

# Data Analytics



# Title & Research Question

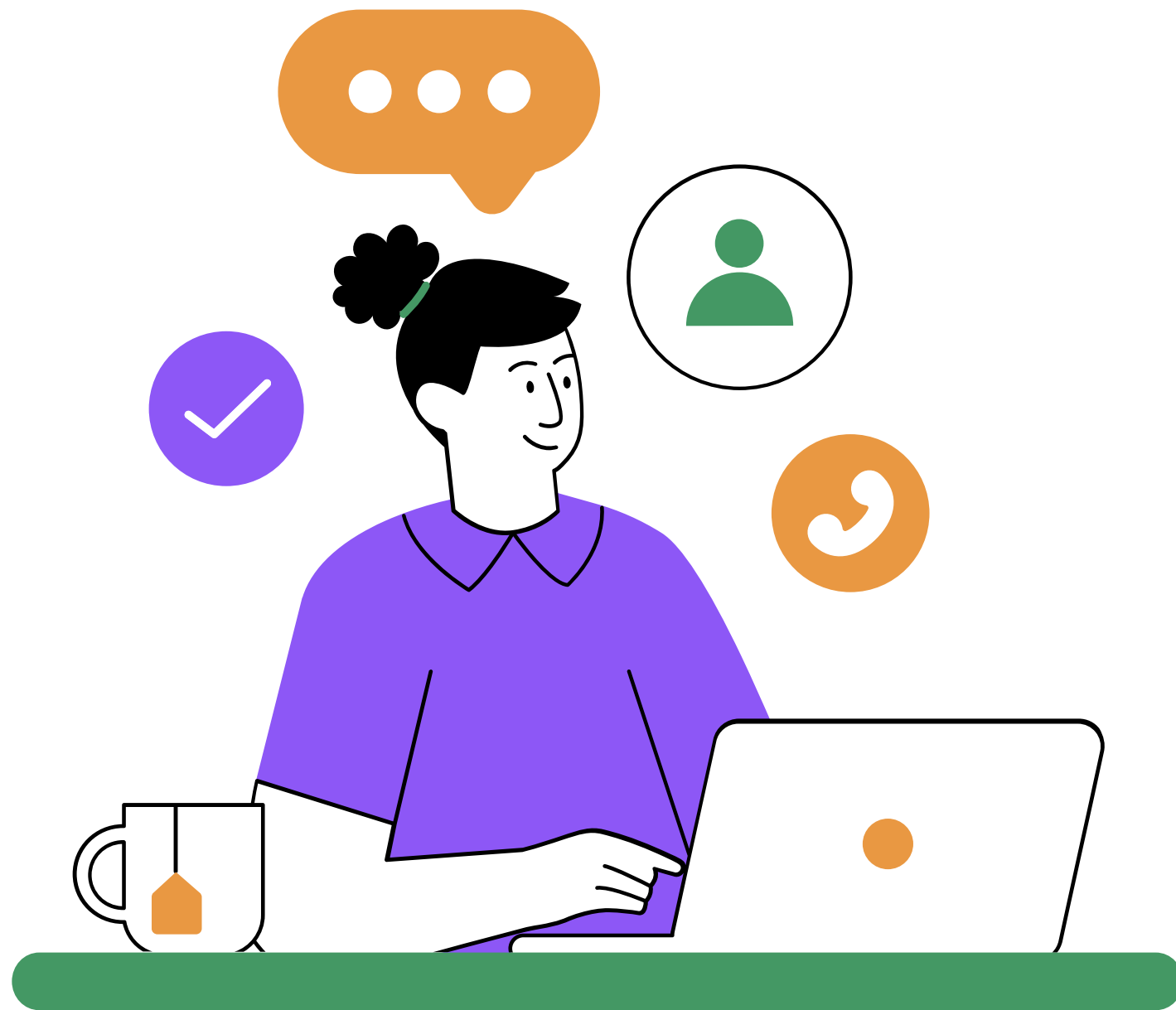
Player Valuation in Professional Football

Which Factors Predict The Market Value of Current Footballers?





# Dataset Overview



- Over 2000 Players analyzed
- Over 100 Variables in Total
- Great Analysis
- Big Dataset with a lot of information
- No Market Value Included



## Data Card

Code (2)

Discussion (0)

Suggestions (0)

## About Dataset

1. Market values ranging from €200M (Haaland, Vinicius Jr.) to €20M
2. Player statistics including goals, assists, and appearances
3. Demographic data including age (17-37) and nationality
4. Club affiliations across major leagues
5. Position-specific information
6. Performance metrics including yellow/red cards and substitution patterns

