Wage Determination in Professional Football: The Role of Performance, Club Affiliation, and Playing Position

Nathalia Garcia   Ema Guiochet   Vanessa Martinez

Valentina Morales   Lucas Simatovic

2026-01-30

1 Introduction

Wages in professional football vary considerably across players, even among those with similar performance profiles. While these differences are often attributed primarily to individual talent and on-field performance, structural factors related to the organization of the sport—particularly club affiliation and playing position—may also play significant roles in determining wages.

From an economic perspective, player wages reflect both individual productivity and market-related factors. Generally, higher performance translates into higher wages. However, football clubs differ substantially in financial resources, market power, and visibility, potentially enabling certain clubs to pay wage premiums not fully explained by individual performance. Similarly, playing positions differ in roles, scarcity, and market valuations, which may generate systematic wage differences even among players with comparable skill levels.

This paper analyzes the factors that drive wage differences among male professional football players. Specifically, we examine whether wages are determined exclusively by individual ability and performance, or whether club affiliation and playing position also exert independent influences after accounting for performance differences. Understanding these mechanisms helps assess the extent to which wages reflect merit-based compensation versus structural advantages associated with clubs and positional roles.

The research question guiding this study is:

What factors explain wage differences among professional football players, and to what extent do club affiliation and playing position matter once individual performance attributes are taken into account?

This question relates to growing literature on wage determination in professional football, which emphasizes institutional and market-related factors beyond individual performance.

2 Literature Review

The economics of sport has extensively examined wage determination in professional football, highlighting specific institutional characteristics of this labor market. Dobson and Goddard (2001) argue that football clubs do not operate as standard profit-maximizing firms but rather face incentives related to sporting success, reputation, and financial sustainability. Within this institutional setting, player wages may depend not only on individual performance and human capital but also on club-specific characteristics and positional organization.

Empirically, Franck and Nüesch (2010) demonstrate that football players' remuneration is not solely driven by sporting talent but is also influenced by popularity and visibility, giving rise to superstar effects. Their findings suggest that players with comparable performance may receive substantially different wages, reinforcing the importance of considering additional determinants beyond individual ability.

Frick (2007) provides further empirical evidence from major European leagues, showing that wages are significantly affected by institutional factors such as club affiliation and playing position. Specifically, Frick documents systematic wage premiums associated with certain clubs and persistent positional differences, even after controlling for sporting performance. These findings confirm that wages in professional football reflect not only individual productivity but also organizational and market-related factors.

Overall, existing studies indicate that player wages are positively related to individual performance while exhibiting systematic differences across clubs and positions. However, much empirical evidence relies on match-based statistics and contractual data, which vary in availability and comparability across leagues. This paper contributes by using standardized FIFA performance attributes as proxies for individual ability, enabling consistent analysis of wage determination across players, clubs, and positions.

3 Data

3.1 Data Description

This study uses player-level data from the FIFA video game series, obtained from SoFIFA (https://sofifa.com). The analysis relies on players_21.csv, which contains detailed attributes for professional football players from FIFA 21.

The dataset provides comprehensive information on players evaluated within a standardized framework, enabling consistent comparison across leagues and countries. It includes over 18,000 player observations and more than 100 variables capturing individual performance characteristics—such as overall rating and detailed skill indicators—as well as positional information, club affiliation, and personal characteristics.

Player wage information, reported at the individual level, constitutes the primary outcome variable. Performance attributes provided by FIFA serve as standardized proxies for individual ability and on-field performance, enabling comparisons across players and positions.

Importantly, both wages and performance measures are based on FIFA's internal evaluations and should be interpreted as standardized estimates rather than observed real-world values. These estimates reflect how players are valued within the FIFA Career Mode environment rather than actual contractual wages or match statistics.

This standardized structure makes the dataset particularly suitable for analyzing wage differences while controlling for individual performance attributes, club affiliation, and playing position, reducing issues related to cross-league comparability and heterogeneous performance metrics.

3.2 Data Cleaning

The dataset was processed to ensure consistency, comparability, and suitability for econometric modeling. Several transformations were required to standardize units, remove redundant variables, and improve interpretability.

Monetary variables were converted into euros by harmonizing values expressed in millions (M) and thousands (K). Physical characteristics were standardized by converting height to centimeters and weight to kilograms. Player positions were recoded from abbreviations to full position names. Contract-related variables and URLs not relevant for wage determination were removed.

Given the absence of match-level statistics such as actual goals scored, the finishing attribute was retained as a proxy for goal-scoring ability, capturing a player's effectiveness in converting chances within FIFA's standardized evaluation framework.

Extreme outliers at the upper tail of the wage distribution were inspected during exploratory analysis. While all observations are retained for regression analysis, selected descriptive visualizations exclude the single highest-paid player to avoid graphical distortions.

3.2.1 Cleaning Functions

```r
library(here)
library(socviz)
library(tidyverse)
library(rvest)
library(gt)
library(kableExtra)
library(readODS)
library(knitr)
library(lubridate)
library(readxl)
library(tibble)
library(knitr)
library(broom)
```

```r
clean_money <- function(x) {
  x <- as.character(x)
  case_when(
    str_detect(x, "M") ~ as.numeric(str_remove_all(x, "[€M]")) * 1000000,
    str_detect(x, "K") ~ as.numeric(str_remove_all(x, "[€K]")) * 1000,
    TRUE ~ as.numeric(str_remove_all(x, "[€]"))
  )
}

clean_weight <- function(x) {
  x <- as.character(x)
  ifelse(str_detect(x, "lbs"),
         as.numeric(str_remove(x, "lbs")) * 0.453592,
         as.numeric(str_remove(x, "kg")))
}

clean_height <- function(x) {
  x <- as.character(x)
  if_else(str_detect(x, "'"),
          as.numeric(str_extract(x, "^\\d+")) * 30.48 +
            as.numeric(str_extract(x, "\\d+(?=\")")) * 2.54,
          as.numeric(str_remove(x, "cm")))
}

clean_hits <- function(x) {
  x <- as.character(x)
  case_when(
    str_detect(x, "K") ~ as.numeric(str_remove(x, "K")) * 1000,
    is.na(x) ~ 0,
    TRUE ~ as.numeric(x)
  )
}

position_map <- c(
  "GK"  = "Goalkeeper",
  "ST"  = "Striker",
  "CF"  = "Center Forward",
  "LW"  = "Left Winger",
  "RW"  = "Right Winger",
  "LF"  = "Left Forward",
  "RF"  = "Right Forward",
  "CAM" = "Central Attacking Midfielder",
  "CM"  = "Central Midfielder",
  "CDM" = "Central Defensive Midfielder",
```

```
  "LM"  = "Left Midfielder",
  "RM"  = "Right Midfielder",
  "CB"  = "Center Back",
  "LB"  = "Left Back",
  "RB"  = "Right Back",
  "LWB" = "Left Wing Back",
  "RWB" = "Right Wing Back"
)

clean_best_pos <- function(x) {
  recode(str_trim(x), !!!position_map, .default = x)
}

clean_positions_list <- function(x) {
  str_replace_all(x, position_map)
}
```

These cleaning functions eliminate duplicates and missing values, preparing the data for analysis.

3.2.2 Data Cleaning Procedures

This section prepares the data for subsequent analysis. Positions are cleaned first, followed by clubs, values, player physical attributes, and other variables such as dates, statistics, and hits. The original variables are then replaced with cleaned versions.

```
raw_data <- read_excel("C:/Users/lucas/OneDrive/Desktop/COURS M2 DADEE/S1/UTILISATION DE

fifa_clean <- raw_data %>%
  mutate(
    `Best Position` = clean_best_pos(`Best Position`),
    Positions = clean_positions_list(Positions)
  ) %>%
  mutate(Club = str_remove_all(Club, "\n")) %>%
  mutate(
    Value_EUR = clean_money(Value),
    Wage_EUR = clean_money(Wage),
    Release_Clause_EUR = clean_money(`Release Clause`)
  ) %>%
  mutate(
    Height_cm = clean_height(Height),
    Weight_kg = round(clean_weight(Weight), 2)
  ) %>%
```

```
  mutate(Hits_Numeric = clean_hits(Hits)) %>%
  mutate(Joined_Date = mdy(Joined)) %>%
  mutate(
    W_F = as.numeric(str_remove(`W/F`, " ")),
    SM = as.numeric(str_remove(`SM`, " ")),
    IR = as.numeric(str_remove(`IR`, " "))
  )

fifa_final <- fifa_clean %>%
  select(-Value, -Wage, -`Release Clause`, -Height, -Weight, -Hits)

write_csv(fifa_final, "FIFA21_Cleaned.csv")
```

