248
physic                      2.327e+02  3.963e+02   0.587 0.557130
```

```
physic                       2.327e+02  3.963e+02   0.587 0.557130
attacking_crossing          -1.760e+02  8.019e+01  -2.194 0.028253 *
attacking_finishing          9.266e+01  1.796e+02   0.516 0.605894
attacking_heading_accuracy  -2.003e+01  4.306e+01  -0.465 0.641780
attacking_short_passing     -3.643e+02  1.435e+02  -2.539 0.011148 *
attacking_volleys            5.123e+01  2.594e+01   1.975 0.048333 *
skill_dribbling              9.244e+01  2.005e+02   0.461 0.644739
skill_curve                 -5.370e+01  2.610e+01  -2.058 0.039643 *
skill_fk_accuracy           -7.796e+01  2.458e+01  -3.172 0.001521 **
skill_long_passing          -2.253e+02  6.460e+01  -3.488 0.000489 ***
skill_ball_control           7.748e+01  1.264e+02   0.613 0.540040
movement_acceleration       -7.488e+01  1.769e+02  -0.423 0.672115
movement_sprint_speed       -8.743e+01  2.171e+02  -0.403 0.687148
movement_agility             8.258e+00  4.429e+01   0.186 0.852100
movement_reactions          -4.630e+01  3.412e+01  -1.357 0.174839
movement_balance             4.547e+01  2.760e+01   1.648 0.099441 .
power_shot_power             9.483e+01  8.199e+01   1.157 0.247476
power_jumping               -5.601e+00  2.301e+01  -0.243 0.807702
power_stamina               -1.173e+02  1.002e+02  -1.171 0.241692
power_strength              -1.600e+02  1.986e+02  -0.805 0.420627
power_long_shots             5.259e+01  8.147e+01   0.645 0.518624
mentality_aggression        -1.544e+01  8.049e+01  -0.192 0.847850
mentality_interceptions     -2.045e+01  8.165e+01  -0.250 0.802250
mentality_positioning        2.675e+01  2.882e+01   0.928 0.353353
mentality_vision            -2.177e+02  8.216e+01  -2.650 0.008057 **
mentality_penalties          5.244e+01  2.590e+01   2.025 0.042887 *
mentality_composure         -1.596e+01  2.159e+01  -0.739 0.459668
defending_marking           -5.886e+01  1.204e+02  -0.489 0.625058
defending_standing_tackle   -4.568e+01  1.224e+02  -0.373 0.709074
defending_sliding_tackle     2.608e+01  4.650e+01   0.561 0.574905
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

-

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10500 on 8461 degrees of freedom
Multiple R-squared:  0.744,     Adjusted R-squared:  0.7427
F-statistic: 599.7 on 41 and 8461 DF,  p-value: < 2.2e-16
```

-

From the model above, I noticed that the market value (value_eur) and the age play a significant part in determining whether a player may continue to play for a longer amount of time or if they would rather retire. As stated above, though this is not the final model I came up with. I determined this by observing the t-Values of the model, and the reader can see that the t-Value for the market value (value_eur) and age t-value play a significant role in the players weekly wage (wage_eur).

Secondly, I did backward elimination on the model as well as checked for multicollinearity, and noticed that some of the variables I had chosen could be removed from the model.

First Backward Elimination and Multicollinearity Check:

- *bstep <- stepAIC(wageprediction, direction = "backward")*
- *summary(bstep)*
- *vif(bstep)*
-

```
> bstep <- stepAIC(wageprediction, direction = "backward")
Start:  AIC=157505.3
wage_eur ~ value_eur + age + height_cm + weight_kg + overall +
    potential + pace + shooting + passing + dribbling + defending +
    physic + attacking_crossing + attacking_finishing + attacking_heading_accuracy +
    attacking_short_passing + attacking_volleys + skill_dribbling +
    skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_control +
    movement_acceleration + movement_sprint_speed + movement_agility +
    movement_reactions + movement_balance + power_shot_power +
    power_jumping + power_stamina + power_strength + power_long_shots +
    mentality_aggression + mentality_interceptions + mentality_positioning +
    mentality_vision + mentality_penalties + mentality_composure +
    defending_marking + defending_standing_tackle + defending_sliding_tackle

                              Df  Sum of Sq        RSS    AIC
- movement_agility             1 3.8337e+06 9.3312e+11 157503
- mentality_aggression         1 4.0600e+06 9.3312e+11 157503
- power_jumping                1 6.5336e+06 9.3312e+11 157503
- mentality_interceptions      1 6.9173e+06 9.3312e+11 157503
- dribbling                    1 8.5781e+06 9.3312e+11 157503
- pace                         1 1.0636e+07 9.3313e+11 157503
- defending_standing_tackle    1 1.5353e+07 9.3313e+11 157503
- defending                    1 1.6577e+07 9.3313e+11 157503
- movement_sprint_speed        1 1.7888e+07 9.3313e+11 157503
- movement_acceleration        1 1.9758e+07 9.3314e+11 157503
- skill_dribbling              1 2.3448e+07 9.3314e+11 157504
- attacking_heading_accuracy   1 2.3870e+07 9.3314e+11 157504
- defending_marking            1 2.6340e+07 9.3314e+11 157504
- attacking_finishing          1 2.9360e+07 9.3314e+11 157504
- defending_sliding_tackle     1 3.4692e+07 9.3315e+11 157504
- physic                       1 3.8018e+07 9.3315e+11 157504
- skill_ball_control           1 4.1411e+07 9.3316e+11 157504
- power_long_shots             1 4.5952e+07 9.3316e+11 157504
- mentality_composure          1 6.0298e+07 9.3318e+11 157504
```

```
Step:  AIC=157476.5
wage_eur ~ value_eur + age + height_cm + weight_kg + shooting +
    passing + physic + attacking_crossing + attacking_short_passing +
    attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy +
    skill_long_passing + movement_sprint_speed + movement_reactions +
    movement_balance + power_shot_power + power_stamina + power_strength +
    mentality_vision + mentality_penalties + defending_sliding_tackle

                           Df  Sum of Sq        RSS    AIC
<none>                                   9.3391e+11 157477
- weight_kg                 1 2.3041e+08 9.3414e+11 157477
- movement_balance          1 4.2410e+08 9.3433e+11 157478
- skill_dribbling           1 4.3249e+08 9.3434e+11 157478
- skill_curve               1 4.8444e+08 9.3439e+11 157479
- attacking_crossing        1 5.3702e+08 9.3444e+11 157479
- movement_reactions        1 5.3721e+08 9.3444e+11 157479
- height_cm                 1 5.7739e+08 9.3448e+11 157480
- attacking_volleys         1 6.1932e+08 9.3452e+11 157480
- power_shot_power          1 6.5453e+08 9.3456e+11 157480
- mentality_penalties       1 6.8595e+08 9.3459e+11 157481
- attacking_short_passing   1 6.9563e+08 9.3460e+11 157481
- mentality_vision          1 7.3466e+08 9.3464e+11 157481
- physic                    1 7.4866e+08 9.3465e+11 157481
- movement_sprint_speed     1 7.7680e+08 9.3468e+11 157482
- passing                   1 8.1566e+08 9.3472e+11 157482
- skill_fk_accuracy         1 1.0651e+09 9.3497e+11 157484
- shooting                  1 1.2308e+09 9.3514e+11 157486
- power_strength            1 1.2409e+09 9.3515e+11 157486
- skill_long_passing        1 1.3503e+09 9.3526e+11 157487
- defending_sliding_tackle  1 1.5016e+09 9.3541e+11 157488
- power_stamina             1 2.5363e+09 9.3644e+11 157498
- age                       1 1.5237e+10 9.4914e+11 157612
- value_eur                 1 1.4466e+12 2.3805e+12 165431
```

```
movement_sprint_speed    -3.912e+01  1.473e+01  -2.656 0.007930 **
movement_reactions       -5.353e+01  2.424e+01  -2.208 0.027237 *
movement_balance          3.581e+01  1.825e+01   1.962 0.049766 *
power_shot_power          5.237e+01  2.148e+01   2.438 0.014800 *
power_stamina            -1.008e+02  2.102e+01  -4.799 1.62e-06 ***
power_strength           -1.205e+02  3.591e+01  -3.357 0.000793 ***
mentality_vision         -2.110e+02  8.170e+01  -2.583 0.009821 **
mentality_penalties       4.073e+01  1.632e+01   2.496 0.012595 *
defending_sliding_tackle  4.290e+01  1.162e+01   3.692 0.000224 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10490 on 8479 degrees of freedom
Multiple R-squared:  0.7438,    Adjusted R-squared:  0.7431
F-statistic:  1070 on 23 and 8479 DF,  p-value: < 2.2e-16
```

-

```
> vif(bstep)
               value_eur                      age                height_cm
                1.843271                 1.864081                 3.192434
               weight_kg                 shooting                  passing
                2.603190                 8.879377               812.929016
                  physic        attacking_crossing  attacking_short_passing
               25.757483                62.106020                96.282449
       attacking_volleys            skill_dribbling              skill_curve
                3.508244                 4.542512                 9.076151
       skill_fk_accuracy         skill_long_passing    movement_sprint_speed
                8.748930                30.079360                 1.910987
      movement_reactions           movement_balance         power_shot_power
                3.500084                 2.792106                 4.187612
           power_stamina             power_strength         mentality_vision
                4.837634                14.513000                50.515568
     mentality_penalties  defending_sliding_tackle
                2.433380                 2.910534
>
```

From the screen captures above the reader will notice that even though I have done backward elimination on the model, there is a lot of multicollinearity on this model, which is not good. This means that any variable with a multicollinearity higher than ten will be removed from the model.

Third of all, I continued the process of backward elimination and remodeling of the original model. After removing the following variables because of their multicollinearity being above ten, I then remodeled the model. Variables removed from the original model are listed down below and the updated model code along with the repeated process is also down below.

- *Removed Variables:*
    - *Passing*
    - *Physic*
    - *Attacking_crossing*
    - *Attacking_short_passing*
    - *Skill_long_passing*
    - *Power_strength*
    - *Mentality_vision*
- *Second Iteration of Model:*
    - *wageprediction2 <- lm(wage_eur ~ value_eur + age + height_cm + weight_kg + potential + pace + shooting + dribbling + attacking_finishing + attacking_heading_accuracy + attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + skill_ball_control + movement_acceleration + movement_sprint_speed + movement_agility + movement_reactions + movement_balance + power_shot_power + power_jumping + power_stamina + power_long_shots + mentality_aggression + mentality_interceptions + mentality_positioning + mentality_penalties + mentality_composure + defending_marking + defending_standing_tackle + defending_sliding_tackle, data = Midfielders)*
    - *summary(wageprediction2)*

```
> summary(wageprediction2)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + weight_kg +
    potential + pace + shooting + dribbling + attacking_finishing +
    attacking_heading_accuracy + attacking_volleys + skill_dribbling +
    skill_curve + skill_fk_accuracy + skill_ball_control + movement_accelera
tion +
    movement_sprint_speed + movement_agility + movement_reactions +
    movement_balance + power_shot_power + power_jumping + power_stamina +
    power_long_shots + mentality_aggression + mentality_interceptions +
    mentality_positioning + mentality_penalties + mentality_composure +
    defending_marking + defending_standing_tackle + defending_sliding_tackl
e,
    data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128397   -2347    -255    1667  194862

Coefficients:
                                Estimate      Std. Error t value
(Intercept)               -24605.48991642  6771.27582430  -3.634
value_eur                      0.00310638     0.00002795 111.139
age                          451.96454495    49.38689735   9.152
height_cm                     69.58135587    34.43418414   2.021
weight_kg                     21.11207250    28.61891535   0.738
potential                     37.91401708    38.56142583   0.983
pace                         152.12001243   391.57407763   0.388
shooting                    -312.72595642   396.97932561  -0.788
dribbling                   -118.09613757   395.93714361  -0.298
attacking_finishing           84.87364925   179.69746299   0.472
attacking_heading_accuracy    -9.87197356    15.57644313  -0.634
attacking_volleys             56.21376996    25.91582899   2.169
skill_dribbling              110.67305510   200.45477354   0.552
```

```
attacking_volleys             56.21376996     25.91582899   2.169
skill_dribbling              110.67305510    200.45477354   0.552
skill_curve                    9.47847713     16.51709643   0.574
skill_fk_accuracy            -27.60803949     14.33277094  -1.926
skill_ball_control            41.32426207    125.15002595   0.330
movement_acceleration        -79.03989781    177.00434758  -0.447
movement_sprint_speed       -102.53007586    217.21633454  -0.472
movement_agility              12.93897763     44.28363575   0.292
movement_reactions           -53.60573477     32.96311354  -1.626
movement_balance              51.99465736     27.46370633   1.893
power_shot_power              79.35812260     82.00005042   0.968
power_jumping                  3.42539127     11.55731749   0.296
power_stamina                -70.15353724     13.17501229  -5.325
power_long_shots              41.61014589     81.48394704   0.511
mentality_aggression          25.26714149     12.89370199   1.960
mentality_interceptions        5.62356591     17.50508585   0.321
mentality_positioning         30.17415408     28.32261604   1.065
mentality_penalties           53.38901249     25.89888467   2.061
mentality_composure          -32.29058295     21.00705770  -1.537
defending_marking            -15.81828576     14.35254364  -1.102
defending_standing_tackle     -4.06983915     25.74281451  -0.158
defending_sliding_tackle      43.54078134     23.45577589   1.856
                                       Pr(>|t|)
(Intercept)                            0.000281 ***
value_eur                  < 0.0000000000000002 ***
age                        < 0.0000000000000002 ***
height_cm                              0.043342 *
weight_kg                              0.460719
potential                              0.325532
pace                                   0.697668
shooting                               0.430857
dribbling                              0.765504
attacking_finishing                    0.636715
attacking_heading_accuracy             0.526244
attacking_volleys                      0.030104 *
```

```
attacking_finishing                      0.636715
attacking_heading_accuracy               0.526244
attacking_volleys                        0.030104 *
skill_dribbling                          0.580888
skill_curve                              0.566079
skill_fk_accuracy                        0.054111 .
skill_ball_control                       0.741259
movement_acceleration                    0.655217
movement_sprint_speed                    0.636926
movement_agility                         0.770153
movement_reactions                       0.103937
movement_balance                         0.058364 .
power_shot_power                         0.333181
power_jumping                            0.766945
power_stamina                         0.000000104 ***
power_long_shots                         0.609606
mentality_aggression                     0.050069 .
mentality_interceptions                  0.748026
mentality_positioning                    0.286737
mentality_penalties                      0.039291 *
mentality_composure                      0.124299
defending_marking                        0.270439
defending_standing_tackle                0.874385
defending_sliding_tackle                 0.063447 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10510 on 8470 degrees of freedom
Multiple R-squared:  0.7431,    Adjusted R-squared:  0.7421
F-statistic: 765.6 on 32 and 8470 DF,  p-value: < 0.00000000000000022
```

- *Second Backward Elimination and Multicollinearity Check:*
  - *bstep2 <- stepAIC(wageprediction2, direction = "backward")*
  - *summary(bstep2)*
  - *vif(bstep2)*

```
> bstep2 <- stepAIC(wageprediction2, direction = "backward")
Start:  AIC=157516.7
wage_eur ~ value_eur + age + height_cm + weight_kg + potential + 
    pace + shooting + dribbling + attacking_finishing + attacking_heading_ac
curacy + 
    attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + 
    skill_ball_control + movement_acceleration + movement_sprint_speed + 
    movement_agility + movement_reactions + movement_balance + 
    power_shot_power + power_jumping + power_stamina + power_long_shots + 
    mentality_aggression + mentality_interceptions + mentality_positioning + 

    mentality_penalties + mentality_composure + defending_marking + 
    defending_standing_tackle + defending_sliding_tackle

                            Df   Sum of Sq          RSS    AIC
- defending_standing_tackle  1     2763081 936344862318 157515
- movement_agility           1     9437664 936351536902 157515
- power_jumping              1     9710854 936351810092 157515
- dribbling                  1     9834902 936351934140 157515
```