## Usability ⓘ

## License

MIT

**Expected update frequency**

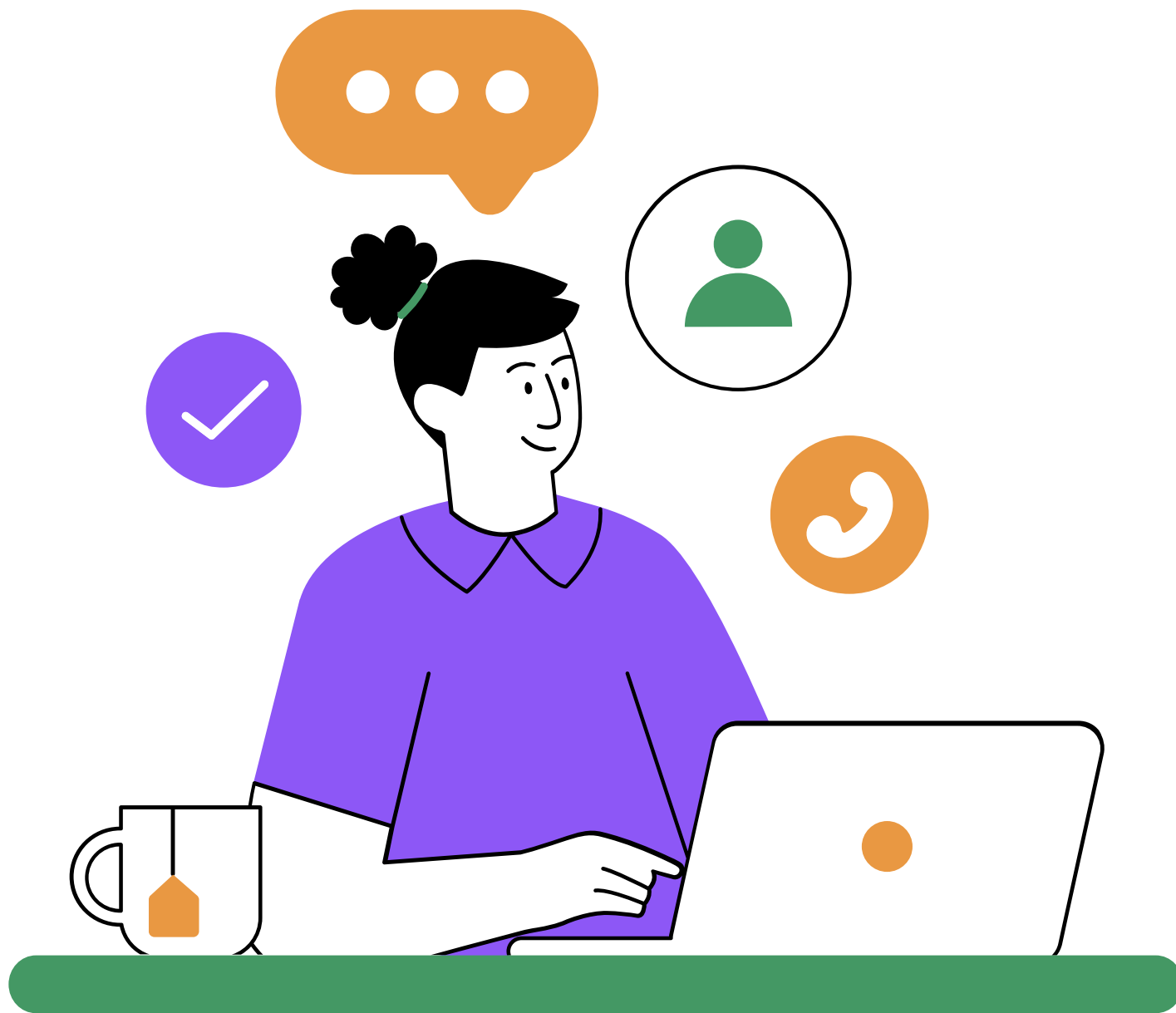
Annually

## Tags

## Football

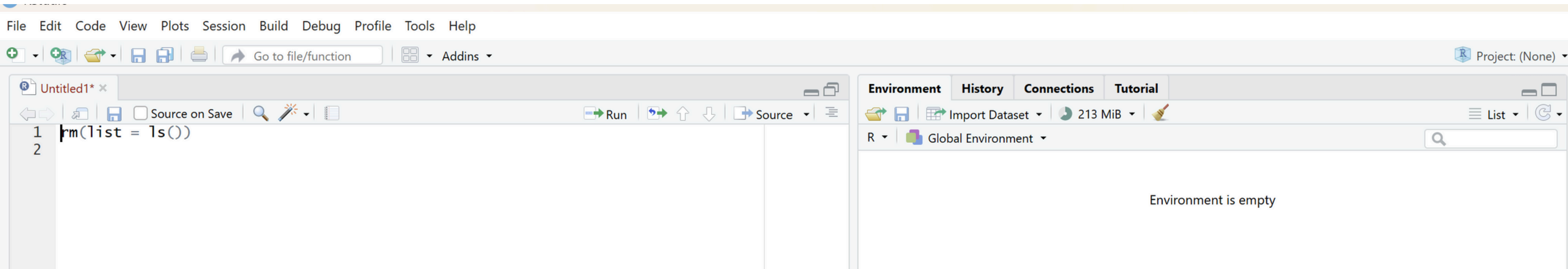


# Dataset Overview

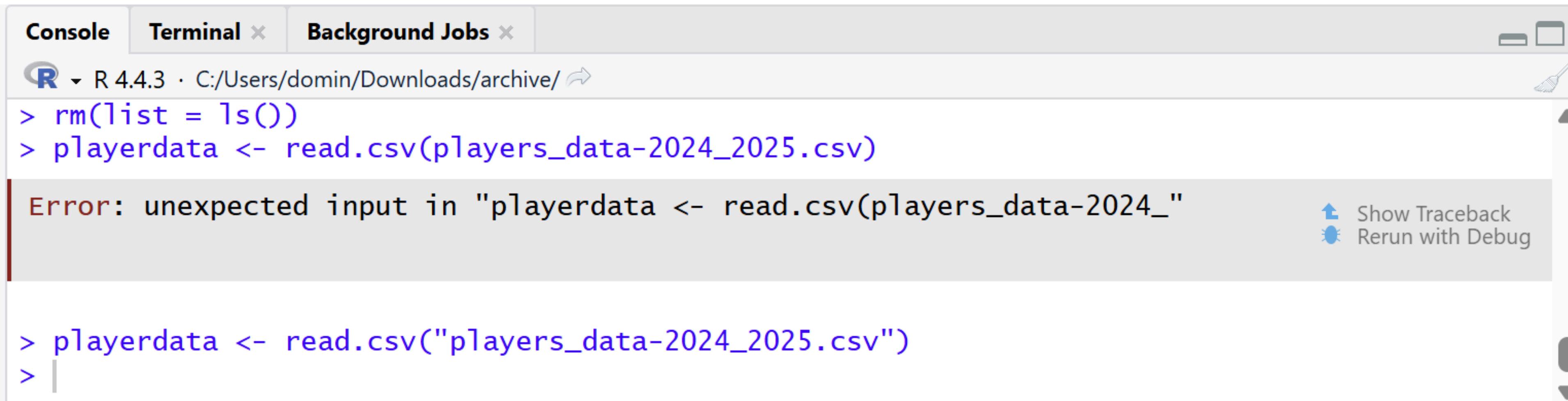


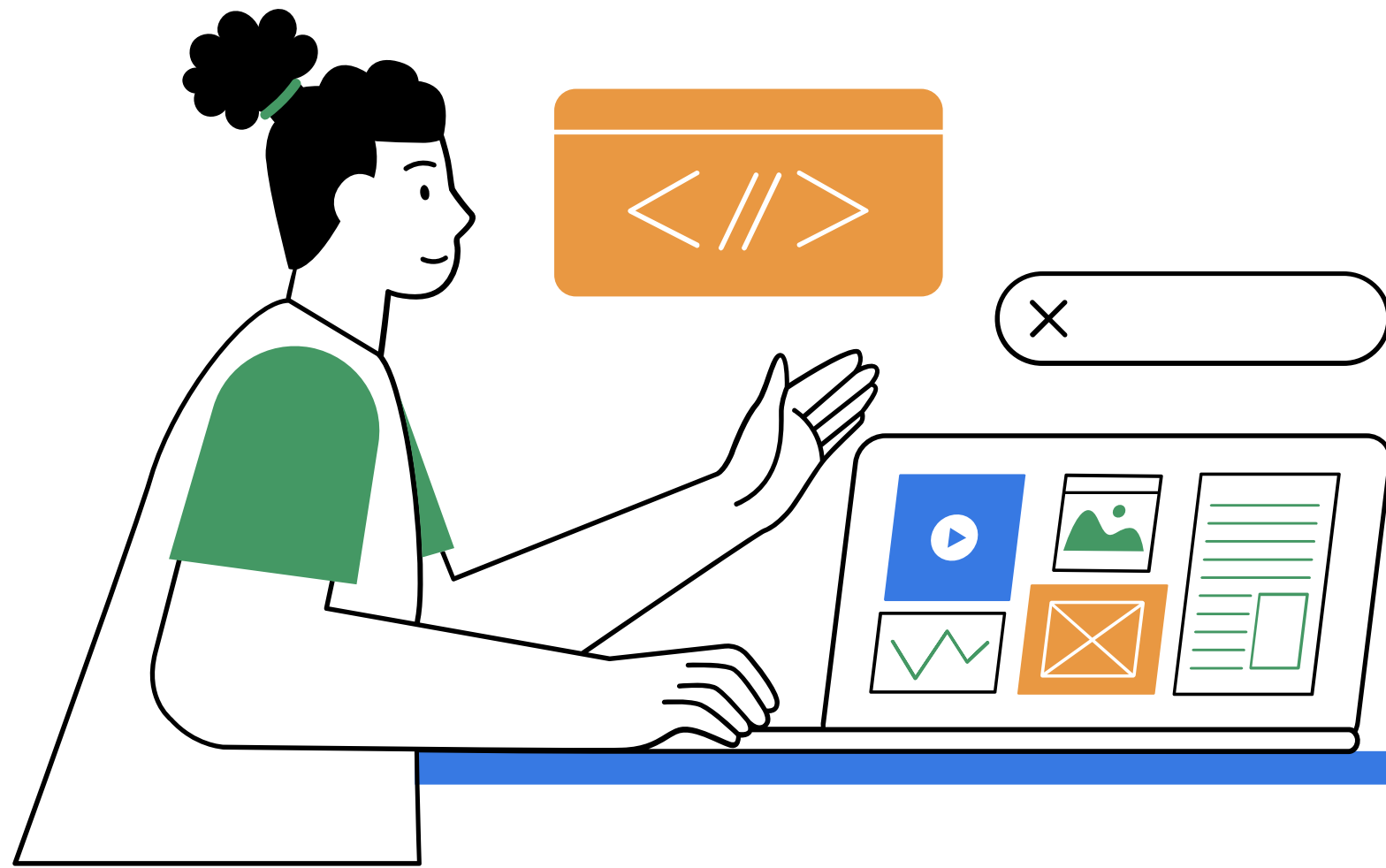
- Top 500 most valuable players (Transfermarkt 2024)
- Variables used:
- Market Value
- Goals
- Assists
- Age
- Elite sample → limited age variation

# Clean up Previous Experiments



## Import the Kaggle Dataset





# Planned Analysis

- Simple regressions:
  1. Market Value ~ Age
  2. Market Value ~ Goals
  3. Market Value ~ Assists
- Multiple regression:
  4. Market Value ~ Goals + Assists + Age