3.3 Variables

3.3.1 Construction of Key Variables

Since the dataset is based on FIFA 21 attributes rather than historical match statistics, a
'Number of Goals' variable was absent. We use the 'Finishing' attribute as a **proxy for
goal-scoring ability**, capturing a player's effectiveness in converting chances within FIFA's
standardized evaluation framework.

```
if(!"Goals" %in% names(fifa_final)) {
  fifa_final <- fifa_final %>%
    mutate(Goals = Finishing)
}
```

After data cleaning, the dataset contained 79 variables covering player characteristics, per-
formance attributes, physical traits, demographic information, and club-related factors. A
subset was selected to construct a parsimonious and economically meaningful model. Player
URLs and 'Loan Date End' (containing only missing values) were removed, resulting in 76
variables.

```
fifa_final <- fifa_final %>%
  rename(
    overall_rating = 8,
    Wage_EUR = Wage_EUR,
    potential = POT,
    base_stats = `Base Stats`,
    position = `Best Position`,
    preferred_foot = `Preferred Foot`,
    height_cm = Height_cm,
    weight_kg = Weight_kg,
```

```r
    finishing = Finishing,
    Club = Club,
    Age = Age
  )

fifa_final <- fifa_final %>%
  select(-photoUrl, playerUrl, -`Loan Date End`)


variable_description <- tibble(
  Variable = c(
    "wage_eur",
    "overall_rating",
    "finishing",
    "base_stats",
    "club",
    "position",
    "age",
    "height_cm",
    "weight_kg",
    "preferred_foot",
    "potential"
  ),
  Description = c(
    "Weekly wage of the player in euros (estimated by FIFA)",
    "Overall performance rating assigned by FIFA",
    "Finishing skill, used as a proxy for goal-scoring ability",
    "Aggregate performance indicator based on FIFA attributes",
    "Club to which the player belongs",
    "Primary playing position of the player",
    "Age of the player in years",
    "Height of the player in centimeters",
    "Weight of the player in kilograms",
    "Preferred foot of the player",
    "Potential rating assigned by FIFA, reflecting expected future ability"
  )
)

variable_description |>
  kable(
    caption = "Variable descriptions",
    col.names = c("Variable", "Description"),
    align = "ll"
  )
```
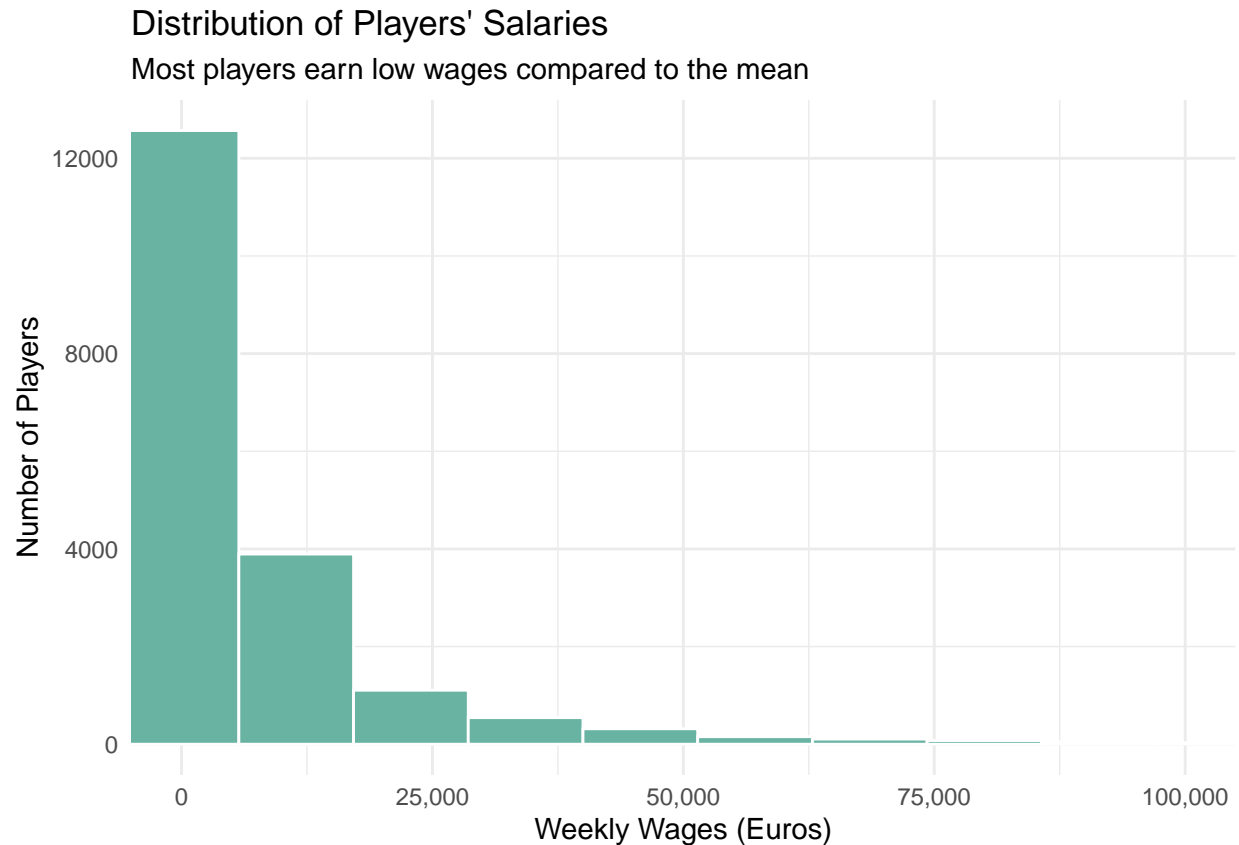
Table 1: Variable descriptions

Variable	Description
wage_eur	Weekly wage of the player in euros (estimated by FIFA)
overall_rating	Overall performance rating assigned by FIFA
finishing	Finishing skill, used as a proxy for goal-scoring ability
base_stats	Aggregate performance indicator based on FIFA attributes
club	Club to which the player belongs
position	Primary playing position of the player
age	Age of the player in years
height_cm	Height of the player in centimeters
weight_kg	Weight of the player in kilograms
preferred_foot	Preferred foot of the player
potential	Potential rating assigned by FIFA, reflecting expected future ability

4 Descriptive Statistics and Exploratory Analysis

Before proceeding to formal econometric modeling, we examine key distributional patterns and relationships in the data. This exploratory analysis provides initial evidence on wage inequality, performance distribution, and the role of club affiliation and age in shaping wage outcomes.