```
                              Df     Sum of Sq           RSS    AIC
- defending_standing_tackle    1       2763081  936344862318 157515
- movement_agility             1       9437664  936351536902 157515
- power_jumping                1       9710854  936351810092 157515
- dribbling                    1       9834902  936351934140 157515
- mentality_interceptions      1      11408965  936353508202 157515
- skill_ball_control           1      12053120  936354152357 157515
- pace                         1      16683839  936358783076 157515
- movement_acceleration        1      22043270  936364142507 157515
- movement_sprint_speed        1      24630245  936366729483 157515
- attacking_finishing          1      24661132  936366760369 157515
- power_long_shots             1      28827402  936370926639 157515
- skill_dribbling              1      33697845  936375797083 157515
- skill_curve                  1      36404986  936378504223 157515
- attacking_heading_accuracy   1      44404043  936386503281 157515
- weight_kg                    1      60159821  936402259059 157515
- shooting                     1      68603018  936410702255 157515
- power_shot_power             1     103539406  936445638644 157516
- potential                    1     106867236  936448966473 157516
- mentality_positioning        1     125474253  936467573490 157516
- defending_marking            1     134280275  936476379512 157516
<none>                                          936342099238 157517
- mentality_composure          1     261199568  936603298806 157517
- movement_reactions           1     292359489  936634458726 157517
- defending_sliding_tackle     1     380928941  936723028179 157518
- movement_balance             1     396232677  936738331914 157518
- skill_fk_accuracy            1     410168170  936752267407 157518
- mentality_aggression         1     424529729  936766628966 157519
- height_cm                    1     451395598  936793494835 157519
- mentality_penalties          1     469778128  936811877365 157519
- attacking_volleys            1     520123352  936862222590 157519
- power_stamina                1    3134355330  939476454567 157543
- age                          1    9258409550  945600508788 157598
- value_eur                    1 1365474422244 2301816521481 165163
```

```
Step:  AIC=157514.7
wage_eur ~ value_eur + age + height_cm + weight_kg + potential + 
    pace + shooting + dribbling + attacking_finishing + attacking_heading_ac
curacy + 
    attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + 
    skill_ball_control + movement_acceleration + movement_sprint_speed + 
    movement_agility + movement_reactions + movement_balance + 
    power_shot_power + power_jumping + power_stamina + power_long_shots + 
    mentality_aggression + mentality_interceptions + mentality_positioning + 

    mentality_penalties + mentality_composure + defending_marking + 
    defending_sliding_tackle

                             Df  Sum of Sq          RSS    AIC
- mentality_interceptions     1    8922639 936353784958 157513
- movement_agility            1    9438344 936354300662 157513
- dribbling                   1    9809663 936354671981 157513
- power_jumping               1   10024709 936354887027 157513
- skill_ball_control          1   11810427 936356672745 157513
- pace                        1   16703602 936361565920 157513
- movement_acceleration       1   22012821 936366875140 157513
- movement_sprint_speed       1   24633970 936369496288 157513
- attacking_finishing         1   24748541 936369610860 157513
- power_long_shots            1   28868673 936373730992 157513
- skill_dribbling             1   33733644 936378595963 157513
- skill_curve                 1   36152259 936381014577 157513
- attacking_heading_accuracy  1   44036134 936388898452 157513
- weight_kg                   1   59607230 936404469549 157513
- shooting                    1   68767476 936413629794 157513
- power_shot_power            1  103905410 936448767728 157514
- potential                   1  106577060 936451439379 157514
- mentality_positioning       1  125798126 936470660444 157514
- defending_marking           1  143567419 936488429737 157514
<none>                                     936344862318 157515
- mentality_composure         1  263095089 936607957408 157515
```

```
Step:  AIC=157489.4
wage_eur ~ value_eur + age + height_cm + shooting + attacking_volleys +
    skill_dribbling + skill_fk_accuracy + movement_sprint_speed +
    movement_reactions + movement_balance + power_shot_power +
    power_stamina + mentality_aggression + mentality_penalties +
    defending_sliding_tackle

                           Df     Sum of Sq           RSS    AIC
<none>                                       937084810150 157489
- movement_sprint_speed     1     296721899  937381532049 157490
- power_shot_power          1     395660301  937480470451 157491
- skill_fk_accuracy         1     399875752  937484685902 157491
- mentality_aggression      1     422486651  937507296800 157491
- movement_balance          1     672365855  937757176005 157494
- height_cm                 1     722549674  937807359824 157494
- mentality_penalties       1     732750821  937817560971 157494
- movement_reactions        1     742734727  937827544876 157494
- skill_dribbling           1     838346449  937923156598 157495
- attacking_volleys         1     854291063  937939101212 157495
- defending_sliding_tackle  1    1077536863  938162347013 157497
- shooting                  1    1370071299  938454881448 157500
- power_stamina             1    3560720460  940645530610 157520
- age                       1   17598229059  954683039208 157646
- value_eur                 1 1495722165493 2432806975642 165600
```

```
> summary(bstep2)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + shooting +
    attacking_volleys + skill_dribbling + skill_fk_accuracy +
    movement_sprint_speed + movement_reactions + movement_balance +
    power_shot_power + power_stamina + mentality_aggression +
    mentality_penalties + defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128743   -2325    -246    1663  195231

Coefficients:
                               Estimate      Std. Error t value
(Intercept)              -22519.50033202   6242.48328367  -3.607
value_eur                     0.00310850      0.00002671 116.389
age                         418.64631353     33.16081045  12.625
height_cm                    75.72929988     29.60345736   2.558
shooting                   -107.92535980     30.63827211  -3.523
attacking_volleys            46.11261810     16.57788784   2.782
skill_dribbling              63.60540217     23.08312410   2.755
skill_fk_accuracy           -22.90313630     12.03496331  -1.903
movement_sprint_speed       -22.49200767     13.72036834  -1.639
movement_reactions          -60.13368492     23.18533102  -2.594
movement_balance             43.90001958     17.78992990   2.468
power_shot_power             39.75310461     21.00013859   1.893
power_stamina               -71.38223879     12.56995065  -5.679
mentality_aggression         24.04462463     12.29203687   1.956
mentality_penalties          42.00197498     16.30436552   2.576
defending_sliding_tackle     33.43113719     10.70156834   3.124
```

```
                                 Pr(>|t|)
(Intercept)                      0.000311 ***
value_eur                   < 0.0000000000000002 ***
age                         < 0.0000000000000002 ***
height_cm                        0.010541 *
shooting                         0.000430 ***
attacking_volleys                0.005422 **
skill_dribbling                  0.005873 **
skill_fk_accuracy                0.057068 .
movement_sprint_speed            0.101185
movement_reactions               0.009514 **
movement_balance                 0.013618 *
power_shot_power                 0.058393 .
power_stamina                 0.000000014 ***
mentality_aggression             0.050484 .
mentality_penalties              0.010008 *
defending_sliding_tackle         0.001790 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10510 on 8487 degrees of freedom
Multiple R-squared:  0.7429,	Adjusted R-squared:  0.7424
F-statistic:  1635 on 15 and 8487 DF,  p-value: < 0.00000000000000022
```

```
> vif(bstep2)
               value_eur                      age
                  1.7757                   1.6555
               height_cm                 shooting
                  2.3884                   8.4478
       attacking_volleys          skill_dribbling
                  3.3751                   3.4085
       skill_fk_accuracy    movement_sprint_speed
                  2.1238                   1.6540
      movement_reactions         movement_balance
                  3.1944                   2.6467
        power_shot_power            power_stamina
                  3.9916                   1.7264
    mentality_aggression      mentality_penalties
                  2.4028                   2.4223
defending_sliding_tackle
                  2.4632
```

From the above the reader can see that even though I removed those variables, I still have more variables to remove, so I repeat the process in order to have the perfect model for the midfielder category.

Lastly, I observed the backward elimination and then removed the following variables from the backward elimination final step model. The variables removed are down below, the updated final model is down below, I also did backward elimination on this final model, and nothing was eliminated, this means that all variables are significant, I then checked for multicollinearity, adjusted R^2 and they all look good.

A. *Variables Removed:*
   - *Skill_fk_accuracy*
   - *Power_shot_power*

B. *Third Iteration of Model:*
   - *wageprediction3 <- lm(wage_eur ~ value_eur + age + height_cm + shooting + attacking_volleys + skill_dribbling + + movement_reactions + movement_balance + power_stamina + mentality_aggression + mentality_penalties + defending_sliding_tackle, data = Midfielders)*
   - *summary(wageprediction3)*

```
> summary(wageprediction3)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + shooting +
    attacking_volleys + skill_dribbling + +movement_reactions +
    movement_balance + power_stamina + mentality_aggression +
    mentality_penalties + defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128772   -2324    -249    1641  195373

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -2.355e+04  6.224e+03  -3.783 0.000156 ***
value_eur                 3.108e-03  2.671e-05 116.359  < 2e-16 ***
age                       4.165e+02  3.261e+01  12.772  < 2e-16 ***
height_cm                 8.025e+01  2.954e+01   2.716 0.006612 **
shooting                 -8.080e+01  2.451e+01  -3.296 0.000983 ***
attacking_volleys         4.296e+01  1.645e+01   2.612 0.009015 **
skill_dribbling           4.885e+01  2.227e+01   2.193 0.028315 *
movement_reactions       -5.704e+01  2.316e+01  -2.463 0.013779 *
movement_balance          3.752e+01  1.755e+01   2.138 0.032548 *
power_stamina            -7.688e+01  1.192e+01  -6.452 1.16e-10 ***
mentality_aggression      3.059e+01  1.199e+01   2.551 0.010772 *
mentality_penalties       3.667e+01  1.574e+01   2.330 0.019809 *
defending_sliding_tackle  3.681e+01  1.021e+01   3.605 0.000314 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10510 on 8490 degrees of freedom
Multiple R-squared:  0.7426,    Adjusted R-squared:  0.7423
F-statistic:  2042 on 12 and 8490 DF,  p-value: < 2.2e-16
```

C. <u>*Third Backward Elimination Check:*</u>

- *bstep3 <- stepAIC(wageprediction3, direction = "backward")*
- *summary(bstep3)*

```
> bstep3 <- stepAIC(wageprediction3, direction = "backward")
Start:  AIC=157492
wage_eur ~ value_eur + age + height_cm + shooting + attacking_volleys +
    skill_dribbling + +movement_reactions + movement_balance +
    power_stamina + mentality_aggression + mentality_penalties +
    defending_sliding_tackle

                           Df  Sum of Sq        RSS    AIC
<none>                                   9.3803e+11 157492
- movement_balance          1 5.0503e+08 9.3854e+11 157495
- skill_dribbling           1 5.3149e+08 9.3856e+11 157495
- mentality_penalties       1 6.0002e+08 9.3863e+11 157495
- movement_reactions        1 6.7052e+08 9.3870e+11 157496
- mentality_aggression      1 7.1876e+08 9.3875e+11 157497
- attacking_volleys         1 7.5386e+08 9.3879e+11 157497
- height_cm                 1 8.1528e+08 9.3885e+11 157497
- shooting                  1 1.2006e+09 9.3923e+11 157501
- defending_sliding_tackle  1 1.4361e+09 9.3947e+11 157503
- power_stamina             1 4.5998e+09 9.4263e+11 157532
- age                       1 1.8023e+10 9.5606e+11 157652
- value_eur                 1 1.4959e+12 2.4340e+12 165598
```

```
> summary(bstep3)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + shooting +
    attacking_volleys + skill_dribbling + +movement_reactions +
    movement_balance + power_stamina + mentality_aggression +
    mentality_penalties + defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128772   -2324    -249    1641  195373

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -2.355e+04  6.224e+03  -3.783 0.000156 ***
value_eur                 3.108e-03  2.671e-05 116.359  < 2e-16 ***
age                       4.165e+02  3.261e+01  12.772  < 2e-16 ***
height_cm                 8.025e+01  2.954e+01   2.716 0.006612 **
shooting                 -8.080e+01  2.451e+01  -3.296 0.000983 ***
attacking_volleys         4.296e+01  1.645e+01   2.612 0.009015 **
skill_dribbling           4.885e+01  2.227e+01   2.193 0.028315 *
movement_reactions       -5.704e+01  2.316e+01  -2.463 0.013779 *
movement_balance          3.752e+01  1.755e+01   2.138 0.032548 *
power_stamina            -7.688e+01  1.192e+01  -6.452 1.16e-10 ***
mentality_aggression      3.059e+01  1.199e+01   2.551 0.010772 *
mentality_penalties       3.667e+01  1.574e+01   2.330 0.019809 *
defending_sliding_tackle  3.681e+01  1.021e+01   3.605 0.000314 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10510 on 8490 degrees of freedom
Multiple R-squared:  0.7426,    Adjusted R-squared:  0.7423
F-statistic:  2042 on 12 and 8490 DF,  p-value: < 2.2e-16
```

D. *Third Multicollinearity Check:*

- *vif(bstep3)*

```
> vif(bstep3)
           value_eur                  age               height_cm
            1.774758             1.600083                2.376995
            shooting     attacking_volleys          skill_dribbling
            5.402888             3.320112                3.171400
  movement_reactions      movement_balance           power_stamina
            3.183999             2.573596                1.550155
mentality_aggression   mentality_penalties defending_sliding_tackle
            2.286343             2.254775                2.241257
> |
```

From the above the reader can see that *wageprediction3* is the best model for this project regarding midfielder positions, as there is no multicollinearity, and backward elimination cannot eliminate any further variables, but it is not a clean model.

In the following, I will show how the best model so far can be improved and look cleaner. I have cleaned up our midfielder model using two Interaction Terms, one Second Order Terms, and how our previous model gave the same results but this model looks cleaner.

A. *Two Interaction Term*

a. The Midfielder Interaction Term consists of a combination of Mentality Variables as well as Movement Variables.

i. The Mentality Interaction Term was produced from the previous variables of mentality_aggression and mentality_penalties. This makes sense because what I am evaluating is how mentaly there are midfield players during a game. How aggressive they may play and how calm they are when taking a penalty.

1. *Midfielders$Mentality = ((Midfielders$mentality_aggression + Midfielders$mentality_penalties)/2)*

```
Midfielders$Mentality = ((Midfielders$mentality_aggression +
                          Midfielders$mentality_penalties)/2)
```

ii. The Movement Interaction Term was produced from the previous variables of movement_reactions and movement_balance. This once again makes sense because as someone with a high market value (value_eur) has to have great reactions as well as balance in order to maintain that high market value (value_eur) as they get older.

1. *Midfielders$Movement = ((Midfielders$movement_reactions + Midfielders$movement_balance)/2)*

```
> Midfielders$Movement = ((Midfielders$movement_reactions +
+                             Midfielders$movement_balance)/2)
```

B. *Second Order Term*

a. The second order term for the midfielder model consists of attacking_volleys being squared. Meaning how much of a difference would it make if the midfielder were twice as consistent in scoring volleys.

i. *Midfielders$VolleySQ = ((Midfielders$attacking_volleys)^2)*

```
> Midfielders$VolleySQ = ((Midfielders$attacking_volleys)^2)
```

C. *Fourth Iteration of Model*

a. The updated model removes previous variables such as attacking_volleys, movement_reactions, movement_balance, mentality_aggression, and mentality_penalties, because to include these with the addition of the two interaction terms and the second order term would be redundant and not produce accurate results.

i. *wageprediction4 <- lm(wage_eur ~ value_eur + age + height_cm + shooting + VolleySQ + skill_dribbling + Movement + power_stamina + Mentality + defending_sliding_tackle, data = Midfielders)*

ii. *summary(wageprediction4)*

```
> wageprediction4 <- lm(wage_eur ~ value_eur + age + height_cm + shooting +
+                         VolleySQ + skill_dribbling + Movement +
+                         power_stamina + Mentality +
+                         defending_sliding_tackle, data = Midfielders)
> summary(wageprediction4)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + shooting +
    VolleySQ + skill_dribbling + Movement + power_stamina + Mentality +
    defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128454   -2283    -230    1678  196116

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -1.246e+04  5.525e+03  -2.256 0.024116 *
value_eur                 3.080e-03  2.625e-05 117.334  < 2e-16 ***
age                       3.817e+02  3.127e+01  12.208  < 2e-16 ***
height_cm                 3.269e+01  2.604e+01   1.255 0.209448
shooting                 -8.561e+01  2.307e+01  -3.711 0.000208 ***
VolleySQ                  4.773e-01  1.566e-01   3.048 0.002312 **
skill_dribbling           4.174e+01  2.209e+01   1.889 0.058870 .
Movement                  3.313e+00  2.782e+01   0.119 0.905234
power_stamina            -7.966e+01  1.166e+01  -6.833 8.89e-12 ***
Mentality                 5.590e+01  1.909e+01   2.928 0.003417 **
defending_sliding_tackle  2.925e+01  9.411e+00   3.108 0.001891 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10520 on 8492 degrees of freedom
Multiple R-squared:  0.7424,    Adjusted R-squared:  0.7421
F-statistic:  2447 on 10 and 8492 DF,  p-value: < 2.2e-16
```

The model was once again run through backward step elimination and Movement, height_cm, and skill_dribbling variables were removed. I also checked the Multicollinearity and there was none.

D. *Fourth Backward Step Elimination Check:*

i. *bstep4 <- stepAIC(wageprediction4, direction = "backward")*

ii. *summary(bstep4)*

iii.

