FIFA 19 Player Analysis

Exploring R

Kareen Ziadat

2024-05-20

Loading the necessary libraries

```
library(tidyverse)
library(ggplot2)
library(readr)
library(skimr)
library(caret)
library(lubridate)
library(dplyr)
library(rnaturalearth)
library(rnaturalearthdata)
library(sf)
library(gridExtra)
```

Loading the dataset

```
FIFA19 <- read_csv("Desktop/ExploringR/FIFA19.csv")
```

Showing the first few rows of the dataset

```
head(FIFA19)
```

```
## # A tibble: 6 x 58
##    ...1 Name               Age Nationality Overall Potential Club   Value Wage
##   <dbl> <chr>            <dbl> <chr>         <dbl>     <dbl> <chr>  <chr> <chr>
## 1     0 L. Messi            31 Argentina        94        94 FC Ba~ €110~ €565K
## 2     1 Cristiano Ronaldo   33 Portugal         94        94 Juven~ €77M  €405K
## 3     2 Neymar Jr           26 Brazil           92        93 Paris~ €118~ €290K
## 4     3 De Gea              27 Spain            91        93 Manch~ €72M  €260K
## 5     4 K. De Bruyne        27 Belgium          91        92 Manch~ €102M €355K
## 6     5 E. Hazard           27 Belgium          91        91 Chels~ €93M  €340K
## # i 49 more variables: `Preferred Foot` <chr>,
## #   `International Reputation` <dbl>, `Weak Foot` <dbl>, `Skill Moves` <dbl>,
## #   `Work Rate` <chr>, Position <chr>, `Jersey Number` <dbl>, Joined <chr>,
## #   `Loaned From` <chr>, `Contract Valid Until` <chr>, Height <chr>,
## #   Weight <chr>, Crossing <dbl>, Finishing <dbl>, HeadingAccuracy <dbl>,
## #   ShortPassing <dbl>, Volleys <dbl>, Dribbling <dbl>, Curve <dbl>,
## #   FKAccuracy <dbl>, LongPassing <dbl>, BallControl <dbl>, ...
```

