

# Final Project

Brandon Barber, Omkar Dixit

## Intoduction

Soccer is one of the most popular sports in the world, being played in almost every country. Of the top leagues in the world, the Spanish *La Liga* is among the most elite. This report aims to discover which factors have the most effect on a clubs success in the top league.

Our analysis focuses on the players that make up each club and is based on these questions:

1. Do clubs see more success from purchasing players in the offseason or fostering talent from their youth acaedmies?
2. Do a team's player stastics in the FIFA 23 videogame have a relationship with how the team performs in the season?

## Data Provenance

### 1. Transfer and Academy Data

**Source:** Transfermarkt - portal for data on soccer transfers, market values, and statistics

**Use:** Collect data on offseason spending, club value, and academy trasnfers for the 2022/23 season

### 2. Fifa Statistics

**Source:** Kaggle - dataset for all players in FIFA 23

**Use:** Gather data on the overalls for each player in La Liga

### 3. La Liga Standings

**Source:** ESPN - sports media company

**Use:** Gather data on the La Liga standings for the 2021/22 and 2022/23 seasons

## League Standings

Table 1: Final LaLiga Standings for 2022-23 [1]

Rank	Club	GP	W	D	L	F	A	GD	P
1	FC Barcelona	38	28	4	6	70	20	50	88
2	Real Madrid CF	38	24	6	8	75	36	39	78
3	Atlético de Madrid	38	23	8	7	70	33	37	77
4	Real Sociedad	38	21	8	9	51	35	16	71
5	Villarreal CF	38	19	7	12	59	40	19	64
6	Real Betis Balompié	38	17	9	12	46	41	5	60
7	CA Osasuna	38	15	8	15	37	42	-5	53
8	Athletic Club de Bilbao	38	14	9	15	47	43	4	51
9	RCD Mallorca	38	14	8	16	37	43	-6	50
10	Girona FC	38	13	10	15	58	55	3	49
11	Rayo Vallecano	38	13	10	15	45	53	-8	49
12	Sevilla FC	38	13	10	15	47	54	-7	49
13	RC Celta de Vigo	38	11	10	17	43	53	-10	43
14	Cádiz CF	38	10	12	16	30	53	-23	42
15	Getafe CF	38	10	12	16	34	45	-11	42
16	Valencia CF	38	11	9	18	42	45	-3	42
17	Unión Deportiva Almería	38	11	8	19	49	65	-16	41
18	Real Valladolid CF	38	11	7	20	33	63	-30	40
19	RCD Espanyol de Barcelona	38	8	13	17	52	69	-17	37
20	Elche CF	38	5	10	23	30	67	-37	25

*Legend:*

GP - Matches Played

W - Wins

D - Draws

L - Losses

F - Goals For

A - Goals Against

GD - Goal Differential

P - Points

## Economic Value vs On Field Success in La Liga

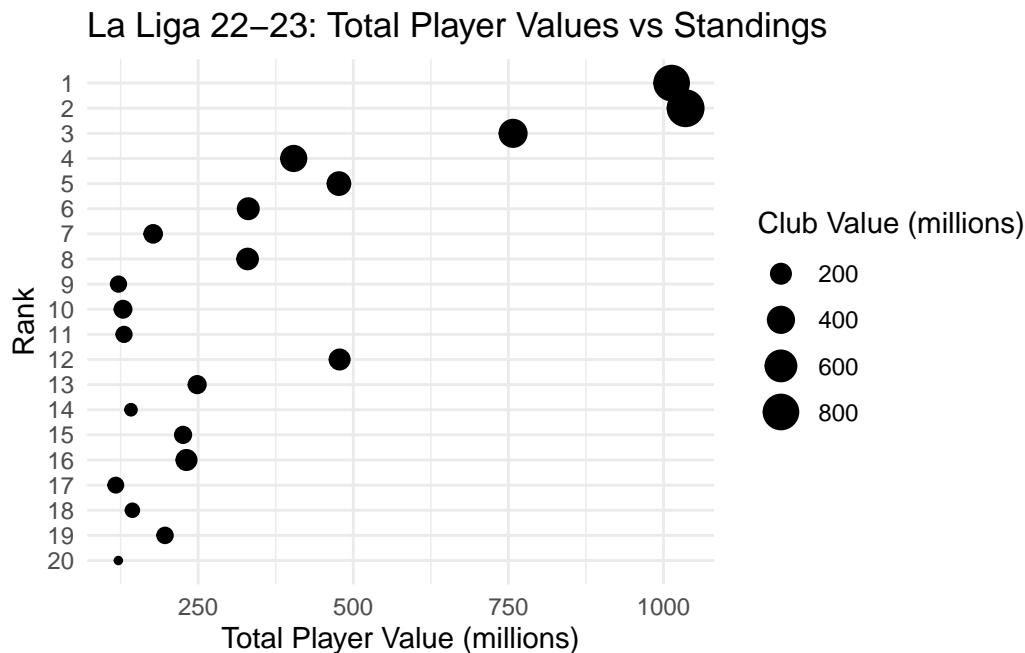


Figure 1: Each La Liga club's combined player market value taken from the FIFA dataset plotted against their 2022/23 final rank. Glyph size indicates overall club value.

