FIFA ANALYSIS

- Solomon Lamkhothang
- Deon S
- Muhammed Hisham T

EXPLORING FIFA WITH DATA ANALYSIS

PROBLEM STATEMENT

To determine the value of a player with respect to his wage, the age of the player, player potential, etc

THE DATASET

- We took our data from www.data.world.
- The dataset contains names, age, nationality, wage, potential, etc. of all the football players in 2018
- The dataset has 17954 rows and 92 columns in which there are a lot of missing datas

```
> # To know the size of data
> dim(fifa)
[1] 17954    92
> #to check the count of null values
> sum(is.na(fifa))
[1] 36560
>
```

CLEANING THE DATASET

01 Removed unnecessary columns that were irrelevant to our analysis.

02 Formatted the fonts and wording errors

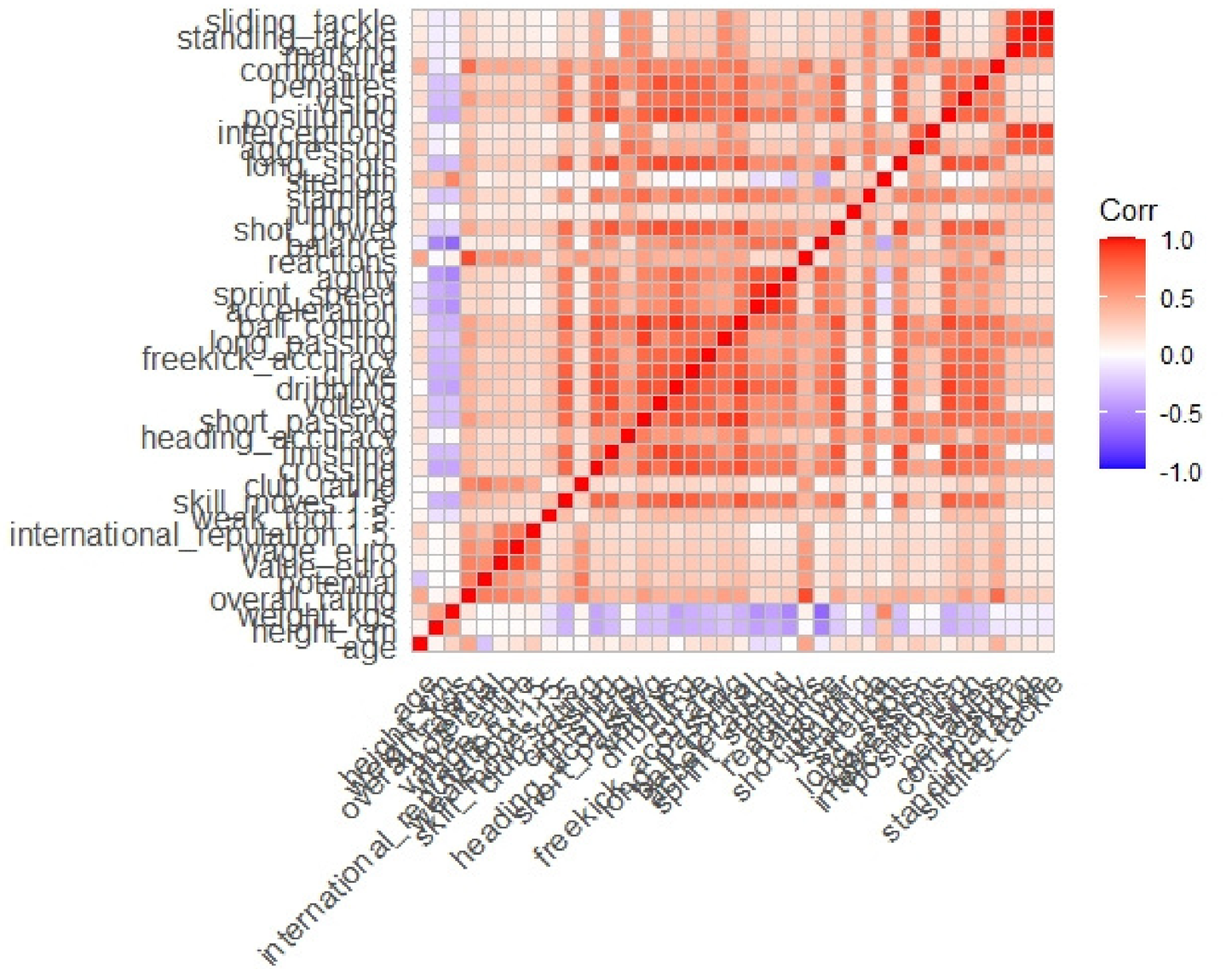
03 Removed Null Values

04 Corrected some error values

```
> dim(fifa_final)
[1] 17699    52
>
```

HEAT MAP CORRELATION

- Here we use heatmap and correlate the numerical variables.
- From this we found the most correlated values using the colours
- Here, the variables which has higher positive correlation are represented by red
- While variables which are negetively correlated are represented by blue