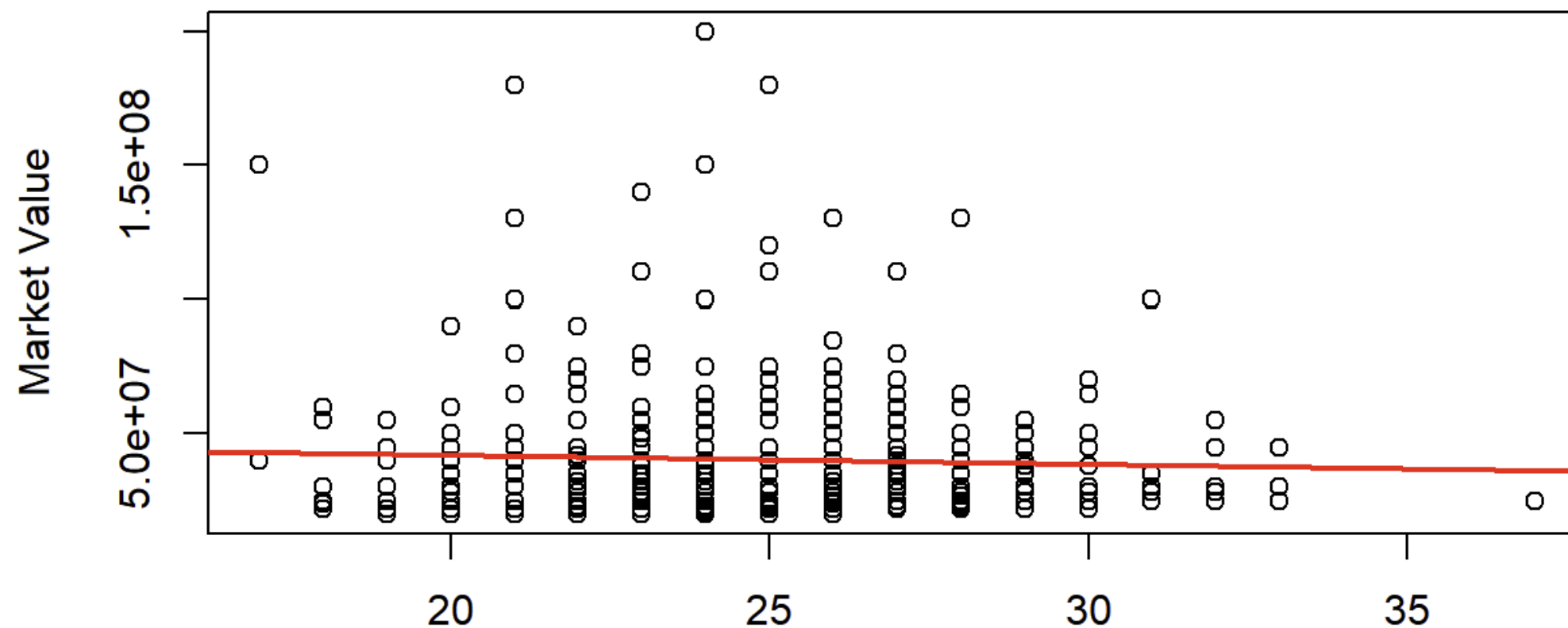
# Variable: Age

- Age coefficient: –€0.38M
- p-value: not significant
- $R^2 = 0.002$

```
2 model_age <- lm(Market.Value ~ Age, data = players)
3 summary(model_age)
```

```
12 plot(players$Age, players$Market.Value,
13       xlab = "Age",
14       ylab = "Market Value",
15       main = "Market Value vs Age")
16 abline(lm(Market.Value ~ Age, data = players), col="red", lwd=2)
```