```
> bstep4 <- stepAIC(wageprediction4, direction = "backward")
Start:  AIC=157497.3
wage_eur ~ value_eur + age + height_cm + shooting + VolleySQ +
    skill_dribbling + Movement + power_stamina + Mentality +
    defending_sliding_tackle

                           Df  Sum of Sq        RSS    AIC
- Movement                  1 1.5674e+06 9.3906e+11 157495
- height_cm                 1 1.7422e+08 9.3924e+11 157497
<none>                                   9.3906e+11 157497
- skill_dribbling           1 3.9477e+08 9.3946e+11 157499
- Mentality                 1 9.4826e+08 9.4001e+11 157504
- VolleySQ                  1 1.0272e+09 9.4009e+11 157505
- defending_sliding_tackle  1 1.0681e+09 9.4013e+11 157505
- shooting                  1 1.5226e+09 9.4059e+11 157509
- power_stamina             1 5.1629e+09 9.4423e+11 157542
- age                       1 1.6481e+10 9.5554e+11 157643
- value_eur                 1 1.5224e+12 2.4615e+12 165689

Step:  AIC=157495.4
wage_eur ~ value_eur + age + height_cm + shooting + VolleySQ +
    skill_dribbling + power_stamina + Mentality + defending_sliding_tackle

                           Df  Sum of Sq        RSS    AIC
<none>                                   9.3906e+11 157495
- height_cm                 1 2.5451e+08 9.3932e+11 157496
- skill_dribbling           1 4.5180e+08 9.3952e+11 157497
- Mentality                 1 9.5780e+08 9.4002e+11 157502
- VolleySQ                  1 1.0451e+09 9.4011e+11 157503
- defending_sliding_tackle  1 1.0924e+09 9.4016e+11 157503
- shooting                  1 1.5218e+09 9.4059e+11 157507
- power_stamina             1 5.3931e+09 9.4446e+11 157542
- age                       1 1.6661e+10 9.5573e+11 157643
- value_eur                 1 1.6147e+12 2.5538e+12 166000
```

iv.

```
> summary(bstep4)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + shooting +
    VolleySQ + skill_dribbling + power_stamina + Mentality +
    defending_sliding_tackle, data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128475   -2279    -232    1680  196116

Coefficients:
                              Estimate     Std. Error t value
(Intercept)              -11999.1719331   3925.3494847  -3.057
value_eur                     0.0030810      0.0000255 120.846
age                         382.0731192     31.1249897  12.275
height_cm                    30.7398940     20.2612916   1.517
shooting                    -85.5808390     23.0681009  -3.710
VolleySQ                      0.4791146      0.1558433   3.074
skill_dribbling              42.5374386     21.0434544   2.021
power_stamina               -79.3460188     11.3611462  -6.984
Mentality                    56.0530144     19.0449532   2.943
defending_sliding_tackle     29.3779506      9.3463343   3.143
                                     Pr(>|t|)
(Intercept)                          0.002244 **
value_eur                < 0.0000000000000002 ***
age                      < 0.0000000000000002 ***
height_cm                            0.129260
shooting                             0.000209 ***
VolleySQ                             0.002116 **
skill_dribbling                      0.043269 *
power_stamina                0.00000000000309 ***
Mentality                            0.003257 **
defending_sliding_tackle             0.001677 **
---
```

```
Coefficients:
                              Estimate      Std. Error t value
(Intercept)              -11999.1719331  3925.3494847  -3.057
value_eur                     0.0030810     0.0000255 120.846
age                         382.0731192    31.1249897  12.275
height_cm                    30.7398940    20.2612916   1.517
shooting                    -85.5808390    23.0681009  -3.710
VolleySQ                      0.4791146     0.1558433   3.074
skill_dribbling              42.5374386    21.0434544   2.021
power_stamina               -79.3460188    11.3611462  -6.984
Mentality                    56.0530144    19.0449532   2.943
defending_sliding_tackle     29.3779506     9.3463343   3.143
                                     Pr(>|t|)
(Intercept)                          0.002244 **
value_eur                < 0.0000000000000002 ***
age                      < 0.0000000000000002 ***
height_cm                            0.129260
shooting                             0.000209 ***
VolleySQ                             0.002116 **
skill_dribbling                      0.043269 *
power_stamina              0.00000000000309 ***
Mentality                            0.003257 **
defending_sliding_tackle             0.001677 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10520 on 8493 degrees of freedom
Multiple R-squared:  0.7424,    Adjusted R-squared:  0.7421
F-statistic:  2719 on 9 and 8493 DF,  p-value: < 0.00000000000000022
```

v.

E. <u>*Fourth Multicollinearity Check:*</u>

i. <u>*vif(bstep4)*</u>

```
> vif(bstep4)
               value_eur                      age
                  1.6159                   1.4565
               height_cm                 shooting
                  1.1172                   4.7822
                VolleySQ          skill_dribbling
                  3.1047                   2.8288
           power_stamina                Mentality
                  1.4083                   2.3303
defending_sliding_tackle
                  1.8762
```

ii.

F. <u>*Fifth Iteration of the Model:*</u>

i. *wageprediction5 <- lm(wage_eur ~ value_eur + age + shooting + VolleySQ + skill_dribbling + power_stamina + Mentality + defending_sliding_tackle, data = Midfielders)*

ii. *summary(wageprediction5)*

```
> summary(wageprediction5)

Call:
lm(formula = wage_eur ~ value_eur + age + shooting + VolleySQ +
    skill_dribbling + power_stamina + Mentality + defending_sliding_
    data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128607   -2258    -221    1673  195830

Coefficients:
                          Estimate      Std. Error t value
(Intercept)            -6367.72477044  1277.16331348  -4.986
value_eur                  0.00308412     0.00002542 121.345
age                      380.03339122    31.09832314  12.220
shooting                 -83.13563030    23.01349392  -3.612
VolleySQ                   0.48057507     0.15585226   3.084
skill_dribbling           35.27834749    20.49387204   1.721
power_stamina            -79.02232146    11.36001316  -6.956
Mentality                 58.47266290    18.97951339   3.081
defending_sliding_tackle  30.86010741     9.29585040   3.320
                                 Pr(>|t|)
(Intercept)                0.00000062915579 ***
value_eur              < 0.0000000000000002 ***
age                    < 0.0000000000000002 ***
shooting                       0.000305 ***
VolleySQ                       0.002052 **
skill_dribbling                0.085213 .
power_stamina              0.00000000000376 ***
Mentality                      0.002071 **
defending_sliding_tackle       0.000905 ***
```

iii.

```
Coefficients:
                              Estimate     Std. Error t value
(Intercept)            -6367.72477044 1277.16331348  -4.986
value_eur                  0.00308412    0.00002542 121.345
age                      380.03339122   31.09832314  12.220
shooting                 -83.13563030   23.01349392  -3.612
VolleySQ                   0.48057507    0.15585226   3.084
skill_dribbling           35.27834749   20.49387204   1.721
power_stamina            -79.02232146   11.36001316  -6.956
Mentality                 58.47266290   18.97951339   3.081
defending_sliding_tackle  30.86010741    9.29585040   3.320
                                   Pr(>|t|)
(Intercept)                0.00000062915579 ***
value_eur              < 0.0000000000000002 ***
age                    < 0.0000000000000002 ***
shooting                           0.000305 ***
VolleySQ                           0.002052 **
skill_dribbling                    0.085213 .
power_stamina              0.00000000000376 ***
Mentality                          0.002071 **
defending_sliding_tackle           0.000905 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10520 on 8494 degrees of freedom
Multiple R-squared:  0.7423,    Adjusted R-squared:  0.742
F-statistic:  3058 on 8 and 8494 DF,  p-value: < 0.00000000000000022
```

iv.

v. *bstep5 <- stepAIC(wageprediction5, direction = "backward")*

vi. *summary(bstep5)*

```
> bstep5 <- stepAIC(wageprediction5, direction = "backward")
Start:  AIC=157495.7
wage_eur ~ value_eur + age + shooting + VolleySQ + skill_dribbling +
    power_stamina + Mentality + defending_sliding_tackle

                           Df     Sum of Sq           RSS    AIC
<none>                                       939319500770 157496
- skill_dribbling           1     327694769  939647195539 157497
- Mentality                 1    1049631076  940369131846 157503
- VolleySQ                  1    1051471123  940370971892 157503
- defending_sliding_tackle  1    1218758610  940538259380 157505
- shooting                  1    1443146274  940762647044 157507
- power_stamina             1    5351098937  944670599707 157542
- age                       1   16514698609  955834199379 157642
- value_eur                 1 1628347094003 2567666594773 166044
```

vii.

```
> summary(bstep5)

Call:
lm(formula = wage_eur ~ value_eur + age + shooting + VolleySQ +
    skill_dribbling + power_stamina + Mentality + defending_sliding_tackl
    data = Midfielders)

Residuals:
    Min      1Q  Median      3Q     Max
-128607   -2258    -221    1673  195830

Coefficients:
                            Estimate     Std. Error t value
(Intercept)             -6367.72477044 1277.16331348  -4.986
value_eur                   0.00308412    0.00002542 121.345
age                       380.03339122   31.09832314  12.220
shooting                  -83.13563030   23.01349392  -3.612
VolleySQ                    0.48057507    0.15585226   3.084
skill_dribbling            35.27834749   20.49387204   1.721
power_stamina             -79.02232146   11.36001316  -6.956
Mentality                  58.47266290   18.97951339   3.081
defending_sliding_tackle   30.86010741    9.29585040   3.320
                                  Pr(>|t|)
(Intercept)                0.00000062915579 ***
value_eur               < 0.0000000000000002 ***
age                     < 0.0000000000000002 ***
shooting                           0.000305 ***
VolleySQ                           0.002052 **
skill_dribbling                    0.085213 .
power_stamina              0.00000000000376 ***
Mentality                          0.002071 **
defending_sliding_tackle           0.000905 ***
```
viii.

```
Residuals:
    Min      1Q  Median      3Q     Max
-128607   -2258    -221    1673  195830

Coefficients:
                              Estimate     Std. Error t value
(Intercept)             -6367.72477044  1277.16331348  -4.986
value_eur                   0.00308412     0.00002542 121.345
age                       380.03339122    31.09832314  12.220
shooting                  -83.13563030    23.01349392  -3.612
VolleySQ                    0.48057507     0.15585226   3.084
skill_dribbling            35.27834749    20.49387204   1.721
power_stamina             -79.02232146    11.36001316  -6.956
Mentality                  58.47266290    18.97951339   3.081
defending_sliding_tackle   30.86010741     9.29585040   3.320
                                    Pr(>|t|)
(Intercept)                  0.00000062915579 ***
value_eur                < 0.0000000000000002 ***
age                      < 0.0000000000000002 ***
shooting                             0.000305 ***
VolleySQ                             0.002052 **
skill_dribbling                      0.085213 .
power_stamina                0.00000000000376 ***
Mentality                            0.002071 **
defending_sliding_tackle             0.000905 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10520 on 8494 degrees of freedom
Multiple R-squared:  0.7423,    Adjusted R-squared:  0.742
F-statistic:  3058 on 8 and 8494 DF,  p-value: < 0.00000000000000022
```

ix.

G. *Fifth Multicollinearity Check:*

  i. *vif(bstep5)*

```
> vif(bstep5)
               value_eur                      age
                  1.6056                   1.4537
                shooting                 VolleySQ
                  4.7588                   3.1046
         skill_dribbling            power_stamina
                  2.6825                   1.4078
               Mentality defending_sliding_tackle
                  2.3140                   1.8557
```

  ii.

H. *Final Iteration of Model:*

  i. *wageprediction6 <- lm(wage_eur ~ value_eur + age + shooting + VolleySQ + power_stamina + Mentality + defending_sliding_tackle, data = Midfielders)*

ii. *summary(wageprediction6)*

```
> summary(wageprediction6)

Call:
lm(formula = wage_eur ~ value_eur + age + shooting + VolleySQ + 
    power_stamina + Mentality + defending_sliding_tackle, data = Midfielder
s)

Residuals:
    Min      1Q  Median      3Q     Max 
-129041   -2257    -226    1607  195705 

Coefficients:
                              Estimate     Std. Error t value
(Intercept)              -5096.36413808  1042.07930123  -4.891
value_eur                    0.00309604     0.00002446 126.588
age                        376.22283253    31.02302257  12.127
shooting                   -63.82919338    20.09794043  -3.176
VolleySQ                     0.49796098     0.15554263   3.201
power_stamina              -77.86913326    11.34155374  -6.866
Mentality                   56.30636771    18.93993491   2.973
defending_sliding_tackle    30.89284223     9.29690505   3.323
                                     Pr(>|t|)    
(Intercept)                0.00000102388580 ***
value_eur                < 0.0000000000000002 ***
age                      < 0.0000000000000002 ***
shooting                           0.001499 ** 
VolleySQ                           0.001372 ** 
power_stamina               0.00000000000708 ***
Mentality                          0.002958 ** 
defending_sliding_tackle           0.000895 ***
```

iii.

```
---
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-129041   -2257    -226    1607  195705 

Coefficients:
                              Estimate     Std. Error t value
(Intercept)             -5096.36413808  1042.07930123  -4.891
value_eur                   0.00309604     0.00002446 126.588
age                       376.22283253    31.02302257  12.127
shooting                  -63.82919338    20.09794043  -3.176
VolleySQ                    0.49796098     0.15554263   3.201
power_stamina             -77.86913326    11.34155374  -6.866
Mentality                  56.30636771    18.93993491   2.973
defending_sliding_tackle   30.89284223     9.29690505   3.323
                                    Pr(>|t|)    
(Intercept)                 0.00000102388580 ***
value_eur               < 0.0000000000000002 ***
age                     < 0.0000000000000002 ***
shooting                            0.001499 ** 
VolleySQ                            0.001372 ** 
power_stamina               0.00000000000708 ***
Mentality                           0.002958 ** 
defending_sliding_tackle            0.000895 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10520 on 8495 degrees of freedom
Multiple R-squared:  0.7422,	Adjusted R-squared:  0.742 
F-statistic:  3494 on 7 and 8495 DF,  p-value: < 0.00000000000000022
```

iv.

I. <u>*Final Backward Elimination of Model:*</u>

   i. *bstep6 <- stepAIC(wageprediction6, direction = "backward")*

   ii. *summary(bstep6*

```
> bstep6 <- stepAIC(wageprediction6, direction = "backward")
Start:  AIC=157496.6
wage_eur ~ value_eur + age + shooting + VolleySQ + power_stamina + 
    Mentality + defending_sliding_tackle

                           Df     Sum of Sq           RSS    AIC
<none>                                       939647195539 157497
- Mentality                 1     977596187  940624791726 157503
- shooting                  1    1115673305  940762868844 157505
- VolleySQ                  1    1133687013  940780882551 157505
- defending_sliding_tackle  1    1221350686  940868546225 157506
- power_stamina             1    5214191897  944861387436 157542
- age                       1   16267601257  955914796795 157641
- value_eur                 1 1772505459367 2712152654905 166508
```

   iii.

```
> summary(bstep6)

Call:
lm(formula = wage_eur ~ value_eur + age + shooting + VolleySQ +
    power_stamina + Mentality + defending_sliding_tackle, data = Midfielder
s)

Residuals:
    Min      1Q  Median      3Q     Max
-129041   -2257    -226    1607  195705

Coefficients:
                             Estimate     Std. Error t value
(Intercept)            -5096.36413808  1042.07930123  -4.891
value_eur                  0.00309604     0.00002446 126.588
age                      376.22283253    31.02302257  12.127
shooting                 -63.82919338    20.09794043  -3.176
VolleySQ                   0.49796098     0.15554263   3.201
power_stamina            -77.86913326    11.34155374  -6.866
Mentality                 56.30636771    18.93993491   2.973
defending_sliding_tackle  30.89284223     9.29690505   3.323
                                     Pr(>|t|)
(Intercept)                0.00000102388580 ***
value_eur              < 0.0000000000000002 ***
age                    < 0.0000000000000002 ***
shooting                           0.001499 **
VolleySQ                           0.001372 **
power_stamina              0.00000000000708 ***
Mentality                          0.002958 **
defending_sliding_tackle           0.000895 ***
```

iv.

