Project Part 2

Drashti Patel (7964120)

11/17/2024

```
fifa <- read.csv("FIFA.csv")
```

Here's a description of my data set. # This project delves into the performance metrics of top FIFA players from the 2024 season. The dataset captures a wide range of statistics for these athletes, including their overall performance ratings, potential ratings, and the number of hits—a measure of their popularity and engagement. By analyzing this data, we can gain a deeper understanding of the factors that influence a player's performance, which is a critical focus for clubs, managers, and analysts looking to evaluate talent and make informed decisions.The response variable for this analysis is Overall, which represents a player's current performance rating. This rating is an essential indicator of a player's skills and abilities at a given moment, offering insight into their value and impact in professional soccer.The explanatory variables are Hits and Potential. Hits reflects the attention a player receives, which might boost their performance through confidence or support. Potential indicates a player's future growth prospects and likely impacts their overall rating by showing how much faith teams have in their development.

Here's the dataset of FIFA of first 10 obsevations from dataset in a tabular form

```
Name = c("Lionel Messi", "Cristiano Ronaldo", "Neymar Jr", "Virgil van Dijk", "Jan Oblak",
           "Kevin De Bruyne", "Robert Lewandowski", "Eden Hazard", "Alisson", "Mohamed Salah")
Hits = c(299, 276, 186, 127, 47, 119, 89, 66, 53, 94)
Potential = c(94, 93, 92, 92, 93, 91, 91, 91, 91, 90)
Overall = c(94, 93, 92, 91, 91, 91, 91, 91, 90, 90)
fifa_data <- data.frame(Name,Hits,Potential,Overall)
fifa_data
```

```
##                  Name Hits Potential Overall
## 1        Lionel Messi  299        94      94
## 2   Cristiano Ronaldo  276        93      93
## 3           Neymar Jr  186        92      92
## 4     Virgil van Dijk  127        92      91
## 5           Jan Oblak   47        93      91
## 6     Kevin De Bruyne  119        91      91
## 7  Robert Lewandowski   89        91      91
## 8         Eden Hazard   66        91      91
## 9             Alisson   53        91      90
## 10      Mohamed Salah   94        90      90
```

```
knitr::kable(fifa_data,"pipe",col.names = c("Name","Hits","Potential","Overall"), align =c("l","c","c")
```

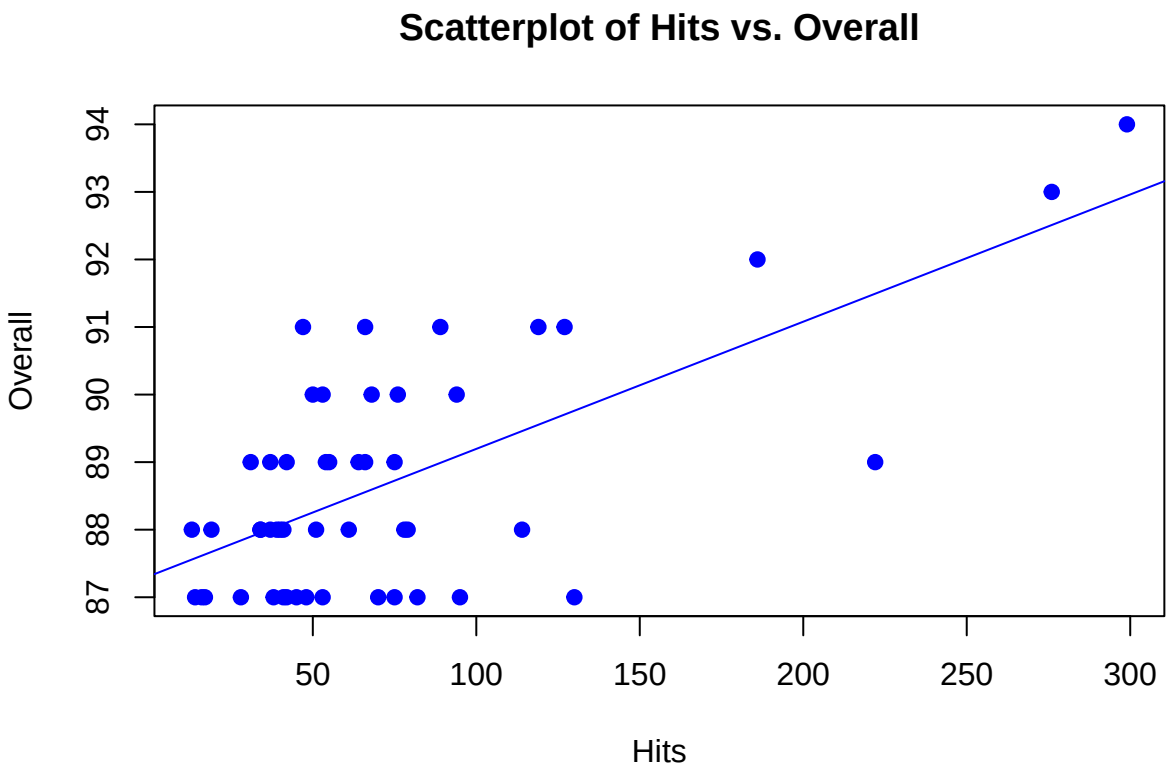
Name	Hits	Potential	Overall
Lionel Messi	299	94	94
Cristiano Ronaldo	276	93	93
Neymar Jr	186	92	92
Virgil van Dijk	127	92	91
Jan Oblak	47	93	91
Kevin De Bruyne	119	91	91
Robert Lewandowski	89	91	91
Eden Hazard	66	91	91
Alisson	53	91	90
Mohamed Salah	94	90	90