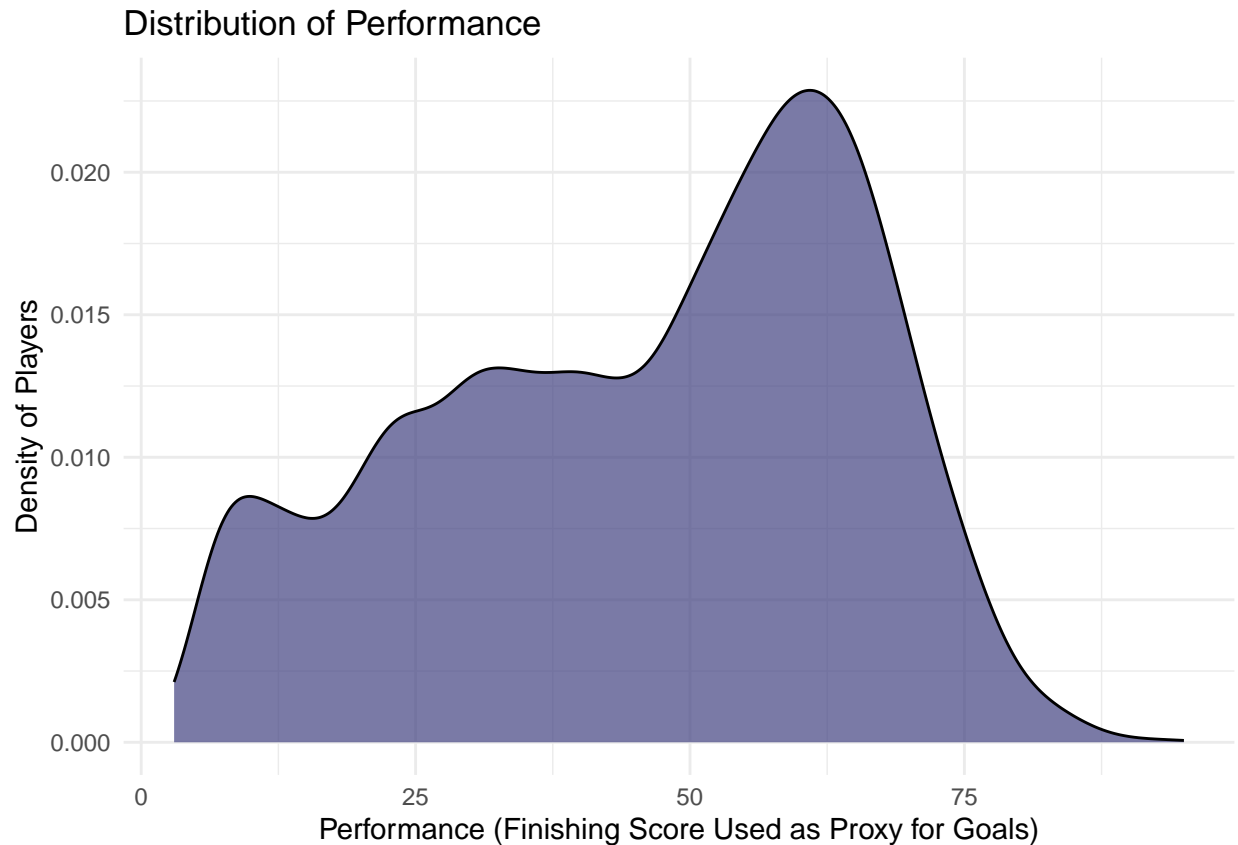## 4.1 Distribution of Wages

```
ggplot(fifa_final, aes(x = Wage_EUR)) +
  geom_histogram(fill = "#69b3a2", color = "white", bins = 50) +
  theme_minimal() +
  coord_cartesian(xlim = c(0, 100000)) +
  scale_x_continuous(labels = scales::comma) +
  labs(title = "Distribution of Players' Salaries",
       subtitle = "Most players earn low wages compared to the mean",
       x = "Weekly Wages (Euros)",
       y = "Number of Players"
  )
```

Distribution of Players' Salaries
Most players earn low wages compared to the mean

The distribution of weekly wages is highly right-skewed, with most players earning relatively low salaries and a small number receiving very high wages. This indicates substantial wage inequality, where the mean wage is influenced by a limited group of top earners. The concentration in lower wage ranges suggests high salaries are not representative of the typical player, motivating analysis of contributing factors.

4.2 Distribution of Performance

```
ggplot(fifa_final, aes(x = Goals)) +
  geom_density(fill = "#404080", alpha = 0.7) +
  theme_minimal() +
  labs(title = "Distribution of Performance",
       x = "Performance (Finishing Score Used as Proxy for Goals)",
       y = "Density of Players")
```

Distribution of Performance

Performance, measured through finishing ability, shows a unimodal distribution with most players concentrated around intermediate levels. Extreme values are relatively rare, indicating that most players exhibit moderate finishing ability. This distribution suggests less dispersion in performance than in wages, reinforcing the idea that large wage differences may not be fully explained by performance differences alone.

4.3 Wage and Performance Relationship

To improve readability, the single highest-paid player (Messi) is excluded from the scatter plot.
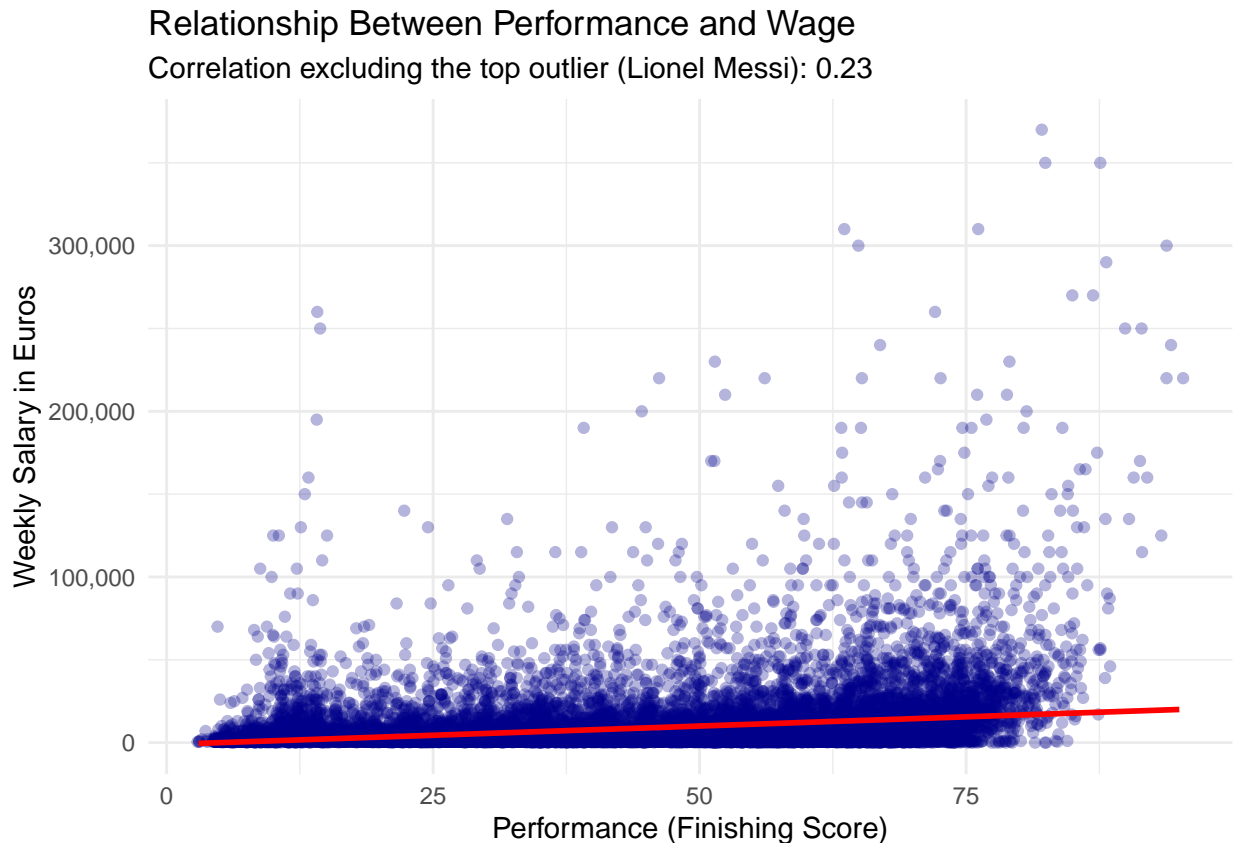
```
fifa_no_messi <- fifa_final %>%
  filter(Wage_EUR < max(Wage_EUR, na.rm = TRUE))

correlation_no_messi <- cor(fifa_no_messi$Wage_EUR, fifa_no_messi$Goals, use = "complete

ggplot(fifa_no_messi, aes(x = Goals, y = Wage_EUR)) +
  geom_jitter(alpha = 0.3, color = "darkblue", width = 0.5, height = 0.1) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma) +
```

```
  labs(title = "Relationship Between Performance and Wage",
       subtitle = paste("Correlation excluding the top outlier (Lionel Messi):", round(c
       x = "Performance (Finishing Score)",
       y = "Weekly Salary in Euros"
  )
```

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

Relationship Between Performance and Wage
Correlation excluding the top outlier (Lionel Messi): 0.23

The scatter plot shows a positive but weak relationship between performance and wages. Although higher finishing scores associate with higher salaries on average, large wage dispersion at similar performance levels indicates that performance alone does not fully explain wage differences. The low correlation coefficient (0.23, excluding the top outlier) suggests additional factors—such as club affiliation or playing position—likely play important roles in wage determination.