Undertanding the dataset

```
skim(FIFA19)
```

Table 1: Data summary

Name	FIFA19
Number of rows	18147
Number of columns	58
Column type frequency:	
character	16
numeric	42
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Name	0	1	2	22	0	17140	0
Nationality	0	1	4	20	0	164	0
Club	0	1	3	35	0	652	0
Value	0	1	2	7	0	217	0
Wage	0	1	2	5	0	144	0
Preferred Foot	0	1	4	5	0	2	0
Work Rate	0	1	8	14	0	9	0
Position	0	1	2	3	0	27	0
Joined	0	1	3	12	0	1737	0
Loaned From	0	1	3	35	0	342	0
Contract Valid Until	0	1	4	12	0	37	0
Height	0	1	3	4	0	21	0
Weight	0	1	6	6	0	57	0
Release Clause	0	1	3	17	0	1245	0
League	0	1	1	25	0	38	0
Speciality	0	1	6	20	0	19	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
…1	0	1	9089.24	5257.92	0	4536.5	9076	13662.5	18206	
Age	0	1	25.12	4.67	16	21.0	25	28.0	45	
Overall	0	1	66.25	6.91	46	62.0	66	71.0	94	
Potential	0	1	71.32	6.13	48	67.0	71	75.0	95	
International Reputation	0	1	1.11	0.39	1	1.0	1	1.0	5	
Weak Foot	0	1	2.95	0.66	1	3.0	3	3.0	5	
Skill Moves	0	1	2.36	0.76	1	2.0	2	3.0	5	
Jersey Number	0	1	19.55	15.95	1	8.0	17	26.0	99	
Crossing	0	1	49.74	18.36	5	38.0	54	64.0	93	
Finishing	0	1	45.55	19.53	2	30.0	49	62.0	95	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
HeadingAccuracy	0	1	52.30	17.38	4	44.0	56	64.0	94	
ShortPassing	0	1	58.70	14.70	7	54.0	62	68.0	93	
Volleys	0	1	42.91	17.70	4	30.0	44	57.0	90	
Dribbling	0	1	55.38	18.91	4	49.0	61	68.0	97	
Curve	0	1	47.18	18.40	6	34.0	48	62.0	94	
FKAccuracy	0	1	42.87	17.48	3	31.0	41	57.0	94	
LongPassing	0	1	52.72	15.33	9	43.0	56	64.0	93	
BallControl	0	1	58.37	16.69	5	54.0	63	69.0	96	
Acceleration	0	1	64.61	14.93	12	57.0	67	75.0	97	
SprintSpeed	0	1	64.73	14.65	12	57.0	67	75.0	96	
Agility	0	1	63.50	14.77	14	55.0	66	74.0	96	
Reactions	0	1	61.84	9.01	21	56.0	62	68.0	96	
Balance	0	1	63.96	14.14	16	56.0	66	74.0	96	
ShotPower	0	1	55.47	17.24	2	45.0	59	68.0	95	
Jumping	0	1	65.09	11.82	15	58.0	66	73.0	95	
Stamina	0	1	63.22	15.90	12	56.0	66	74.0	96	
Strength	0	1	65.32	12.55	17	58.0	67	74.0	97	
LongShots	0	1	47.11	19.26	3	33.0	51	62.0	94	
Aggression	0	1	55.88	17.37	11	44.0	59	69.0	95	
Interceptions	0	1	46.70	20.70	3	26.0	52	64.0	92	
Positioning	0	1	49.96	19.53	2	38.0	55	64.0	95	
Vision	0	1	53.41	14.15	10	44.0	55	64.0	94	
Penalties	0	1	48.55	15.70	5	39.0	49	60.0	92	
Composure	0	1	58.65	11.44	3	51.0	60	67.0	96	
Marking	0	1	47.29	19.90	3	30.0	53	64.0	94	
StandingTackle	0	1	47.70	21.66	2	27.0	55	66.0	93	
SlidingTackle	0	1	45.67	21.29	3	24.0	52	64.0	91	
GKDiving	0	1	16.62	17.70	1	8.0	11	14.0	90	
GKHandling	0	1	16.39	16.91	1	8.0	11	14.0	92	
GKKicking	0	1	16.23	16.50	1	8.0	11	14.0	91	
GKPositioning	0	1	16.39	17.04	1	8.0	11	14.0	90	
GKReflexes	0	1	16.71	17.96	1	8.0	11	14.0	94	

```
summary(FIFA19)
```

```
##       ...1           Name                 Age        Nationality
##  Min.   :    0   Length:18147       Min.   :16.00   Length:18147
##  1st Qu.: 4536   Class :character   1st Qu.:21.00   Class :character
##  Median : 9076   Mode  :character   Median :25.00   Mode  :character
##  Mean   : 9089                      Mean   :25.12
##  3rd Qu.:13662                      3rd Qu.:28.00
##  Max.   :18206                      Max.   :45.00
##     Overall        Potential         Club              Value
##  Min.   :46.00   Min.   :48.00   Length:18147       Length:18147
##  1st Qu.:62.00   1st Qu.:67.00   Class :character   Class :character
##  Median :66.00   Median :71.00   Mode  :character   Mode  :character
##  Mean   :66.25   Mean   :71.32
##  3rd Qu.:71.00   3rd Qu.:75.00
##  Max.   :94.00   Max.   :95.00
##      Wage           Preferred Foot     International Reputation   Weak Foot
##  Length:18147       Length:18147       Min.   :1.000           Min.   :1.000
##  Class :character   Class :character   1st Qu.:1.000           1st Qu.:3.000
```

```
##  Mode  :character   Mode  :character   Median :1.000            Median :3.000
##                                        Mean   :1.113            Mean   :2.947
##                                        3rd Qu.:1.000            3rd Qu.:3.000
##                                        Max.   :5.000            Max.   :5.000
##   Skill Moves     Work Rate           Position         Jersey Number
##  Min.   :1.000   Length:18147       Length:18147       Min.   : 1.00
##  1st Qu.:2.000   Class :character   Class :character   1st Qu.: 8.00
##  Median :2.000   Mode  :character   Mode  :character   Median :17.00
##  Mean   :2.361                                         Mean   :19.55
##  3rd Qu.:3.000                                         3rd Qu.:26.00
##  Max.   :5.000                                         Max.   :99.00
##     Joined          Loaned From        Contract Valid Until    Height
##  Length:18147       Length:18147       Length:18147         Length:18147
##  Class :character   Class :character   Class :character     Class :character
##  Mode  :character   Mode  :character   Mode  :character     Mode  :character
##
##
##
##     Weight             Crossing       Finishing     HeadingAccuracy
##  Length:18147       Min.   : 5.00   Min.   : 2.00   Min.   : 4.0
##  Class :character   1st Qu.:38.00   1st Qu.:30.00   1st Qu.:44.0
##  Mode  :character   Median :54.00   Median :49.00   Median :56.0
##                     Mean   :49.74   Mean   :45.55   Mean   :52.3
##                     3rd Qu.:64.00   3rd Qu.:62.00   3rd Qu.:64.0
##                     Max.   :93.00   Max.   :95.00   Max.   :94.0
##   ShortPassing     Volleys        Dribbling         Curve          FKAccuracy
##  Min.   : 7.0   Min.   : 4.00   Min.   : 4.00   Min.   : 6.00   Min.   : 3.00
##  1st Qu.:54.0   1st Qu.:30.00   1st Qu.:49.00   1st Qu.:34.00   1st Qu.:31.00
##  Median :62.0   Median :44.00   Median :61.00   Median :48.00   Median :41.00
##  Mean   :58.7   Mean   :42.91   Mean   :55.38   Mean   :47.18   Mean   :42.87
##  3rd Qu.:68.0   3rd Qu.:57.00   3rd Qu.:68.00   3rd Qu.:62.00   3rd Qu.:57.00
##  Max.   :93.0   Max.   :90.00   Max.   :97.00   Max.   :94.00   Max.   :94.00
##   LongPassing     BallControl     Acceleration    SprintSpeed       Agility
##  Min.   : 9.00   Min.   : 5.00   Min.   :12.00   Min.   :12.00   Min.   :14.0
##  1st Qu.:43.00   1st Qu.:54.00   1st Qu.:57.00   1st Qu.:57.00   1st Qu.:55.0
##  Median :56.00   Median :63.00   Median :67.00   Median :67.00   Median :66.0
##  Mean   :52.72   Mean   :58.37   Mean   :64.61   Mean   :64.73   Mean   :63.5
##  3rd Qu.:64.00   3rd Qu.:69.00   3rd Qu.:75.00   3rd Qu.:75.00   3rd Qu.:74.0
##  Max.   :93.00   Max.   :96.00   Max.   :97.00   Max.   :96.00   Max.   :96.0
##    Reactions        Balance        ShotPower        Jumping
##  Min.   :21.00   Min.   :16.00   Min.   : 2.00   Min.   :15.00
##  1st Qu.:56.00   1st Qu.:56.00   1st Qu.:45.00   1st Qu.:58.00
##  Median :62.00   Median :66.00   Median :59.00   Median :66.00
##  Mean   :61.84   Mean   :63.96   Mean   :55.47   Mean   :65.09
##  3rd Qu.:68.00   3rd Qu.:74.00   3rd Qu.:68.00   3rd Qu.:73.00
##  Max.   :96.00   Max.   :96.00   Max.   :95.00   Max.   :95.00
##     Stamina         Strength       LongShots       Aggression    Interceptions
##  Min.   :12.00   Min.   :17.00   Min.   : 3.00   Min.   :11.00   Min.   : 3.0
##  1st Qu.:56.00   1st Qu.:58.00   1st Qu.:33.00   1st Qu.:44.00   1st Qu.:26.0
##  Median :66.00   Median :67.00   Median :51.00   Median :59.00   Median :52.0
##  Mean   :63.22   Mean   :65.32   Mean   :47.11   Mean   :55.88   Mean   :46.7
##  3rd Qu.:74.00   3rd Qu.:74.00   3rd Qu.:62.00   3rd Qu.:69.00   3rd Qu.:64.0
##  Max.   :96.00   Max.   :97.00   Max.   :94.00   Max.   :95.00   Max.   :92.0
##   Positioning        Vision        Penalties       Composure
```

```
##  Min.   : 2.00   Min.   :10.00   Min.   : 5.00   Min.   : 3.00
##  1st Qu.:38.00   1st Qu.:44.00   1st Qu.:39.00   1st Qu.:51.00
##  Median :55.00   Median :55.00   Median :49.00   Median :60.00
##  Mean   :49.96   Mean   :53.41   Mean   :48.55   Mean   :58.65
##  3rd Qu.:64.00   3rd Qu.:64.00   3rd Qu.:60.00   3rd Qu.:67.00
##  Max.   :95.00   Max.   :94.00   Max.   :92.00   Max.   :96.00
##     Marking      StandingTackle SlidingTackle      GKDiving       GKHandling
##  Min.   : 3.00   Min.   : 2.0   Min.   : 3.00   Min.   : 1.00   Min.   : 1.00
##  1st Qu.:30.00   1st Qu.:27.0   1st Qu.:24.00   1st Qu.: 8.00   1st Qu.: 8.00
##  Median :53.00   Median :55.0   Median :52.00   Median :11.00   Median :11.00
##  Mean   :47.29   Mean   :47.7   Mean   :45.67   Mean   :16.62   Mean   :16.39
##  3rd Qu.:64.00   3rd Qu.:66.0   3rd Qu.:64.00   3rd Qu.:14.00   3rd Qu.:14.00
##  Max.   :94.00   Max.   :93.0   Max.   :91.00   Max.   :90.00   Max.   :92.00
##    GKKicking     GKPositioning     GKReflexes    Release Clause
##  Min.   : 1.00   Min.   : 1.00   Min.   : 1.00   Length:18147
##  1st Qu.: 8.00   1st Qu.: 8.00   1st Qu.: 8.00   Class :character
##  Median :11.00   Median :11.00   Median :11.00   Mode  :character
##  Mean   :16.23   Mean   :16.39   Mean   :16.71
##  3rd Qu.:14.00   3rd Qu.:14.00   3rd Qu.:14.00
##  Max.   :91.00   Max.   :90.00   Max.   :94.00
##     League           Speciality
##  Length:18147       Length:18147
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
colSums(is.na(FIFA19))
```

```
##                      ...1                      Name                       Age
##                         0                         0                         0
##               Nationality                   Overall                 Potential
##                         0                         0                         0
##                      Club                     Value                      Wage
##                         0                         0                         0
##           Preferred Foot International Reputation                 Weak Foot
##                         0                         0                         0
##              Skill Moves                 Work Rate                  Position
##                         0                         0                         0
##            Jersey Number                    Joined               Loaned From
##                         0                         0                         0
##     Contract Valid Until                    Height                    Weight
##                         0                         0                         0
##                 Crossing                 Finishing         HeadingAccuracy
##                         0                         0                         0
##              ShortPassing                   Volleys                 Dribbling
##                         0                         0                         0
##                     Curve                FKAccuracy              LongPassing
##                         0                         0                         0
##               BallControl              Acceleration              SprintSpeed
##                         0                         0                         0
##                   Agility                 Reactions                   Balance
##                         0                         0                         0
##                ShotPower                   Jumping                   Stamina
```

```
##                    0                    0                    0
##             Strength            LongShots           Aggression
##                    0                    0                    0
##        Interceptions          Positioning               Vision
##                    0                    0                    0
##            Penalties           Composure              Marking
##                    0                    0                    0
##       StandingTackle       SlidingTackle             GKDiving
##                    0                    0                    0
##           GKHandling           GKKicking        GKPositioning
##                    0                    0                    0
##           GKReflexes       Release Clause               League
##                    0                    0                    0
##          Speciality
##                    0
```