The citaion for the dataset is: Mishra, A. (2024). FIFA 2021 complete player data [Data set]. Kaggle. https://www.kaggle.com/datasets/aayushmishra1512/fifa-2021-complete-player-data

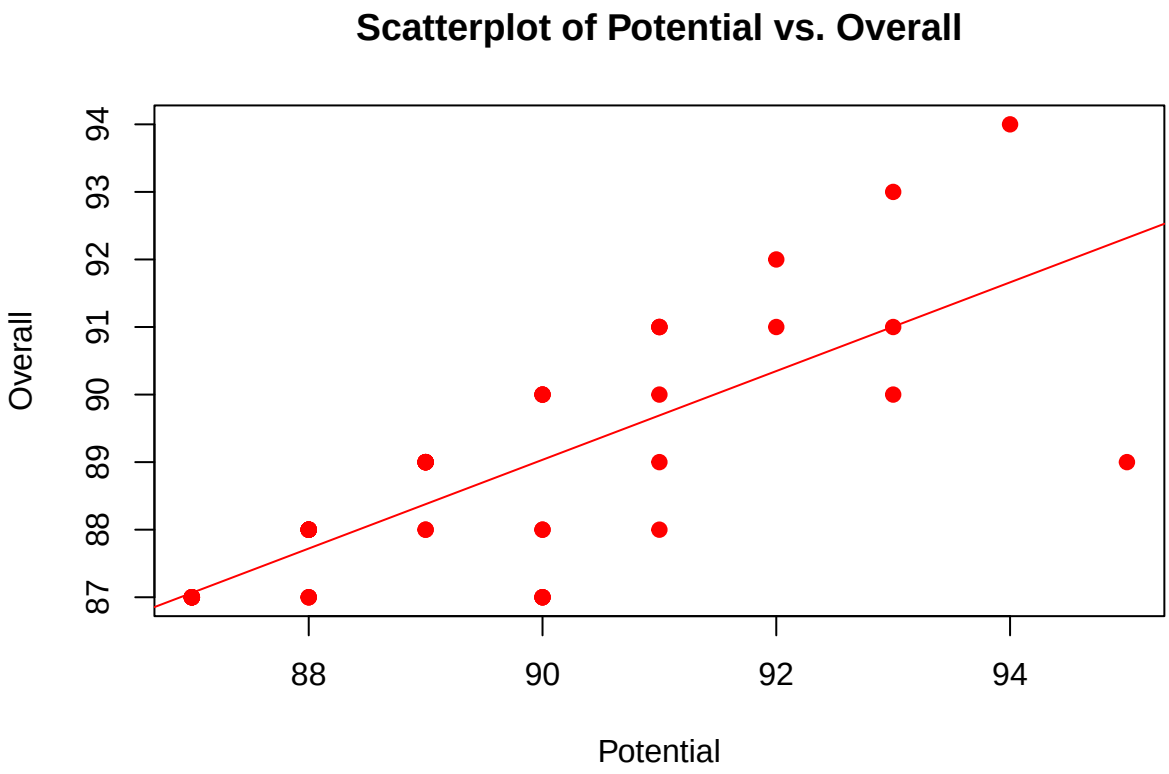
Here's an explanation of each variable in your dataset: # Overall: This variable represents a player's current performance rating on a scale from 1 to 100. It shows how skilled and valuable the player is in professional soccer during the 2024 season. # Hits: Hits refer to the amount of attention or interest a player attracts, measured by the number of interactions they receive.A higher number of hits suggests that the player is gaining more popularity or catching the attention of fans and analysts alike. # Potential: Potential represents the highest level of performance a player is projected to achieve during their career. This rating is given on a scale of 1 to 100 and offers a glimpse into the player's future growth, highlighting their untapped abilities and what they are expected to accomplish as they develop in their career.