4.4 Wage Differences Across Clubs

```
fifa_final <- fifa_final %>%
  mutate(
```

```
    Club = stringr::str_replace_all(Club, "\r", ""),
    Club = stringr::str_replace_all(Club, "\n", ""),
    Club = stringr::str_squish(Club)
  )

club_stats <- fifa_final %>%
  group_by(Club) %>%
  summarise(
    Avg_Perf_Per_Player = mean(Goals, na.rm = TRUE),
    Total_Goals_Potential = sum(Goals, na.rm = TRUE),
    Avg_Wage = mean(Wage_EUR, na.rm = TRUE),
    Nb_Players = n()
  ) %>%
  filter(Nb_Players > 10) %>%
  arrange(desc(Avg_Wage))

print("Top 10 Clubs by Mean Wage:")
```

```
## [1] "Top 10 Clubs by Mean Wage:"
```

```
print(head(club_stats %>% select(Club, Avg_Wage, Avg_Perf_Per_Player), 10))
```

```
## # A tibble: 10 x 3
##    Club                 Avg_Wage Avg_Perf_Per_Player
##    <chr>                   <dbl>               <dbl>
##  1 Real Madrid           156233.                60.1
##  2 FC Barcelona          123727.                57.8
##  3 Manchester City       110273.                53.1
##  4 Liverpool              91773.                55.2
##  5 Manchester United      90485.                54.7
##  6 Inter                  90192.                55.3
##  7 FC Bayern München      86043.                59.1
##  8 Chelsea                76788.                53.0
##  9 Paris Saint-Germain    75885                 54.9
## 10 Tottenham Hotspur      75606.                52.8
```

```
head(club_stats %>% select(Club, Avg_Wage, Avg_Perf_Per_Player), 10) %>%
  kable(
    caption = "Top 10 Clubs by Average Wage",
    col.names = c("Club", "Average Weekly Wage (€)", "Average Performance"),
    digits = 2,
    booktabs = TRUE
  ) %>%
```
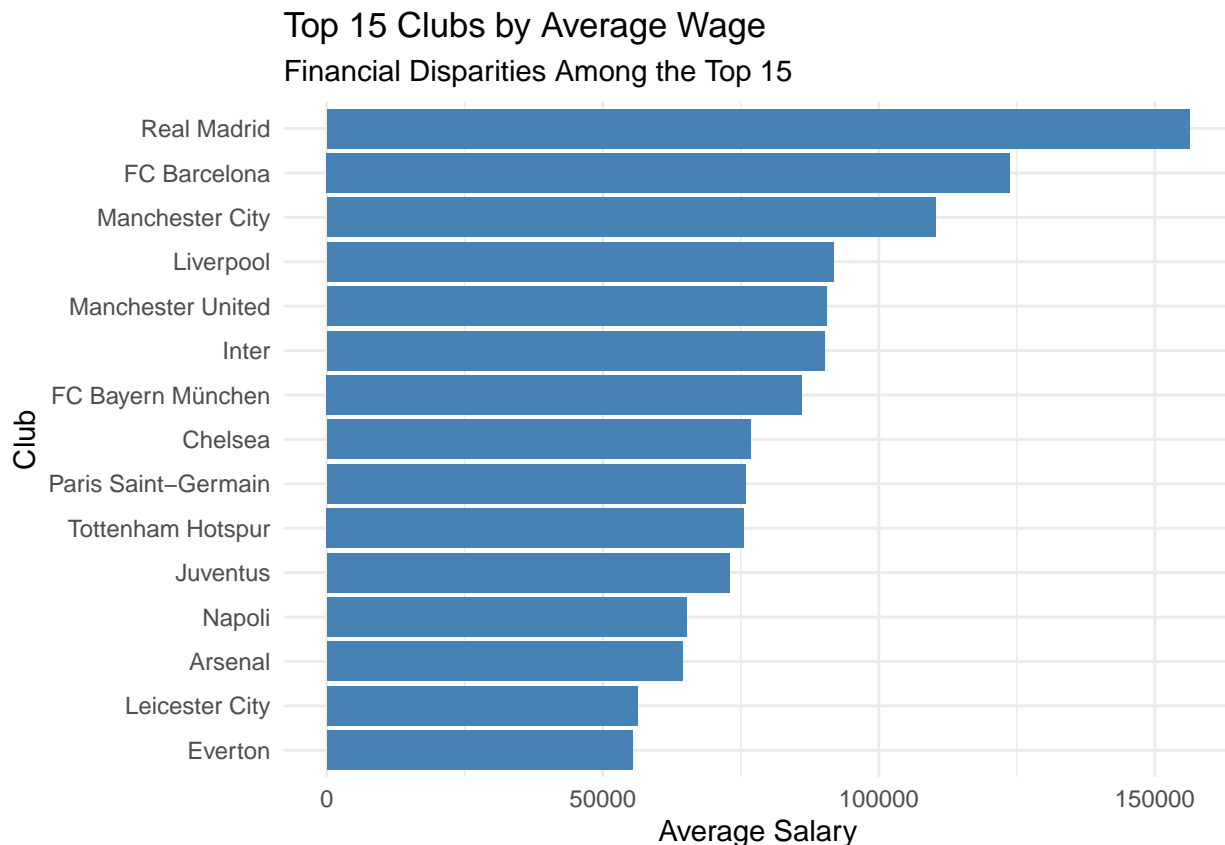
```
kable_styling(
  full_width = FALSE,
  position = "center"
)
```

Table 2: Top 10 Clubs by Average Wage

Club	Average Weekly Wage (€)	Average Performance
Real Madrid	156233.33	60.10
FC Barcelona	123727.27	57.79
Manchester City	110272.73	53.09
Liverpool	91772.73	55.24
Manchester United	90484.85	54.67
Inter	90192.31	55.31
FC Bayern München	86043.48	59.13
Chelsea	76787.88	52.97
Paris Saint-Germain	75885.00	54.87
Tottenham Hotspur	75606.06	52.85

```
ggplot(head(club_stats, 15), aes(x = reorder(Club, Avg_Wage), y = Avg_Wage)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Top 15 Clubs by Average Wage",
       subtitle = "Financial Disparities Among the Top 15",
       x = "Club",
       y = "Average Salary")
```

Top 15 Clubs by Average Wage
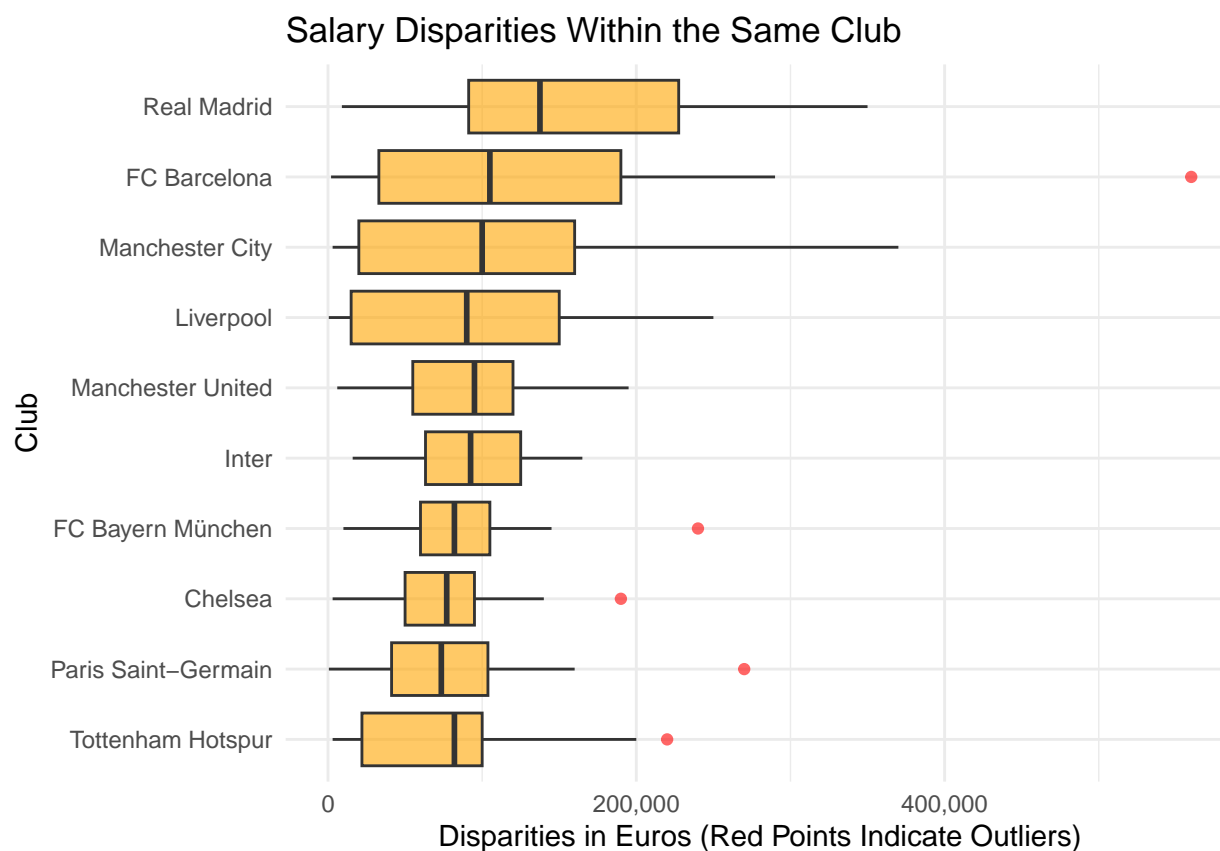Financial Disparities Among the Top 15

Substantial differences in average player salaries exist among top clubs. Elite clubs such as Real Madrid, FC Barcelona, and Manchester City exhibit the highest average wages, while other well-known clubs display significantly lower levels. These disparities suggest club affiliation plays an important role in wage determination, reflecting differences in financial resources, market power, and revenue-generating capacity. This provides descriptive evidence that wages vary systematically by club, even before accounting for individual performance differences.

