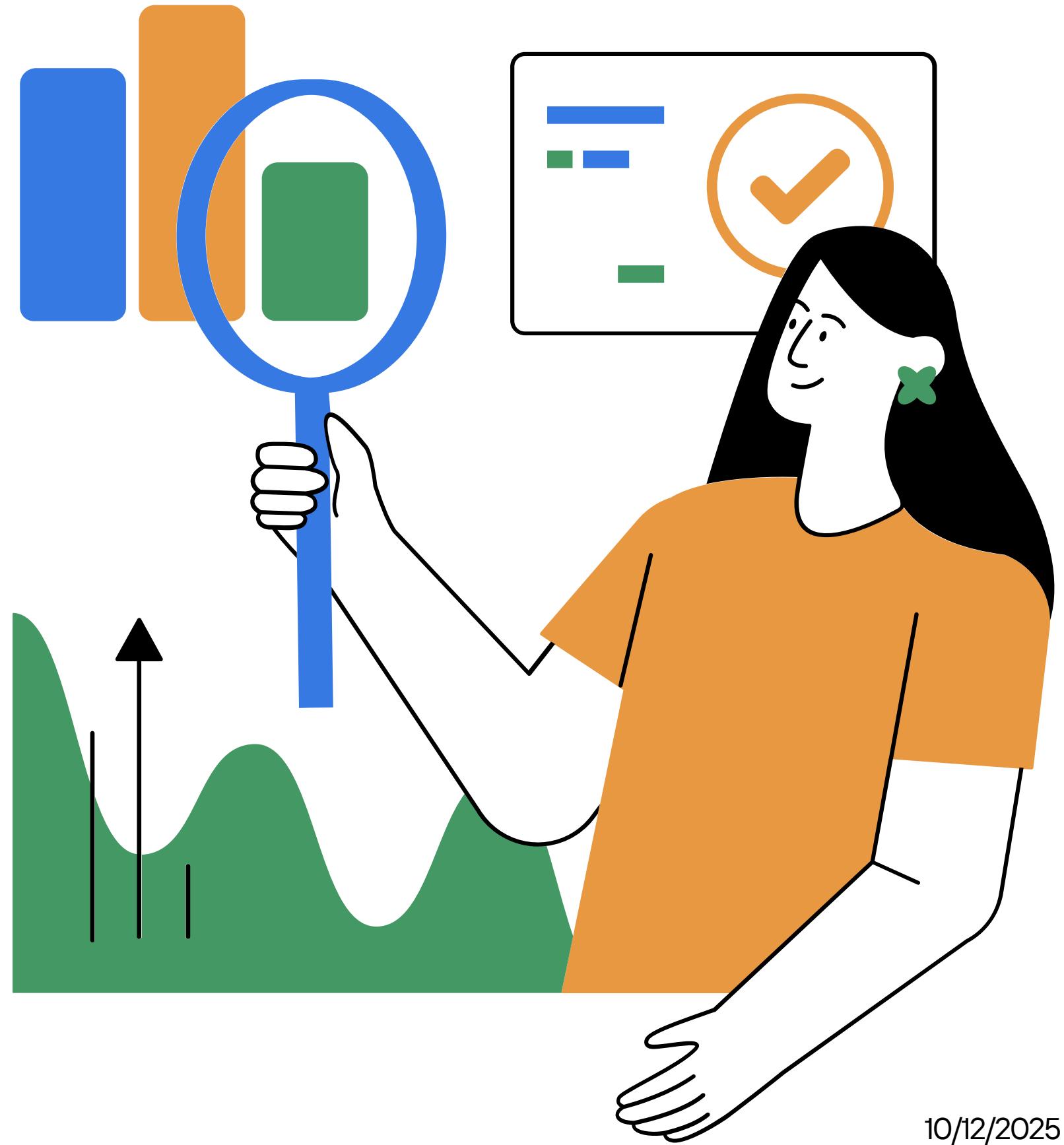


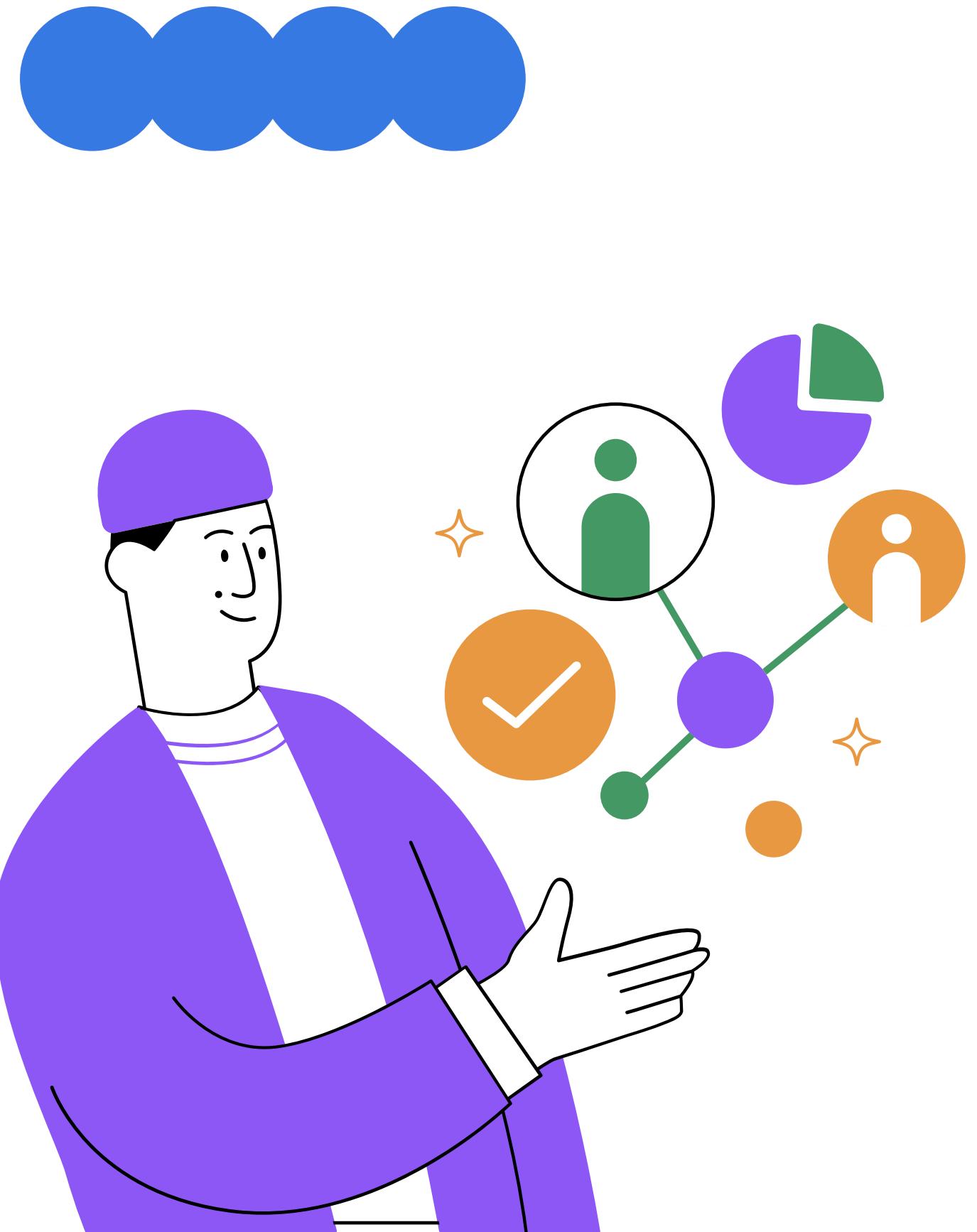
Data Analytics



Title & Research Question

Player Valuation in Professional Football

Which Factors Predict The Market Value of Current Footballers?



+ Create

Home

Competitions

Datasets

Models

Benchmarks

Game Arena

Code

Discussions

Learn

More

View Active Events

Football Players Stats (2024-2025)

78

Code

Download

:

Data Card Code (19) Discussion (6) Suggestions (0)

View more

players_data-2024_2025.csv (2.65 MB)

Download

Detail Compact Column

10 of 267 columns

About this file

File consists of over 200+ columns, providing comprehensive football statistics sourced from FBref. It includes detailed player and team data from the Big 5 European leagues. For a complete breakdown of column meanings, refer to: <https://fbref.com/en/comps/Big5/stats/players/Big-5-European-Leagues-Stats>

#	Rk	Player	Nation	Pos	Squad
		Player full name	Nationality	Position	Player's team
1	2854	2702 unique values	es ESP fr FRA Other (2093)	15% 12% 73%	DF MF Other (1406)
1	2854	Max Aarons	eng ENG	30% 21% 49%	Como Valladolid Other (2778)
					Bournemouth