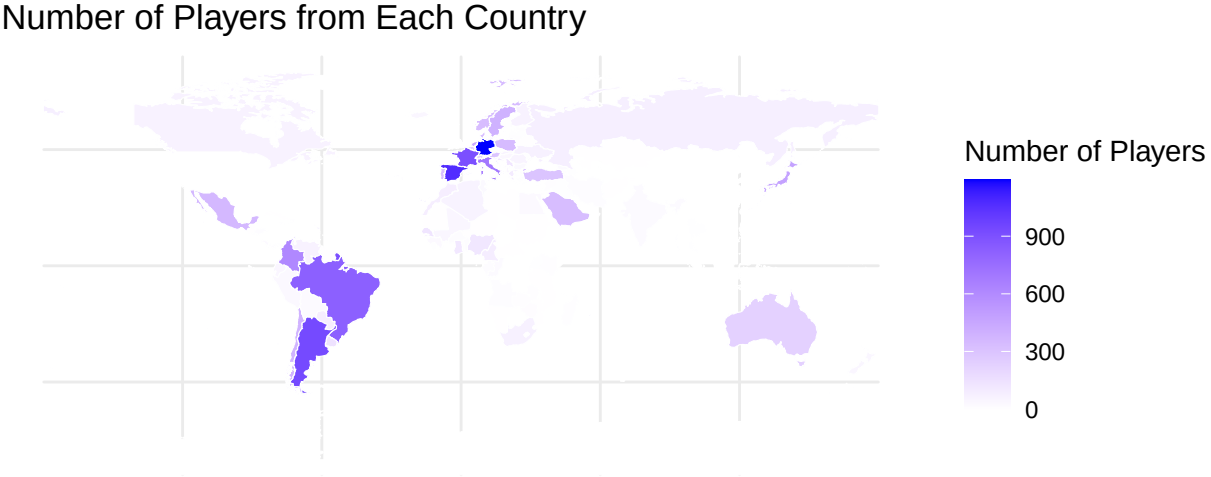
No missing values. Wage, Value are characters that need to be made numeric for further analysis.