The graph shows a relationship between higher total player value and better league performance. The two clubs ranked highest in the league are seen in the top right corner with large glyphs, possibly indicating that more spending is indicative of better performance.

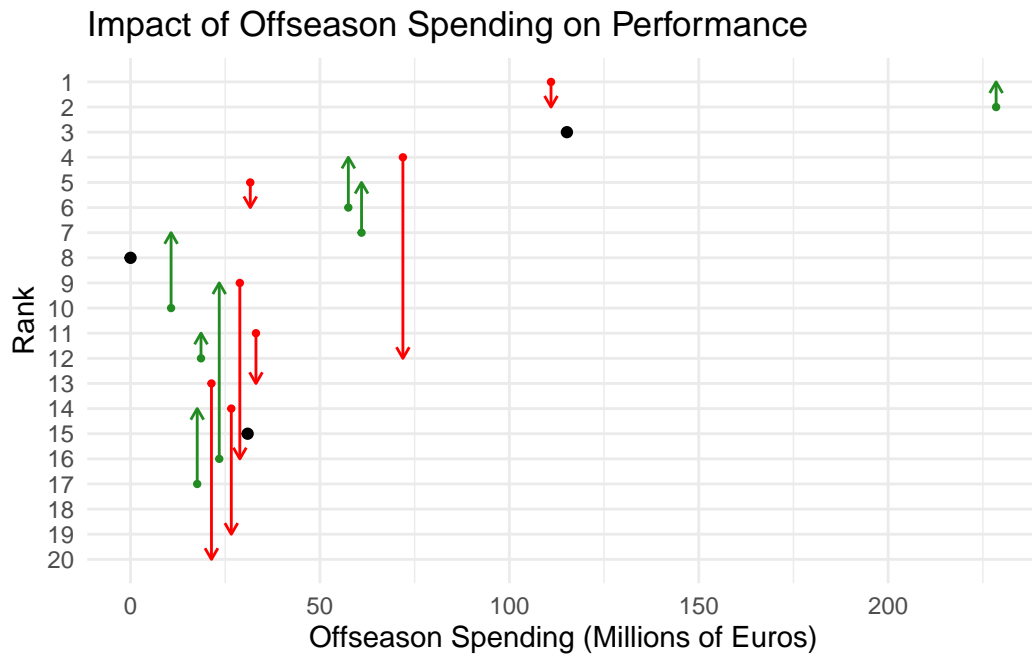
On the other hand, outside of the top left and bottom right corners, club value seems to show no relationship with rank, especially in the teams ranked in the bottom half of the league. Club value is calculated from a club's annual revenue and infrastructure value. This graph suggests that the players have much more of an impact on team performance than the overall wealth of the club.

Upon further analysis, the overall shape of the graph is shown to be downward sloping, and the relationship between player value and rank flattens out at around €350 million. This suggests that spending is important for smaller clubs ranked lower in the league but becomes less important for the larger clubs towards the middle and top of the league.

[1][4][5][6]

## Impact of the Offseason

Soccer clubs can acquire players by any one of three methods, a purchase of a player from another club also known as a transfer, a player being loaned from one club to another for a set amount of time, or training players from a young age to be eventually promoted to their first team. These plots will analyze the impact of purchasing players in the offseason on team performance versus promoting academy players.