4.5 Within-Club Wage Disparities

```r
top_10_names <- head(club_stats$Club, 10)
data_top_10 <- fifa_final %>% filter(Club %in% top_10_names)

ggplot(data_top_10, aes(x = reorder(Club, Wage_EUR), y = Wage_EUR)) +
  geom_boxplot(fill = "orange", alpha = 0.6, outlier.colour = "red") +
  coord_flip() +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Salary Disparities Within the Same Club",
```

```
        x = "Club",
        y = "Disparities in Euros (Red Points Indicate Outliers)")
```

Salary Disparities Within the Same Club

The boxplots reveal substantial wage disparities within clubs, indicating players within the same team can earn very different salaries. Even among top clubs, wage spreads are wide, with several clubs displaying extreme outliers corresponding to very highly paid players. This intra-club variation suggests wage inequality is driven not only by differences across clubs but also by heterogeneity in player roles, performance, and bargaining power within clubs.
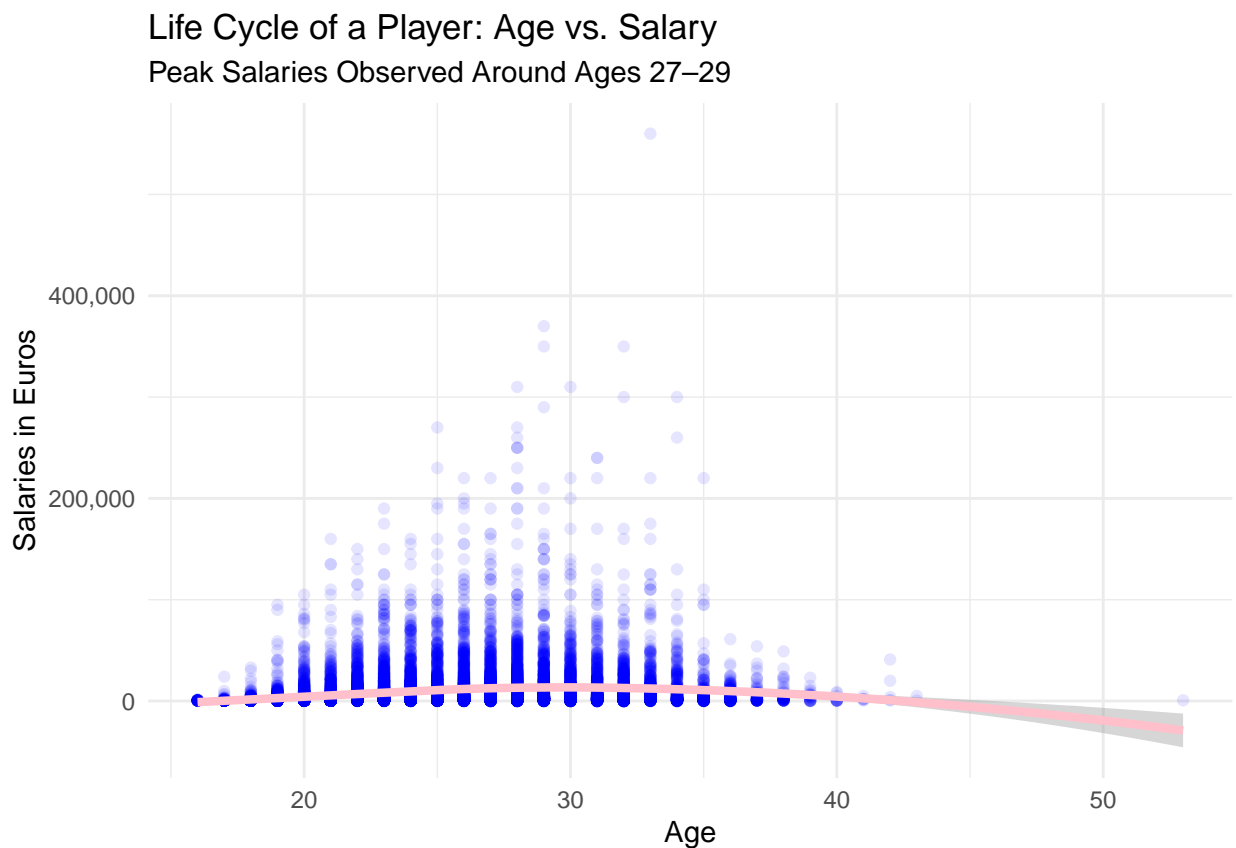
4.6   Age and Wage Life Cycle

```
ggplot(fifa_final, aes(x = Age, y = Wage_EUR)) +
  geom_point(alpha = 0.1, color = "blue") +
  geom_smooth(method = "loess", color = "pink", size = 1.5) +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Life Cycle of a Player: Age vs. Salary",
       subtitle = "Peak Salaries Observed Around Ages 27-29",
```

```
    x = "Age",
    y = "Salaries in Euros")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

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

Life Cycle of a Player: Age vs. Salary
Peak Salaries Observed Around Ages 27–29

A clear life-cycle pattern exists between age and salary. Player wages increase with age, reaching highest levels around the late twenties (approximately 27–29 years old), before declining at older ages. This pattern is consistent with players reaching peak performance and bargaining power during their prime years, followed by decline as performance and market value decrease. Wide salary dispersion at peak ages suggests age alone does not fully explain wage differences, reinforcing the role of additional factors such as performance, club affiliation, and position.

4.7 Summary

The exploratory analysis reveals key patterns underlying wage determination in professional football. Wage distribution is highly unequal, with most players earning relatively low salaries while a small group receives extremely high wages. Performance measures are more concentrated, and the relationship between performance and wages, while positive, is weak and characterized by substantial dispersion.

Analysis by club reveals significant wage differences across teams and pronounced disparities within clubs, suggesting club affiliation plays an important role in shaping wage outcomes. The age–salary relationship follows a life-cycle pattern, with wages peaking during prime years and declining thereafter, while displaying considerable heterogeneity at similar ages.

These descriptive patterns indicate wage differences cannot be fully explained by individual performance alone, motivating formal analysis of the extent to which club affiliation and playing position independently contribute to wage determination.

5 Methodology

Having established the key descriptive patterns in the data, we now turn to formal econometric analysis. This section outlines the empirical specifications and variable construction used to assess the determinants of player wages.

```r
fifa_model <- fifa_final %>%
  select(
    Wage_EUR,
    overall_rating,
    finishing,
    base_stats,
    Club,
    position,
    Age,
    height_cm,
    weight_kg,
    preferred_foot,
    potential
  )
```

5.1 Empirical Specifications

The analysis is based on linear regressions estimated using Ordinary Least Squares (OLS). Three main specifications are estimated to progressively incorporate additional determinants:

5.1.1 Model 1 (Performance)

$$\log(wage_i) = \alpha + \beta X_i + \varepsilon_i$$

Wages depend solely on individual performance and control variables, assessing the extent to which salaries reflect individual merit.

5.1.2 Model 2 (Performance + Position)

$$\log(wage_i) = \alpha + \beta X_i + \delta Position_i + \varepsilon_i$$

Position group variables examine whether systematic wage differences across positions persist once performance and individual characteristics are controlled for.

5.1.3 Model 3 (Performance + Position + Club)

$$\log(wage_i) = \alpha + \beta X_i + \delta Position_i + \theta Club_i + \varepsilon_i$$

Club fixed effects identify whether club affiliation has additional impact on wages, beyond player performance and position.

where X_i represents performance and control variables, $Position_i$ corresponds to position group, and $Club_i$ captures club fixed effects.

5.2 Model Variables

5.2.1 Dependent Variable: Wages (Estimated FIFA21)

The dependent variable is the natural logarithm of weekly wage estimated by FIFA (`log(wage_eur)`). The logarithmic transformation reduces the strong right skewness observed in wage distribution and allows coefficient interpretation in percentage terms.