Changing wage and value to numeric

```
clean_wage <- function(wage) {
  as.numeric(gsub("€|K", "", wage)) * 1000
}

clean_value <- function(value) {
  value <- gsub("€", "", value)
  if (grepl("M", value)) {
    as.numeric(gsub("M", "", value)) * 1e6
  } else if (grepl("K", value)) {
    as.numeric(gsub("K", "", value)) * 1e3
  } else {
    as.numeric(value)
  }
}

FIFA19 <- FIFA19 %>%
  mutate(
    Wage = sapply(Wage, clean_wage),
    Value = sapply(Value, clean_value)
  )
```

Number of players from each country visualized in a map

```
nationality_counts <- FIFA19 %>%
  group_by(Nationality) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count))

# Get world map data
world <- ne_countries(scale = "medium", returnclass = "sf")

world <- left_join(world, nationality_counts, by = c("name" = "Nationalit

world$Count[is.na(world$Count)] <- 0

heatmap <- ggplot(data = world) +
  geom_sf(aes(fill = Count), color = "white") +
  scale_fill_gradient(low = "white", high = "blue", na.value = "grey50")
  labs(title = "Number of Players from Each Country",
       fill = "Number of Players") +
  theme_minimal()

print(heatmap)
```



Key Attributes Summary

```
summary(FIFA19[, c('Age', 'Overall','Wage','Value')])
```

```
##       Age             Overall          Wage             Value
##  Min.   :16.00   Min.   :46.00   Min.   :     0   Min.   :        0
##  1st Qu.:21.00   1st Qu.:62.00   1st Qu.:  1000   1st Qu.:   300000
##  Median :25.00   Median :66.00   Median :  3000   Median :   675000
##  Mean   :25.12   Mean   :66.25   Mean   :  9759   Mean   :  2417729
##  3rd Qu.:28.00   3rd Qu.:71.00   3rd Qu.:  9000   3rd Qu.:  2000000
##  Max.   :45.00   Max.   :94.00   Max.   :565000   Max.   :118500000
```

Conclusions:
Age: Players range from 16 to 45 years old, with a median age of 25.

Overall Rating: Ratings range from 46 to 94, with a median of 66.

Potential: Potential ratings range from 48 to 95, with a median of 71.

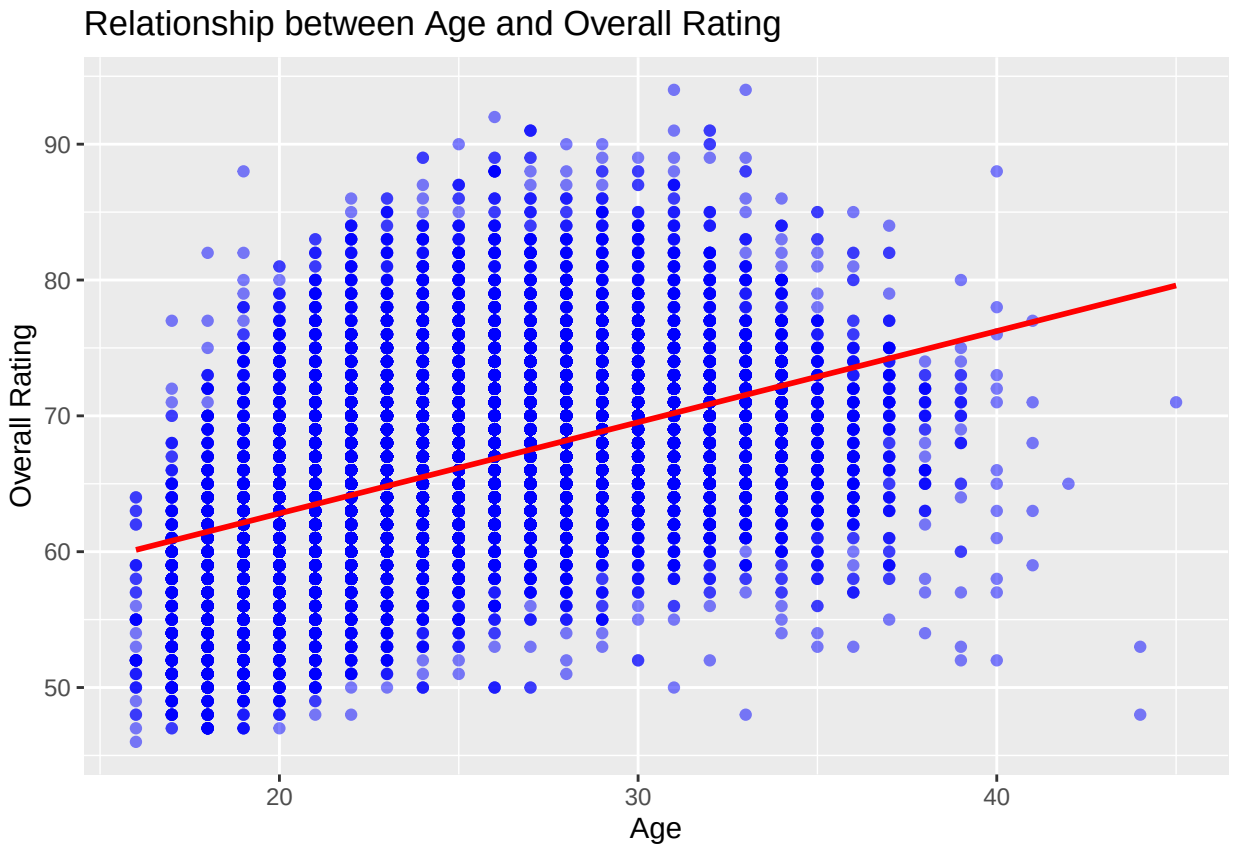
Value: Value of players range from 0 to 110.5 Million Euros, with a median of 675 Thousand Euros.

Wage: Wage of player from 0 to 565 Million Euros, with a median of 3 Million Euros.