The scatterplot for the FIFA dataset:

```r
# Scatterplot for Hits vs. Overall
plot(fifa$Hits, fifa$Overall,
     main = "Scatterplot of Hits vs. Overall",
     xlab = "Hits",
     ylab = "Overall",
     pch = 19,
     col = "blue")
# Adding a linear regression line
abline(lm(Overall ~ Hits, data = fifa), col = "blue")
```




```r
# Scatterplot for Potential vs. Overall
plot(fifa$Potential, fifa$Overall,
     main = "Scatterplot of Potential vs. Overall",
     xlab = "Potential",
     ylab = "Overall",
     pch = 19,
     col = "red")
# Adding a linear regression line
abline(lm(Overall ~ Potential, data = fifa), col = "red")
```




```
fifa1 <- lm(Overall ~ Hits, data = fifa)
summary(fifa1)
```

```
##
## Call:
## lm(formula = Overall ~ Hits, data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.76074 -0.98147  0.01853  0.97198  2.80200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 87.313066   0.288573 302.569  < 2e-16 ***
## Hits         0.018828   0.003075   6.123 1.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.293 on 48 degrees of freedom
## Multiple R-squared:  0.4385, Adjusted R-squared:  0.4268
## F-statistic: 37.49 on 1 and 48 DF,  p-value: 1.629e-07
```

```
# R^2 = 0.4385, Adj_r^2 = 0.4268

fifa2 <- lm(Overall ~ Potential, data = fifa)
summary(fifa2)
```

```
##
## Call:
## lm(formula = Overall ~ Potential, data = fifa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3181 -0.6130  0.2788  0.6221  2.3386
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.93265    7.20616   4.154 0.000134 ***
## Potential    0.65669    0.08053   8.154 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.117 on 48 degrees of freedom
## Multiple R-squared:  0.5808, Adjusted R-squared:  0.572
## F-statistic: 66.49 on 1 and 48 DF,  p-value: 1.288e-10
```

```
# R^2 = 0.5808, Adj_r^2 = 0.572
```

For $Y \sim X1$, $r^2 = 0.4385$ and $R^2_{adj} = 0.4268$ For $Y \sim X2$ $r^2 = 0.5808$ and $R^2_{adj} = 0.572$

Summary for the combined model

```
# Combined model R^2
fifa_combined <- lm(Overall ~ Hits + Potential, data = fifa)
summary(fifa_combined)
```

```
##
## Call:
## lm(formula = Overall ~ Hits + Potential, data = fifa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6311 -0.4591  0.1151  0.7563  1.6093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.723110   8.975289   4.871 1.30e-05 ***
## Hits         0.008054   0.003369   2.391   0.0209 *
## Potential    0.496000   0.102092   4.858 1.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.066 on 47 degrees of freedom
## Multiple R-squared:  0.6262, Adjusted R-squared:  0.6103
## F-statistic: 39.37 on 2 and 47 DF,  p-value: 9.048e-11
```

```
# Adjusted R^2 = 0.6103
```

For $Y \sim X1 + X2$ $R^2_{adj} = 0.6103$

Summary for the above calculations: The scatterplot of Hits vs. Overall shows a moderate positive correlation (R^2 =0.4385), suggesting that higher engagement leads to better performance. The Potential vs. Overall plot shows a stronger correlation (R^2 =0.5808), indicating that a player's potential is a more significant predictor of performance. Combining both variables improves the model's explanatory power, with an adjusted $R^{2,Adj\_r}2$ =0.6103, providing a more accurate view of overall performance.