Data Explorer

4.02 MB

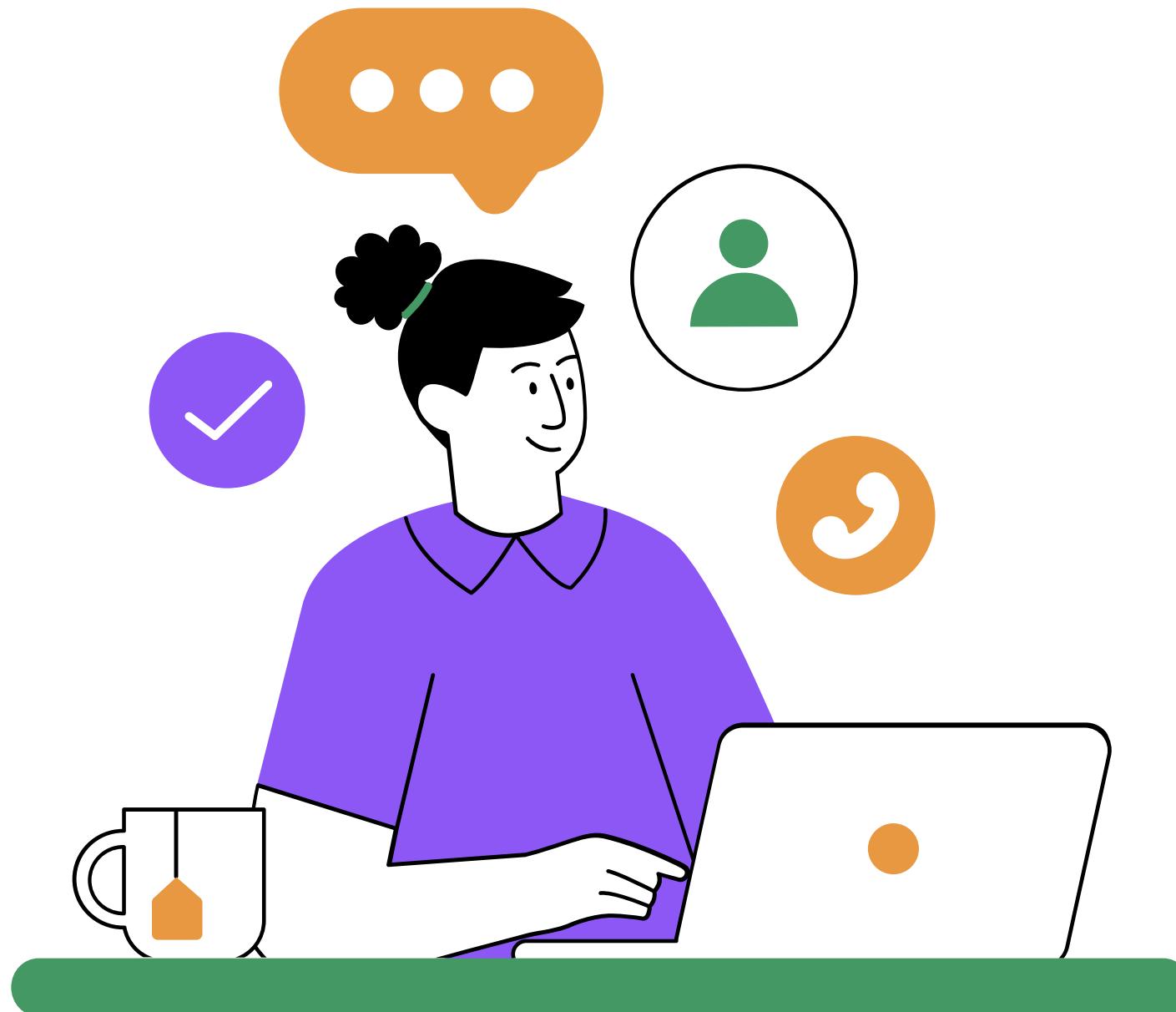
players_data-2024_2025.csv

players_data_light-2024_2025

Summary

- 2 files
- 432 columns

Dataset Overview



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

Create

Home

Competitions

Datasets

Models

Benchmarks

Game Arena

Code

Discussions

Learn

More

View Active Events



SANJEEET SINGH NAIK · UPDATED A YEAR AGO

3

Code

Download

⋮

Most Valuable Footballers 2024

Top 500 Football Players 2024 : Comprehensive Statistics and Valuations

[Data Card](#) [Code \(2\)](#) [Discussion \(0\)](#) [Suggestions \(0\)](#)

About Dataset

This dataset provides detailed information on the top 500 football players in 2024, including their market values, performance statistics, and demographics.

Key features include:

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 ⓘ

10.00

License

MIT

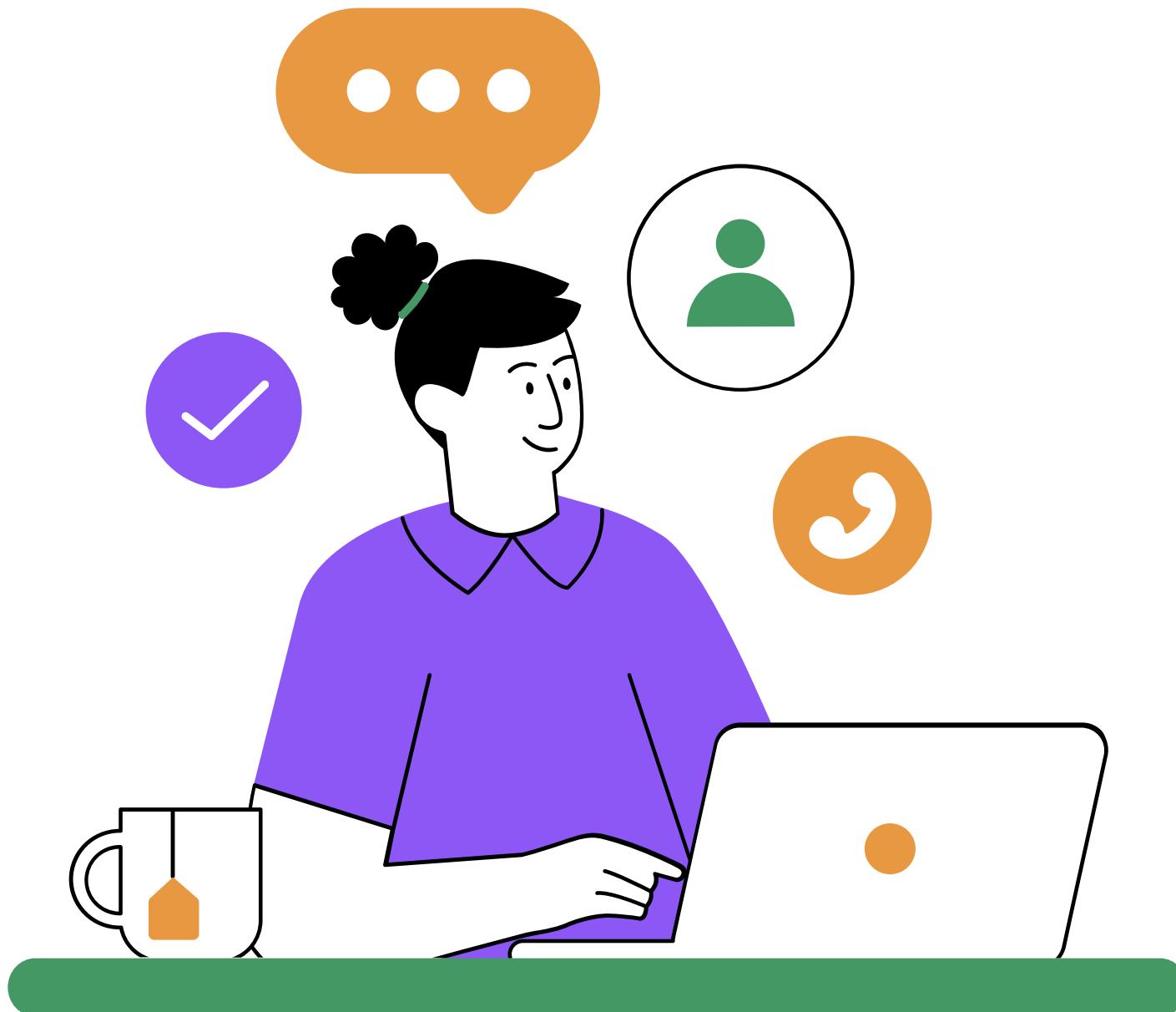
Expected update frequency

Annually

Tags

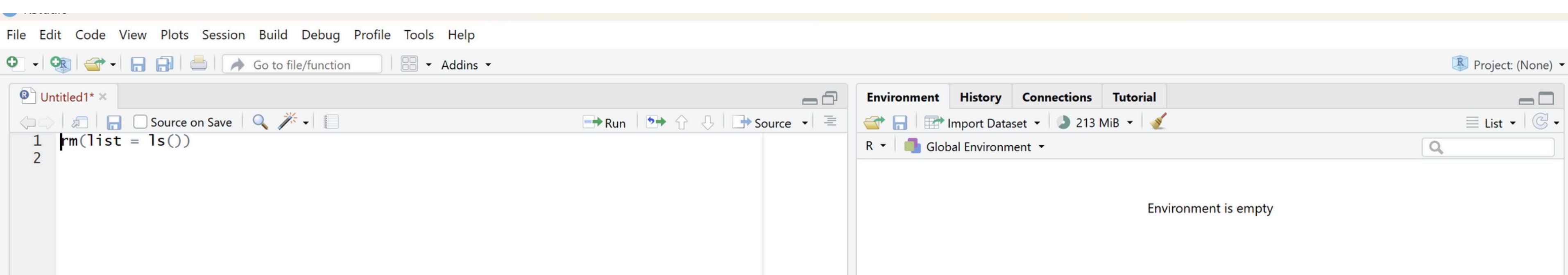
Football

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

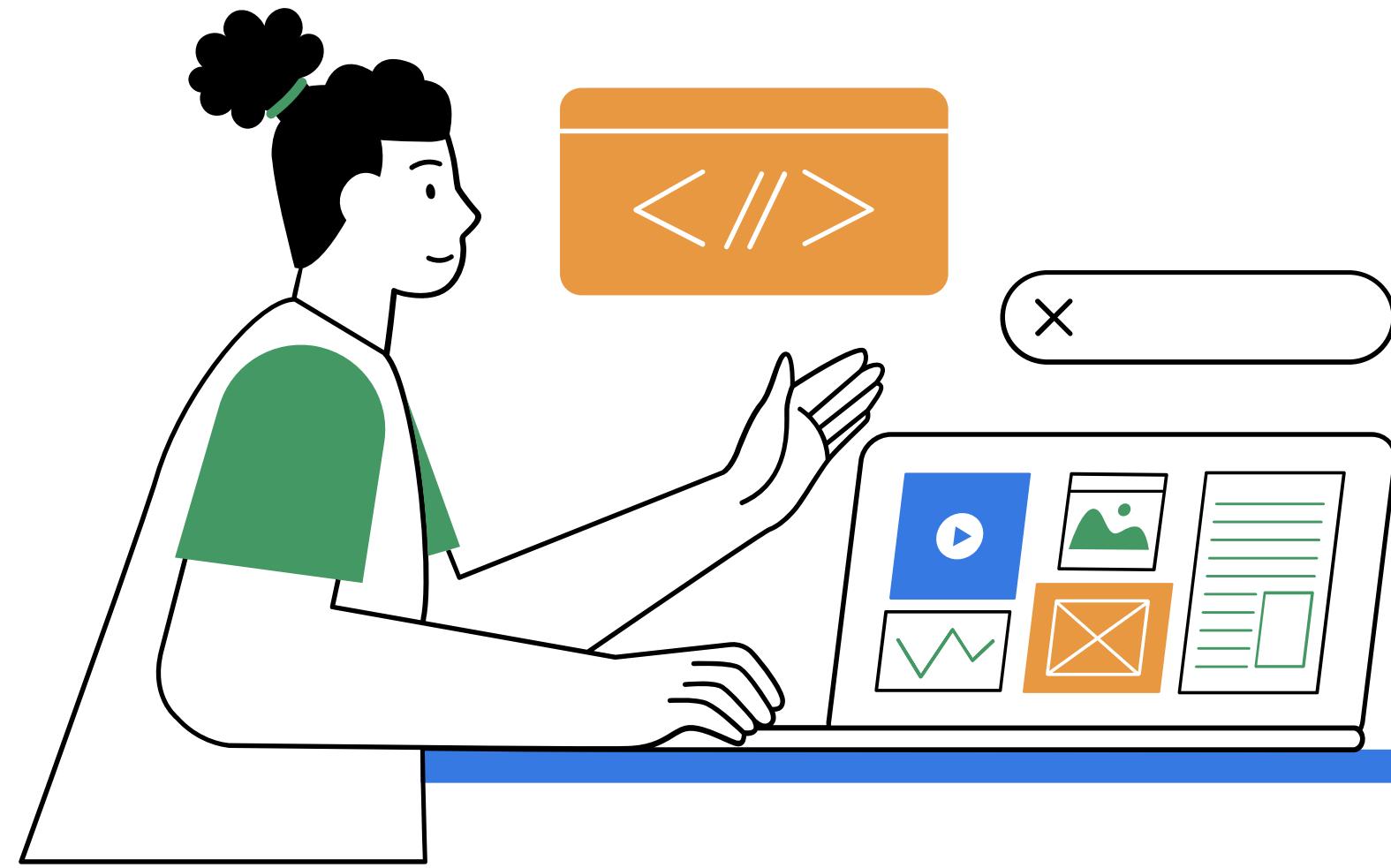


Import the Kaggle Dataset

The screenshot shows the RStudio Console tab. The session environment is R 4.4.3. The console history shows the command 'rm(list = ls())' followed by 'playerdata <- read.csv(players_data-2024_2025.csv)'. An error message 'Error: unexpected input in "playerdata <- read.csv(players_data-2024_2025.csv)"' is displayed. Below the error, the command 'playerdata <- read.csv("players_data-2024_2025.csv")' is shown again. The top right of the console has buttons for 'Show Traceback' and 'Rerun with Debug'.

```
Console Terminal × Background Jobs ×
R 4.4.3 · C:/Users/domin/Downloads/archive/ ↗
> rm(list = ls())
> playerdata <- read.csv(players_data-2024_2025.csv)
Error: unexpected input in "playerdata <- read.csv(players_data-2024_2025.csv)"
Show Traceback
Rerun with Debug
```

```
> playerdata <- read.csv("players_data-2024_2025.csv")
>
```