Exploring relationship between age and overall rating

```
ggplot(FIFA19, aes(x = Age, y = Overall)) +
  geom_point(alpha = 0.5, color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relationship between Age and Overall Rating",
       x = "Age",
       y = "Overall Rating")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



What does this mean?

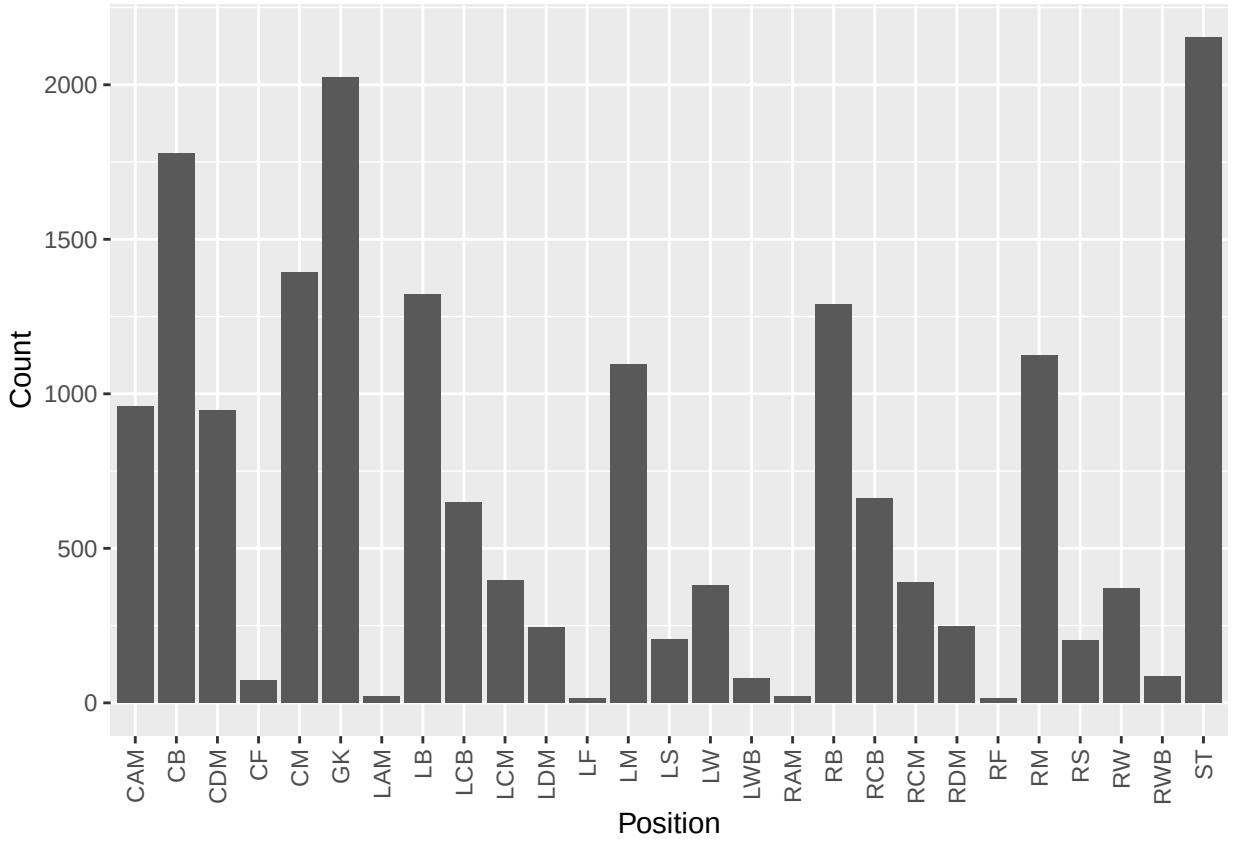
```
model <- lm(Overall ~ Age, data = FIFA19)
summary(model)
```

```
##
## Call:
## lm(formula = Overall ~ Age, data = FIFA19)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.9247  -4.1726  -0.4745   3.8139  25.8543
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.393615   0.250302  197.34   <2e-16 ***
## Age          0.671161   0.009796   68.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.162 on 18145 degrees of freedom
## Multiple R-squared:  0.2055, Adjusted R-squared:  0.2055
## F-statistic:  4694 on 1 and 18145 DF,  p-value: < 2.2e-16
```

These results suggest a substantial correlation between a player's age and overall rating. According to the model, each additional year of age results in an anticipated 0.67 point increase in total rating. The R^2 value of 0.2055 shows that age accounts for 20.55% of the variance in overall rating, indicating a moderate but significant correlation. The F-statistic's exceptionally low p-value emphasizes the model's overall statistical significance, indicating that age is a strong predictor of overall rating in FIFA 19 players.

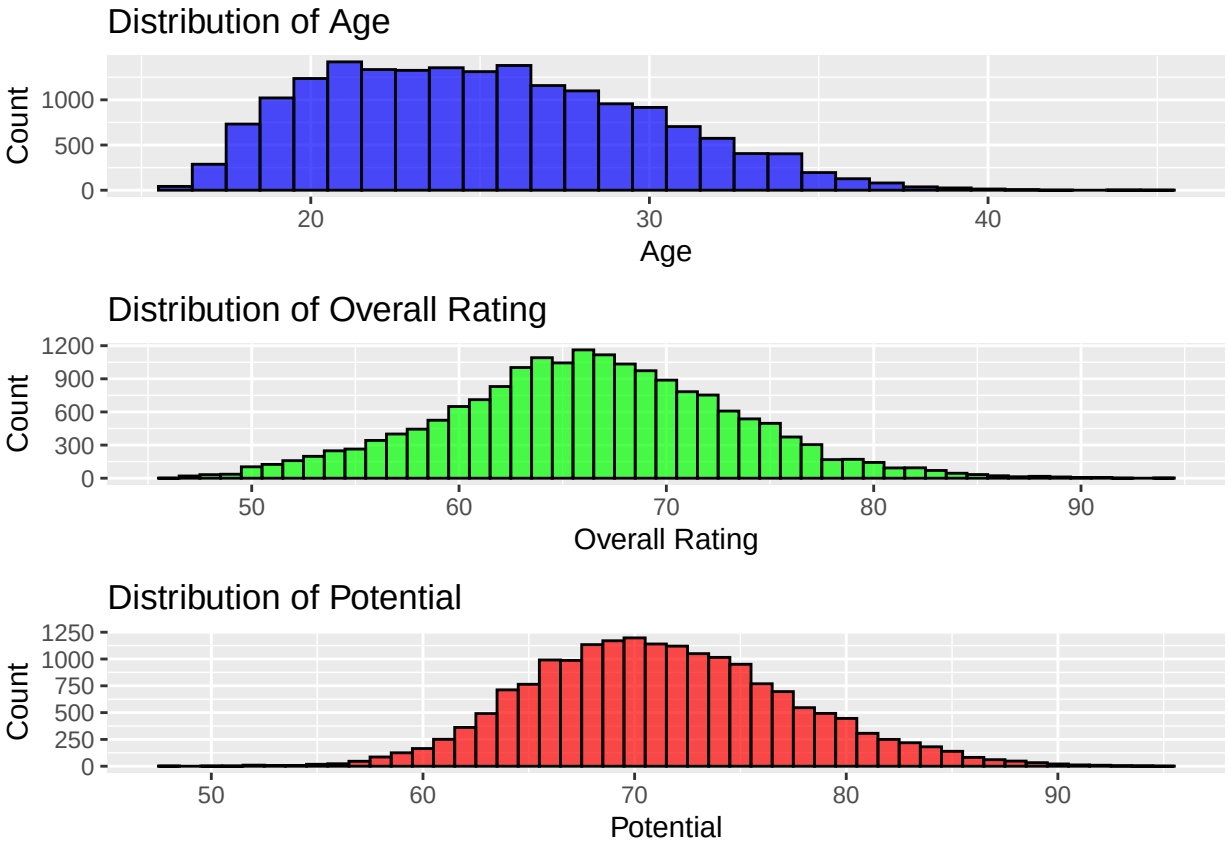
Distribution of player positions