```
fifa_model <- fifa_model %>%
  filter(Wage_EUR > 0) %>%
  mutate(log_wage = log(Wage_EUR))
```

5.2.2 Performance Variables

Individual performance is captured using several FIFA indicators. The **overall rating** reflects general ability level, while **base stats** represent aggregate performance based on multiple in-game attributes. The **finishing** attribute is included as a proxy for offensive ability and goal-scoring capacity.

5.2.3 Control Variables

The model controls for individual characteristics that may affect wages independently of performance: **age**, **height**, **weight**, and **potential** (FIFA's assessment of expected future ability).

5.2.4 Playing Positions

Playing positions are grouped into four categories: goalkeepers, defenders, midfielders, and forwards. This aggregation simplifies interpretation and improves readability. Using the original 15 positions would generate numerous coefficients, making it difficult to interpret positional effects clearly. Grouping positions captures systematic differences in positional roles while preserving economic intuition behind wage differentials across playing positions.

```r
fifa_model <- fifa_model %>%
  mutate(position_group = case_when(
    position == "Goalkeeper" ~ "Goalkeeper",
    position %in% c("Center Back", "Left Back", "Right Back",
                    "Left Wing Back", "Right Wing Back") ~ "Defender",
    position %in% c("Central Midfielder", "Central Defensive Midfielder",
                    "Central Attacking Midfielder", "Left Midfielder",
                    "Right Midfielder") ~ "Midfielder",
    position %in% c("Striker", "Center Forward",
                    "Left Winger", "Right Winger") ~ "Forward",
    TRUE ~ NA_character_
  ))

fifa_model %>%
  count(position_group) %>%
  mutate(
    Percentage = round(100 * n / sum(n), 2)
  ) %>%
  kable(
    caption = "Distribution of Players by Position Group",
    col.names = c("Position Group", "Number of Players", "Percentage (%)"),
    align = "lcc"
  )
```

Table 3: Distribution of Players by Position Group

Position Group	Number of Players	Percentage (%)
Defender	6309	33.66
Forward	3206	17.11

Position Group	Number of Players	Percentage (%)
Goalkeeper	2041	10.89
Midfielder	7186	38.34

5.2.5 Club (Fixed Effects)

Club affiliation is controlled through **club fixed effects**. Without fixed effects, wage comparisons across players from different clubs may introduce bias due to substantial differences in financial resources, market power, and visibility.

By incorporating club fixed effects, the analysis compares **players within the same club**, holding constant the institutional and financial environment. This approach prevents wage differences from being driven simply by belonging to a wealthier or more prestigious club and allows clearer identification of individual characteristics' impact.

Club fixed effects isolate individual attributes' effects by absorbing unobserved club-level heterogeneity. The focus is not on interpreting individual club coefficients but rather on assessing whether, once club-specific factors are controlled for, systematic wage differences across players can still be explained by performance and other individual characteristics.

6 Results

We now present the estimation results from the three OLS specifications, progressively incorporating playing position and club affiliation as additional determinants of player wages.

6.1 Model 1: Baseline Specification – Individual Performance

```
model_1 <- lm(
  log_wage ~ overall_rating + finishing  +
    Age + height_cm + weight_kg + potential,
  data = fifa_model
)


tidy(model_1) |>
  mutate(
    estimate = round(estimate, 4),
    std.error = round(std.error, 4),
    statistic = round(statistic, 2),
    p.value = round(p.value, 4)
  ) |>
```

```
select(term, estimate, std.error, statistic, p.value) |>
kable(
  caption = "Baseline OLS Results: Performance and Wages",
  col.names = c("Variable", "Coefficient", "Std. Error", "t-value", "p-value"),
  align = "lcccc"
)
```

Table 4: Baseline OLS Results: Performance and Wages

Variable	Coefficient	Std. Error	t-value	p-value
(Intercept)	-4.6101	0.2204	-20.92	0.0000
overall_rating	0.1516	0.0022	69.05	0.0000
finishing	0.0054	0.0004	15.10	0.0000
Age	-0.0166	0.0026	-6.45	0.0000
height_cm	0.0178	0.0014	12.45	0.0000
weight_kg	-0.0088	0.0014	-6.35	0.0000
potential	0.0053	0.0022	2.36	0.0181

This baseline model establishes the relationship between wages and individual-level characteristics, excluding positional and club-specific effects.

Overall rating is by far the most important wage determinant. A one-point increase in overall rating associates with approximately 15% higher weekly wages, holding other factors constant. This confirms that FIFA's aggregated performance evaluation is a strong proxy for players' market valuation.

The finishing coefficient is positive and highly significant even after controlling for overall rating, suggesting specific offensive skills are rewarded with an additional wage premium beyond general performance.

Age enters with a negative and significant coefficient, indicating older players earn lower wages conditional on performance. This pattern is consistent with depreciation of football-specific human capital or reduced contractual value at later career stages.

Regarding physical characteristics, height has a positive and significant effect on wages, while weight is negatively associated with wages. Although smaller in magnitude compared to performance measures, these effects suggest certain physical attributes are valued within FIFA's evaluation framework.

Finally, potential has a positive but relatively small effect, indicating expectations about future ability play a secondary role compared to current performance.

The model explains a substantial share of wage variation, with an adjusted R^2 of approximately 0.63, highlighting the central role of individual performance attributes.

6.2 Model 2: The Role of Playing Position

```r
model_2 <- lm(
  log_wage ~ overall_rating + finishing +
    Age + height_cm + weight_kg + potential + position_group,
  data = fifa_model
)

tidy(model_2) |>
  mutate(
    estimate = round(estimate, 4),
    std.error = round(std.error, 4),
    statistic = round(statistic, 2),
    p.value = round(p.value, 4)
  ) |>
  select(term, estimate, std.error, statistic, p.value) |>
  kable(
    caption = "Regression Results: Performance and Playing Position",
    col.names = c("Variable", "Coefficient", "Std. Error", "t-value", "p-value"),
    align = "lcccc"
  )
```

Table 5: Regression Results: Performance and Playing Position

Variable	Coefficient	Std. Error	t-value	p-value
(Intercept)	-4.6934	0.2265	-20.72	0.0000
overall_rating	0.1534	0.0023	68.00	0.0000
finishing	0.0024	0.0007	3.31	0.0009
Age	-0.0156	0.0026	-6.06	0.0000
height_cm	0.0180	0.0014	12.46	0.0000
weight_kg	-0.0088	0.0014	-6.30	0.0000
potential	0.0056	0.0022	2.52	0.0118
position_groupForward	0.0967	0.0286	3.38	0.0007
position_groupGoalkeeper	-0.1520	0.0268	-5.67	0.0000
position_groupMidfielder	0.0395	0.0203	1.95	0.0517

The second model extends the baseline by including dummy variables for playing positions, assessing whether positional wage differences persist after controlling for individual performance.

Position inclusion slightly improves model fit, indicating playing position provides additional explanatory power beyond individual attributes.

Clear positional wage premiums emerge. Relative to defenders (the reference category), forwards earn approximately 9–10% higher wages, conditional on performance. Midfielders also display a positive wage premium, although smaller in magnitude and marginally significant. Conversely, goalkeepers experience a substantial wage discount of roughly 15%, even after controlling for overall rating and other characteristics.

Notably, the finishing coefficient decreases markedly compared to the baseline model, suggesting part of the finishing premium in the first model was capturing systematic differences between offensive and defensive positions.

These findings indicate playing position matters for wage determination, although its contribution remains secondary relative to overall performance.

6.3   Model 3: The Role of Club Affiliation