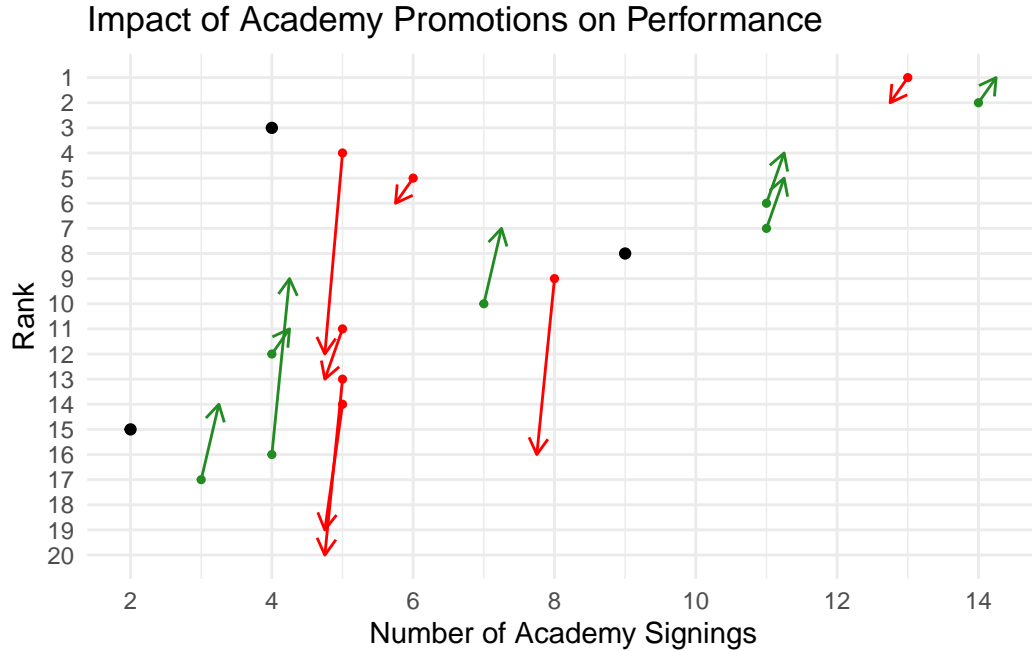


Figure 3: Number of academy players promoted to the first team through for each club the 2022/23 La Liga season plotted against end of season rank. Each point represents a club, with arrows indicating changes in rank from the 2021/22 season. Green arrows indicate improvement, red indicate decline, black points indicate no change, and arrow length indicates the magnitude of change.

Unlike the graph on offseason spending, there is a slightly stronger trend of clubs with higher numbers of academy signings showing improvement in rank. This suggests that developing players with the sole purpose of being on the first team may lead to more success than purchasing big name players used to playing in other systems. Although a few clubs with many academy promotions declined, the overall graph shows a benefit in academy signings.

[2][3][4] # Does FIFA Rating Help Predict Success?

FIFA 23 is a soccer video game made by EA Sports that contains real world players, teams, and leagues. Each player is assigned an *overall* rating, which is a score from 1 to 99 based on their general skill level. The rating is based on individual attributes such as: shooting, speed, tackling, and strength.

The following plot aims to determine if the rankings of the players in the video game, released before every season, can be used to predict a team's league performance.

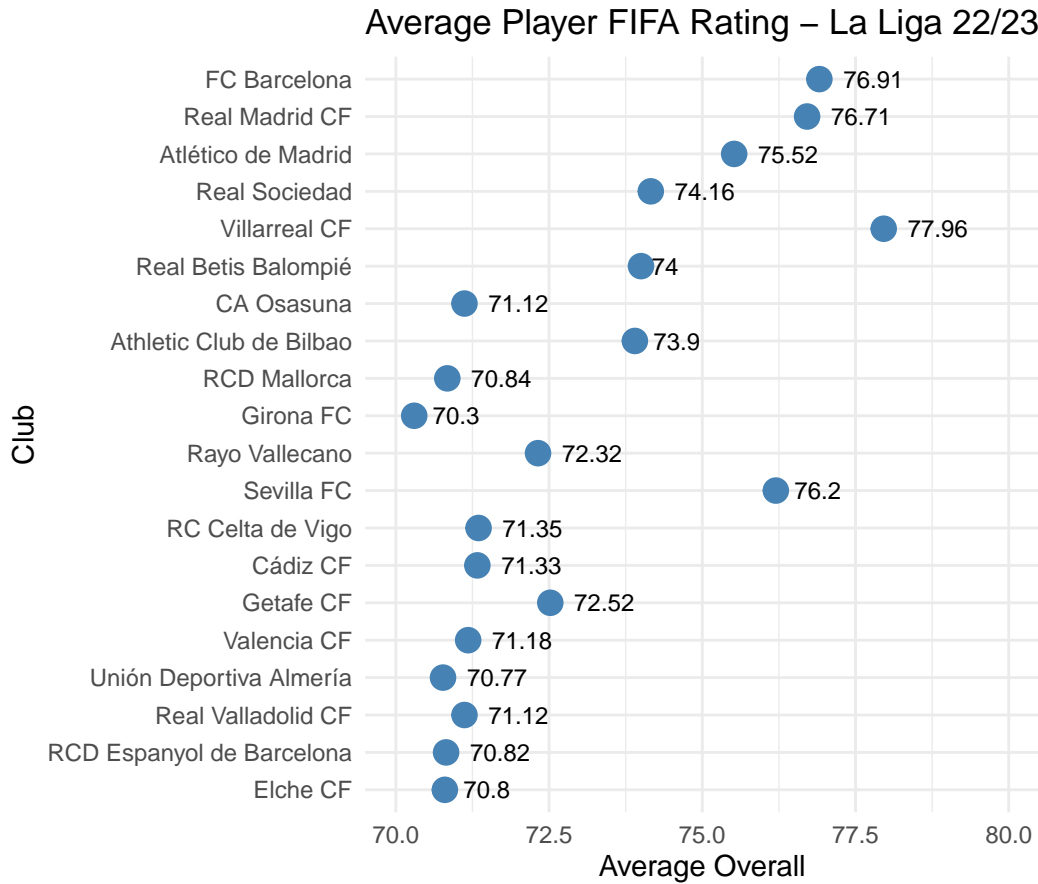


Figure 4: Average FIFA 23 player rating for each La Liga club, based on mean overall ratings of all rostered players. Clubs are ordered by their final rank in the 2022/23 season, with the top representing the highest ranked team and the bottom the lowest.