```
position_distribution <- FIFA19 %>%
  group_by(Position) %>%
  summarise(Count = n())

ggplot(position_distribution, aes(x = Position, y = Count)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Distribution of player overall ratings

```
p1 <- ggplot(FIFA19, aes(x=Age)) +
  geom_histogram(binwidth=1, fill='blue', color='black', alpha=0.7) +
  labs(title='Distribution of Age', x='Age', y='Count')

p2 <- ggplot(FIFA19, aes(x=Overall)) +
  geom_histogram(binwidth=1, fill='green', color='black', alpha=0.7) +
  labs(title='Distribution of Overall Rating', x='Overall Rating', y='Cou

p3 <- ggplot(FIFA19, aes(x=Potential)) +
  geom_histogram(binwidth=1, fill='red', color='black', alpha=0.7) +
  labs(title='Distribution of Potential', x='Potential', y='Count')

grid.arrange(p1, p2, p3, ncol=1)
```



Normally distributed

Exploring Correlation between these attributes

```
corr <- cor(FIFA19[c('Overall','Potential','Wage','Value')])
corr
```

```
##               Overall Potential      Wage     Value
## Overall   1.0000000 0.6609468 0.5717952 0.6270254
## Potential 0.6609468 1.0000000 0.4866083 0.5766062
## Wage      0.5717952 0.4866083 1.0000000 0.8583764
## Value     0.6270254 0.5766062 0.8583764 1.0000000
```

Size of Correlation	Interpretation
.90 to 1.00 (−.90 to −1.00)	Very high positive (negative) correlation
.70 to .90 (−.70 to −.90)	High positive (negative) correlation
.50 to .70 (−.50 to −.70)	Moderate positive (negative) correlation
.30 to .50 (−.30 to −.50)	Low positive (negative) correlation
.00 to .30 (.00 to −.30)	negligible correlation

The strongest correlation is between Wage and Value (0.858), which means that players with higher market values earn higher wages, which is understandable given that more valued players are often more sought after and better compensated. Moreover, potential has a moderate positive correlation with Overall (0.661), suggesting that players with better potential tend to have higher present ratings.I find the correlation between Potential and Value (0.577) particularly noteworthy, implying that players with higher potential have a higher market values, likely due to their anticipated future development and contribution to the club.

Which country should we scout from ?

Average overall rating by nationality

```
top_10_nationalities <- FIFA19 %>%
  group_by(Nationality) %>%
  summarise(Avg_Overall = mean(Overall, na.rm = TRUE)) %>%
  arrange(desc(Avg_Overall)) %>%
  head(10)%>%
  pull(Nationality)

print(top_10_nationalities)
```

```
##  [1] "United Arab Emirates" "Central African Rep." "Israel"
##  [4] "Dominican Republic"   "Oman"                 "São Tomé & Príncip
##  [7] "Cape Verde"           "Portugal"             "Togo"
## [10] "Brazil"
```

This doesn't make a lot of sense let's explore the count of players from these nationalities.

```
players_by_top_10_nationalities <- FIFA19 %>%
  filter(Nationality %in% top_10_nationalities) %>%
  group_by(Nationality) %>%
  summarise(Player_Count = n()) %>%
  arrange(desc(Player_Count))

players_by_top_10_nationalities
```

```
## # A tibble: 10 x 2
##    Nationality          Player_Count
##    <chr>                       <int>
##  1 Brazil                        825
##  2 Portugal                      322
##  3 Cape Verde                     19
##  4 Israel                         14
##  5 Togo                           12
##  6 Central African Rep.            3
```

```
##  7 Dominican Republic              2
##  8 Oman                            1
##  9 São Tomé & Príncipe             1
## 10 United Arab Emirates            1
```

Let's narrow the search down to only the nationalities with more than 300 players (significant and a more fair approach).

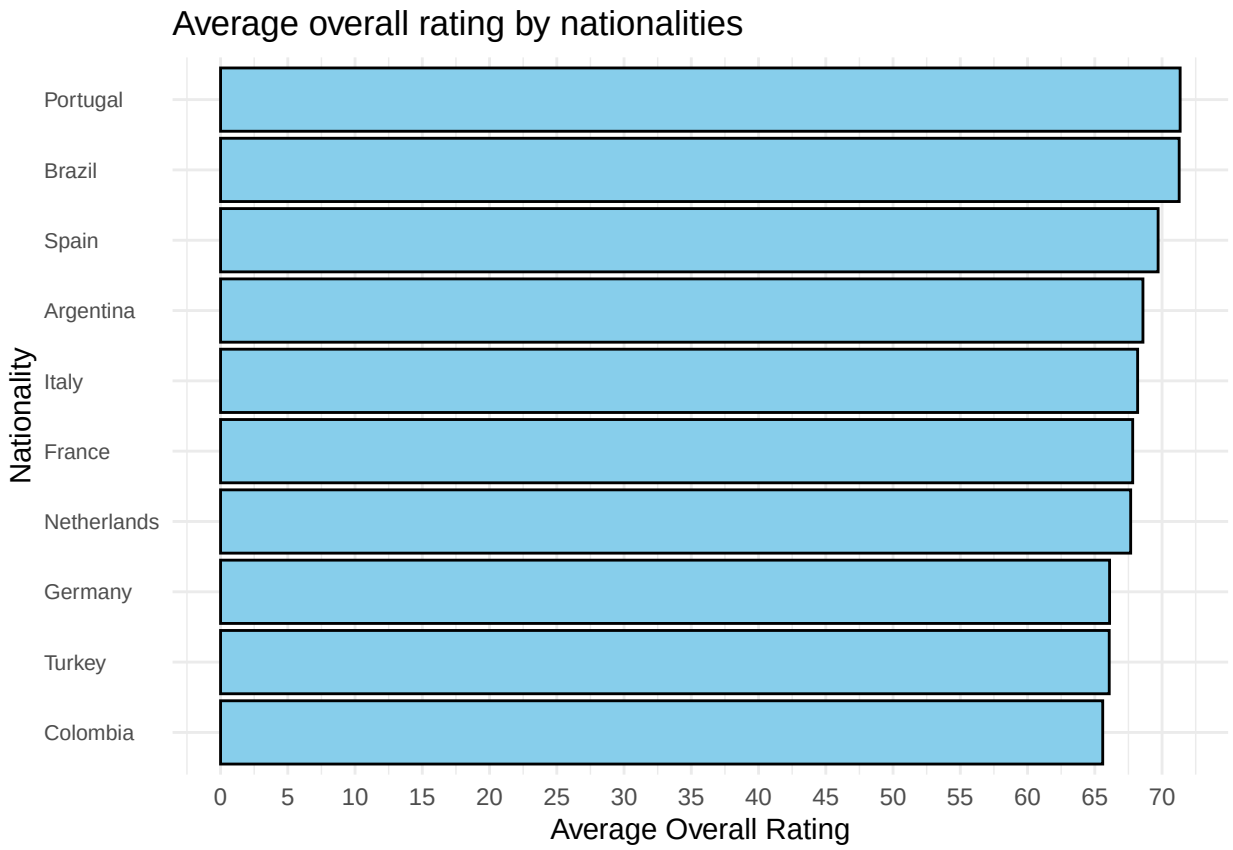
```
player_counts <- FIFA19 %>%
  group_by(Nationality) %>%
  summarise(Player_Count = n())

nationalities_with_300_plus_players <- player_counts %>%
  filter(Player_Count > 300) %>%
  pull(Nationality)
```