```
model_3 <- lm(log_wage ~ overall_rating + finishing + Age + height_cm +
   weight_kg + potential + position_group + factor(Club),
   data = fifa_model)
```

6.3.1   Regression Results: Determinants of Players' Wages with Club Fixed Effects

Variable	Coefficient	Std. Error	Significance
Overall Rating	0.1370	0.0011	***
Finishing (Performance)	0.0017	0.0003	***
Age	-0.0177	0.0013	***
Height (cm)	0.0015	0.0007	**
Weight (kg)	0.0004	0.0007	
Potential	-0.0363	0.0012	***
Forward (ref: Defender)	0.1184	0.0133	***
Midfielder (ref: Defender)	0.0604	0.0094	***
Goalkeeper (ref: Defender)	-0.1553	0.0124	***

Observations: 18,742
R²: 0.637

Notes:
- Dependent variable: log of estimated weekly wages (FIFA 21).
- All models include **club fixed effects**.
- Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The third model introduces club fixed effects, enabling comparison of players within the same club and isolating individual characteristics' and playing position's impact from club-level wage policies.

Once club affiliation is controlled for, overall rating remains large and highly significant, though slightly smaller than in previous models. This suggests part of the overall rating effect operates through sorting into higher-paying clubs, but performance remains a key wage determinant even within clubs.

The finishing effect remains positive but is further attenuated, reinforcing that specialized skills matter less once institutional factors are considered.

Age continues to exhibit a negative and robust wage association. Interestingly, potential now enters with a negative coefficient, indicating that within the same club, players with higher potential but similar current performance tend to earn lower wages. This pattern is consistent with younger players who have high expected future ability but are still early in their contractual progression.

Importantly, positional wage differentials persist even after controlling for club affiliation. Forwards and midfielders continue to earn significant wage premiums relative to defenders, while goalkeepers face a substantial wage discount. This suggests positional valuation is not merely driven by club composition differences but reflects systematic role-based wage hierarchies.

The model achieves the highest explanatory power, with an R^2 of approximately 0.64, highlighting the importance of controlling for club-level heterogeneity.

7   Robustness Check: Elastic Net Regularization

While the OLS specifications provide clear and interpretable results, the FIFA dataset contains numerous highly correlated performance attributes that may lead to multicollinearity concerns. To address this issue and ensure the robustness of our findings, we complement the baseline analysis with Elastic Net regularization, which combines variable selection with coefficient shrinkage to handle high-dimensional, correlated data more effectively.

An Elastic Net regression is estimated to formally select relevant performance attributes among a large set of correlated FIFA skill indicators. The model combines LASSO and Ridge penalties and uses 10-fold cross-validation to select the optimal regularization parameter.

```r
# 1. Relevant variable selection
fifa_elastic <- fifa_final %>%
  select(
    # Outcome
    Wage_EUR,

    # Demographics & Physical
    Age,
    height_cm,
    weight_kg,
    preferred_foot,
```

```
# Performance: Aggregate Measures
overall_rating,
potential,
base_stats,
`Total Stats`,

# Attacking
Attacking,
Crossing,
finishing,
`Heading Accuracy`,
`Short Passing`,
Volleys,

# Skill / Dribbling
Skill,
Dribbling,
Curve,
`FK Accuracy`,
`Long Passing`,
`Ball Control`,

# Movement
Movement,
Acceleration,
`Sprint Speed`,
Agility,
Reactions,
Balance,

# Power
Power,
`Shot Power`,
Jumping,
Stamina,
Strength,
`Long Shots`,

# Mentality
Mentality,
Aggression,
Interceptions,
Positioning,
Vision,
```

```r
    Penalties,
    Composure,

    # Defending
    Defending,
    Marking,
    `Standing Tackle`,
    `Sliding Tackle`,

    # Goalkeeping
    Goalkeeping,
    `GK Diving`,
    `GK Handling`,
    `GK Kicking`,
    `GK Positioning`,
    `GK Reflexes`,

    # Position
    position
  )
```

```r
# 2. Transformations
fifa_elastic <- fifa_elastic %>%
  mutate(
    log_wage = log(Wage_EUR),
    preferred_foot = factor(preferred_foot),
    position = factor(position)
  ) %>%
  select(-Wage_EUR) %>%
  na.omit()


fifa_elastic <- fifa_elastic %>%
  filter(!is.na(log_wage), is.finite(log_wage))
```

```r
library(glmnet)
```

```
## Warning: le package 'glmnet' a été compilé avec la version R 4.4.3
```

```
## Le chargement a nécessité le package : Matrix
```

```
## Warning: le package 'Matrix' a été compilé avec la version R 4.4.3
```

```
##
## Attachement du package : 'Matrix'

## Les objets suivants sont masqués depuis 'package:tidyr':
##
##      expand, pack, unpack

## Loaded glmnet 4.1-10
```

```r
# Regressor matrix
X <- model.matrix(
  log_wage ~ .,
  data = fifa_elastic
)[, -1]    # Remove intercept

y <- fifa_elastic$log_wage

set.seed(123)

cv_enet <- cv.glmnet(
  x = X,
  y = y,
  alpha = 0.5,          # Elastic Net
  standardize = TRUE,
  nfolds = 10
)

cv_enet$lambda.min      # Minimizing the error
```

```
## [1] 0.001040417
```

```r
cv_enet$lambda.1se      # Parsimonious model
```

```
## [1] 0.01407733
```

```r
enet_model <- glmnet(
  x = X,
  y = y,
  alpha = 0.5,
  lambda = cv_enet$lambda.1se,
  standardize = TRUE
)
```

7.1   Why Elastic Net Is Preferred to LASSO

Elastic Net is particularly well suited to this application due to strong correlations among performance-related predictors. Many FIFA attributes—such as overall rating, finishing, dribbling, shot power, and reactions—capture overlapping dimensions of player ability.

While LASSO tends to select only one variable from a group of highly correlated predictors, potentially discarding relevant information, Elastic Net combines   and   penalties and exhibits the grouping effect. This allows correlated variables to be selected jointly, leading to more stable and economically interpretable coefficient estimates.
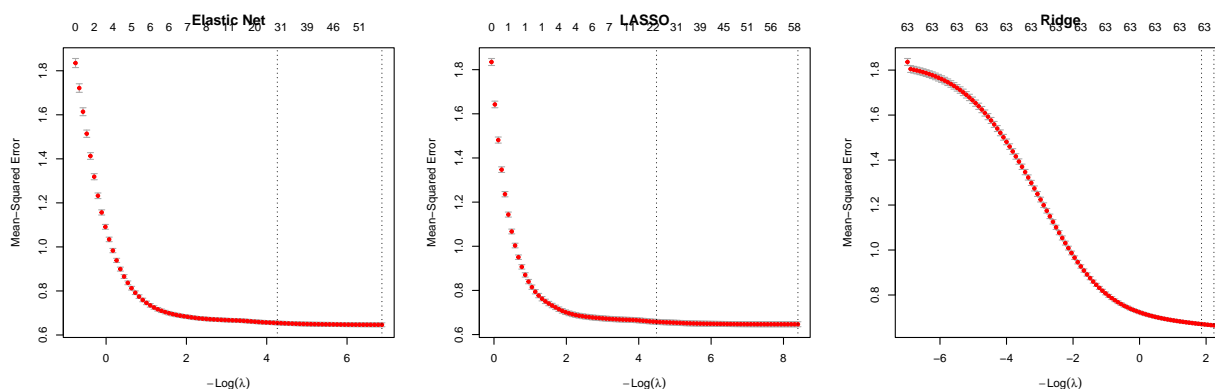
Moreover, Elastic Net provides better balance between variable selection and coefficient shrinkage. Compared to LASSO's aggressive sparsity, Elastic Net reduces estimation instability without retaining all predictors, as Ridge would. As a result, the model achieves comparable or lower prediction error while maintaining robustness to small data changes.

Overall, Elastic Net offers a more reliable representation of the wage determination process in professional football, where salaries depend on a bundle of correlated performance attributes rather than isolated skills.

```r
cv_lasso <- cv.glmnet(X, y, alpha = 1)
cv_ridge <- cv.glmnet(X, y, alpha = 0)

par(mfrow = c(1, 3))  # 1 row, 3 columns

plot(cv_enet, main = "Elastic Net")
plot(cv_lasso, main = "LASSO")
plot(cv_ridge, main = "Ridge")
```


The cross-validation plots display mean squared error (MSE) as a function of regularization parameter   for Elastic Net, LASSO, and Ridge regression.

For **Elastic Net** and **LASSO**, the error curve decreases sharply as   is reduced. This rapid decay reflects the strong shrinkage effect of the   penalty, which quickly drives many coefficients exactly to zero. As a result, the model transitions rapidly from an over-regularized

specification to a more flexible one, performing implicit variable selection. This behavior is typical when signal is concentrated in a relatively small subset of predictors.

In contrast, the **Ridge** regression plot exhibits a smoother and more gradual decline in error. Because Ridge relies solely on an penalty, coefficients are shrunk continuously but never set to zero. Consequently, all predictors remain in the model, and the bias–variance trade-off evolves more progressively as changes.

Overall, the different curve shapes are consistent with the theoretical properties of the penalties: LASSO and Elastic Net favor sparsity and abrupt model simplification, while Ridge emphasizes coefficient stabilization without variable elimination.

Now we examine the non-zero coefficients selected by the Elastic Net model.