```
---
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-129041   -2257    -226    1607  195705

Coefficients:
                             Estimate     Std. Error t value
(Intercept)            -5096.36413808  1042.07930123  -4.891
value_eur                  0.00309604     0.00002446 126.588
age                      376.22283253    31.02302257  12.127
shooting                 -63.82919338    20.09794043  -3.176
VolleySQ                   0.49796098     0.15554263   3.201
power_stamina            -77.86913326    11.34155374  -6.866
Mentality                 56.30636771    18.93993491   2.973
defending_sliding_tackle  30.89284223     9.29690505   3.323
                                    Pr(>|t|)
(Intercept)                 0.00000102388580 ***
value_eur              < 0.0000000000000002 ***
age                    < 0.0000000000000002 ***
shooting                            0.001499 **
VolleySQ                            0.001372 **
power_stamina               0.00000000000708 ***
Mentality                           0.002958 **
defending_sliding_tackle            0.000895 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10520 on 8495 degrees of freedom
Multiple R-squared:  0.7422,    Adjusted R-squared:  0.742
F-statistic:  3494 on 7 and 8495 DF,  p-value: < 0.00000000000000022
```

v.

J. <u>*Final Multicollinearity Check:*</u>

i. *vif(bstep6)*

```
> vif(bstep6)
               value_eur                      age
                  1.4864                   1.4464
                shooting                 VolleySQ
                  3.6286                   3.0915
           power_stamina                Mentality
                  1.4029                   2.3038
defending_sliding_tackle
                  1.8557
```

ii.

K. <u>*Final Model:*</u>

a. *wageprediction6 <- lm(wage_eur ~ value_eur + age + shooting + VolleySQ + power_stamina + Mentality + defending_sliding_tackle, data = Midfielders)*

b. *summary(wageprediction6)*

Conclusion

To Conclude, there was improvement in the midfielder model as now it looks cleaner, and more irrelevant variables were removed, there is no multicollinearity for the model as well. The findings are that mainly the midfield soccer player has to maintain a high market value (value_eur) as they get older, while also being able to score volleys, maintain stamina, and a leadership like mentality. This was determined through the regression models above and the following of the procedure step by step. Whether the variables were removed by backward elimination, having a high multicollinearity, or have a *P* value that was not significant, while also the constant updating of the model.

c. Hasmitha Chatla - Defenders

The salary of a soccer player depends on several specific things, including the abilities of the team, past season results, age, development trend and personality etc. On this basis, football players' wages are decided through agreements among the team managers and the lawyers. In our project, the biggest concern with today's Fifa is that as a player becomes old, they choose to quit instead of wanting to pursue a new role in the team that is easier for them. We are trying to work out what significant variables cause the salary (wage eur) change for our data, which implies the wage eur is a dependent variable.

Players in Defense play according to their own target. The players ought to clear the ball, stop a strike from the opposition, and have the skill, heading capacity and pace to transfer. Depending on the main job of the defender and the abilities or skills they need to obtain the salary that is possibly determined by the quality of their defense skills.

What I figured out while building this model is that the first iteration might not be the final iteration. Despite that being said, the phase I took part in was a series of trials and errors. The first model I developed was a complete model, with the dependent variable being wage_eur and the other numerical variables being explanatory variables. The first iteration model code is shown below, with a screen recording of the results:

First Iteration Model (full model):

- *wprediction1 <- lm(wage_eur ~ value_eur + age + height_cm + weight_kg + overall + potential + pace + shooting + passing + dribbling + defending + physic + attacking_crossing + attacking_finishing + attacking_heading_accuracy + attacking_short_passing + attacking_volleys + skill_dribbling + skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_control + movement_acceleration + movement_sprint_speed + movement_agility + movement_reactions + movement_balance + power_shot_power + power_jumping + power_stamina + power_strength + power_long_shots + mentality_aggression + mentality_interceptions + mentality_positioning + mentality_vision + mentality_penalties + mentality_composure + defending_marking + defending_standing_tackle + defending_sliding_tackle, data = defenders_final)*
- *summary(wprediction1)*

```
Residuals:
   Min     1Q Median     3Q    Max
-89090  -2504   -343   1807 187511

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -4.358e+04  7.459e+03  -5.842 5.40e-09 ***
value_eur                   3.205e-03  3.898e-05  82.228  < 2e-16 ***
age                         4.265e+02  6.235e+01   6.840 8.65e-12 ***
height_cm                   1.308e+02  3.774e+01   3.465 0.000533 ***
weight_kg                   1.017e+01  3.123e+01   0.326 0.744812
overall                    -4.700e+01  8.671e+01  -0.542 0.587793
potential                   1.358e+02  5.132e+01   2.647 0.008145 **
pace                        3.367e+02  4.050e+02   0.831 0.405870
shooting                    2.162e+02  4.077e+02   0.530 0.595879
passing                     2.075e+02  4.097e+02   0.506 0.612633
dribbling                  -2.136e+02  4.120e+02  -0.518 0.604136
defending                  -9.494e+02  4.124e+02  -2.302 0.021358 *
physic                      3.970e+02  4.155e+02   0.956 0.339326
attacking_crossing         -4.626e+00  8.371e+01  -0.055 0.955935
attacking_finishing        -1.066e+02  1.842e+02  -0.579 0.562649
attacking_heading_accuracy  1.118e+02  4.711e+01   2.373 0.017666 *
attacking_short_passing    -8.917e+01  1.461e+02  -0.610 0.541638
attacking_volleys           4.800e+01  2.618e+01   1.834 0.066752 .
skill_dribbling             9.917e+01  2.067e+02   0.480 0.631468
skill_curve                 1.555e+01  2.540e+01   0.612 0.540402
skill_fk_accuracy          -6.037e+01  2.524e+01  -2.391 0.016813 *
skill_long_passing         -7.140e+01  6.472e+01  -1.103 0.269977
skill_ball_control          7.104e+01  1.268e+02   0.560 0.575323
movement_acceleration      -1.860e+02  1.829e+02  -1.017 0.309314
movement_sprint_speed      -1.862e+02  2.245e+02  -0.829 0.406862
movement_agility            3.428e+01  4.487e+01   0.764 0.444906
movement_reactions          1.564e+01  3.738e+01   0.418 0.675732
movement_balance            2.718e+01  2.750e+01   0.988 0.323104
power_shot_power           -6.074e+01  8.311e+01  -0.731 0.464910
power_jumping               9.709e+00  2.442e+01   0.398 0.690877
power_stamina              -1.709e+02  1.049e+02  -1.629 0.103289
power_strength             -2.681e+02  2.091e+02  -1.282 0.199795
power_long_shots           -3.767e+01  8.345e+01  -0.451 0.651690
mentality_aggression       -7.900e+01  8.458e+01  -0.934 0.350301
mentality_interceptions     1.809e+02  8.811e+01   2.053 0.040131 *
mentality_positioning       1.569e+01  2.623e+01   0.598 0.549819
mentality_vision            9.133e+00  8.393e+01   0.109 0.913358
mentality_penalties         2.017e+01  2.602e+01   0.775 0.438295
mentality_composure        -1.396e+01  2.202e+01  -0.634 0.526265
defending_marking           2.349e+02  1.276e+02   1.841 0.065594 .
defending_standing_tackle   3.572e+02  1.325e+02   2.696 0.007026 **
defending_sliding_tackle    1.917e+02  5.884e+01   3.258 0.001129 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9505 on 6479 degrees of freedom
Multiple R-squared:  0.7214,	Adjusted R-squared:  0.7196
F-statistic: 409.2 on 41 and 6479 DF,  p-value: < 2.2e-16

>
```

After this I have done backward elimination on the model to check the multicollinearity. The code used is mentioned below.

<u>First Backward Elimination and Multicollinearity Check:</u>

- *bstep1 <- stepAIC(wprediction1, direction = "backward")*
- *summary(bstep1)*
- *vif(bstep1)*

```
> bstep1 <- stepAIC(wprediction1, direction = "backward")
Start:  AIC=119501.1
wage_eur ~ value_eur + age + height_cm + weight_kg + overall +
    potential + pace + shooting + passing + dribbling + defending +
    physic + attacking_crossing + attacking_finishing + attacking_heading_accuracy +
    attacking_short_passing + attacking_volleys + skill_dribbling +
    skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_control +
    movement_acceleration + movement_sprint_speed + movement_agility +
    movement_reactions + movement_balance + power_shot_power +
    power_jumping + power_stamina + power_strength + power_long_shots +
    mentality_aggression + mentality_interceptions + mentality_positioning +
    mentality_vision + mentality_penalties + mentality_composure +
    defending_marking + defending_standing_tackle + defending_sliding_tackle

                             Df  Sum of Sq        RSS    AIC
- attacking_crossing          1 2.7586e+05 5.8535e+11 119499
- mentality_vision            1 1.0696e+06 5.8536e+11 119499
- weight_kg                   1 9.5723e+06 5.8536e+11 119499
- power_jumping               1 1.4288e+07 5.8537e+11 119499
- movement_reactions          1 1.5809e+07 5.8537e+11 119499
- power_long_shots            1 1.8412e+07 5.8537e+11 119499
- skill_dribbling             1 2.0788e+07 5.8538e+11 119499
- passing                     1 2.3163e+07 5.8538e+11 119499
- dribbling                   1 2.4288e+07 5.8538e+11 119499
- shooting                    1 2.5413e+07 5.8538e+11 119499
- overall                     1 2.6547e+07 5.8538e+11 119499
- skill_ball_control          1 2.8359e+07 5.8538e+11 119499
- attacking_finishing         1 3.0282e+07 5.8538e+11 119499
- mentality_positioning       1 3.2315e+07 5.8539e+11 119499
- attacking_short_passing     1 3.3659e+07 5.8539e+11 119499
- skill_curve                 1 3.3865e+07 5.8539e+11 119499
- mentality_composure         1 3.6287e+07 5.8539e+11 119500
- power_shot_power            1 4.8255e+07 5.8540e+11 119500
- movement_agility            1 5.2732e+07 5.8541e+11 119500
- mentality_penalties         1 5.4282e+07 5.8541e+11 119500
- movement_sprint_speed       1 6.2162e+07 5.8542e+11 119500
- pace                        1 6.2425e+07 5.8542e+11 119500
- mentality_aggression        1 7.8827e+07 5.8543e+11 119500
- physic                      1 8.2496e+07 5.8544e+11 119500
- movement_balance            1 8.8222e+07 5.8544e+11 119500
- movement_acceleration       1 9.3397e+07 5.8545e+11 119500
- skill_long_passing          1 1.0996e+08 5.8546e+11 119500
- power_strength              1 1.4855e+08 5.8550e+11 119501
<none>                                     5.8535e+11 119501
- power_stamina               1 2.3985e+08 5.8559e+11 119502
- attacking_volleys           1 3.0377e+08 5.8566e+11 119503
- defending_marking           1 3.0637e+08 5.8566e+11 119503
- mentality_interceptions     1 3.8072e+08 5.8574e+11 119503
- defending                   1 4.7883e+08 5.8583e+11 119504
- attacking_heading_accuracy  1 5.0881e+08 5.8586e+11 119505
- skill_fk_accuracy           1 5.1667e+08 5.8587e+11 119505
- potential                   1 6.3294e+08 5.8599e+11 119506
- defending_standing_tackle   1 6.5690e+08 5.8601e+11 119506
- defending_sliding_tackle    1 9.5876e+08 5.8631e+11 119510
- height_cm                   1 1.0850e+09 5.8644e+11 119511
- age                         1 4.2267e+09 5.8958e+11 119546
- value_eur                   1 6.1087e+11 1.1962e+12 124160
```

```
Step:  AIC=119470.3
wage_eur ~ value_eur + age + height_cm + potential + passing +
    defending + physic + attacking_heading_accuracy + attacking_short_passing +
    attacking_volleys + skill_fk_accuracy + skill_long_passing +
    movement_acceleration + power_stamina + power_strength +
    mentality_aggression + mentality_interceptions + mentality_positioning +
    mentality_penalties + defending_marking + defending_standing_tackle +
    defending_sliding_tackle

                             Df  Sum of Sq        RSS    AIC
<none>                                     5.8600e+11 119470
- mentality_positioning       1 1.8984e+08 5.8619e+11 119470
- mentality_penalties         1 2.8060e+08 5.8628e+11 119471
- defending_marking           1 2.8378e+08 5.8628e+11 119471
- movement_acceleration       1 3.6270e+08 5.8636e+11 119472
- mentality_interceptions     1 3.6734e+08 5.8636e+11 119472
- attacking_heading_accuracy  1 4.4336e+08 5.8644e+11 119473
- defending                   1 4.5613e+08 5.8645e+11 119473
- mentality_aggression        1 6.1038e+08 5.8661e+11 119475
- defending_standing_tackle   1 6.1166e+08 5.8661e+11 119475
- potential                   1 6.6398e+08 5.8666e+11 119476
- physic                      1 6.8913e+08 5.8668e+11 119476
- attacking_short_passing     1 8.3350e+08 5.8683e+11 119478
- defending_sliding_tackle    1 9.5230e+08 5.8695e+11 119479
- power_strength              1 1.0326e+09 5.8703e+11 119480
- skill_long_passing          1 1.1165e+09 5.8711e+11 119481
- attacking_volleys           1 1.2707e+09 5.8727e+11 119482
- height_cm                   1 1.2894e+09 5.8728e+11 119483
- power_stamina               1 1.3623e+09 5.8736e+11 119483
- skill_fk_accuracy           1 1.6333e+09 5.8763e+11 119486
- passing                     1 1.9307e+09 5.8793e+11 119490
- age                         1 5.6503e+09 5.9165e+11 119531
- value_eur                   1 6.8207e+11 1.2681e+12 124502
>
```

```
> vif(bstep1)
                value_eur                       age                height_cm
                 1.902104                  3.736008                 2.571949
                potential                   passing                defending
                 4.353137                 20.491024               725.285593
                   physic attacking_heading_accuracy   attacking_short_passing
               192.081701                 16.649329                 8.222923
        attacking_volleys            skill_fk_accuracy        skill_long_passing
                 2.212060                  2.705626                 5.164842
    movement_acceleration              power_stamina           power_strength
                 2.460743                 25.938553               100.801574
     mentality_aggression   mentality_interceptions     mentality_positioning
                17.887481                 40.970186                 3.753001
      mentality_penalties          defending_marking  defending_standing_tackle
                 1.982808                 86.203532                73.624339
  defending_sliding_tackle
                15.234947
> |
```

From the above, we can see that there is a lot of multicollinearity in the model which is not a good model. I have done backward elimination on the model. Any variable which is higher than 10 was removed from the model.

After this I have eliminated few variables which are:

passing	*weight_kg*	*pace*
dribbling	*potential*	*shooting*
defending	*attacking_crossing*	*attacking_finishing*
physic	*defending_sliding_tackle*	*defending_standing_tackle*
overall	*mentality_composure*	*defending_marking*
mentality_vision	*mentality_interceptions*	*mentality_aggression*
power_long_shots	*power_strength*	*power_stamina*
power_jumping	*power_shot_power*	*movement_balance*

movement_reactions	*movement_agility*	*movement_sprint_speed*
skill_ball_control	*skill_dribbling*	*skill_curve*

I have built the second model as there were unnecessary variables which I have eliminated due to multicollinearity and also due to p-value. The code used for second iteration of the model is mentioned below:

<u>*Second Iteration of the Model:*</u>

- *wprediction2 <- lm(wage_eur ~ value_eur + age + height_cm + potential + attacking_short_passing + attacking_volleys +skill_fk_accuracy + skill_long_passing + movement_acceleration + mentality_positioning +mentality_penalties , data = defenders_final)*
- *summary(wprediction2)*

```
>
> wprediction2 <- lm(wage_eur ~ value_eur + age + height_cm + potential +
+                        attacking_short_passing + attacking_volleys +
+                        skill_fk_accuracy + skill_long_passing +
+                        movement_acceleration + mentality_positioning +
+                        mentality_penalties , data = defenders_final)
> summary(wprediction2)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + potential +
    attacking_short_passing + attacking_volleys + skill_fk_accuracy +
    skill_long_passing + movement_acceleration + mentality_positioning +
    mentality_penalties, data = defenders_final)

Residuals:
   Min     1Q Median     3Q    Max
-89674  -2439   -378   1691 188096

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -2.509e+04  5.343e+03  -4.697 2.70e-06 ***
value_eur                3.227e-03  3.598e-05  89.700  < 2e-16 ***
age                      4.221e+02  3.761e+01  11.222  < 2e-16 ***
height_cm                1.889e+01  2.467e+01   0.766 0.443879
potential                1.748e+02  3.283e+01   5.324 1.05e-07 ***
attacking_short_passing -4.352e+00  2.513e+01  -0.173 0.862500
attacking_volleys        5.849e+01  1.450e+01   4.032 5.59e-05 ***
skill_fk_accuracy       -3.412e+01  1.265e+01  -2.698 0.006995 **
skill_long_passing      -3.625e+01  1.903e+01  -1.904 0.056906 .
movement_acceleration   -2.894e+01  1.350e+01  -2.144 0.032095 *
mentality_positioning    4.956e+01  1.292e+01   3.835 0.000127 ***
mentality_penalties      3.468e+01  1.562e+01   2.220 0.026437 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9570 on 6509 degrees of freedom
Multiple R-squared:  0.7163,	Adjusted R-squared:  0.7158
F-statistic:  1494 on 11 and 6509 DF,  p-value: < 2.2e-16

> |
```

After this I have again done backward elimination on the model to check the multicollinearity. The code used is mentioned below. The following results are:

<u>Second Backward Elimination and Multicollinearity Check:</u>

- *bstep2 <- stepAIC(wprediction2, direction = "backward")*
- *summary(bstep2)*
- *vif(bstep2)*

```
> bstep2 <- stepAIC(wprediction2, direction = "backward")
Start:  AIC=119560.4
wage_eur ~ value_eur + age + height_cm + potential + attacking_short_passing +
    attacking_volleys + skill_fk_accuracy + skill_long_passing +
    movement_acceleration + mentality_positioning + mentality_penalties