Market Value vs Age

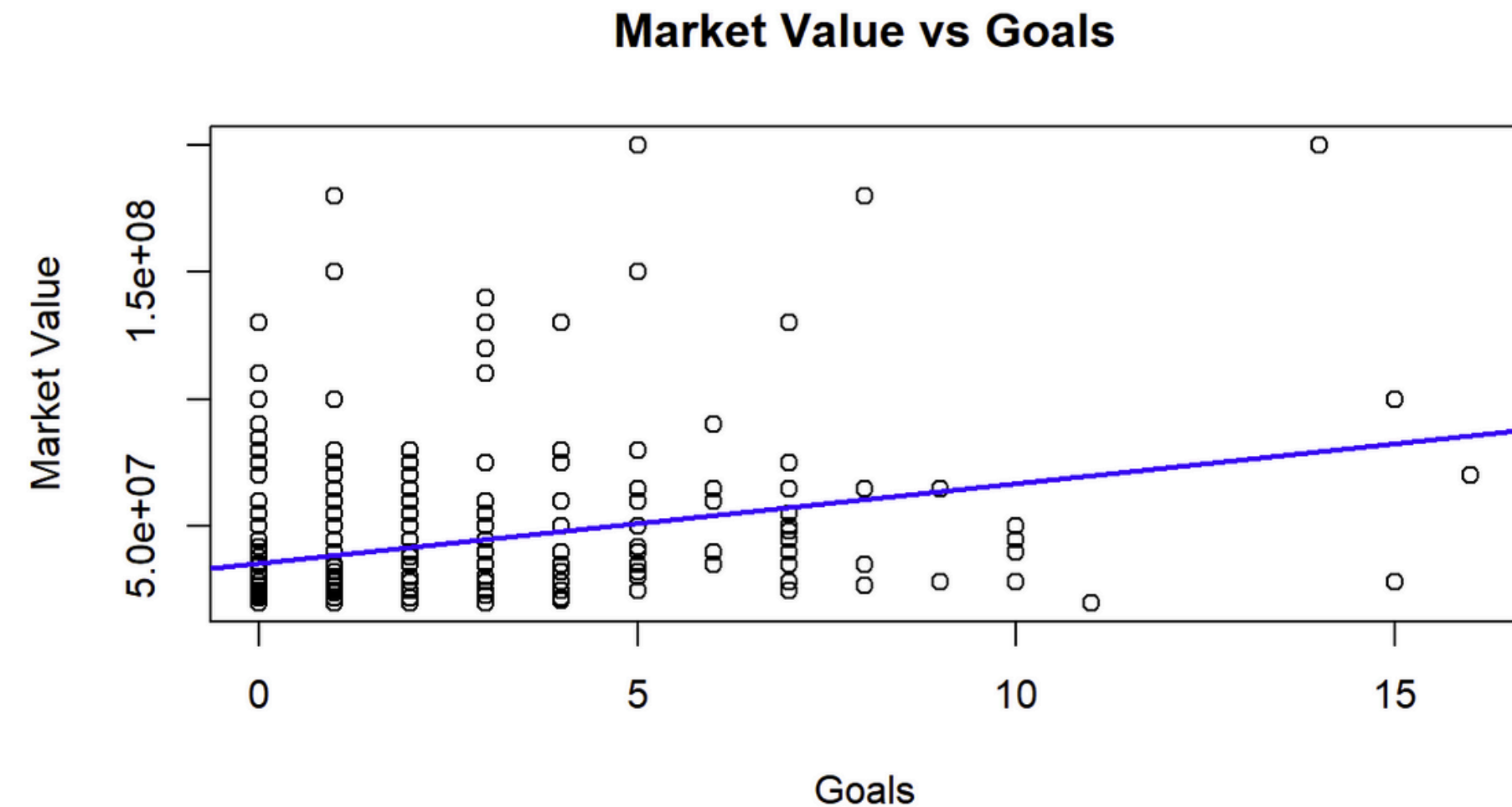


# Variable: Goals

- Goal Coefficient: +€3.15M per goal
- p-value < 0.001
- $R^2 = 0.094$

```
4 model_goals <- lm(Market.Value ~ Goals, data = players)
5 summary(model_goals)
```

```
17 plot(players$Goals, players$Market.Value,
18       xlab = "Goals",
19       ylab = "Market Value",
20       main = "Market Value vs Goals")
21 abline(lm(Market.Value ~ Goals, data = players), col="blue", lwd=2)
```

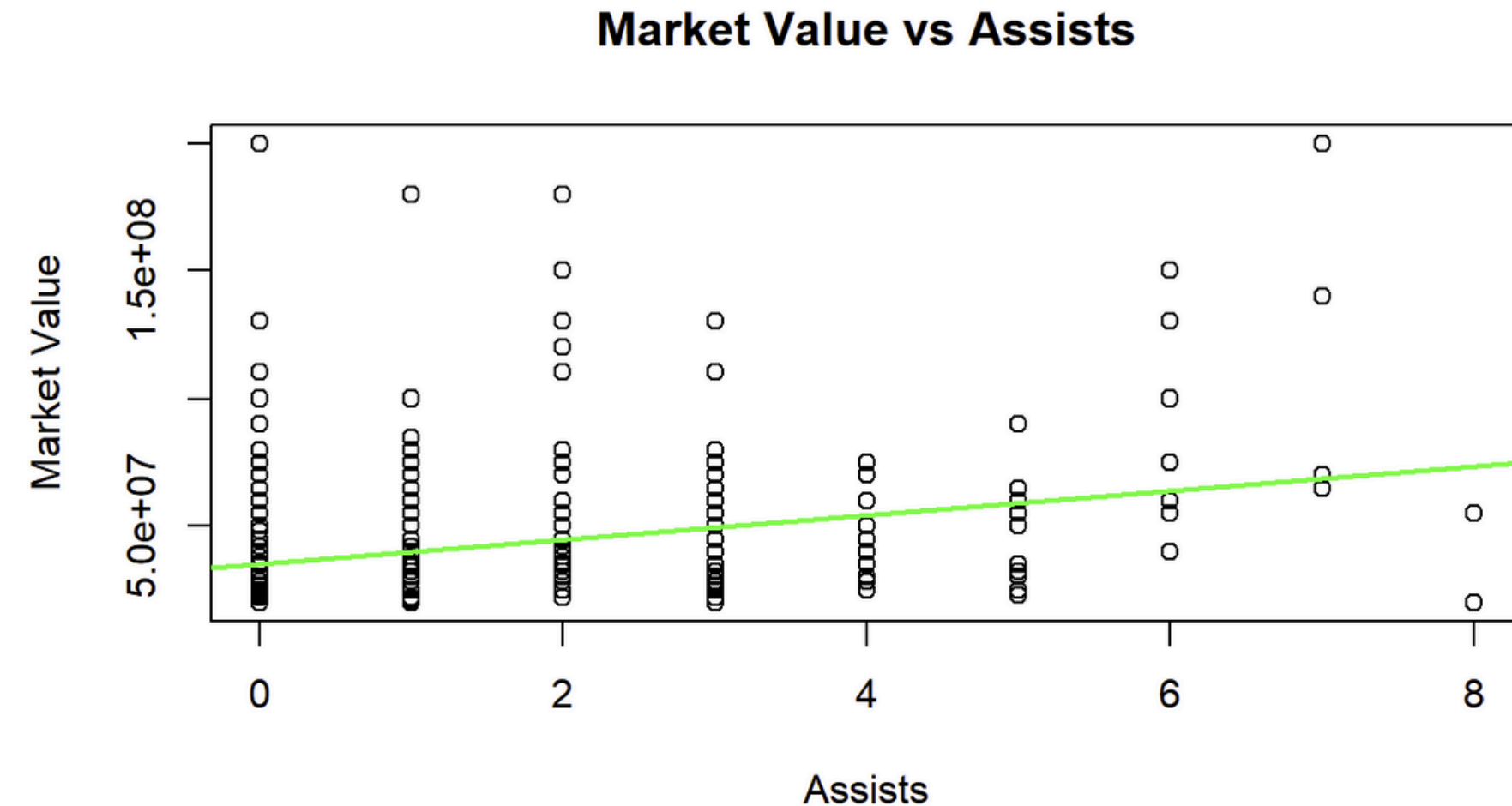


# Variable: Assists

- Assist Coefficient: +€4.81M per assist
- p-value < 0.001
- $R^2 = 0.093$

```
6 model_assists <- lm(Market.Value ~ Assists, data = players)
7 summary(model_assists)
```

```
22 plot(players$Assists, players$Market.Value,
23       xlab = "Assists",
24       ylab = "Market Value",
25       main = "Market Value vs Assists")
26 abline(lm(Market.Value ~ Assists, data = players), col="green", lwd=2)
```