```r
# Coefficients retained after regularization
coef_enet <- coef(enet_model)
coef_df <- as.matrix(coef_enet)
coef_df <- coef_df[coef_df[,1] != 0, , drop = FALSE]
coef_df
```

```
##                                               s0
## (Intercept)                         -4.8290490817
## Age                                 -0.0115710200
## height_cm                            0.0170895370
## weight_kg                           -0.0007537812
## overall_rating                       0.1276334016
## potential                            0.0093385574
## finishing                           -0.0004163434
## `Heading Accuracy`                   0.0020666636
## Dribbling                            0.0024521547
## Curve                                0.0008338880
## Reactions                            0.0059271041
## Balance                              0.0025614842
## `Shot Power`                         0.0060637479
## Jumping                              0.0002665259
## Stamina                             -0.0040845936
## Strength                            -0.0041341235
## Aggression                           0.0016452649
## Vision                               0.0004962356
## Penalties                            0.0008627106
## Composure                            0.0051661033
## `Sliding Tackle`                     0.0012178369
## `GK Reflexes`                        0.0007490731
## positionCentral Attacking Midfielder -0.0519071656
## positionCentral Defensive Midfielder -0.0760122204
## positionCentral Midfielder           0.0021228613
```

```
## positionLeft Back             -0.0019507002
## positionLeft Midfielder       -0.0021800972
## positionLeft Winger           -0.0610250378
## positionRight Wing Back       -0.0583380241
## positionStriker                0.0365060005
```

7.2 Elastic Net Results and Interpretation

The Elastic Net estimation confirms that player wages are primarily driven by overall performance indicators rather than isolated technical skills. Variables such as overall rating, potential, reactions, composure, and shot power remain positively associated with wages after regularization, highlighting the importance of general playing quality and psychological attributes in salary determination. In contrast, several individual skill metrics exhibit heavily shrunk coefficients, suggesting their marginal contribution to wages is limited once global performance measures are considered.

Age displays a negative coefficient, consistent with wages peaking and then declining as players approach later career stages, even after controlling for performance. Physical characteristics such as height and weight show relatively small effects, indicating physical build plays a secondary role compared to skill-related attributes. Positional effects persist after regularization: attacking roles, such as striker, are associated with higher wages, while several defensive and wide positions are penalized relative to the reference category.

Overall, Elastic Net produces a parsimonious yet stable model that retains economically meaningful predictors while mitigating multicollinearity among highly correlated performance variables. This supports the interpretation that professional football wages reflect a bundle of correlated abilities rather than a single dominant skill, making Elastic Net particularly well suited for this application.

7.3 Comparison with Baseline OLS Results

Compared to the baseline OLS specification, the Elastic Net model delivers a more conservative and stable interpretation of player wage determinants. While OLS assigns statistical significance to a relatively large set of performance variables, Elastic Net shrinks many coefficients toward zero, indicating several individual skills identified as significant under OLS likely reflect multicollinearity rather than independent explanatory power. This is particularly relevant given the strong correlation structure among FIFA performance attributes.

Both approaches agree on the central role of overall rating and potential as key wage drivers, confirming that global performance evaluations dominate salary determination. However, Elastic Net reduces the influence of narrowly defined technical skills, reallocating explanatory weight toward broader indicators such as reactions, composure, and shot power. This suggests OLS may overstate the importance of granular skill measures when included alongside highly aggregated performance metrics.

Regarding positional effects, OLS estimates tend to exhibit larger and less stable coefficients across playing positions, whereas Elastic Net retains only the most economically meaningful positional differences. Attacking positions remain positively associated with wages, while several defensive and wide roles are penalized, but effect magnitudes are attenuated under regularization. This indicates part of the positional wage premium captured by OLS may be driven by correlated performance characteristics rather than pure role-specific valuation.

Overall, Elastic Net improves upon OLS by balancing interpretability and predictive stability. By explicitly addressing multicollinearity and overfitting, it provides a more credible assessment of which factors truly explain wage differences once individual performance, club affiliation, and playing position are jointly considered.

7.4 Validity of the OLS Baseline Models

The baseline OLS models provide a solid and economically intuitive starting point for analysis. The estimated coefficients exhibit signs and magnitudes fully consistent with theoretical expectations in professional football labor markets. Higher overall ratings and potential associate with higher wages, age carries a negative effect consistent with career dynamics, and playing position plays a meaningful role in wage determination. These results suggest the core economic intuition underlying the OLS specifications is sound.

However, the FIFA dataset is characterized by high-dimensional structure with many strongly correlated performance attributes. In this context, OLS estimates may overstate individual skills' independent contribution, even when overall model explanatory power remains strong. This does not invalidate OLS results but rather highlights their limitations when interpretation shifts from global effects to granular performance metrics.

The regularized models, particularly Elastic Net, do not overturn OLS conclusions. Instead, they reinforce them by filtering out redundant information and stabilizing coefficient estimates. The fact that key determinants identified under OLS—such as overall rating, age, and positional effects—remain relevant under Elastic Net provides additional confidence in the baseline models' robustness. Regularization therefore acts as a complementary tool that refines interpretation rather than correcting flawed intuition.

8 Discussion and Limitations

The results provide consistent evidence that individual performance, age, and playing position are key determinants of wage differences among professional football players. The stability of main coefficients across OLS and regularized specifications suggests core economic mechanisms underlying wage formation are well captured by the proposed models. In particular, the persistence of overall performance measures and positional effects supports the interpretation of wages as a function of both individual productivity and role-specific market valuation.

Several limitations should nevertheless be acknowledged. First, the analysis relies on a linear modeling framework, which captures average marginal effects but may not fully reflect potential nonlinearities in wage determination. In practice, wage-setting may involve threshold effects or disproportionately high premia for top-performing or highly visible players, particularly at the upper tail of the wage distribution. Similarly, the age-wage relationship may follow a nonlinear career profile only approximated by a linear specification.

Second, the wage variable is based on FIFA estimates rather than observed contractual salaries. While these estimates are widely used in empirical work, they may be subject to measurement error, which could attenuate estimated effects and limit causal interpretation. The findings should therefore be interpreted as descriptive and explanatory rather than causal.

Third, potential endogeneity cannot be fully ruled out. Performance indicators and wages are likely jointly determined, as higher wages may influence playing time, visibility, or perceived performance, while performance affects wage negotiations. Although club fixed effects help control for unobserved club-level heterogeneity—such as financial resources, league competitiveness, and market exposure—they do not fully address individual-level bargaining dynamics or contract-specific features.

Finally, the analysis is based on cross-sectional data from a single season, limiting its external validity over time. The models therefore explain wage dispersion at a given point in time but cannot capture dynamic adjustments in wages or performance across players' careers.

Despite these limitations, the combination of baseline OLS models and regularized techniques provides a coherent and robust framework for understanding wage differentials in professional football, balancing interpretability with statistical stability in a high-dimensional setting.

9 Conclusion

This study examines wage determinants among professional football players using standardized FIFA 21 data, focusing on individual performance, playing position, and club affiliation. The results consistently show player wages are primarily explained by performance-related attributes, most notably overall rating, while age and physical characteristics also play meaningful roles.

Baseline OLS estimates indicate a strong positive association between performance measures and wages, alongside a negative age effect, consistent with standard economic intuition regarding productivity and career dynamics in professional sports. Playing position reveals systematic wage differentials across roles, with forwards earning wage premiums relative to defenders and goalkeepers earning significantly less, even after controlling for performance.

Controlling for club affiliation through fixed effects, core performance variables remain statistically significant, although some coefficients are attenuated. This suggests that while clubs differ substantially in overall wage levels—reflecting financial capacity, league status, and market exposure—individual performance continues to be a central wage determinant within clubs.

Regularized models, particularly Elastic Net, reinforce these findings by identifying a broader set of relevant performance attributes while mitigating multicollinearity in a high-dimensional setting. Importantly, main variables highlighted by OLS remain influential, supporting the validity of parsimonious baseline specifications and confirming that the economic intuition guiding variable selection was well founded.

Overall, the results suggest wage disparities in professional football reflect a combination of individual productivity, positional market valuation, and club-specific factors. While the analysis is not causal, it provides robust evidence that performance and role specialization are key components of wage formation, even within a standardized and highly competitive evaluation framework.

10   References

Dobson, S., & Goddard, J. (2001). *The Economics of Football.* Cambridge: Cambridge University Press.

Franck, E., & Nüesch, S. (2010). Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, 48(4), 853–873. https://doi.org/10.1111/j.1465-7295.2009.00265.x

Frick, B. (2007). The football players' labor market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, 54(3), 422–446. https://doi.org/10.1111/j.1467-9485.2007.00423.x

SoFIFA. (2021). *FIFA 21 Player Attributes and Wages Dataset.* Retrieved from https://sofifa.com