Average overall rating by nationalities with 300+ players.

```
top_10_nationalities_by_avg_rating <- FIFA19 %>%
  filter(Nationality %in% nationalities_with_300_plus_players) %>%
  group_by(Nationality) %>%
  summarise(Avg_Overall = mean(Overall, na.rm = TRUE)) %>%
  arrange(desc(Avg_Overall)) %>%
  head(10)

top_10_nationalities_by_avg_rating
```

```
## # A tibble: 10 x 2
##    Nationality Avg_Overall
##    <chr>             <dbl>
##  1 Portugal           71.3
##  2 Brazil             71.3
##  3 Spain              69.7
##  4 Argentina          68.6
##  5 Italy              68.2
##  6 France             67.8
##  7 Netherlands        67.7
##  8 Germany            66.1
##  9 Turkey             66.1
## 10 Colombia           65.6
```

```
ggplot(top_10_nationalities_by_avg_rating, aes(x = Avg_Overall, y = reord
  geom_bar(stat = "identity", fill = "skyblue", color = "black") +
  labs(x = "Average Overall Rating", y = "Nationality", title = "Average
  scale_x_continuous(breaks = seq(0, max(top_10_nationalities_by_avg_rati
  theme_minimal() +
  theme(axis.text.y = element_text(size = 8, hjust = 0))
```



Conclusions:
As expected, European countries, Brazil and Argentina lead the list in terms of average player ratings. This may be attributed to existing structures for developing young talent, such as strong academies. Additionally, the presence of football legends from these regions is likely to inspire players. There is a chance to capitalize on this by proactively establishing academies to scout and attract people from these high-performing countries, thereby increasing the talent pool accessible for growth and success.

Players to be on the lookout for

My theory is players with a high difference between the potential and their currrent rating are worth looking into.

```
FIFA19$Difference <- FIFA19$Potential - FIFA19$Overall

fifa_sorted <- FIFA19[order(-FIFA19$Difference),]

result <- fifa_sorted[c('Name', 'Age', 'Nationality', 'Overall', 'Potenti

print(n=30, result)
```

```
## # A tibble: 30 x 9
##    Name            Age Nationality Overall Potential  Wage `Contract Va
##    <chr>         <dbl> <chr>         <dbl>     <dbl> <dbl> <chr>
##  1 J. von Moos      17 Switzerland      58        84  2000 2021
##  2 D. Campbell      17 Scotland         50        76  1000 2019
##  3 Y. Lenze         17 Germany          57        82  1000 2020
##  4 B. Mumba         16 England          55        80  1000 2021
##  5 K. Askildsen     17 Norway           52        77  1000 2021
##  6 A. Dabo          17 France           62        86  2000 2022
##  7 G. Azzinnari     17 Italy            59        83  1000 2020
##  8 I. Sauter        17 Switzerland      56        80  1000 2021
##  9 K. Lara          16 Colombia         50        74  1000 2021
## 10 E. Destanog~     17 Turkey           50        74  1000 2020
## 11 N. Ayéva         16 Sweden           48        72  1000 2020
## 12 R. Griffiths     18 England          61        84  5000 2022
## 13 B. Gilmour       17 Scotland         61        84  7000 2021
## 14 J. Romero        17 Argentina        58        81  1000 2022
## 15 A. Appiah        17 England          58        81  4000 2020
## 16 A. Wilson        18 England          57        80  3000 2020
## 17 B. Kavanagh      17 Republic o~      57        80  1000 2020
## 18 Y. Toure         17 Switzerland      57        80  3000 2021
## 19 S. Giménez       17 Mexico           56        79  2000 2021
## 20 O. Aga           17 Norway           55        78  1000 2020
## 21 A. Sønderga~     17 Denmark          54        77  2000 2020
```

```
## 22 T. Renaud       17 France            54        77  1000 2021
## 23 E. Ceide        16 Norway            52        75  1000 2021
## 24 J. Dunn         17 Canada            51        74  1000 2021
## 25 E. McCue        17 Sweden            51        74  1000 2021
## 26 S. Sarri        17 Italy             51        74  1000 2021
## 27 D. Asonganyi    17 England           50        73  1000 2020
## 28 F. Ranocchia    17 Italy             49        72  1000 2021
## 29 J. Livesey      18 England           47        70  1000 2021
## 30 W. Geubbels     16 France            64        86  5000 2023
## # i 2 more variables: `Release Clause` <chr>, Speciality <chr>
```

We can test this theory since its 2024 and I can see if I recognize any of the above players. The results show that this isn't a promising strategy.