# Multiple Regression (Goals + Assists + Age)

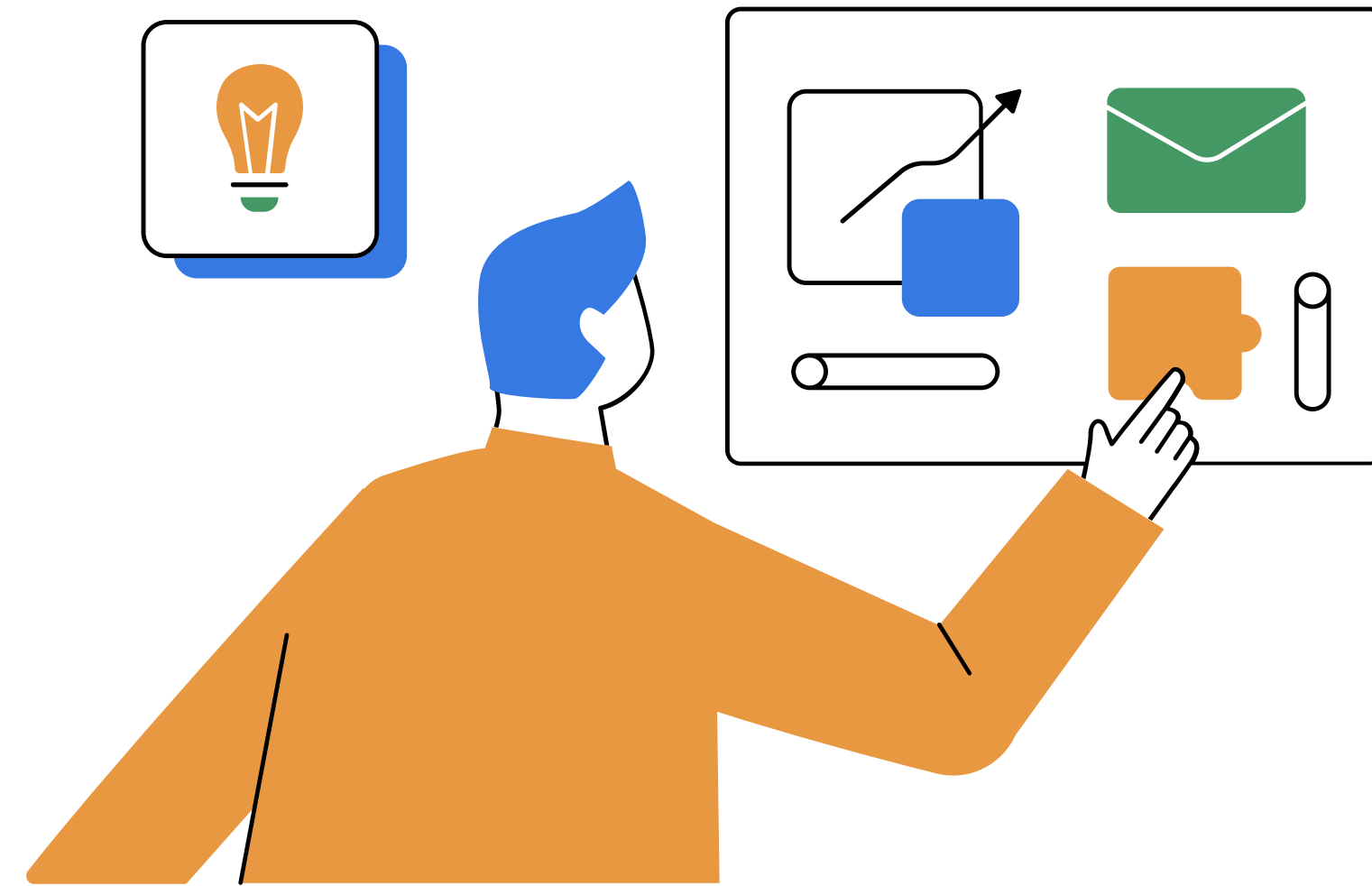
Goals: +€2.23M (significant)

Assists: +€3.33M (significant)

Age: not significant

$R^2 = 0.133$

```
10 model_full <- lm(Market.Value ~ Goals + Assists + Age, data = players)
11 summary(model_full)
```