Planned Analysis

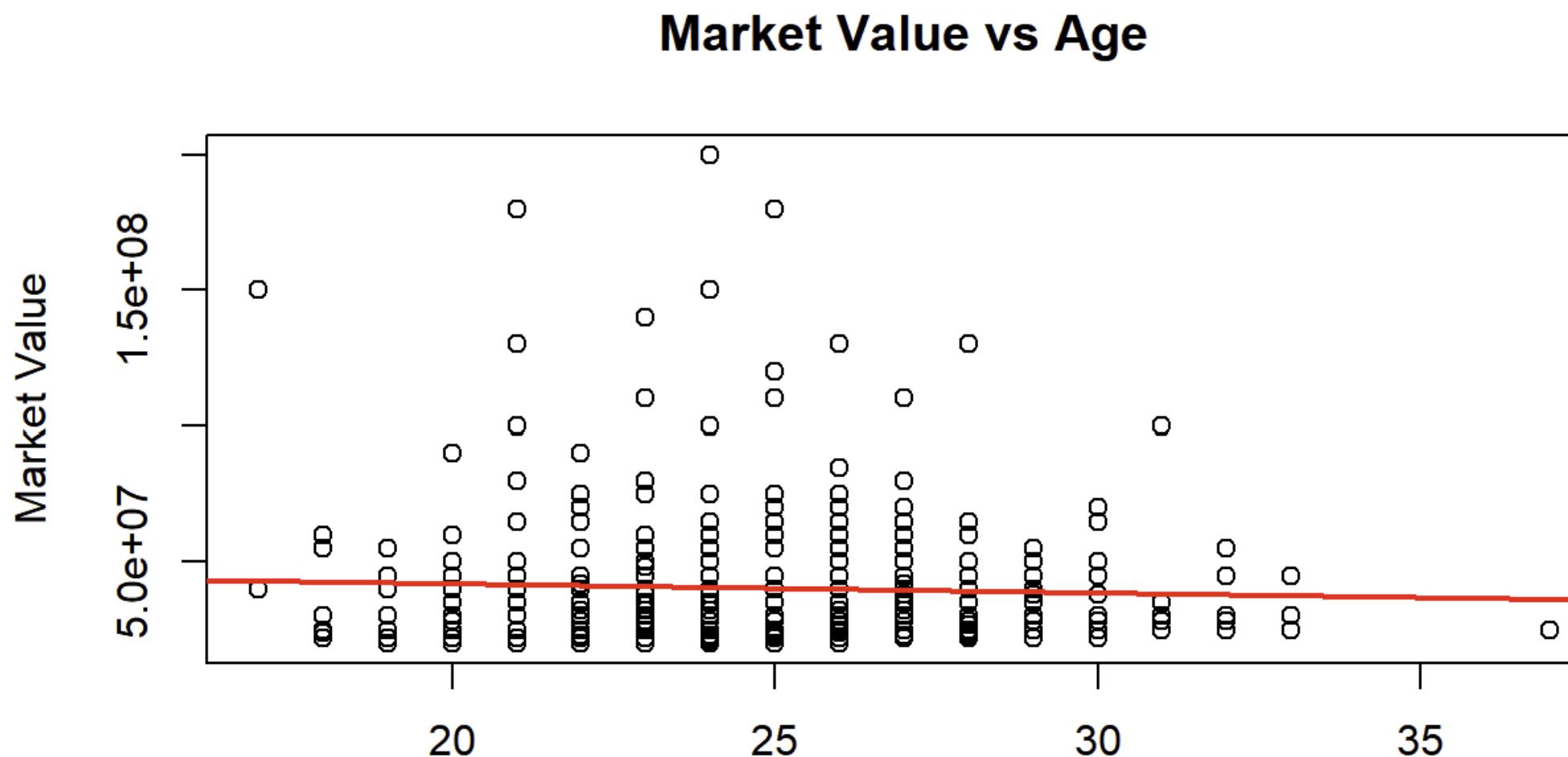
- Simple regressions:
 1. Market Value ~ Age
 2. Market Value ~ Goals
 3. Market Value ~ Assists
- Multiple regression:
 4. Market Value ~ Goals + Assists + Age