                          Df  Sum of Sq        RSS    AIC
- attacking_short_passing  1 2.7475e+06 5.9617e+11 119558
- height_cm                1 5.3700e+07 5.9622e+11 119559
<none>                                  5.9616e+11 119560
- skill_long_passing       1 3.3216e+08 5.9650e+11 119562
- movement_acceleration    1 4.2090e+08 5.9659e+11 119563
- mentality_penalties      1 4.5149e+08 5.9662e+11 119563
- skill_fk_accuracy        1 6.6668e+08 5.9683e+11 119566
- mentality_positioning    1 1.3472e+09 5.9751e+11 119573
- attacking_volleys        1 1.4892e+09 5.9765e+11 119575
- potential                1 2.5964e+09 5.9876e+11 119587
- age                      1 1.1535e+10 6.0770e+11 119683
- value_eur                1 7.3694e+11 1.3331e+12 124806

Step:  AIC=119558.5
wage_eur ~ value_eur + age + height_cm + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    mentality_positioning + mentality_penalties

                        Df  Sum of Sq        RSS    AIC
- height_cm              1 5.5514e+07 5.9622e+11 119557
<none>                                5.9617e+11 119558
- movement_acceleration  1 4.2526e+08 5.9659e+11 119561
- mentality_penalties    1 4.5125e+08 5.9662e+11 119561
- skill_long_passing     1 6.6323e+08 5.9683e+11 119564
- skill_fk_accuracy      1 6.6711e+08 5.9683e+11 119564
- mentality_positioning  1 1.3649e+09 5.9753e+11 119571
- attacking_volleys      1 1.4875e+09 5.9766e+11 119573
- potential              1 2.8643e+09 5.9903e+11 119588
- age                    1 1.2186e+10 6.0835e+11 119688
- value_eur              1 7.3714e+11 1.3333e+12 124805

Step:  AIC=119557.1
wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    mentality_positioning + mentality_penalties

                        Df  Sum of Sq        RSS    AIC
<none>                                5.9622e+11 119557
- mentality_penalties    1 4.4821e+08 5.9667e+11 119560
- skill_long_passing     1 6.4853e+08 5.9687e+11 119562
- movement_acceleration  1 6.6687e+08 5.9689e+11 119562
- skill_fk_accuracy      1 6.9325e+08 5.9692e+11 119563
- mentality_positioning  1 1.3096e+09 5.9753e+11 119569
- attacking_volleys      1 1.5039e+09 5.9773e+11 119572
- potential              1 3.0178e+09 5.9924e+11 119588
- age                    1 1.2197e+10 6.0842e+11 119687
- value_eur              1 7.4108e+11 1.3373e+12 124823
> summary(bstep2
```

```
> vif(bstep2)
            value_eur                   age             potential
             1.785967              1.912901              2.312111
    attacking_volleys     skill_fk_accuracy    skill_long_passing
             2.130340              2.179739              2.127128
movement_acceleration mentality_positioning   mentality_penalties
             1.678222              2.367975              1.954633
>
```

From the above, we can see that the backward elimination is still eliminating variables and that wprediction2 is not the final model for this portion of the project, this means we continue the process of trying to figure out the best model for the defender portion of the data.
I have built a third and final model for the data, as there were unnecessary variables removed from our second model due to the backward elimination stage. The code and third model along with a backward elimination test as well as variance inflation test of multicollinearity is found down below.

Third Iteration of Model:

- *wprediction3 <- lm(wage_eur ~ value_eur + age + potential + attacking_volleys + skill_fk_accuracy + skill_long_passing + movement_acceleration + mentality_positioning + mentality_penalties, data = defenders_final)*
- *summary(wprediction3)*

```
Residuals:
   Min     1Q Median     3Q    Max 
-89636  -2436   -374   1701 188139 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)           -2.145e+04  2.606e+03  -8.232  < 2e-16 ***
value_eur              3.229e-03  3.589e-05  89.961  < 2e-16 ***
age                    4.207e+02  3.645e+01  11.541  < 2e-16 ***
potential              1.760e+02  3.066e+01   5.741 9.85e-09 ***
attacking_volleys      5.875e+01  1.450e+01   4.052 5.13e-05 ***
skill_fk_accuracy     -3.473e+01  1.262e+01  -2.751 0.005949 ** 
skill_long_passing    -3.796e+01  1.427e+01  -2.661 0.007804 ** 
movement_acceleration -3.329e+01  1.233e+01  -2.699 0.006981 ** 
mentality_positioning  4.706e+01  1.244e+01   3.782 0.000157 ***
mentality_penalties    3.455e+01  1.562e+01   2.212 0.026975 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9569 on 6511 degrees of freedom
Multiple R-squared:  0.7162,	Adjusted R-squared:  0.7158 
F-statistic:  1826 on 9 and 6511 DF,  p-value: < 2.2e-16
```

Third Backward Elimination check:

- *bstep3 <- stepAIC(wprediction3, direction = "backward")*
- *summary(bstep3)*

```
> bstep3 <- stepAIC(wprediction3, direction = "backward")
Start:  AIC=119557.1
wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    mentality_positioning + mentality_penalties

                        Df  Sum of Sq        RSS    AIC
<none>                                5.9622e+11 119557
- mentality_penalties    1 4.4821e+08 5.9667e+11 119560
- skill_long_passing     1 6.4853e+08 5.9687e+11 119562
- movement_acceleration  1 6.6687e+08 5.9689e+11 119562
- skill_fk_accuracy      1 6.9325e+08 5.9692e+11 119563
- mentality_positioning  1 1.3096e+09 5.9753e+11 119569
- attacking_volleys      1 1.5039e+09 5.9773e+11 119572
- potential              1 3.0178e+09 5.9924e+11 119588
- age                    1 1.2197e+10 6.0842e+11 119687
- value_eur              1 7.4108e+11 1.3373e+12 124823

Residuals:
   Min     1Q Median     3Q    Max
-89636  -2436   -374   1701 188139

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           -2.145e+04  2.606e+03  -8.232  < 2e-16 ***
value_eur              3.229e-03  3.589e-05  89.961  < 2e-16 ***
age                    4.207e+02  3.645e+01  11.541  < 2e-16 ***
potential              1.760e+02  3.066e+01   5.741 9.85e-09 ***
attacking_volleys      5.875e+01  1.450e+01   4.052 5.13e-05 ***
skill_fk_accuracy     -3.473e+01  1.262e+01  -2.751 0.005949 **
skill_long_passing    -3.796e+01  1.427e+01  -2.661 0.007804 **
movement_acceleration -3.329e+01  1.233e+01  -2.699 0.006981 **
mentality_positioning  4.706e+01  1.244e+01   3.782 0.000157 ***
mentality_penalties    3.455e+01  1.562e+01   2.212 0.026975 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9569 on 6511 degrees of freedom
Multiple R-squared:  0.7162,    Adjusted R-squared:  0.7158
F-statistic:  1826 on 9 and 6511 DF,  p-value: < 2.2e-16
```

Third Multicollinearity Test:

- *vif(bstep3)*

```
> vif(bstep3)
            value_eur                   age             potential
             1.785967              1.912901              2.312111
     attacking_volleys     skill_fk_accuracy    skill_long_passing
             2.130340              2.179739              2.127128
movement_acceleration mentality_positioning    mentality_penalties
             1.678222              2.367975              1.954633
```

In summary, it appears that the significant variables which will cause a change in the weekly salary (wage_eur) for defenders as they get older are market value (value_eur), age, the potential they have for growth, abilities to score a bicycle kick or similar volley goals (attacking_volleys), how skilled they are as a freekick taker (skill_fk_accuracy), how skilled they are as a long passer which can change the play from defense to attack (skill_long_passing), their movement_acceleration, and the mentalities they have while they are positioning themselves in defense and taking penalties, meaning they have to keep cool calm and collected as they get older and show that they are experienced leader. A great example of such a player is Sergio Ramos, as he has been playing in defense for over 15 years and is able to provide all of the above, while also maintaining a constant great market value (value_eur) as he gets older. Through the process of building a model, doing backward elimination on the model at hand, checking for multicollinearity, and then building a better model, I have built the perfect model for this portion of our project, that model being *wprediction3*.

To improve the final model I have built, I cleaned up our defender model using two Interaction Terms, one Second Order Terms, and how the previous defender model produced the same results, but this model looks cleaner.

INTERACTION TERM:
#Interaction Term added to model and removed mentality variables

Two Interaction terms:

- The defender interaction term consists of a combination of mentality variables as well as movement variables.The mentality interaction was produced by previous variables of mentality_positioning, mentality_penalties.
    - *Defenders$ingame_mentality = ((Defenders$mentality_positioning + Defenders$mentality_penalties)/2)*

```
> Defenders$ingame_mentality = ((Defenders$mentality_positioning + Defenders$mentalit
alties)/2)
```

- The movement interaction was produced by pervious variables of defendersmovement_acceleration
    - *Defenders$MovementSQ <- (Defenders$movement_acceleration)^2*

```
> Defenders$MovementSQ <- (Defenders$movement_acceleration)^2
```

SECOND ORDER ANALYSIS:

I have checked for the second order terms. I think defenders movement can be used as a second order term as I think there is a relationship between defenders movement wage. The second order term for the defenders consists of defenders movement acceleration being squared.

#Second order term being skilled long passing

- *Defenders$MovementSQ <- (Defenders$movement_acceleration)^2*

MULTICOLLINEARITY CHECK:

I have checked for the multicollinearity in the bstep4 and bstep5.

- vif(bstep4)
- vif(bstep5)

```
> vif(bstep4)
            value_eur                   age             potential
                 1.79                  1.91                  2.29
      attacking_volleys      skill_fk_accuracy     skill_long_passing
                 2.12                  2.12                  2.10
movement_acceleration        ingame_mentality
                 1.49                  3.01
```

```
> vif(bstep5)
         value_eur                age          potential   attacking_volleys
              1.78               1.87               2.28                2.12
 skill_fk_accuracy skill_long_passing         MovementSQ    ingame_mentality
              2.12               2.10               1.50                3.05
```

Final Model:

- *wprediction5 <- lm(wage_eur ~ value_eur + age + potential + attacking_volleys + skill_fk_accuracy + skill_long_passing + MovementSQ + ingame_mentality, data= Defenders)*
- *summary(wprediction5)*

```
> #Interaction Term added to model and removed mentality variables
> Defenders$ingame_mentality = ((Defenders$mentality_positioning + Defenders$mentality_pen
alties)/2)
> wprediction4 <- lm(wage_eur ~ value_eur + age + potential +
+                      attacking_volleys + skill_fk_accuracy +
+                      skill_long_passing + movement_acceleration +
+                      ingame_mentality,
+                    data = Defenders)
> summary(wprediction4)

Call:
lm(formula = wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    ingame_mentality, data = Defenders)

Residuals:
   Min     1Q Median     3Q    Max
-89765  -2434   -390   1706 187963

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -2.15e+04   2.60e+03   -8.28  < 2e-16 ***
value_eur              3.23e-03   3.59e-05   90.00  < 2e-16 ***
age                    4.20e+02   3.64e+01   11.53  < 2e-16 ***
potential              1.74e+02   3.05e+01    5.71  1.2e-08 ***
attacking_volleys      5.82e+01   1.45e+01    4.02  5.8e-05 ***
skill_fk_accuracy     -3.60e+01   1.24e+01   -2.90   0.0038 **
skill_long_passing    -3.70e+01   1.42e+01   -2.61   0.0090 **
movement_acceleration -3.08e+01   1.16e+01   -2.65   0.0082 **
ingame_mentality       8.43e+01   1.90e+01    4.44  9.2e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9570 on 6512 degrees of freedom
Multiple R-squared:  0.716,     Adjusted R-squared:  0.716
F-statistic: 2.05e+03 on 8 and 6512 DF,  p-value: <2e-16
```

In this I have included the variables like wage_eur, age, potential, value_eur, attacking_volleys, sill_fk_accuracy, sill_passing, movement_acceleration etc. in this we got the adjusted R2 71.6% which is high and the t value of value_eur is extremely high which says that the transfer value of the player depends the wage of the player.

```
> bstep4 <- stepAIC(wprediction4, direction = "backward")
Start:  AIC=119555
wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    ingame_mentality

                        Df Sum of Sq      RSS    AIC
<none>                               5.96e+11 119555
- skill_long_passing     1  6.24e+08 5.97e+11 119560
- movement_acceleration  1  6.41e+08 5.97e+11 119560
- skill_fk_accuracy      1  7.68e+08 5.97e+11 119562
- attacking_volleys      1  1.48e+09 5.98e+11 119570
- ingame_mentality       1  1.80e+09 5.98e+11 119573
- potential              1  2.98e+09 5.99e+11 119586
- age                    1  1.22e+10 6.08e+11 119685
- value_eur              1  7.42e+11 1.34e+12 124824
```

```
> summary(bstep4)

Call:
lm(formula = wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + movement_acceleration +
    ingame_mentality, data = Defenders)

Residuals:
   Min     1Q Median     3Q    Max
-89765  -2434   -390   1706 187963

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -2.15e+04   2.60e+03   -8.28  < 2e-16 ***
value_eur              3.23e-03   3.59e-05   90.00  < 2e-16 ***
age                    4.20e+02   3.64e+01   11.53  < 2e-16 ***
potential              1.74e+02   3.05e+01    5.71  1.2e-08 ***
attacking_volleys      5.82e+01   1.45e+01    4.02  5.8e-05 ***
skill_fk_accuracy     -3.60e+01   1.24e+01   -2.90   0.0038 **
skill_long_passing    -3.70e+01   1.42e+01   -2.61   0.0090 **
movement_acceleration -3.08e+01   1.16e+01   -2.65   0.0082 **
ingame_mentality       8.43e+01   1.90e+01    4.44  9.2e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9570 on 6512 degrees of freedom
Multiple R-squared:  0.716,     Adjusted R-squared:  0.716
F-statistic: 2.05e+03 on 8 and 6512 DF,  p-value: <2e-16

> vif(bstep4)
            value_eur                   age             potential
                 1.79                  1.91                  2.29
    attacking_volleys     skill_fk_accuracy    skill_long_passing
                 2.12                  2.12                  2.10
movement_acceleration      ingame_mentality
                 1.49                  3.01
```

```
> Defenders$MovementSQ <- (Defenders$movement_acceleration)^2
> wprediction5 <- lm(wage_eur ~ value_eur + age + potential +
+                     attacking_volleys + skill_fk_accuracy +
+                     skill_long_passing + MovementSQ +
+                     ingame_mentality,
+                   data = Defenders)
> summary(wprediction5)

Call:
lm(formula = wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + MovementSQ + ingame_mentality,
    data = Defenders)

Residuals:
   Min     1Q Median     3Q    Max
-89707  -2435   -384   1700 187891

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -2.24e+04   2.48e+03   -9.03  < 2e-16 ***
value_eur           3.23e-03   3.59e-05   90.04  < 2e-16 ***
age                 4.18e+02   3.60e+01   11.61  < 2e-16 ***
potential           1.75e+02   3.04e+01    5.74  1.0e-08 ***
attacking_volleys   5.83e+01   1.45e+01    4.03  5.6e-05 ***
skill_fk_accuracy  -3.63e+01   1.24e+01   -2.92   0.0035 **
skill_long_passing -3.74e+01   1.42e+01   -2.64   0.0083 **
MovementSQ         -2.85e-01   9.45e-02   -3.01   0.0026 **
ingame_mentality    8.77e+01   1.91e+01    4.60  4.4e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9570 on 6512 degrees of freedom
Multiple R-squared:  0.716,	Adjusted R-squared:  0.716
F-statistic: 2.06e+03 on 8 and 6512 DF,  p-value: <2e-16
```

```
> bstep5 <- stepAIC(wprediction5, direction = "backward")
Start:  AIC=119553
wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + MovementSQ + ingame_mentality

                     Df Sum of Sq      RSS    AIC
<none>                            5.96e+11 119553
- skill_long_passing  1  6.38e+08 5.97e+11 119558
- skill_fk_accuracy   1  7.79e+08 5.97e+11 119560
- MovementSQ          1  8.31e+08 5.97e+11 119560
- attacking_volleys   1  1.49e+09 5.98e+11 119568
- ingame_mentality    1  1.93e+09 5.98e+11 119573
- potential           1  3.01e+09 5.99e+11 119584
- age                 1  1.23e+10 6.08e+11 119685
- value_eur           1  7.42e+11 1.34e+12 124825
```

```
> summary(bstep5)

Call:
lm(formula = wage_eur ~ value_eur + age + potential + attacking_volleys +
    skill_fk_accuracy + skill_long_passing + MovementSQ + ingame_mentality,
    data = Defenders)