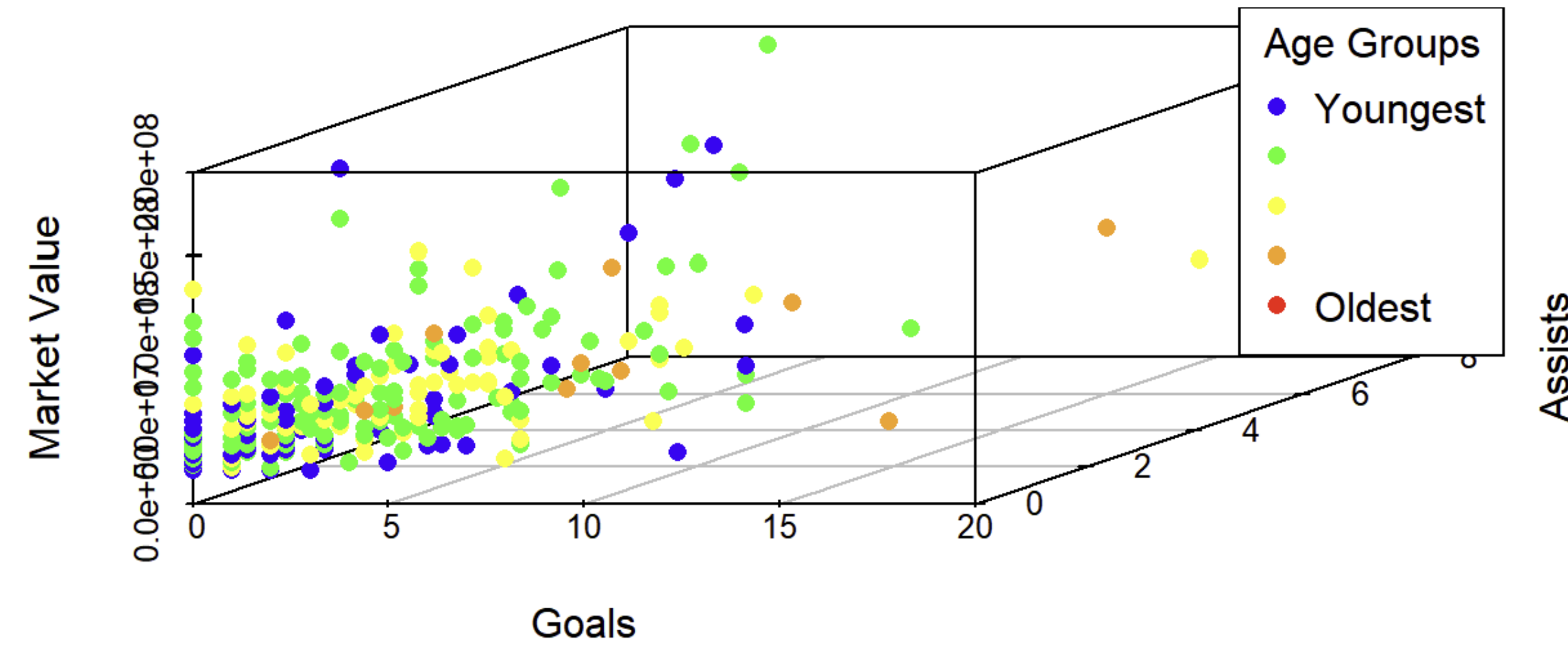


# 3D Visualization

```
install.packages("scatterplot3d")
```

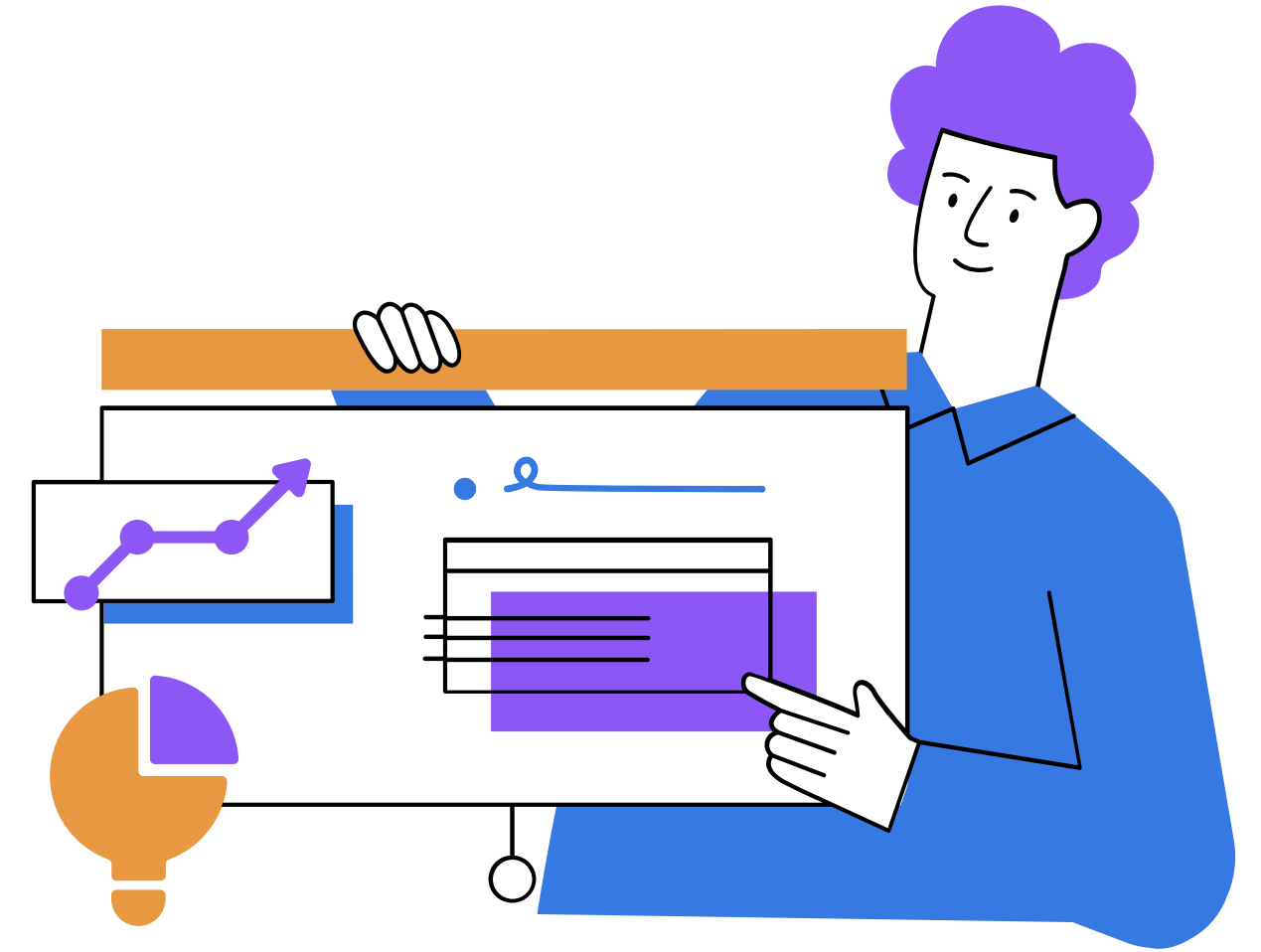
```
27 library(scatterplot3d)
28
29 install.packages("scatterplot3d")
30 library(scatterplot3d)
31 # Color scale based on Age
32 color_age <- cut(players$Age,
33                 breaks = 5,
34                 labels = c("blue", "green", "yellow", "orange", "red"))
35
36 scatterplot3d(
37   x = players$Goals,
38   y = players$Assists,
39   z = players$Market.Value,
40   pch = 16,
41   color = as.character(color_age),
42   xlab = "Goals",
43   ylab = "Assists",
44   zlab = "Market Value",
45   main = "Market Value Explained by Goals, Assists, and Age"
46 )
47
48 legend("topright",
49       legend = c("Youngest", "", "", "", "Oldest"),
50       col = c("blue", "green", "yellow", "orange", "red"),
51       pch = 16,
52       title = "Age Groups")
```