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)
```



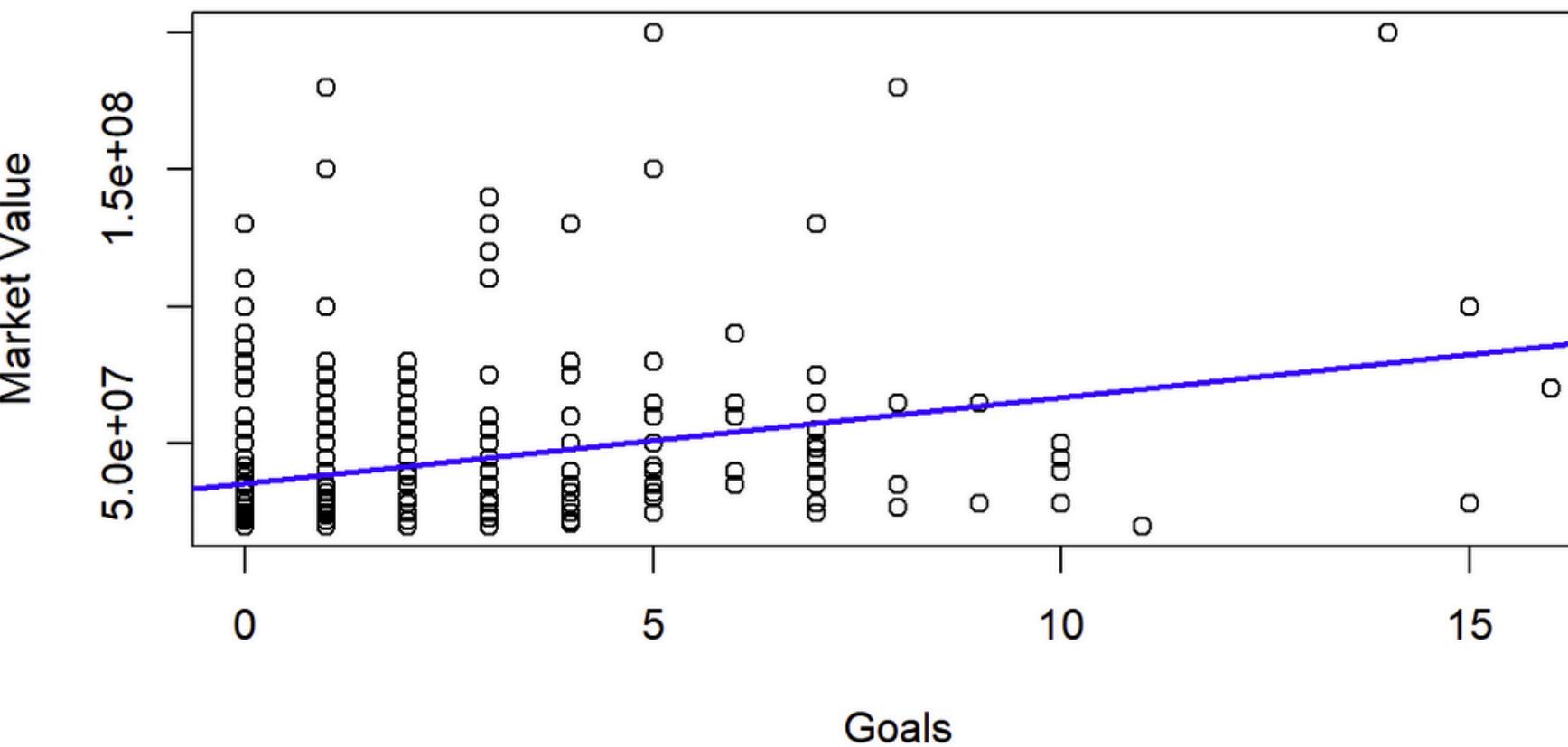
Variable: Goals

- 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)
```