Residuals:
   Min     1Q Median     3Q    Max
-89707  -2435   -384   1700 187891

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -2.24e+04   2.48e+03   -9.03  < 2e-16 ***
value_eur           3.23e-03   3.59e-05   90.04  < 2e-16 ***
age                 4.18e+02   3.60e+01   11.61  < 2e-16 ***
potential           1.75e+02   3.04e+01    5.74  1.0e-08 ***
attacking_volleys   5.83e+01   1.45e+01    4.03  5.6e-05 ***
skill_fk_accuracy  -3.63e+01   1.24e+01   -2.92   0.0035 **
skill_long_passing -3.74e+01   1.42e+01   -2.64   0.0083 **
MovementSQ         -2.85e-01   9.45e-02   -3.01   0.0026 **
ingame_mentality    8.77e+01   1.91e+01    4.60  4.4e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9570 on 6512 degrees of freedom
Multiple R-squared:  0.716,     Adjusted R-squared:  0.716
F-statistic: 2.06e+03 on 8 and 6512 DF,  p-value: <2e-16

> vif(bstep5)
         value_eur                age          potential  attacking_volleys
              1.78               1.87               2.28               2.12
 skill_fk_accuracy skill_long_passing         MovementSQ   ingame_mentality
              2.12               2.10               1.50               3.05
```

In the second order term I have included the variables like wage_eur, age, potential, value_eur, attacking_volleys, sill_fk_accuracy, sill_passing, movement_acceleration etc. in this we got the adjusted R2 71.6% which is high and the t value of value_eur is extremely high which says that the transfer value of the player depends the wage of the player. There is no much difference in between wprediction4 and wprediction5 in terms of all the variables it is because we added the variables and removed them when they were individually in the model. I have cleaned up the model and created interaction terms, second order terms, checked for multicollinearity etc.

R factor is high which tells us that there is high dependency between the two variables. If a defender's mentality is aggressive, there are higher chances of giving a penalty to the opposition by fouling the opponent.

Conclusion:

Therefore, the contract values boost the defender's estimation of wage values due to the essence of the distribution of transfer value. After transforming the value it seems that only age and value eur and potential statistics have now been significant in determining the transfer value. None of the terms of interaction or second order were critical, nor did they considerably improve the model's modified r squared.

d. Alyssa Robinson - Goalkeepers

When building the Goalkeepers model, I'll be focusing on the Goalkeepers part of the data, evaluating their attributes, adjusted R, F-Value, T-Test, multicollinearity, backward and forward elimination.
the problem I've encountered within this project is determining which variables should be removed from the dataset and if the Goalkeepers get old, they usually retire and not play a different position of their liking. My group and I will be predicting which important variables cause the change of weekly salary within our data. Wage_euro will be used as the dependent variable while the numerical data will be.

Building the models for this project, originally, I had 2 models, but a third model was needed to determine the backward elimination. The first model included a full dataset which I've shown below for Goalkeepers

Full Model:

- *wage1 <- lm(wage_eur ~ value_eur + age + height_cm + weight_kg +*
- *overall + potential + gk_diving + gk_handling*
- *+ gk_reflexes + gk_speed + gk_positioning +*
- *attacking_crossing + attacking_finishing +*
- *attacking_heading_accuracy + attacking_short_passing +*
- *skill_dribbling + skill_curve + skill_fk_accuracy +*
- *skill_long_passing + skill_ball_control +*
- *movement_acceleration + movement_sprint_speed +*
- *movement_agility + movement_reactions + movement_balance*
- *+ power_shot_power + power_jumping + power_stamina +*
- *+ power_strength + power_long_shots +*
- *mentality_aggression + mentality_interceptions +*
- *mentality_positioning + mentality_vision +*
- *mentality_penalties + mentality_composure +*
- *defending_marking + defending_standing_tackle +*
- *defending_sliding_tackle + goalkeeping_kicking*
- *, data = GoalKeepers)*
- *summary(wage1)*

```
> summary(wage1)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + weight_kg +
    overall + potential + gk_diving + gk_handling + gk_reflexes +
    gk_speed + gk_positioning + attacking_crossing + attacking_finishing +
    attacking_heading_accuracy + attacking_short_passing + skill_dribbling +

    skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_contro
l +
    movement_acceleration + movement_sprint_speed + movement_agility +
    movement_reactions + movement_balance + power_shot_power +
    power_jumping + power_stamina + +power_strength + power_long_shots +
    mentality_aggression + mentality_interceptions + mentality_positioning +

    mentality_vision + mentality_penalties + mentality_composure +
    defending_marking + defending_standing_tackle + defending_sliding_tackle
 +
    goalkeeping_kicking, data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-96227  -2456   -197   1817 102922

Coefficients:
                           Estimate     Std. Error t value
(Intercept)           -54837.55214589 10184.66387573  -5.384
value_eur                  0.00280242     0.00005023  55.792
age                      480.22498352    79.97637649   6.005
height_cm                217.75435003    55.70006944   3.909
weight_kg                 -1.87242766    40.18288318  -0.047
overall                 1975.30799423   614.97304144   3.212
potential                132.27832119    75.43600310   1.754
gk_diving               -428.85033210   145.13524382  -2.955
gk_handling             -381.66225611   138.81470489  -2.749
```

gk_diving	-428.85033210	145.13524382	-2.955
gk_handling	-381.66225611	138.81470489	-2.749
gk_reflexes	-433.07631961	141.22906737	-3.066
gk_speed	249.85109743	639.77729812	0.391
gk_positioning	-480.92711038	140.27726119	-3.428
attacking_crossing	-98.56715302	57.39656688	-1.717
attacking_finishing	-47.88038571	70.68470310	-0.677
attacking_heading_accuracy	55.99067390	56.24531759	0.995
attacking_short_passing	6.05399408	33.92920306	0.178
skill_dribbling	-105.02069799	54.14849163	-1.939
skill_curve	-60.96934178	52.83776550	-1.154
skill_fk_accuracy	11.59174755	50.36707376	0.230
skill_long_passing	116.96525010	32.01134391	3.654
skill_ball_control	77.42459741	39.83668294	1.944
movement_acceleration	-82.90417070	289.54316291	-0.286
movement_sprint_speed	-138.18625957	353.76847430	-0.391
movement_agility	-24.89884133	21.38032048	-1.165
movement_reactions	-260.63617111	73.25624795	-3.558
movement_balance	14.32185760	23.40824496	0.612
power_shot_power	-1810.58458609	659.98871858	-2.743
power_jumping	30.89533148	22.65993877	1.363
power_stamina	-14.89303088	29.84263292	-0.499
power_strength	-57.32987487	21.26099556	-2.696
power_long_shots	-45.67383661	65.04527335	-0.702
mentality_aggression	10.25628800	28.83227268	0.356
mentality_interceptions	75.55318910	44.90092005	1.683
mentality_positioning	-125.93815870	65.98971714	-1.908
mentality_vision	-11.91558032	16.82765607	-0.708
mentality_penalties	126.68936494	33.51183767	3.780
mentality_composure	18.50197219	19.05747362	0.971
defending_marking	-47.39369014	35.90525509	-1.320
defending_standing_tackle	-81.14814847	69.62942110	-1.165
defending_sliding_tackle	-25.66192606	68.06824899	-0.377
goalkeeping_kicking	1239.22220877	497.31908239	2.492

```
                                      Pr(>|t|)
(Intercept)                   0.00000008147 ***
value_eur              < 0.0000000000000002 ***
age                           0.00000000228 ***
height_cm                     0.00009567672 ***
weight_kg                          0.962839
overall                            0.001339 **
potential                          0.079670 .
gk_diving                          0.003166 **
gk_handling                        0.006025 **
gk_reflexes                        0.002196 **
gk_speed                           0.696189
gk_positioning                     0.000620 ***
attacking_crossing                 0.086083 .
attacking_finishing                0.498245
attacking_heading_accuracy         0.319630
attacking_short_passing            0.858404
skill_dribbling                    0.052585 .
skill_curve                        0.248683
skill_fk_accuracy                  0.818003
skill_long_passing                 0.000265 ***
skill_ball_control                 0.052093 .
movement_acceleration              0.774658
movement_sprint_speed              0.696127
movement_agility                   0.244336
movement_reactions                 0.000383 ***
movement_balance                   0.540722
power_shot_power                   0.006137 **
power_jumping                      0.172903
power_stamina                      0.617799
power_strength                     0.007068 **
power_long_shots                   0.482647
mentality_aggression               0.722087
mentality_interceptions            0.092600 .
```

```
power_strength                      0.007068 **
power_long_shots                    0.482647
mentality_aggression                0.722087
mentality_interceptions             0.092600 .
mentality_positioning               0.056479 .
mentality_vision                    0.478971
mentality_penalties                 0.000161 ***
mentality_composure                 0.331742
defending_marking                   0.187001
defending_standing_tackle           0.243988
defending_sliding_tackle            0.706212
goalkeeping_kicking                 0.012792 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8197 on 1955 degrees of freedom
Multiple R-squared:  0.7711,    Adjusted R-squared:  0.7664
F-statistic: 164.7 on 40 and 1955 DF,  p-value: < 0.00000000000000022
```

-

From this model the adjusted R square is 0.7675 which is good and the t-values are good as well. In this model we can see that age and market value has a correlation with each other.

When performing a backwards elimination on the model, and searching for multicollinearity, certain variables were removed from the model which is shown below.

<u>Backward Elimination on Full Model:</u>

- *bstep1 <- stepAIC(wage1, direction = "backward")*
- *summary(bstep1)*

```
> bstep1 <- stepAIC(wage1, direction = "backward")
Start:  AIC=36014.47
wage_eur ~ value_eur + age + height_cm + weight_kg + overall +
    potential + gk_diving + gk_handling + gk_reflexes + gk_speed +
    gk_positioning + attacking_crossing + attacking_finishing +
    attacking_heading_accuracy + attacking_short_passing + skill_dribbling +

    skill_curve + skill_fk_accuracy + skill_long_passing + skill_ball_contro
l +
    movement_acceleration + movement_sprint_speed + movement_agility +
    movement_reactions + movement_balance + power_shot_power +
    power_jumping + power_stamina + +power_strength + power_long_shots +
    mentality_aggression + mentality_interceptions + mentality_positioning +

    mentality_vision + mentality_penalties + mentality_composure +
    defending_marking + defending_standing_tackle + defending_sliding_tackle
 +
    goalkeeping_kicking
```

-

```
Step:  AIC=35988.56
wage_eur ~ value_eur + age + height_cm + overall + potential +
    gk_diving + gk_handling + gk_reflexes + gk_positioning +
    attacking_crossing + skill_dribbling + skill_long_passing +
    skill_ball_control + movement_reactions + power_shot_power +
    power_jumping + power_strength + mentality_interceptions +
    mentality_positioning + mentality_penalties + defending_marking +
    defending_standing_tackle + goalkeeping_kicking

                            Df    Sum of Sq          RSS   AIC
<none>                                      131884547143 35989
- defending_marking          1    138482829 132023029973 35989
- defending_standing_tackle  1    144716064 132029263208 35989
- power_jumping              1    190595610 132075142753 35989
- mentality_interceptions    1    231136629 132115683772 35990
- skill_ball_control         1    266759708 132151306851 35991
- potential                  1    270347667 132154894810 35991
- skill_dribbling            1    319868135 132204415278 35991
- mentality_positioning      1    333725276 132218272419 35992
- attacking_crossing         1    356020120 132240567263 35992
- goalkeeping_kicking        1    427831055 132312378198 35993
- power_strength             1    471032382 132355579526 35994
- power_shot_power           1    522383963 132406931106 35994
- gk_handling                1    546460939 132431008082 35995
- gk_diving                  1    605508528 132490055671 35996
- gk_reflexes                1    659945057 132544492200 35997
- overall                    1    709944315 132594491459 35997
- gk_positioning             1    810930238 132695477381 35999
- movement_reactions         1    885883774 132770430917 36000
- mentality_penalties        1    943947378 132828494521 36001
- height_cm                  1   1189467470 133074014613 36004
- skill_long_passing         1   1453080007 133337627151 36008
- age                        1   2810604771 134695151915 36029
- value_eur                  1 212887896272 344772443416 37905
```

-
```
> summary(bstep1)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + overall +
    potential + gk_diving + gk_handling + gk_reflexes + gk_positioning +
    attacking_crossing + skill_dribbling + skill_long_passing +
    skill_ball_control + movement_reactions + power_shot_power +
    power_jumping + power_strength + mentality_interceptions +
    mentality_positioning + mentality_penalties + defending_marking +
    defending_standing_tackle + goalkeeping_kicking, data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-97037  -2378   -189   1802 102694

Coefficients:
                              Estimate      Std. Error t value
(Intercept)             -52220.68056139  9430.15042469  -5.538
value_eur                    0.00280767     0.00004976  56.420
age                        491.95704898    75.88759666   6.483
height_cm                  202.39070287    47.99077627   4.217
overall                   1986.50826761   609.70777728   3.258
potential                  147.88049699    73.55176342   2.011
gk_diving                 -432.23217269   143.64828946  -3.009
gk_handling               -393.03941660   137.49916010  -2.858
gk_reflexes               -439.03522777   139.76195917  -3.141
gk_positioning            -483.80885451   138.93944699  -3.482
attacking_crossing        -123.55080569    53.54905509  -2.307
skill_dribbling           -113.09735717    51.71427831  -2.187
skill_long_passing         119.30787398    25.59574908   4.661
skill_ball_control          77.51740118    38.81348047   1.997
movement_reactions        -263.71688973    72.45912286  -3.640
power_shot_power         -1831.86105563   655.45233204  -2.795
power_jumping               33.68832919    19.95567980   1.688
power_strength             -52.94756485    19.95097431  -2.654
```

```
power_strength              -52.94756485     19.95097431  -2.654
mentality_interceptions      80.17992403     43.12953283   1.859
mentality_positioning      -137.21638884     61.42638203  -2.234
mentality_penalties         120.79697836     32.15330940   3.757
defending_marking           -51.00446325     35.44489697  -1.439
defending_standing_tackle   -92.03466241     62.56573534  -1.471
goalkeeping_kicking        1249.63823460    494.07360490   2.529
                                       Pr(>|t|)
(Intercept)                   0.000000034763 ***
value_eur                   < 0.0000000000000002 ***
age                           0.000000000114 ***
height_cm                     0.000025845412 ***
overall                             0.001141 **
potential                           0.044507 *
gk_diving                           0.002655 **
gk_handling                         0.004301 **
gk_reflexes                         0.001707 **
gk_positioning                      0.000508 ***
attacking_crossing                  0.021144 *
skill_dribbling                     0.028862 *
skill_long_passing            0.000003353542 ***
skill_ball_control                  0.045943 *
movement_reactions                  0.000280 ***
power_shot_power                    0.005243 **
power_jumping                       0.091539 .
power_strength                      0.008021 **
mentality_interceptions             0.063169 .
mentality_positioning               0.025606 *
mentality_penalties                 0.000177 ***
defending_marking                   0.150315
defending_standing_tackle           0.141449
goalkeeping_kicking                 0.011508 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

-

```
                                       Pr(>|t|)
(Intercept)                       0.000000034763 ***
value_eur                     < 0.0000000000000002 ***
age                               0.000000000114 ***
height_cm                         0.000025845412 ***
overall                                 0.001141 **
potential                               0.044507 *
gk_diving                               0.002655 **
gk_handling                             0.004301 **
gk_reflexes                             0.001707 **
gk_positioning                          0.000508 ***
attacking_crossing                      0.021144 *
skill_dribbling                         0.028862 *
skill_long_passing                0.000003353542 ***
skill_ball_control                      0.045943 *
movement_reactions                      0.000280 ***
power_shot_power                        0.005243 **
power_jumping                           0.091539 .
power_strength                          0.008021 **
mentality_interceptions                 0.063169 .
mentality_positioning                   0.025606 *
mentality_penalties                     0.000177 ***
defending_marking                       0.150315
defending_standing_tackle               0.141449
goalkeeping_kicking                     0.011508 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8178 on 1972 degrees of freedom
Multiple R-squared:  0.7702,    Adjusted R-squared:  0.7675
F-statistic: 287.4 on 23 and 1972 DF,  p-value: < 0.00000000000000022
```

-

<u>*Multicollinearity Check:*</u>

- *vif(bstep1)*

```
> vif(bstep1)
              value_eur                     age
                 1.8594                  4.9897
              height_cm                 overall
                 1.4451                644.2800
              potential               gk_diving
                 6.6678                 36.9540
            gk_handling             gk_reflexes
                29.6250                 38.9580
         gk_positioning      attacking_crossing
                41.2900                  1.3070
         skill_dribbling      skill_long_passing
                 1.5315                  1.3166
      skill_ball_control      movement_reactions
                 1.5620                 18.6470
        power_shot_power           power_jumping
               409.6200                  1.6038
          power_strength mentality_interceptions
                 1.4765                  2.0037
    mentality_positioning     mentality_penalties
                 2.0305                  1.3919
       defending_marking defending_standing_tackle
                 1.2386                  1.2644
      goalkeeping_kicking
               412.1900
```

Here the adjusted R is 0.7675 which is good, the t-values are good so I kept value_euro, but the multicollinearity for this model wasn't so good, there are values listed that were higher than 10. So I had to create another model with the variables that were removed.