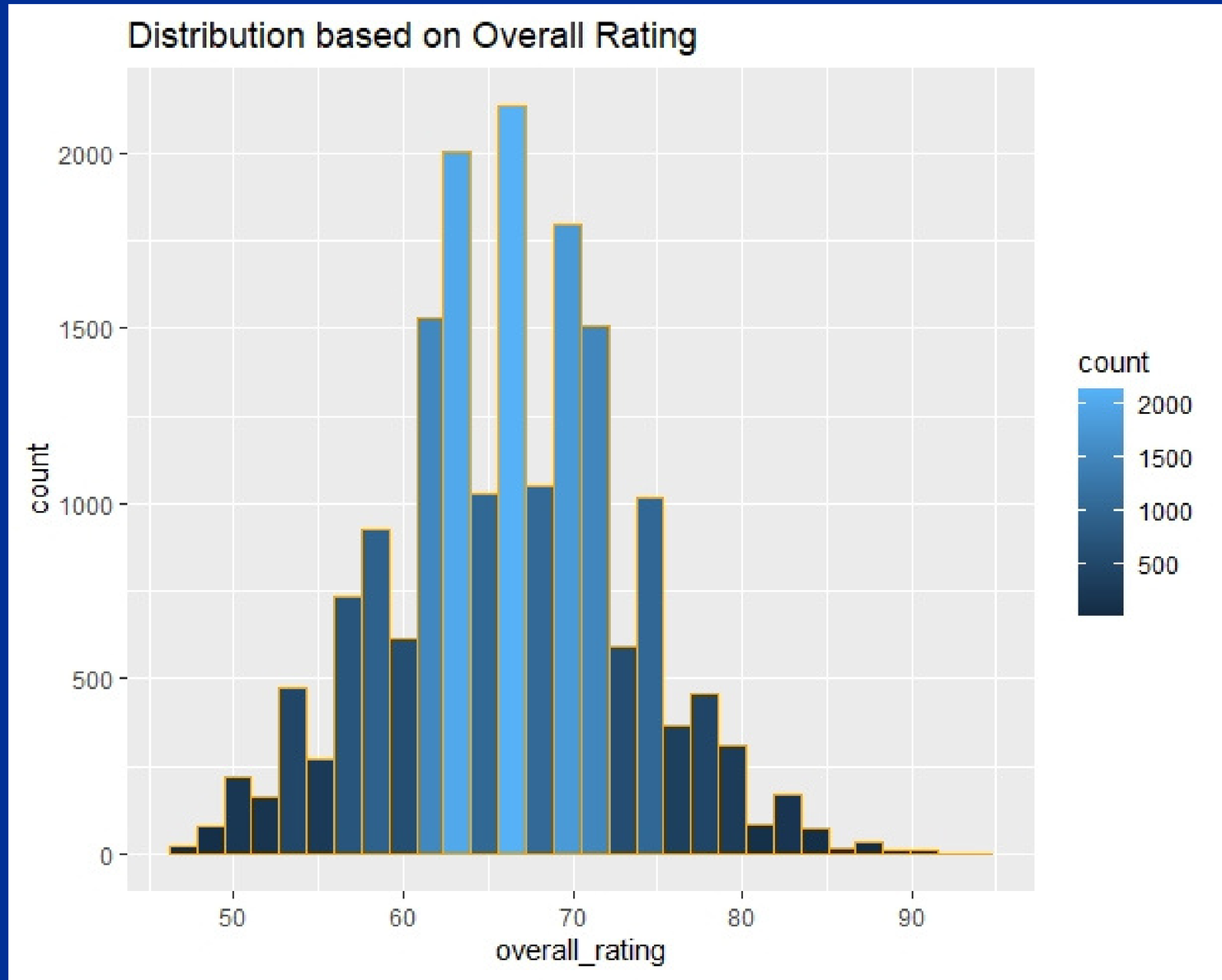


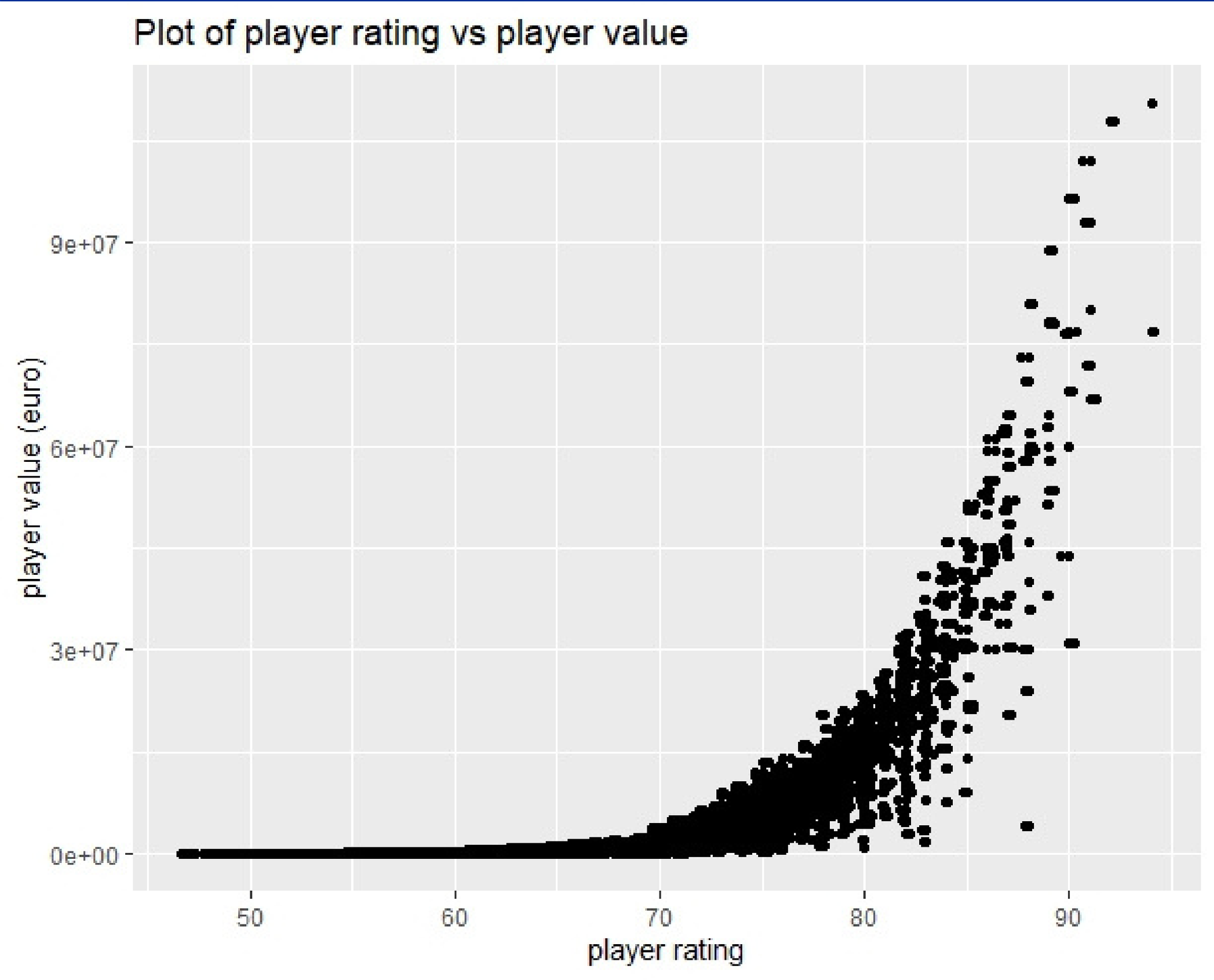
NOTE: THE HIGHER THE INTENSITY, THE HIGHER THE CORRELATION

OVERALL RATING DISTRIBUTION

- The numbers of players with rating between 60 and 75 are found to contribute the most.

- Players with ratings higher than 80 are comparatively less than players with ratings less than 60.