Market Value vs Goals



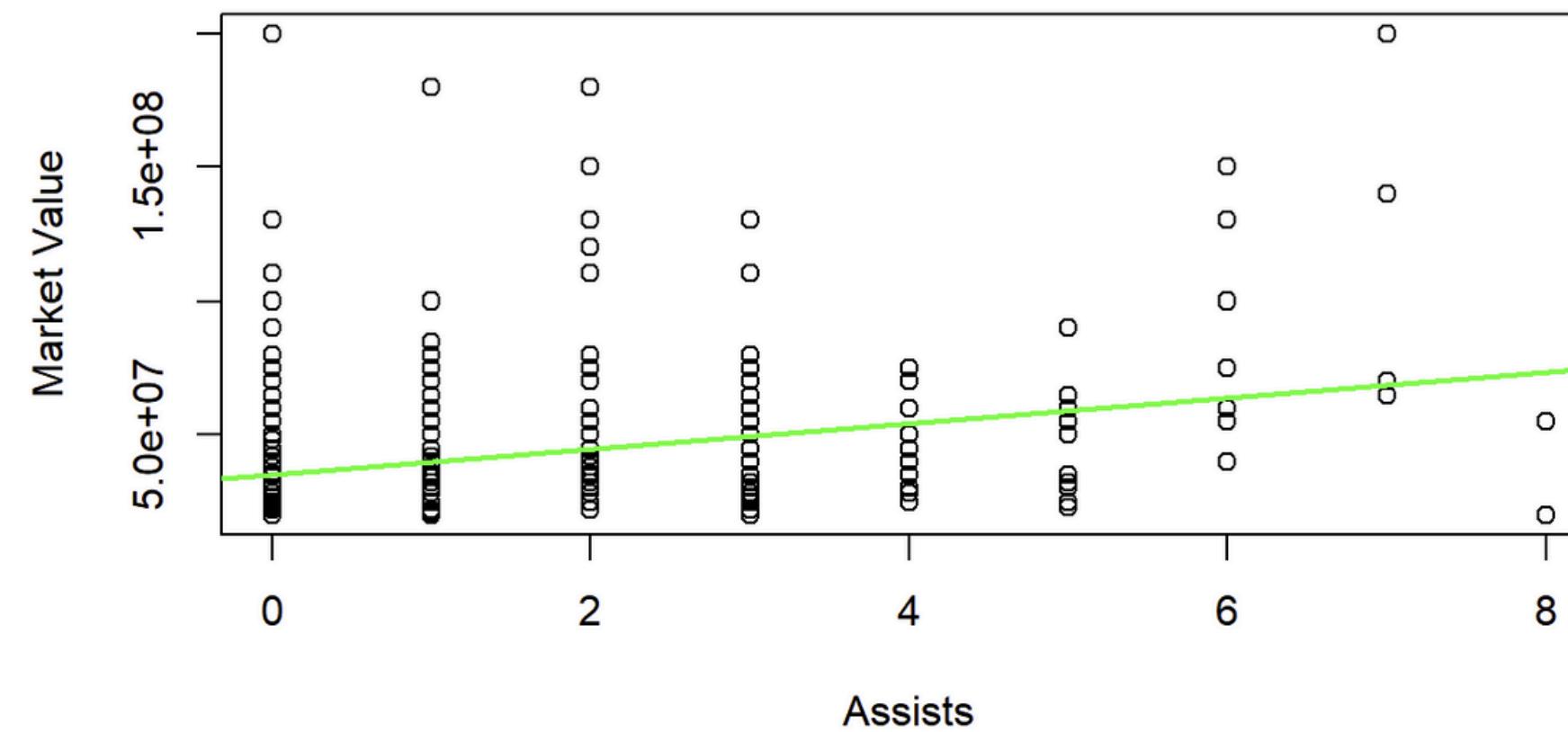
Variable: Assists

- 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)
```

Market Value vs Assists



Multiple Regression (Goals + Assists + Age)

Goals: +€2.23M (significant)

Assists: +€3.33M (significant)

Age: not significant

$R^2 = 0.133$