The graph exhibits a clear linear relationship between a club's average FIFA 23 player rating and its final league standing in the 2022/23 La Liga season. Clubs ranked higher in the league, such as Barcelona and Real Madrid, tend to have higher average player ratings. Conversely, teams with lower rank, such as Elche and Espanyol show lower average ratings.

This positive correlation suggests that FIFA 23 player ratings can reflect the quality of real teams, despite being part of a video game.

[1][4]

## Conclusion

This report examined key factors that can have influence on team success in La Liga, including total player market value, offseason transfer spending, academy player promotions, and FIFA 23 player ratings. The graphs showed a strong linear relationship between both total squad value and average player ratings with a team's final league ranking. Clubs like Barcelona and Real Madrid that ranked highest in both player value and average ratings, also finished at the top of the league. This suggests that roster quality is key in team success. This also shows that FIFA's video game ratings may reasonably reflect team strength.

On the other hand, offseason transfer spending alone did not have as strong a correlation with improved performance as academy spending. Some clubs spent large amounts and declined in rank, while others, with lesser investments, improved. In contrast, clubs that promoted more players from their academies showed a pattern of improvement, suggesting that internal development of players can be both cost effective, and can lead to future success. Overall, the visuals indicate that, while financial value and player quality are important to success in La Liga, strategies like youth promotion can provide a meaningful competitive edge.

## FAIR Principles

- **Findable:** Data is all accessible by either public url or csv file
- **Interoperable:** Data is stored as csv files or html tables

## Works Cited

1. cashncarry. *FIFA 23 Complete Player Dataset*. Kaggle, 2023.  
<https://www.kaggle.com/datasets/cashncarry/fifa-23-complete-player-dataset>. Accessed 15 June 2023.
2. Rayhan32. *La Liga Player Status Season 2022-2023*. Kaggle, 2023.  
<https://www.kaggle.com/datasets/rayhan32/la-liga-player-status-season-2022-2023>. Accessed 15 June 2023.
3. ESPN. *La Liga Standings - 2021/22 Season*. 2022.  
[https://www.espn.com/soccer/standings/\\_/league/ESP.1/season/2021](https://www.espn.com/soccer/standings/_/league/ESP.1/season/2021). Accessed 15 June 2023.
4. ESPN. *La Liga Standings - 2022/23 Season*. 2023.  
[https://www.espn.com/soccer/standings/\\_/league/ESP.1/season/2022](https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022). Accessed 15 June 2023.

5. Transfermarkt. *LaLiga Market Values as of June 15, 2023*. 2023.  
<https://www.transfermarkt.com/laliga/marktwerteverein/wettbewerb/ES1/plus/?stichtag=2023-06-15>. Accessed 15 June 2023.
6. Transfermarkt. *LaLiga2 Market Values as of July 15, 2023*. 2023.  
<https://www.transfermarkt.com/laliga2/marktwerteverein/wettbewerb/ES2/plus/?stichtag=2023-07-15>. Accessed 15 June 2023.
7. Transfermarkt. *LaLiga Spending and Earnings - 2021/22 Season*. 2022.  
[https://www.transfermarkt.com/laliga/einnahmenausgaben/wettbewerb/ES1/ids/a/sa/saison\\_id/2021/saison\\_id\\_bis/2022/nat/0/pos//w\\_s//intern/0](https://www.transfermarkt.com/laliga/einnahmenausgaben/wettbewerb/ES1/ids/a/sa/saison_id/2021/saison_id_bis/2022/nat/0/pos//w_s//intern/0). Accessed 15 June 2023.
8. Transfermarkt. *LaLiga2 Spending and Earnings - 2021/22 Season*. 2022.  
[https://www.transfermarkt.us/laliga2/einnahmenausgaben/wettbewerb/ES2/plus/0?ids=a&sa=&saison\\_id=2021&saison\\_id\\_bis=2022](https://www.transfermarkt.us/laliga2/einnahmenausgaben/wettbewerb/ES2/plus/0?ids=a&sa=&saison_id=2021&saison_id_bis=2022). Accessed 15 June 2023.