Once the variables were removed, I observed the backwards elimination on this model (wage2) and for the third model.

<u>*Second Iteration:*</u>

- *wage2 <- lm(wage_eur ~ value_eur + age + height_cm + potential +*
- *attacking_crossing + skill_dribbling + skill_long_passing +*
- *skill_ball_control + movement_reactions +*
- *power_jumping + power_strength + mentality_interceptions +*
- *mentality_positioning + mentality_penalties + defending_marking +*
- *defending_standing_tackle*
- *, data = GoalKeepers)*
- *summary(wage2)*

```
> summary(wage2)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + potential +
    attacking_crossing + skill_dribbling + skill_long_passing +
    skill_ball_control + movement_reactions + power_jumping +
    power_strength + mentality_interceptions + mentality_positioning +
    mentality_penalties + defending_marking + defending_standing_tackle,
    data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-96794  -2272    -99   1679 104854

Coefficients:
                               Estimate      Std. Error t value
(Intercept)               -47008.85613757   9177.14783971  -5.122
value_eur                      0.00282737      0.00004655  60.738
age                          425.49461221     53.92746107   7.890
height_cm                    192.78097871     47.71916313   4.040
potential                     79.14572671     46.55512616   1.700
attacking_crossing          -129.38648978     53.50252819  -2.418
skill_dribbling             -114.29636970     51.80346110  -2.206
skill_long_passing           119.54361291     25.34716140   4.716
skill_ball_control            72.41092683     38.87746933   1.863
movement_reactions           -65.49786570     31.03328807  -2.111
power_jumping                 32.97196234     19.86599724   1.660
power_strength               -56.73222159     19.91313557  -2.849
mentality_interceptions       79.94678910     43.16355146   1.852
mentality_positioning       -148.36714282     61.33512316  -2.419
mentality_penalties          120.68946320     32.14530932   3.754
defending_marking            -51.02407222     35.52430851  -1.436
defending_standing_tackle    -92.09970516     62.61543654  -1.471
```

```
                                   Pr(>|t|)
(Intercept)                0.00000033100511931 ***
value_eur                < 0.0000000000000002 ***
age                        0.00000000000000495 ***
height_cm                  0.00005551506009183 ***
potential                             0.089280 .
attacking_crossing                    0.015682 *
skill_dribbling                       0.027474 *
skill_long_passing         0.00000257011705691 ***
skill_ball_control                    0.062675 .
movement_reactions                    0.034935 *
power_jumping                         0.097130 .
power_strength                        0.004431 **
mentality_interceptions               0.064148 .
mentality_positioning                 0.015655 *
mentality_penalties                   0.000179 ***
defending_marking                     0.151071
defending_standing_tackle             0.141483
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8209 on 1979 degrees of freedom
Multiple R-squared:  0.7676,    Adjusted R-squared:  0.7657
F-statistic: 408.5 on 16 and 1979 DF,  p-value: < 0.00000000000000022
```

<u>*Second Backward Step Elimination*</u>:

- *bstep2 <- stepAIC(wage2, direction = "backward")*
- *summary(bstep2)*

```
> bstep2 <- stepAIC(wage2, direction = "backward")
Start:  AIC=35997
wage_eur ~ value_eur + age + height_cm + potential + attacking_crossing +
    skill_dribbling + skill_long_passing + skill_ball_control +
    movement_reactions + power_jumping + power_strength + mentality_intercep
tions +
    mentality_positioning + mentality_penalties + defending_marking +
    defending_standing_tackle

                            Df    Sum of Sq          RSS   AIC
<none>                                      133375625440 35997
- defending_marking          1    139036761 133514662201 35997
- defending_standing_tackle  1    145808999 133521434439 35997
- power_jumping              1    185651954 133561277393 35998
- potential                  1    194782799 133570408238 35998
- mentality_interceptions    1    231205619 133606831059 35998
- skill_ball_control         1    233799142 133609424581 35998
- movement_reactions         1    300212925 133675838365 35999
- skill_dribbling            1    328078669 133703704108 36000
- attacking_crossing         1    394148545 133769773984 36001
- mentality_positioning      1    394355278 133769980718 36001
- power_strength             1    547029761 133922655201 36003
- mentality_penalties        1    950023106 134325648545 36009
- height_cm                  1   1099951152 134475576592 36011
- skill_long_passing         1   1499079857 134874705297 36017
- age                        1   4195646183 137571271623 36057
- value_eur                  1 248632615810 382008241250 38095
```

```
> summary(bstep2)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + potential + 
    attacking_crossing + skill_dribbling + skill_long_passing + 
    skill_ball_control + movement_reactions + power_jumping + 
    power_strength + mentality_interceptions + mentality_positioning + 
    mentality_penalties + defending_marking + defending_standing_tackle, 
    data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max 
-96794  -2272    -99   1679 104854 

Coefficients:
                               Estimate      Std. Error t value
(Intercept)              -47008.85613757   9177.14783971  -5.122
value_eur                     0.00282737      0.00004655  60.738
age                         425.49461221     53.92746107   7.890
height_cm                   192.78097871     47.71916313   4.040
potential                    79.14572671     46.55512616   1.700
attacking_crossing         -129.38648978     53.50252819  -2.418
skill_dribbling            -114.29636970     51.80346110  -2.206
skill_long_passing          119.54361291     25.34716140   4.716
skill_ball_control           72.41092683     38.87746933   1.863
movement_reactions          -65.49786570     31.03328807  -2.111
power_jumping                32.97196234     19.86599724   1.660
power_strength              -56.73222159     19.91313557  -2.849
mentality_interceptions      79.94678910     43.16355146   1.852
mentality_positioning      -148.36714282     61.33512316  -2.419
mentality_penalties         120.68946320     32.14530932   3.754
defending_marking           -51.02407222     35.52430851  -1.436
defending_standing_tackle   -92.09970516     62.61543654  -1.471
```

```
                              Pr(>|t|)
(Intercept)               0.00000033100511931 ***
value_eur               < 0.0000000000000002 ***
age                       0.00000000000000495 ***
height_cm                 0.00005551506009183 ***
potential                            0.089280 .
attacking_crossing                   0.015682 *
skill_dribbling                      0.027474 *
skill_long_passing        0.00000257011705691 ***
skill_ball_control                   0.062675 .
movement_reactions                   0.034935 *
power_jumping                        0.097130 .
power_strength                       0.004431 **
mentality_interceptions              0.064148 .
mentality_positioning                0.015655 *
mentality_penalties                  0.000179 ***
defending_marking                    0.151071
defending_standing_tackle            0.141483
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8209 on 1979 degrees of freedom
Multiple R-squared:  0.7676,    Adjusted R-squared:  0.7657
F-statistic: 408.5 on 16 and 1979 DF,  p-value: < 0.00000000000000022
```

Second Multicollinearity Check:

- *vif(bstep2)*

```
> vif(bstep2)
                value_eur                       age
                   1.6145                    2.5004
                height_cm                 potential
                   1.4178                    2.6509
       attacking_crossing           skill_dribbling
                   1.2947                    1.5250
       skill_long_passing        skill_ball_control
                   1.2813                    1.5552
       movement_reactions             power_jumping
                   3.3941                    1.5772
           power_strength   mentality_interceptions
                   1.4596                    1.9915
    mentality_positioning       mentality_penalties
                   2.0089                    1.3806
        defending_marking defending_standing_tackle
                   1.2346                    1.2567
>
```

Third Iteration of Model:

- *wage3 <- lm(wage_eur ~ value_eur + age + height_cm +*
- *attacking_crossing + skill_long_passing +*

- *power_strength + mentality_positioning + mentality_penalties*
- *, data = GoalKeepers)*
- *summary(wage3)*

```
> summary(wage3)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + attacking_crossing +
    skill_long_passing + power_strength + mentality_positioning +
    mentality_penalties, data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-96966  -2284   -158   1650 103754

Coefficients:
                          Estimate      Std. Error t value
(Intercept)           -36792.14175706   8110.22509710  -4.537
value_eur                  0.00284090      0.00003871  73.383
age                      365.59369941     39.33988357   9.293
height_cm                158.54652534     43.93149769   3.609
attacking_crossing      -162.06141195     51.30288379  -3.159
skill_long_passing       131.20210118     24.22229721   5.417
power_strength           -47.56114602     19.30151718  -2.464
mentality_positioning   -134.39258017     51.96050566  -2.586
mentality_penalties      120.25995857     30.83379326   3.900
                                  Pr(>|t|)
(Intercept)                   0.0000060606 ***
value_eur             < 0.0000000000000002 ***
age                   < 0.0000000000000002 ***
height_cm                         0.000315 ***
attacking_crossing                0.001607 **
skill_long_passing            0.0000000681 ***
power_strength                    0.013819 *
mentality_positioning             0.009768 **
mentality_penalties           0.0000992815 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

-

```
Residuals:
   Min     1Q Median     3Q    Max 
-96966  -2284   -158   1650 103754 

Coefficients:
                           Estimate     Std. Error t value
(Intercept)          -36792.14175706  8110.22509710  -4.537
value_eur                 0.00284090     0.00003871  73.383
age                     365.59369941    39.33988357   9.293
height_cm               158.54652534    43.93149769   3.609
attacking_crossing     -162.06141195    51.30288379  -3.159
skill_long_passing      131.20210118    24.22229721   5.417
power_strength          -47.56114602    19.30151718  -2.464
mentality_positioning  -134.39258017    51.96050566  -2.586
mentality_penalties     120.25995857    30.83379326   3.900
                                 Pr(>|t|)    
(Intercept)                  0.0000060606 ***
value_eur             < 0.0000000000000002 ***
age                   < 0.0000000000000002 ***
height_cm                        0.000315 ***
attacking_crossing               0.001607 ** 
skill_long_passing           0.0000000681 ***
power_strength                   0.013819 *  
mentality_positioning            0.009768 ** 
mentality_penalties          0.0000992815 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8244 on 1987 degrees of freedom
Multiple R-squared:  0.7647,	Adjusted R-squared:  0.7637 
F-statistic: 807.2 on 8 and 1987 DF,  p-value: < 0.00000000000000022
```

<u>*Third Backward Step Check:*</u>

- *bstep3 <- stepAIC(wage3, direction = "backward")*
- *summary(bstep3)*

- ```
  > bstep3 <- stepAIC(wage3, direction = "backward")
  Start:  AIC=36005.82
  wage_eur ~ value_eur + age + height_cm + attacking_crossing +
      skill_long_passing + power_strength + mentality_positioning +
      mentality_penalties

                          Df    Sum of Sq          RSS   AIC
  <none>                                  135044261627 36006
  - power_strength         1    412667240 135456928867 36010
  - mentality_positioning  1    454655149 135498916776 36011
  - attacking_crossing     1    678194008 135722455635 36014
  - height_cm              1    885196537 135929458164 36017
  - mentality_penalties    1   1033871241 136078132868 36019
  - skill_long_passing     1   1994018534 137038280161 36033
  - age                    1   5869613869 140913875496 36089
  - value_eur              1 365988667281 501032928908 38621
  ```

- ```
  > summary(bstep3)

  Call:
  lm(formula = wage_eur ~ value_eur + age + height_cm + attacking_crossing +
      skill_long_passing + power_strength + mentality_positioning +
      mentality_penalties, data = GoalKeepers)

  Residuals:
     Min     1Q Median     3Q    Max
  -96966  -2284   -158   1650 103754

  Coefficients:
                             Estimate      Std. Error t value
  (Intercept)         -36792.14175706   8110.22509710  -4.537
  value_eur                0.00284090      0.00003871  73.383
  age                    365.59369941     39.33988357   9.293
  height_cm              158.54652534     43.93149769   3.609
  attacking_crossing    -162.06141195     51.30288379  -3.159
  skill_long_passing     131.20210118     24.22229721   5.417
  power_strength         -47.56114602     19.30151718  -2.464
  mentality_positioning -134.39258017     51.96050566  -2.586
  mentality_penalties    120.25995857     30.83379326   3.900
                                    Pr(>|t|)
  (Intercept)               0.0000060606 ***
  value_eur             < 0.0000000000000002 ***
  age                   < 0.0000000000000002 ***
  height_cm                     0.000315 ***
  attacking_crossing            0.001607 **
  skill_long_passing        0.0000000681 ***
  power_strength                0.013819 *
  mentality_positioning         0.009768 **
  mentality_penalties       0.0000992815 ***
  ---
  ```

```
Residuals:
   Min     1Q Median     3Q    Max
-96966  -2284   -158   1650 103754

Coefficients:
                          Estimate     Std. Error t value
(Intercept)           -36792.14175706  8110.22509710  -4.537
value_eur                  0.00284090     0.00003871  73.383
age                      365.59369941    39.33988357   9.293
height_cm                158.54652534    43.93149769   3.609
attacking_crossing      -162.06141195    51.30288379  -3.159
skill_long_passing       131.20210118    24.22229721   5.417
power_strength           -47.56114602    19.30151718  -2.464
mentality_positioning   -134.39258017    51.96050566  -2.586
mentality_penalties      120.25995857    30.83379326   3.900
                                Pr(>|t|)
(Intercept)                 0.0000060606 ***
value_eur             < 0.0000000000000002 ***
age                   < 0.0000000000000002 ***
height_cm                       0.000315 ***
attacking_crossing              0.001607 **
skill_long_passing          0.0000000681 ***
power_strength                  0.013819 *
mentality_positioning           0.009768 **
mentality_penalties         0.0000992815 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8244 on 1987 degrees of freedom
Multiple R-squared:  0.7647,    Adjusted R-squared:  0.7637
F-statistic: 807.2 on 8 and 1987 DF,  p-value: < 0.00000000000000022
```

-

Third Multicollinearity Check:

- *vif(bstep3)*

```
> vif(bstep3)
            value_eur                   age             height_cm
               1.1074                1.3195                1.1916
   attacking_crossing    skill_long_passing        power_strength
               1.1805                1.1603                1.3599
mentality_positioning   mentality_penalties
               1.4297                1.2596
```

-

The adjusted R square is .7637 which is good, the t-value is good and p values are good as well. The multicollinearity is under 10 which is a good sign as well. Based on my finding, model 3 can be used to make a better prediction since there wasn't any multicollinearity. I've learned that

Age and Market Value plays a significant role for weekly wages. My plans for the next milestone is to find ways to improve the forward and backwards elimination.

Once the variables with high multicollinearity were removed from the data, second order terms and interactions terms were added to the model to help determine the wage salary for Goalkeepers.

The Interaction Term for GoalKeeper's included movements such as crossing the ball. The Second Order Term consisted of GoalKeepers Strength being doubled.

<u>Interaction Terms:</u>

- *GoalKeepers$Crossing = ((GoalKeepers$attacking_crossing + GoalKeepers$skill_long_passing)/2)*

<u>Second Order Term:</u>

- *GoalKeepers$Strength = ((GoalKeepers$power_strength)^2)*

The final model included prediction of the salary for GoalKeepers using their wages. The independent variables consisted of; value_eur, age, height_cm, crossing, strength, and mentality, these variables help to determine the Goalkeepers salary and if the player will keep their position. Once again, I did a final check for backward elimination, as well as multicollinearity, and the model did not remove any variables, as well as the model was not multicollinear.

<u>Final Model:</u>

- *wage4 <- lm(wage_eur ~ value_eur + age + height_cm + Crossing +*
- *Strength, data = GoalKeepers)*
- *summary(wage4)*

-
```
> summary(wage4)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + Crossing +
    Strength, data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-98180  -2238   -313   1529 104758

Coefficients:
                  Estimate    Std. Error t value             Pr(>|t|)
(Intercept) -39422.6639831  8373.4502581  -4.708           0.00000267
value_eur        0.0028587     0.0000389  73.493 < 0.0000000000000002
age            348.5928016    37.3152076   9.342 < 0.0000000000000002
height_cm      162.4756439    44.4767348   3.653             0.000266
Crossing       149.3945115    40.4054558   3.697             0.000224
Strength        -0.4130368     0.1673972  -2.467             0.013693

(Intercept) ***
value_eur   ***
age         ***
height_cm   ***
Crossing    ***
Strength    *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8328 on 1990 degrees of freedom
Multiple R-squared:  0.7595,    Adjusted R-squared:  0.7589
F-statistic:  1257 on 5 and 1990 DF,  p-value: < 0.00000000000000022
```

*Final Backward Elimination Check:*

- *bstep4 <- stepAIC(wage4, direction = "backward")*
- *summary(bstep4)*
-
```
> bstep4 <- stepAIC(wage4, direction = "backward")
Start:  AIC=36043.23
wage_eur ~ value_eur + age + height_cm + Crossing + Strength

            Df    Sum of Sq          RSS   AIC
<none>                       138013387447 36043
- Strength   1    422229899 138435617346 36047
- height_cm  1    925505236 138938892683 36055
- Crossing   1    948107112 138961494559 36055
- age        1   6052478350 144065865797 36127
- value_eur  1 374594836520 512608223967 38660
```