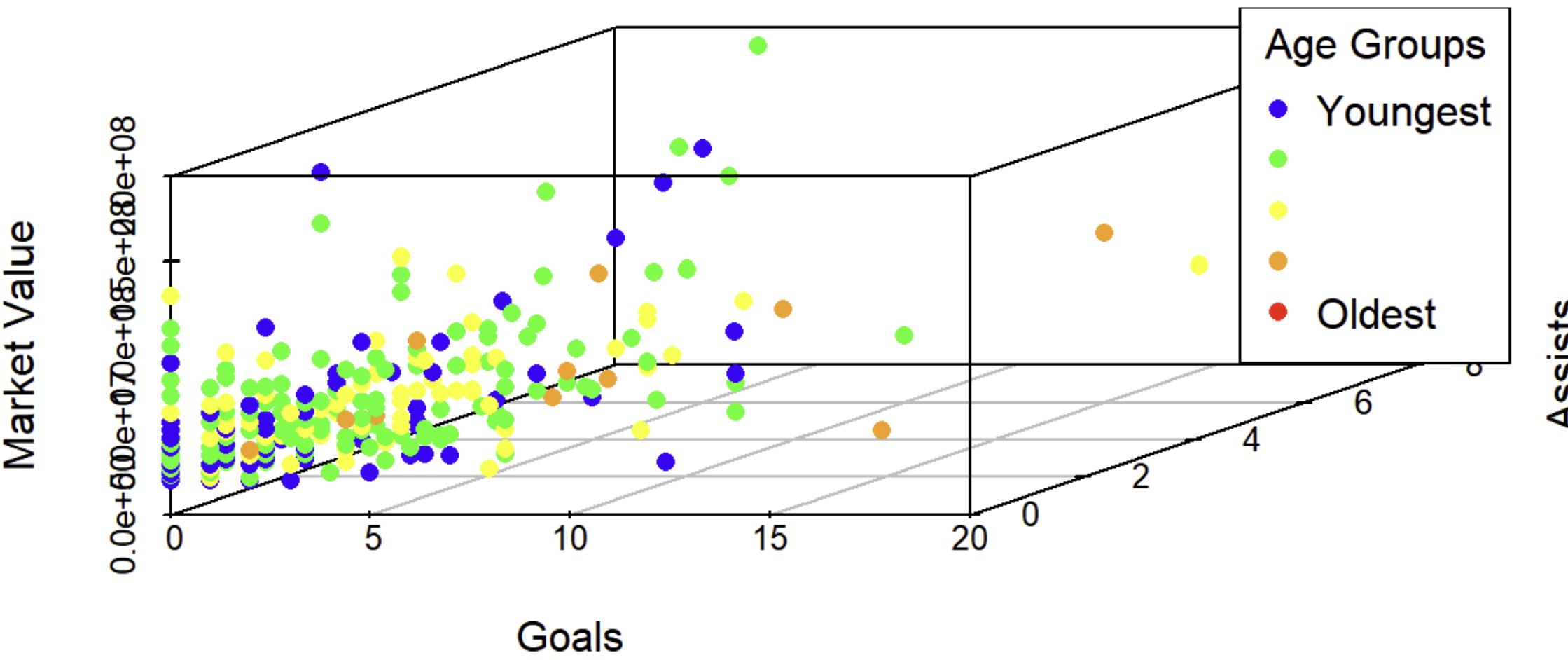
```
10 model_full <- lm(Market.Value ~ Goals + Assists + Age, data = players)
11 summary(model_full)
```



3D Visualization

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

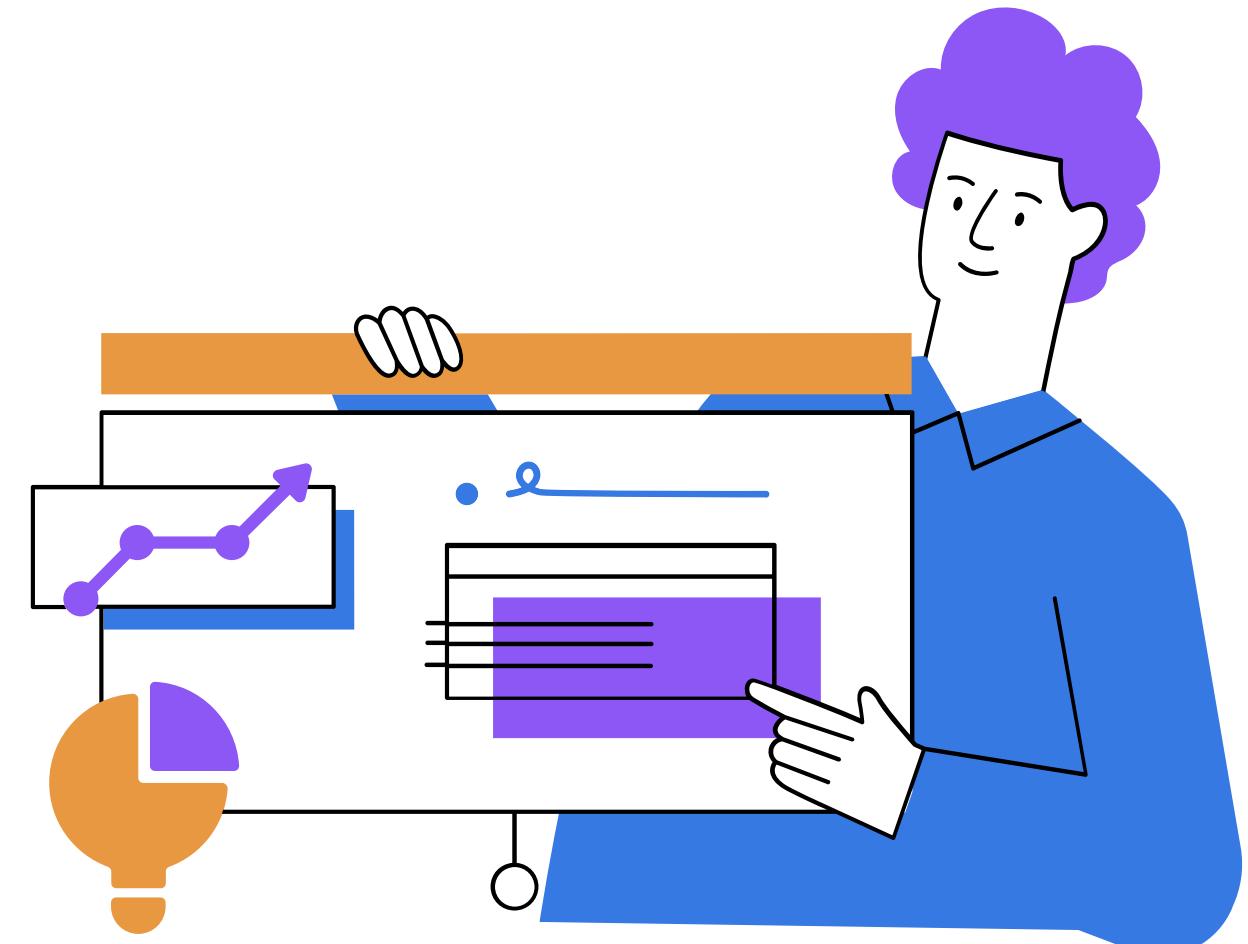
Market Value Explained by Goals, Assists, and Age