lets narrow it down to players with a rating of 65+ as their level is already considered decent.

```
FIFA19_filtered <- FIFA19[FIFA19$Overall >= 65, ]

FIFA19_filtered$Difference <- FIFA19_filtered$Potential - FIFA19_filtered

fifa_sorted <- FIFA19_filtered[order(-FIFA19_filtered$Difference), ]

result <- fifa_sorted[c('Name', 'Age', 'Nationality', 'Overall', 'Potenti

print(n = 30, result)
```

```
## # A tibble: 30 x 9
##    Name           Age Nationality Overall Potential  Wage `Contract Va
##    <chr>        <dbl> <chr>         <dbl>     <dbl> <dbl> <chr>
##  1 S. Tonali       18 Italy            68        89  1000 2021
##  2 P. Pellegri     17 Italy            67        88 11000 2022
##  3 S. Diop         18 France           66        87  8000 2021
##  4 C. Früchtl      18 Germany          65        86  3000 2020
##  5 E. Ampadu       17 Wales            65        86  8000 2023
##  6 Riqui Puig      18 Spain            69        89 24000 2021
```

```
##  7 A. Bastoni      19 Italy           67        87  7000 Jun 30, 2019
##  8 A. Gomes        17 England         67        87 15000 2021
##  9 T. Chong        18 Netherlands     65        85 13000 2019
## 10 J. Arp          18 Germany         69        88  4000 2020
## 11 Sergio Gómez    17 Spain           68        87  8000 2021
## 12 A. Gouiri       18 France          67        86 10000 2021
## 13 M. Gibbs-Wh~    18 England         67        86 10000 2022
## 14 M. Boadu        17 Netherlands     66        85  3000 2021
## 15 M. Sylla        18 France          65        84  8000 2020
## 16 V. Supriaga     18 Ukraine         65        84  1000 2023
## 17 Y. Adli         17 France          65        84  7000 2021
## 18 Kangin Lee      17 Korea Repu~     70        88  8000 2022
## 19 G. Itter        19 Germany         67        85  5000 2021
## 20 K. Jóźwiak      20 Poland          66        84  2000 2019
## 21 D. Jastrzem~    18 Germany         66        84  4000 2020
## 22 C. Gomes        17 France          66        84 10000 2023
## 23 A. Plizzari     18 Italy           66        84  4000 2019
## 24 M. Edwards      19 England         66        84 10000 Jun 30, 2019
## 25 R. Guitane      19 France          66        84  4000 2022
## 26 N. Zaniolo      18 Italy           66        84  7000 2023
## 27 J. Carranza     18 Argentina       65        83  2000 2022
## 28 D. Szoboszl~    17 Hungary         65        83  4000 2021
## 29 C. Hudson-O~    17 England         70        87 25000 2020
## 30 Brahim Díaz     18 Spain           69        86 24000 2019
## # i 2 more variables: `Release Clause` <chr>, Speciality <chr>
```

These results are far more promising as the players in this list currently play in top leagues performing at top level.

Based on this information here is a list of promising players in 2024.

```
FIFA24 <- read_csv("Desktop/ExploringR/FIFA24.csv")
```

```
## Rows: 180021 Columns: 109
## -- Column specification --------------------------------------------
## Delimiter: ","
## chr  (43): player_url, short_name, long_name, player_positions, club_n
## dbl  (63): player_id, fifa_version, fifa_update, overall, potential, v
## date  (3): update_as_of, dob, club_joined_date
##
## i Use `spec()` to retrieve the full column specification for this data
## i Specify the column types or set `show_col_types = FALSE` to quiet th
```

```
FIFA24_filtered <- FIFA24[FIFA24$overall >= 65, ]

FIFA24_filtered$Difference <- FIFA24_filtered$potential - FIFA24_filtered

fifa_sorted <- FIFA24_filtered[order(-FIFA24_filtered$Difference), ]

result <- fifa_sorted[c('short_name', 'age', 'nationality_name', 'overall

print(n = 30, result)
```

```
## # A tibble: 30 x 8
##    short_name          age nationality_name    overall potential wage
##    <chr>             <dbl> <chr>                 <dbl>     <dbl>    <
##  1 S. Tonali            18 Italy                    66        88
##  2 M. Edwards           18 England                  65        87    1
##  3 J. Duranville        17 Belgium                  66        87
##  4 N. Lahmadi           18 France                   65        86
##  5 Kayky                18 Brazil                   66        87    1
##  6 R. Cherki            16 France                   67        88
##  7 B. Saka              17 England                  65        86
##  8 A. Gomes             17 England                  66        87    1
##  9 C. Früchtl           18 Germany                  65        86
## 10 Y. Adli              17 France                   65        86
## 11 S. Diop              18 France                   65        86
```

```
## 12 C. Früchtl            17 Germany                 65        86
## 13 T. Bischof            18 Germany                 66        86
## 14 N. Weiper             18 Germany                 66        86
## 15 A. van Axel Dongen    18 Netherlands             65        85
## 16 A. Nusa               17 Norway                  68        88
## 17 J. Bynoe-Gittens      17 England                 67        87
## 18 Kayky                 19 Brazil                  66        86    1
## 19 António Silva         18 Portugal                66        86
## 20 D. Scarlett           18 England                 65        85
## 21 M. Lazetić            18 Serbia                  65        85
## 22 C. Clark              18 United States           66        86
## 23 F. Wirtz              17 Germany                 68        88
## 24 S. Esposito           17 Italy                   66        86
## 25 A. Vranckx            17 Belgium                 66        86
## 26 X. Simons             17 Netherlands             65        85
## 27 T. Parrott            18 Republic of Ireland     65        85
## 28 H. Elliott            17 England                 65        85
## 29 M. Kana               17 Belgium                 65        85
## 30 M. Greenwood          17 England                 67        87
## # i 2 more variables: club_contract_valid_until_year <dbl>, club_posit
```

It is recommended to invest in young players: Given the correlation between age and overall rating and the promising players being in the age range of 16-20, teams should.focus on scouting and investing in younger players with great potential in order to maximize long-term success.

Note that the data sets are from Kaggle, available on these links:
https://www.kaggle.com/datasets/chaitanyahivlekar/fifa-19-player-dataset/data
https://www.kaggle.com/datasets/rehandl23/fifa-24-player-stats-dataset/data