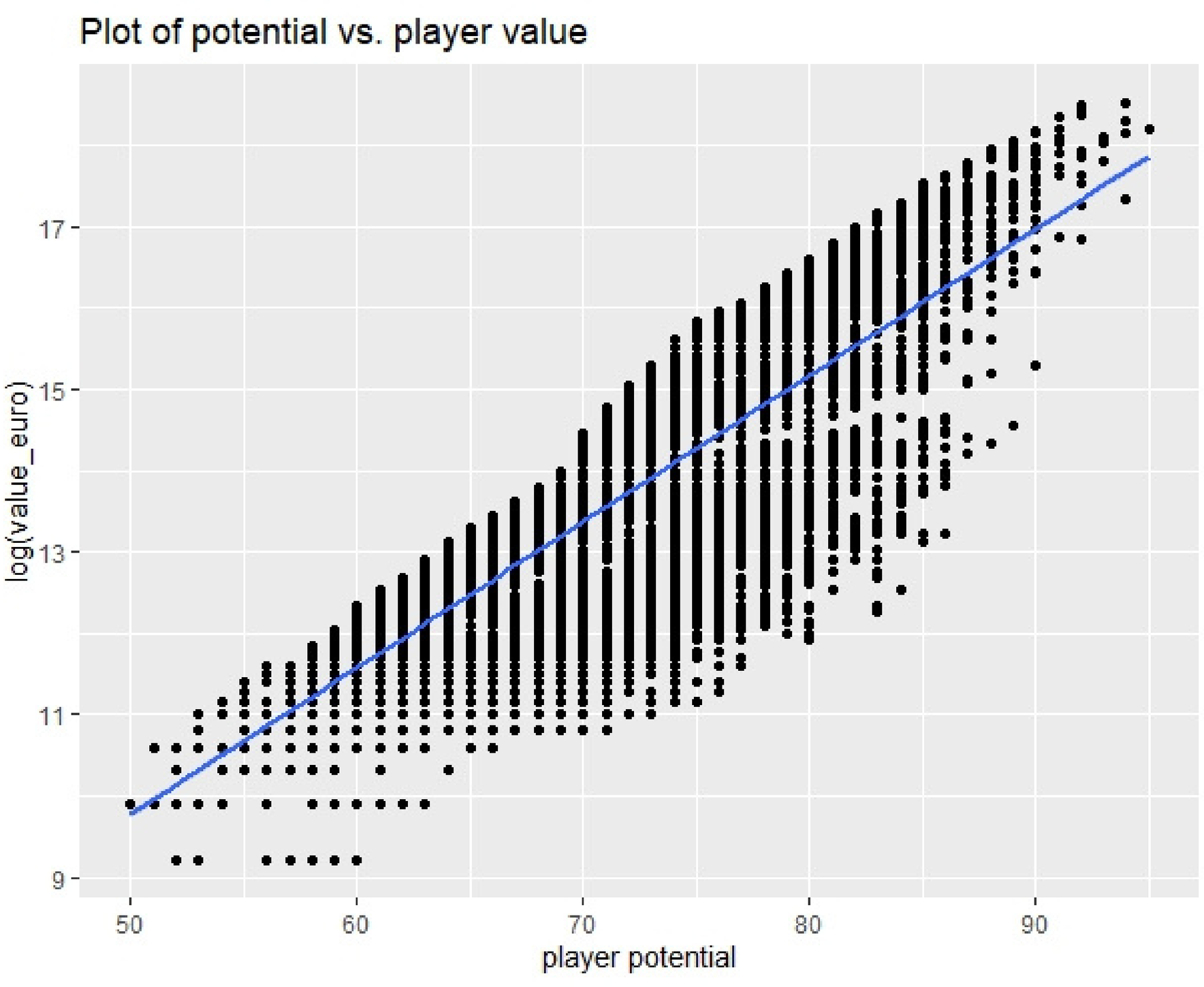
RATING VS VALUE

- It can be seen from the graph that players with ratings more than 75 have higher value in the market

POTENTIAL VS VALUE

- It can be seen that the players potential and their values are positively correlated with each other

- Here, the players with higher potential tend to have higher values.



MULTIPLE REGRESSION MODEL

- It was observe that the R squared value for the model that we tried out had an accuracy of 79.29 %

- The dependent variable that we used was value_euro

- The independent variables were age, overall_rating, potential, wage_euro, international reputation, stamina, sliding_tackle, standing_tackle.

```
Call:
lm(formula = value_euro ~ age + overall_rating + potential +
    wage_euro + international_reputation.1.5. + stamina + sliding_tackle +
    standing_tackle, data = fifa_final)

Residuals:
      Min        1Q    Median        3Q       Max
-19467079   -826406   -100513    705412  53542093

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -9.684e+06  3.769e+05 -25.692  < 2e-16 ***
age                          -2.508e+05  8.743e+03 -28.688  < 2e-16 ***
overall_rating                2.641e+05  7.828e+03  33.733  < 2e-16 ***
potential                    -3.637e+04  7.696e+03  -4.725 2.32e-06 ***
wage_euro                     1.637e+02  1.321e+00 123.969  < 2e-16 ***
international_reputation.1.5. 1.893e+06  6.943e+04  27.258  < 2e-16 ***
stamina                       9.116e+03  1.624e+03   5.613 2.02e-08 ***
sliding_tackle               -2.198e+04  4.154e+03  -5.291 1.23e-07 ***
standing_tackle               7.988e+03  4.174e+03   1.914   0.0557 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2589000 on 17690 degrees of freedom
Multiple R-squared:  0.7929,    Adjusted R-squared:  0.7928
F-statistic:  8466 on 8 and 17690 DF,  p-value: < 2.2e-16
```

THANK YOU