```
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")
```

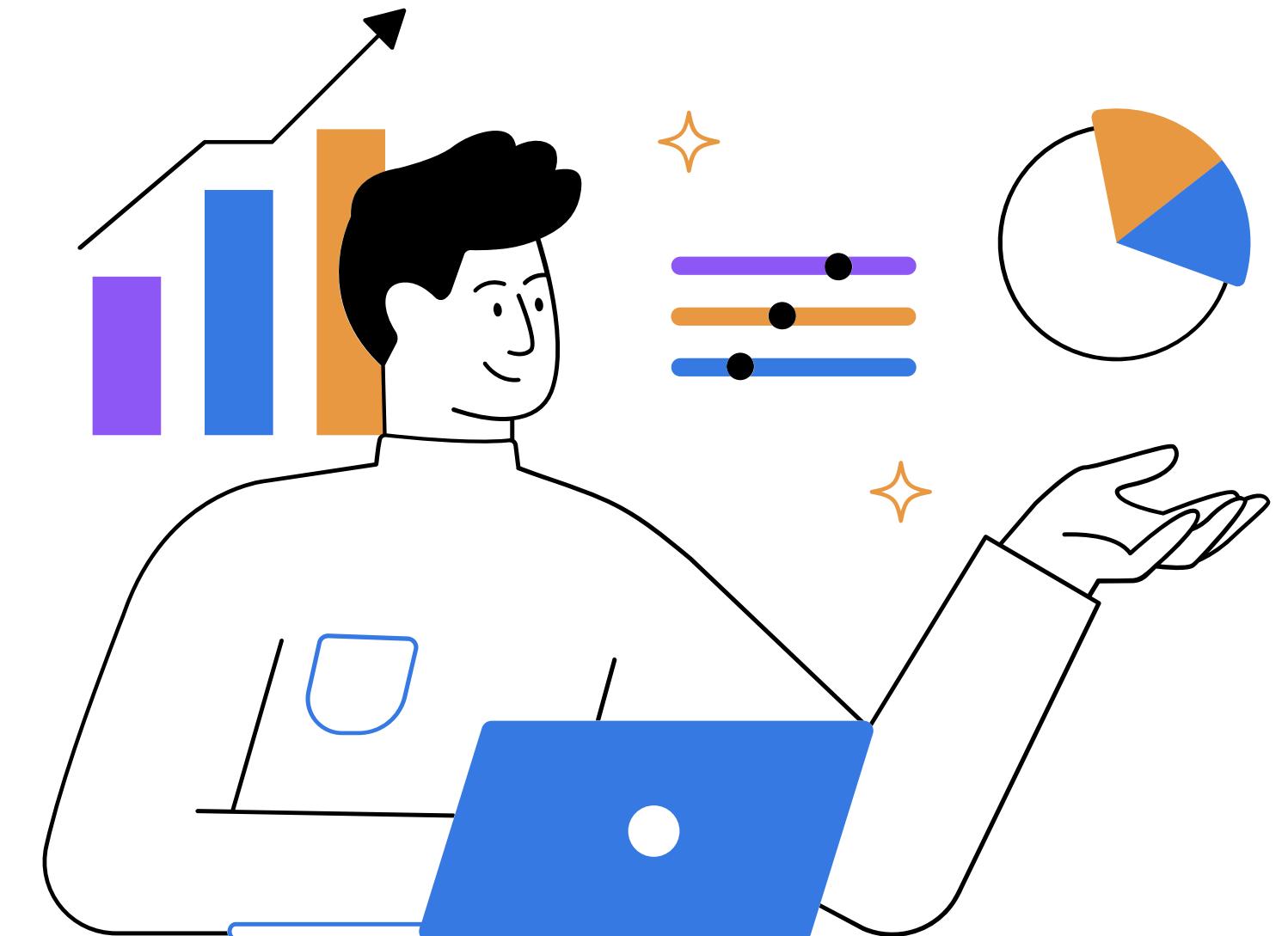
Challenges

- 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