Market Value Explained by Goals, Assists, and Age



# Challenges

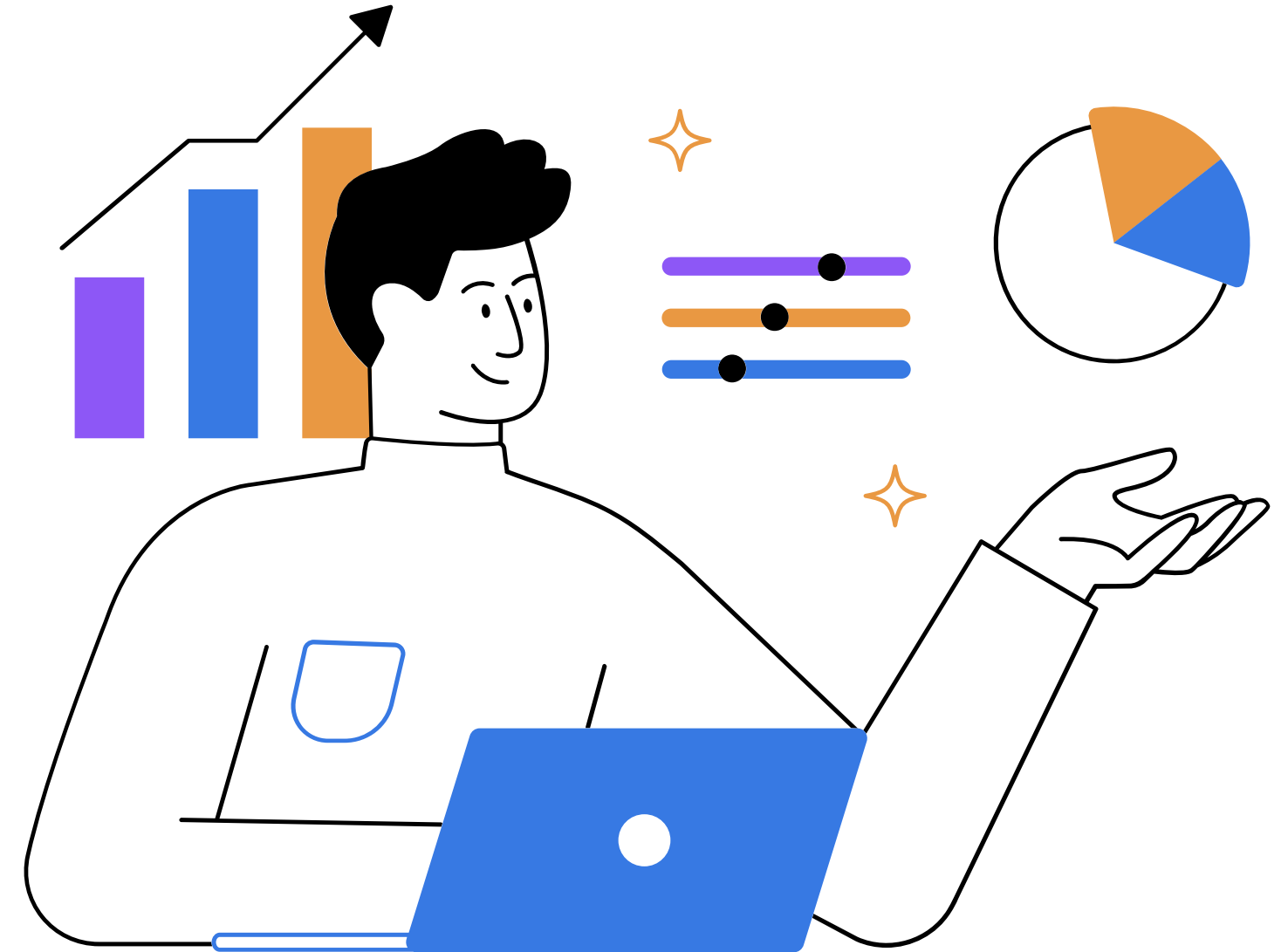
- Limited dataset (top 500 only)
- No defensive metrics (blocks, clearances)
- Valuation depends on many unobserved factors
- Model explains only ~13%
- Trial and error experimentation with R
- Last 3D scatterplot was a bit confusing to analyse





# Next Steps

- Add descriptive statistics
- Add correlation matrix
- Improve interpretation
- Expand discussion & limitations
- Write report in Quarto



# Thank You