## Code Appendix

### League Standings

```
#Tidyverse Style Guide
```

```
library(knitr)
library(kableExtra)
library(readr)
library(rvest)
```

```
#import data
url <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022"
standings_2023 <- read_html(url) %>%
  html_table() %>%
  as.data.frame()
```

```
# La Liga team names used to standardize across datasets
```

```
la_liga_2023<- c(
  "FC Barcelona", "Real Madrid CF", "Atlético de Madrid", "Real Sociedad",
  "Villarreal CF", "Real Betis Balompie", "CA Osasuna", "Athletic Club de Bilbao",
  "RCD Mallorca", "Girona FC", "Rayo Vallecano", "Sevilla FC", "RC Celta de Vigo",
  "Cádiz CF", "Getafe CF", "Valencia CF", "Unión Deportiva Almería", "Real Valladolid CF", "I
)
```



```

standings_2023 <-
  standings_2023 %>%

  rename(Club = 1) %>%

  #add universal club names from datasets
  mutate(Club = la_liga_2023)%>%

  #add season end rank
  mutate(Rank = seq(1, nrow(standings_2023)))

standings_2023 <-
  standings_2023[c(10, 1, 2, 3, 4, 5, 6, 7, 8, 9)]

kable(standings_2023, caption = "Final LaLiga Standings for 2022-23 [1] ") %>%
  kable_styling(font_size = 8) %>%
  footnote(general =
    c("GP - Matches Played",
      "W - Wins",
      "D - Draws",
      "L - Losses",
      "F - Goals For",
      "A - Goals Against",
      "GD - Goal Differential",
      "P - Points"),
    general_title = "Legend:")

```

## Club and Player Value

```

library(dplyr)
library(tidyr)
library(readr)
library(rvest)
library(ggplot2)
library(scales)

##Player value

#Load Data
player_value_raw <- read_csv("players_fifa23.csv")

```

```

#La Liga teams list to filter for La Liga players
la_liga<- c(
  "FC Barcelona", "Real Madrid CF", "Atlético de Madrid", "Real Sociedad",
  "Villarreal CF", "Real Betis Balompié", "CA Osasuna", "Athletic Club de Bilbao",
  "RCD Mallorca", "Girona FC", "Rayo Vallecano", "Sevilla FC", "RC Celta de Vigo",
  "Cádiz CF", "Getafe CF", "Valencia CF", "Unión Deportiva Almería", "Real Valladolid CF", "I
)

#Filter for La Liga players
#Group by club
#Sum total player value

player_value <-
  player_value_raw %>%

  #select relevant columns
  dplyr::filter(Club %in% la_liga) %>%
  dplyr::select(Club, ValueEUR) %>%

  group_by(Club) %>%
  summarize(`Total Player Value (millions)` = sum(ValueEUR, na.rm = TRUE))%>%

  #Show number in number of millions for clarity
  dplyr::mutate(`Total Player Value (millions)` =
    round(`Total Player Value (millions)`/ 1000000, 1))

## Standings Data

#import data
url <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022"
standings_2023 <- read_html(url) %>%
  html_table() %>%
  as.data.frame()

#isolate each team's final position in the league
standings_2023 <-
  standings_2023 %>%
  dplyr::select(1) %>%
  rename(Club = 1) %>%

  mutate(Club = la_liga)%>%

```

```

mutate(Rank = seq(1, nrow(standings_2023)))

## Club value data

#LaLiga has promotion and relegation and this
#contains the 3 teams that were promoted at the end of the year in the top division dataset
#and the 3 teams that were demoted in the second division dataset
#we have to join the top division table with the second division to get all the teams
#that played in the 2022-2023 season

#load data for top division
url2 <- "https://www.transfermarkt.com/laliga/marktwerteverein/wettbewerb/ES1/plus/?stichtag="
page <- read_html(url2) %>%
  html_table()

#Isolate clubs and values for top division
club_value_l1 <-
  #select market value table with columns for club and their value
  page[[2]][-1, c(3,5)] %>%

  #parse number
  dplyr::mutate(across(2, parse_number)) %>%

  dplyr::rename(`Club Value (millions)` = League)

# load data for second division
url3 <- "https://www.transfermarkt.com/laliga2/marktwerteverein/wettbewerb/ES2/plus/?stichtag="
page <- read_html(url3) %>%
  html_table()

#Isolate clubs and value from second division
club_value_l2 <-
  #select market value table with columns for club and their value
  page[[2]][-1, c(3,5)] %>%

  #parse number
  dplyr::mutate(across(2, parse_number)) %>%

  dplyr::rename(`Club Value (millions)` = League)

#combine tables to join with standings
club_values <-

```

```

bind_rows(club_value_l1, club_value_l2)

#Clean data so all team names are equal for joining

standings_names <- standings_2023[[1]]
player_val_names <- player_value[[1]]
club_values_names <- club_values[[1]]

#remove all prefixes and other strings that are unique to each dataset
prefix <- c("Unión Deportiva", "FC", "CF", "RCD", "RC", "de", "Club", "CA", " ", "UD", "CD",

for (term in prefix){
  standings_names <- str_replace_all(standings_names, term, "")
  player_val_names <- str_replace_all(player_val_names, term, "")
  club_values_names <- str_replace_all(club_values_names, term, "")
}

#replace dataset club names with univiersal cclub names
standings_2023 <-
  standings_2023 %>%
  dplyr::mutate(Club = standings_names)

player_value <-
  player_value %>%
  dplyr::mutate(Club = player_val_names)

club_values <-
  club_values %>%
  dplyr::mutate(Club = club_values_names)

#join tables standings, club value, palyer value

value_insight <-
  standings_2023 %>%
  left_join(club_values, by = "Club") %>%
  left_join(player_value, by = "Club" ) %>%

  dplyr::mutate(Club = la_liga)

##Plot