```
> summary(bstep4)

Call:
lm(formula = wage_eur ~ value_eur + age + height_cm + Crossing +
    Strength, data = GoalKeepers)

Residuals:
   Min     1Q Median     3Q    Max
-98180  -2238   -313   1529 104758

Coefficients:
                  Estimate    Std. Error t value            Pr(>|t|)
(Intercept) -39422.6639831  8373.4502581  -4.708          0.00000267
value_eur        0.0028587     0.0000389  73.493 < 0.0000000000000002
age            348.5928016    37.3152076   9.342 < 0.0000000000000002
height_cm      162.4756439    44.4767348   3.653            0.000266
Crossing       149.3945115    40.4054558   3.697            0.000224
Strength        -0.4130368     0.1673972  -2.467            0.013693

(Intercept) ***
value_eur   ***
age         ***
height_cm   ***
Crossing    ***
Strength    *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8328 on 1990 degrees of freedom
Multiple R-squared:  0.7595,    Adjusted R-squared:  0.7589
F-statistic:  1257 on 5 and 1990 DF,  p-value: < 0.00000000000000022
```

Final Multicollinearity Check:

- *vif(bstep4)*

```
> vif(bstep4)
value_eur       age height_cm  Crossing  Strength
   1.0955    1.1634    1.1969    1.1221    1.3524
```

Conclusion:

In conclusion, when analyzing the models for the interaction and second term, there was a need for Goalkeepers to have a high market value, a positive mentality as well as strength, and movement in order for the Goalkeepers to keep their position and not retire. These factors also gave an outline to predict the Goalkeepers weekly salary.

Summarization

It has been suggested that once a soccer player gets older, they will most likely retire. This brings to question the professional soccer players who do not want to retire. For our project we have analyzed the Fifa 20 soccer players list, to analyze what may be the main effects to a soccer player retiring. We have done this to prove a point, that point being just because a soccer player may get old does not necessarily mean they should automatically think about retirement. We have done this through the process of predicting what specific attributes may play a role in a soccer player's continuation of their career.

Our report uses the prediction methods of regression analysis. The prediction method we have used is multiple regression analysis, meaning we have second-order term and interaction terms. We have consistently improved our models to be the best models they can be, and we have done so through the process of constantly remodeling our model, using the backward elimination method.

Soccer players may have a great deal of wage depending on their age and their skills. The simplest way to assess the ability of a player is to evaluate the output results. The purpose of this research is to discover the independent variables that have the greatest impact on the wage of the player. Due to the need for players from different locations to determine their performances specific independent variables, the project is split into four sections depending on the position of the player: goalkeepers, defenders, midfielders and forwards, and the purpose for each division is to predict the wage using the position-related independent variables. This will show us what the players will have to focus on in order to maintain the same or better wage as they get older.

The midfielder's position was analyzed by Endri. I first started off with a full model of our midfielder data, while also using backward elimination to remove unnecessary variables. After the removal of these unnecessary variables, I then consistently checked for multicollinearity. After removing all these variables which were unnecessary, I noticed that my model could also be simplified easier by having second-order terms, as well as interaction terms. The interaction terms for my model consisted of Mentality and Movement of the player, and the second-order term, which I came up with was the number of volleys these players could score. This helped out a lot, as a midfielder myself, I can relate to this because as a midfielder you are consistently organizing the game, and if your movement and mentality is not as good, you will lower your market value. This means that your wage will also be lowered. The final model for the prediction of the wage consisted of the independent variables: market value (value_eur), the age of the player (age), the shooting ability of the player (shooting), the second-order term (VolleySQ), the interaction term (mentality), the endurance of a player (power_stamina) and how well of a defensive mentality they had (defending_sliding_tackle). This is good because for midfielders, these are the main independent variables that affect a players wage, and if whether or not they may decide to continue playing the game. These are the necessary things that professional players will have to maintain high.

The Forward position was analyzed by Srinilay. For the forward position, I used backward elimination and multicollinearity to identify the variables that have good positive results towards the dependent variable. We determined the variables compatibility by checking the p-value and multicollinearity. After eliminating the variables, I wanted to look into more options that could help to determine the wage of the players by using second-order terms and interaction terms. For the interaction term, I created three attack terms, skills, and power2, among those attack and power2 has made into the final model. The power2 term has variable power _stamina which determines the stamina of the player and power _strength is the ability to fend off defenders to score the goals. The term attack has variables attack_ crossing and attack_ short _passing helps the forwards to score goal picking up passes. For the second-order term, a potential variable is used which is to determine the future growth of the player. Potential variables had a positive Correlation towards the wage of the player. I found out that independent variables age, value, and mentality aggression have a positive effect on the wage of the salary. The interaction term(attack, power 2) and second-order model(poten) was used to determine the best variables which affect the salary of the player. As a forward the player needs to receive the pass at the right time and place and also should be able to score with power and accuracy so that it would not be easy to block the incoming attack.

Defenders are usually at the back of the field to stop or interfere with the attack of the opponent. Many of the defenders, in particular those on the left and right wings, are also responsible for helping to attack the opponent squad. Some strategies in modern games require defenders to attack from the defense by passing the ball forth without shooting long-range shots. The Defenders position was analyzed by Hasmitha. I have started by eliminating the unnecessary variables by using backward elimination and constantly checked for multicollinearity and also p-value. In the individual milestone1, I have developed a complete model, with the dependent variable being wage_eur and the other numerical variables being explanatory variables. In the ending I have found that the significant variables which will cause a change in the weekly salary (wage_eur) for defenders as they get older are market value (value_eur), age, the potential they have for growth, abilities to score a bicycle kick or similar volley goals (attacking_volleys), how skilled they are as a freekick taker (skill_fk_accuracy), how skilled they are as a long passer which can change the play from defense to attack (skill_long_passing), their movement_acceleration, and the mentalities they have while they are positioning themselves in defense and taking penalties, meaning they have to keep cool calm and collected as they get older and show that they are experienced, leader. I've shown how I cleaned up our defender model using two Interaction Terms, one Second-Order Term. Good defenders are aggressive. They aren't afraid to make strong tackles and use their body. They attack balls with their heads and get in front of shots. Strong defenders push offensive players off the ball, win headers, and shield the ball well. Defenders need good heading ability to defend against aerial passes, crosses, and set pieces.

The Goalkeeper position was analyzed by (Alyssa) consisted of using the goalkeeper's portion of the dataset. Backward elimination was used to eliminate unnecessary variables within the Goalkeepers dataset. Once these unnecessary variables were eliminated from the dataset, multicollinearity was analyzed within the dataset. Once variables with high multicollinearity were removed from the data, second-order terms and interaction terms were added to the model to help determine the wage salary for Goalkeepers. The Interaction Term for GoalKeeper's included Mentality and Movement (crossing the ball). The Second Order Term consisted of GoalKeepers Strength being doubled. The final model included a prediction of the salary for GoalKeepers using their wages. The independent variables consisted of; value_eur, age, height_cm, crossing, strength, and mentality, these variables help to determine the Goalkeepers salary and if the player will keep their position. This makes sense as the high of the goalkeeper would be beneficial when jumping for a save, the strength of the goalkeeper would help the goalkeeper win a ball which is coming from a corner kick, the mentality of a goalkeeper would have to remain calm and in the game after conceding a goal, and their ability to set up a counter-attack from a cross would also help maintain a high market value, which is one of the main predictors for the salary, and maintain their position as they get older.

Conclusion

In conclusion, we have analyzed the FIFA 20 dataset in order to determine what variables are truly relevant in determining what a players wage_euro (salary) would be and furthermore relevant in what variables are important predictors for any soccer players who would like to continue to play the beautiful game rather than retire. Even though we split our data into four subsets of data: midfielder, forwards, defenders, and goalkeepers, we analyzed our models and we can confidently agree that some of these predictors apply to all our models. One main predictor for salary (wage_euro) is the maintaining of a high market value of a player (value_eur). Furthermore, even though we worked individually on these models, we all had a similar mindset when creating these models. This mindset consisted of a second-order term and interaction term, and further cleaning up our models to the best forms through the process of repetition. From our analysis, we can honestly say that in order for a player of any position to continue playing rather than retiring they would have to maintain a high market value and a positive mentality, movement, power, and strength. This being stated our analysis and methods can be used in the full dataset. Even though some other methods did not make it into the report, this was due to the type of data, and those other methods did not work well for this data type.

Appendix:

Name	**Type (Categorical / Numerical / Rank)**	**Description**
short_name	Nominal	Name Most Commonly Known as to refer to a Player.
long_name	Nominal	Full Legal Name of the Player.
age	Numeric	Age of the Player.
club	Categorical	Current Club of The Player.
overall	Numeric	Rating out of a scale of 100 for the player based on career performance.
potential	Numeric	Potential Growth of a player out of a scale of 100.
Value_euro (market value)	Numeric	The market value of a Player in Euros.
Wage_euro (weekly salary)	Numeric	Weekly Salary of a Player in Euros.
player_positions	Categorical	Preferred Positions of a Player.
release_clause_euro	Numeric	Buyout Clause of a Player in Euros.
team_position	Categorical	The position assigned by the Team Manager.
joined	Numeric	Beginning of Contract at Current Club.
contract_valid_until	Numeric	Ending of a contract at the current club.
pace	Numeric	How fast a player is out of 100 during an in-game performance.

shooting	Numeric	How good of a shot a player is out of 100 during an in-game performance.
passing	Numeric	How good of a passer the player is out of 100 during an in-game performance.
dribbling	Numeric	How good of a trickster/freestyler/dribbler a player is out of 100, during an in-game performance.
defending	Numeric	How good of a defender a player is out of 100 during an in-game performance.
physical	Numeric	How rough or physical a player plays out of 100 during an in-game performance.
gk_diving	Numeric	How far the goalkeeper can reach while diving to block a ball out of 100 during an in-game performance.
gk_handling	Numeric	How well the goalkeeper can handle a powerful shot coming toward them out of 100 during an in-game performance.
gk_kicking	Numeric	How well and for a goalkeeper can kick a ball out of a goal kick or clearing it to prevent a goal from being scored out 100 during the in-game performance.
gk_speed	Numeric	How fast a goalkeeper is out of 100 during an in-game performance.
gk_positioning	Numeric	How well a goalkeeper positions themselves during a game to interpret the

		trajectory of a ball out of 100.
LS (Left Striker)	Numeric	This is an overall if the player is to play in Left Striker Position and overall may change and the formation of the team may change.
ST (Striker)	Numeric	This is an overall if the player is to play in Striker Position and overall may change and the formation of the team may change.
RS (Right Striker)	Numeric	This is an overall if the player is to play in Right Striker Position and overall may change and the formation of the team may change.
LW (Left Winger)	Numeric	This is an overall if the player is to play in Left Winger Position and overall may change and the formation of the team may change.
LF (Left Forward)	Numeric	This is an overall if the player is to play in Left Forward Position and overall may change and the formation of the team may change.
CF (Center Forward)	Numeric	This is an overall if the player is to play in Center Forward Position and overall may change and the formation of the team may change.
RF (Right Forward)	Numeric	This is an overall if the player is to play in Right Forward Position and overall may change and the formation of the team may change.
RW (Right Winger)	Numeric	This is an overall if the player is to play in Right Winger

		Position and overall may change and the formation of the team may change.
LAM (Left Attacking Midfielder)	Numeric	This is an overall if the player is to play in Left Attacking Midfielder Position and overall may change and the formation of the team may change.
CAM (Central Attacking Midfielder)	Numeric	This is an overall if the player is to play in Central Attacking Midfielder and overall may change and the formation of the team may change.
RAM (Right Attacking Midfielder)	Numeric	This is an overall if the player is to play in Right Attacking Midfielder Position and overall may change and the formation of the team may change.
LCM (Left Central Midfielder)	Numeric	This is an overall if the player is to play in Left Central Midfielder Position and overall may change and the formation of the team may change.
CM (Central Midfielder)	Numeric	This is an overall if the player is to play in Central Midfielder Position and overall may change and the formation of the team may change.
RCM (Right Central Midfielder)	Numeric	This is an overall if the player is to play in Right Central Midfielder Position and overall may change and the formation of the team may change.
RM (Right Midfielder)	Numeric	This is an overall if the player

		is to play in Right Midfielder Position and overall may change and the formation of the team may change.
LDM (Left Defensive Midfielder)	Numeric	This is an overall if the player is to play in Left Defensive Midfielder Position and overall may change and the formation of the team may change.
CDM (Central Defensive Midfielder)	Numeric	This is an overall if the player is to play in Central Defensive Midfielder Position and overall may change and the formation of the team may change.
RDM (Right Defensive Midfielder)	Numeric	This is an overall if the player is to play in Right Defensive Midfielder Position and overall may change and the formation of the team may change.
LWB (Left Wing Back)	Numeric	This is an overall if the player is to play in Left Wing Back Position and overall may change and the formation of the team may change.
RWB (Right Wing Back)	Numeric	This is an overall if the player is to play in Right-Wing Back Position and overall may change and the formation of the team may change.
LB (Left Back)	Numeric	This is an overall if the player is to play in Left Back Position and overall may change and the formation of the team may change.
LCB (Left Center Back)	Numeric	This is an overall if the player is to play in Left Center Back

		Position and overall may change and the formation of the team may change.
CB (Center Back)	Numeric	This is an overall if the player is to play in Center Back Position and overall may change and the formation of the team may change.
RCB (Right Center Back)	Numeric	This is an overall if the player is to play in Right Center Back Position and overall may change and the formation of the team may change.
RB (Right Back)	Numeric	This is an overall if the player is to play in Right Back Position and overall may change and the formation of the team may change.
attacking_crossing	Numeric	How well a soccer player can set up an attacking cross, for his fellow teammates to score a goal.
attacking_finishing	Numeric	How well a soccer player can finish an opportunity when attempting to score a goal.
attacking_heading_accuracy	Numeric	How well a soccer player can header a soccer ball, when attempting to score a goal.
attacking_short_passing	Numeric	How well a soccer player can set up a short pass, for his fellow teammates to score a goal.
attacking_volleys	Numeric	How well a soccer player can volley a ball into the back of the net when the opportunity arises.

skill_dribbling	Numeric	How well a soccer player can dribble when attempting to juke the opponent.
skill_curve	Numeric	How well a soccer player can curve the ball when, attempting to score a goal or cross a ball to a teammate so they can score the goal.
skill_fk_accuracy	Numeric	How accurate a soccer player scores from a freekick position.
skill_long_passing	Numeric	How well of a long ball passer a soccer player is.
skill_ball_control	Numeric	How good a soccer player's ball control is.
movement_acceleration	Numeric	What starting speed does a soccer player have during a game.
movement_sprint_speed	Numeric	How fast a soccer player's sprint speed is.
movement_agility	Numeric	How agile a soccer player can move during a soccer game.
movement_reactions	Numeric	How great of a reaction time a soccer player has during a game when the game is changing.
movement_balance	Numeric	How balanced a soccer player moves.
power_shot_power	Numeric	How much power does a soccer player shot have, how lethal is it.
power_jumping	Numeric	How powerfully a soccer player can jump when attempting to win the soccer ball during a game.

power_stamina	Numeric	How powerful a soccer player's stamina is during a soccer game.
power_strength	Numeric	How strong a soccer player is during a soccer game.
power_long_shots	Numeric	How lethal a soccer player's shot is from long range.
mentality_aggression	Numeric	How aggressive a soccer player may play during the game.
mentality_interceptions	Numeric	How often a soccer player has the mentality to intercept the soccer ball from the opponent.
mentality_positioning	Numeric	How well a soccer player positions himself during a soccer game.
mentality_vision	Numeric	How alert a soccer player is of his surroundings during a soccer game in order to make the perfect pass.
mentality_penalties	Numeric	How cool, calm and collected a soccer player is when taking a penalty or blocking a penalty.
mentality_composure	Numeric	How composed a soccer player is during a soccer game.
defending_marking	Numeric	When in a defensive position, how well a soccer player can mark his opponent.
defending_standing_tackle	Numeric	When in a defensive position, how often a soccer player may win the ball with a standing tackle.

defending_sliding_tackle	Numeric	When in a defensive position, how often a soccer player may win the ball with a sliding tackle.
goalkeeping_diving	Numeric	When in a goalkeeping position, how often and well a soccer player may dive for the ball when attempting an opponent from scoring.
goalkeeping_handling	Numeric	When in a goalkeeping position, how well a soccer player can handle a powerful shot coming towards them from an opponent.
goalkeeping_kicking	Numeric	When in a goalkeeping position, how well a soccer player can kick the ball forward for a counter attack on the opponent.
goalkeeping_positioning	Numeric	When in a goalkeeping position, how well a soccer player can position themselves while they are in goal.
goalkeeping_reflexes	Numeric	When in a goalkeeping position, how agile a soccer player is.