```

```

ggplot(value_insight, aes(y= Rank,
                          x = `Total Player Value (millions)`,
                          size = `Club Value (millions)`)) + #different club value
                                                                #has different size

geom_point() +

labs(title = "La Liga 22-23: Total Player Values vs Standings") +

scale_x_continuous(n.breaks = 5)+

scale_y_reverse(n.breaks = 20, minor_breaks = NULL)+ #make rank be top to bottom

theme_minimal()

```

## Offseason Spending

```

### OFF SEASON SPENDING

library(readr)
library(rvest)

library(dplyr)
library(tidyr)
library(stringr)

library(ggplot2)
library(scales)

##STANDINGS TABLES

#2021-2022 standings

#load data
url1 <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2021"
standings_2022 <- read_html(url1) %>%
  html_table() %>%
  as.data.frame()

#standard club names for the 2022 season

```

```

la_liga_2022 <- c("Real Madrid CF", "FC Barcelona", "Atlético de Madrid", "Sevilla FC",
                 "Real Betis Balompié", "Real Sociedad", "Villarreal CF", "Athletic Club de
                 "Valencia CF", "CA Osasuna", "RC Celta de Vigo", "Rayo Vallecano",
                 "Elche CF", "RCD Espanyol de Barcelona", "Getafe CF", "RCD Mallorca",
                 "Cádiz CF", "Granada CF", "Levante UD", "Deportivo Alavés"
                 )

standings_2022 <-
  standings_2022 %>%
  dplyr::select(1) %>%
  rename(Club = 1) %>%

  #add universal club names from datasets
  mutate(Club = la_liga_2022) %>%

  #add season ending rank
  mutate(`Rank 2022` = seq(1, nrow(standings_2022)))

#2022-2023 standings

#load data
url2 <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022"
standings_2023 <- read_html(url2) %>%
  html_table() %>%
  as.data.frame()

la_liga_2023 <- c(
  "FC Barcelona", "Real Madrid CF", "Atlético de Madrid", "Real Sociedad",
  "Villarreal CF", "Real Betis Balompié", "CA Osasuna", "Athletic Club de Bilbao",
  "RCD Mallorca", "Girona FC", "Rayo Vallecano", "Sevilla FC", "RC Celta de Vigo",
  "Cádiz CF", "Getafe CF", "Valencia CF", "Unión Deportiva Almería", "Real Valladolid CF",
  "RCD Espanyol de Barcelona", "Elche CF"
)

standings_2023 <-
  standings_2023 %>%
  dplyr::select(1) %>%
  rename(Club = 1) %>%

  #add universal club names from datasets
  mutate(Club = la_liga_2023) %>%

```

```

#add season ending rank
mutate(`Rank 2023` = seq(1, nrow(standings_2023)))

#Joint standings

#due to promotion and relegation, the same teams are not in
#the top division of Spanish soccer every year. We want to only see
#the teams that were in LaLiga for both years

joint_standings <-
  standings_2023 %>%
  inner_join(standings_2022, by = "Club")

##OFFSEASON SPENDING

#import data
url3 <- "https://www.transfermarkt.com/laliga/einnahmenausgaben/wettbewerb/ES1/ids/a/sa//sais"
page1 <- read_html(url3) %>%
  html_table()
liga1_expenditures<- page1[[2]]

#clean column names (n/a columns were included)
variables <- unique(names(liga1_expenditures))
liga1_expenditures <- liga1_expenditures[-c(1,2,9)]
names(liga1_expenditures) <- variables[-1]

#select club and offseason expenditures
liga1_expenditures <-
  liga1_expenditures %>%
  dplyr::select(1,2)

## Expenditures Comparison

#Clean data so expenditures club names are the same as standings

standings_names <- joint_standings[[1]]
expenditures_names <- liga1_expenditures[[1]]
final_names <- joint_standings[[1]] #copy of our output club names

```

```

#remove all prefixes and other strings that are unique to each dataset
prefix <- c("FC", "CF", "RCD", "RC", "de", "Club", "CA", " ", "LA")

for (term in prefix){
  standings_names <- str_replace_all(standings_names, term, "")
  expenditures_names <- str_replace_all(expenditures_names, term, "")
}

#replace dataset club names with univiersal names
joint_standings <-
  joint_standings %>%
  dplyr::mutate(Club = standings_names)

liga1_expenditures <-
  liga1_expenditures %>%
  dplyr::mutate(Club = expenditures_names)

#join standings with expenditures, parsing for numbers
expenditure_rank <-
  joint_standings %>%
  left_join(liga1_expenditures, by = 'Club') %>%
  dplyr::mutate(Club = final_names) %>%

  dplyr::mutate(Expenditure = str_replace_all(Expenditure, '-', '0')) %>%
  rename(`Expenditure (millions)` = Expenditure) %>%
  dplyr::mutate(across(`Expenditure (millions)`, parse_number))

#split table into teams that stayed the same vs changed
#teams that stayed the same will use a point and not an arrow

rank_up <-
  expenditure_rank %>%
  dplyr::filter(`Rank 2023` < `Rank 2022`)

rank_down <-
  expenditure_rank %>%
  dplyr::filter(`Rank 2023` > `Rank 2022`)

rank_same <-
  expenditure_rank %>%
  dplyr::filter(`Rank 2022` == `Rank 2023`)

```



```

#Plot
ggplot()+

  geom_segment(data = rank_up, aes(
    x = `Expenditure (millions)`,
    xend = `Expenditure (millions)`,
    y = `Rank 2022`,
    yend = `Rank 2023`),

    arrow = arrow(length = unit(0.17, "cm")),
    color = "forestgreen"
  ) +

  geom_point(data = rank_up, aes(
    x = `Expenditure (millions)`,
    y = `Rank 2022`,
    color = "forestgreen",
    size = 0.8
  ) +

  geom_segment(data = rank_down, aes(
    x = `Expenditure (millions)`,
    xend = `Expenditure (millions)`,
    y = `Rank 2022`,
    yend = `Rank 2023`),

    arrow = arrow(length = unit(0.17, "cm")),
    color = "red"
  ) +

  geom_point(data = rank_down, aes(
    x = `Expenditure (millions)`,
    y = `Rank 2022`,
    color = "red",
    size = 0.8
  ) +

  geom_point(data = rank_same, aes(
    x = `Expenditure (millions)`,
    y = `Rank 2023`)) +

```

```

labs(title = "Impact of Offseason Spending on Performance",
     x = "Offseason Spending (Millions of Euros)",
     y = "Rank") +

scale_y_reverse(breaks = 1:20, minor_breaks= NULL)+ #make rank top to bottom

theme_minimal()

```

## Offseason Academy Promotion

```

library(readr)
library(rvest)

library(dplyr)
library(tidyr)
library(stringr)

library(ggplot2)
library(scales)

##STANDINGS TABLES

#2021-2022 standings

#load data
url1 <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2021"
standings_2022 <- read_html(url1) %>%
  html_table() %>%
  as.data.frame()

la_liga_2022 <- c("Real Madrid CF", "FC Barcelona", "Atlético de Madrid", "Sevilla FC",
                  "Real Betis Balompié", "Real Sociedad", "Villarreal CF", "Athletic Club de
                  "Valencia CF", "CA Osasuna", "RC Celta de Vigo", "Rayo Vallecano",
                  "Elche CF", "RCD Espanyol de Barcelona", "Getafe CF", "RCD Mallorca",
                  "Cádiz CF", "Granada CF", "Levante UD", "Deportivo Alavés"
)

standings_2022 <-
  standings_2022 %>%
  dplyr::select(1) %>%

```

```

rename(Club = 1) %>%

#add universal club names from datasets
mutate(Club = la_liga_2022) %>%

#add season ending rank
mutate(`Rank 2022` = seq(1, nrow(standings_2022)))

##2022-2023 standings

#load data
url2 <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022"
standings_2023 <- read_html(url2) %>%
  html_table() %>%
  as.data.frame()

la_liga_2023 <- c(
  "FC Barcelona", "Real Madrid CF", "Atlético de Madrid", "Real Sociedad",
  "Villarreal CF", "Real Betis Balompié", "CA Osasuna", "Athletic Club de Bilbao",
  "RCD Mallorca", "Girona FC", "Rayo Vallecano", "Sevilla FC", "RC Celta de Vigo",
  "Cádiz CF", "Getafe CF", "Valencia CF", "Unión Deportiva Almería", "Real Valladolid CF",
  "RCD Espanyol de Barcelona", "Elche CF"
)

standings_2023 <-
  standings_2023 %>%
  dplyr::select(1) %>%
  rename(Club = 1) %>%

#add universal club names from datasets
mutate(Club = la_liga_2023) %>%

#add season ending rank
mutate(`Rank 2023` = seq(1, nrow(standings_2023)))

##Joint standings

#due to promotion and relegation, the same teams are not in
#the top division of Spanish soccer every year. We want to only see
#the teams that were in LaLiga for both years

```

```

joint_standings <-
  standings_2023 %>%
  inner_join(standings_2022, by = "Club")

##Academy Transfer Data

#load data
academy_callups_raw <- read_csv("La_Liga_Academy_Promotion_2022_23.csv")

#group by club and count how many players called up
academy_callups <-
  academy_callups_raw %>%
  group_by(Club) %>%
  summarise(Count = n())

#Clean data so expenditures club names are the same as standings

standings_names <- joint_standings[[1]]
callups_names <- academy_callups[[1]]
final_names <- joint_standings[[1]] #copy of our output club names

#remove all prefixes and other strings that are unique to each dataset
prefix <- c("FC", "CF", "RCD", "RC", "de", "Club", "CA", " ", "LA")

for (term in prefix){
  standings_names <- str_replace_all(standings_names, term, "")
  callups_names <- str_replace_all(callups_names, term, "")
}

#replace dataset club names with univiersal names
joint_standings <-
  joint_standings %>%
  dplyr::mutate(Club = standings_names)

academy_callups <-
  academy_callups %>%
  dplyr::mutate(Club = callups_names)

#join standings with number of callups
callups_rank <-
  joint_standings %>%

```

```

left_join(academy_callups, by = 'Club') %>%
  dplyr::mutate(Club = final_names)

#split table into teams that stayed the same vs changed
#teams that stayed the same will use a point and not an arrow

rank_up <-
  callups_rank %>%
  dplyr::filter(`Rank 2023` < `Rank 2022`)

rank_down <-
  callups_rank %>%
  dplyr::filter(`Rank 2023` > `Rank 2022`)

rank_same <-
  callups_rank %>%
  dplyr::filter(`Rank 2022` == `Rank 2023`)

#Plot
ggplot()+

  geom_segment(data = rank_up, aes(
    x = Count,
    xend = Count + 0.25, #add angle to distinguish between teams with the same number
    y = `Rank 2022`,
    yend = `Rank 2023`),

    arrow = arrow(length = unit(0.25, "cm")),
    color = "forestgreen"
  ) +

  geom_point(data = rank_up, aes(
    x = Count,
    y = `Rank 2022`,
    color = "forestgreen",
    size = 0.95
  ) +

  geom_segment(data = rank_down, aes(
    x = Count,
    xend = Count - 0.25, #add angle to distinguish between teams with the same number

```

```

    y = `Rank 2022`,
    yend = `Rank 2023`),

    arrow = arrow(length = unit(0.25, "cm")),
    color = "red"
  ) +

  geom_point(data = rank_down, aes(
    x = Count,
    y = `Rank 2022`),
    color = "red",
    size = 0.95
  ) +

  geom_point(data = rank_same, aes(
    x = Count,
    y = `Rank 2023`)) +

  labs(title = "Impact of Academy Promotions on Performance",
    x = "Number of Academy Signings",
    y = "Rank") +

  scale_y_reverse(breaks = 1:20, minor_breaks= NULL)+

  scale_x_continuous(n.breaks = 10) +

  theme_minimal()

```

## Fifa Rating

```

##FIFA OVERALL VISUALIZATION

library(dplyr)
library(tidyr)
library(readr)
library(rvest)

##Team Overall

```

```

#Load Data
fifa_rating <- read_csv("players_fifa23.csv")

#Filter for LaLiga Players
la_liga<- c(
  "FC Barcelona", "Real Madrid CF", "Atlético de Madrid", "Real Sociedad",
  "Villarreal CF", "Real Betis Balompié", "CA Osasuna", "Athletic Club de Bilbao",
  "RCD Mallorca", "Girona FC", "Rayo Vallecano", "Sevilla FC", "RC Celta de Vigo",
  "Cádiz CF", "Getafe CF", "Valencia CF", "Unión Deportiva Almería", "Real Valladolid CF",
  "RCD Espanyol de Barcelona", "Elche CF"
)

fifa_rating <-
  fifa_rating %>%
  dplyr::filter(Club %in% la_liga)%>%
  dplyr::filter(!is.na(Overall))

#Calculate average player rating per club
club_avg_rating <-
  fifa_rating %>%
  group_by(Club) %>%
  summarize(Average_Overall = round(mean(Overall, na.rm = TRUE), 2)) %>%
  arrange(desc(Average_Overall))

##2022-2023 standings

#load data
url <- "https://www.espn.com/soccer/standings/_/league/ESP.1/season/2022"
standings_2023 <- read_html(url) %>%
  html_table() %>%
  as.data.frame()

standings_2023 <-
  standings_2023 %>%
  dplyr::select(1) %>%
  rename(Club = 1) %>%

#add universal club names from datasets
mutate(Club = la_liga) %>%

```

```

#add season ending rank
mutate(`Rank 2023` = seq(1, nrow(standings_2023)))

## Join overall with standings

rating_standings <-
  standings_2023 %>%
  left_join(club_avg_rating, by = "Club")

#Plot bar chart

ggplot(rating_standings, aes(x = Average_Overall, y = reorder(Club, -`Rank 2023`))) +

  geom_point(size = 4, color = "steelblue") +

  xlim(70,80)+

  labs(title = "Average Player FIFA Rating - La Liga 22/23", x = "Average Overall", y = "Club") +

  geom_text(aes(label = Average_Overall), hjust = -0.4, size = 3) +

  theme_minimal()

```