# promising_players

Tasnim Mahat - 2469411

2025-05-29

- This plot shows the improvement in overall rating over time with the most recent overall rating which reflects the current skill level
- For each point in the plot, it shows a player under the age of 23.
- The players like; Ilkay Gundogan, Eden Hazard, Paul Pogba, Richardo Rodriguez, Thibaut Courtois are in the top quadrant who are already highly ranked players and show that they improve rapidly.This players are up and coming players to book.
- The other players on the left quadrant like Mats Hummels and Marc-Andre ter Stegen show high rating but a negative improvement. It indicates that this players are performing well but dont have much further growth for long-term development.
- The other players at the bottom-right show moderate improvement in their rating and show future improvement
- The final players at the bottom left are not suitable for booking with negative growth.
- From this plot, the players at the top-right section are high priority to book because they show improvement are well as skilled and young.

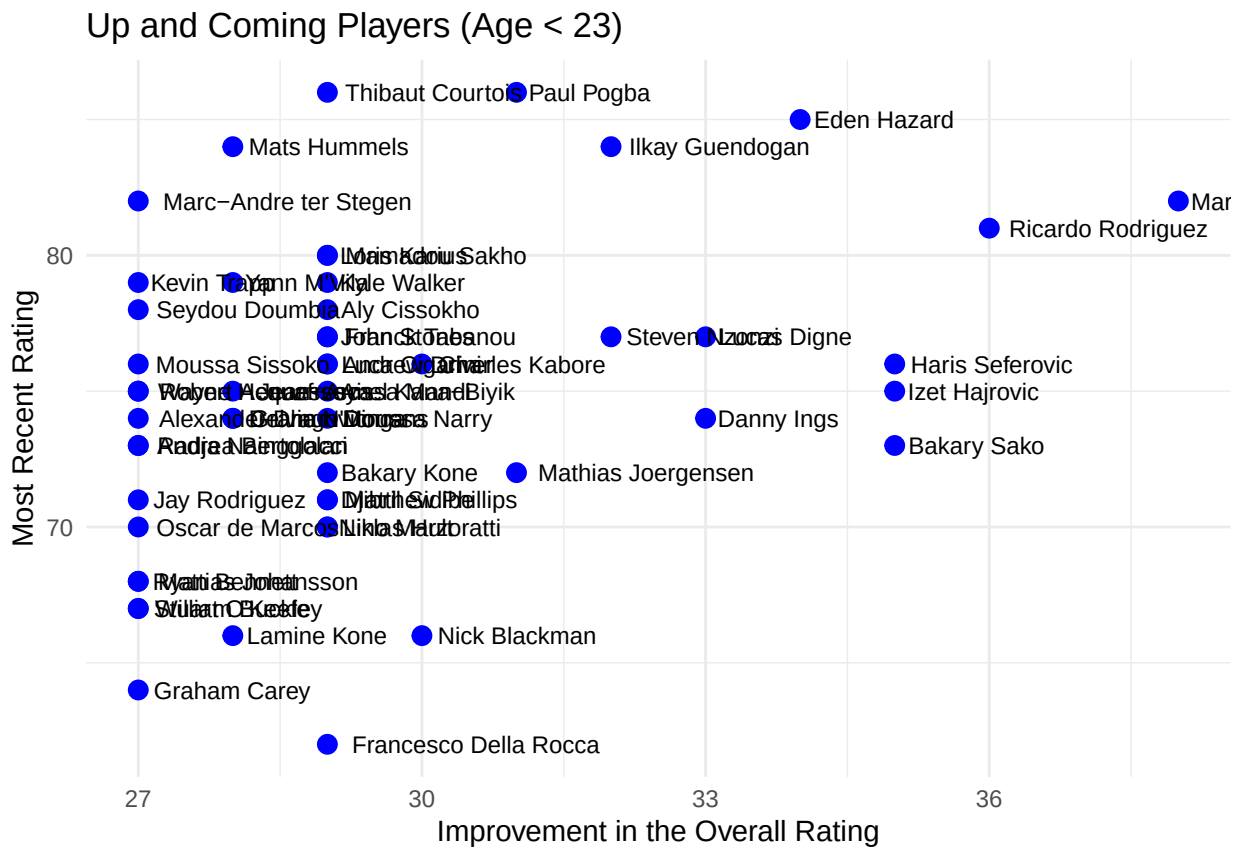
```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
player_att <- read.csv("~/VDS/VDS2425 Football/Player_Attributes.csv")
player <- read.csv("~/VDS/VDS2425 Football/Player.csv")

# Merging the datasets
df <- merge(player_att, player[, c("player_api_id", "player_name", "birthday")], by = "player_api_id")


df$date <- as.Date(df$date)
df$birthday <- as.Date(df$birthday)
df$age <- as.numeric(difftime(df$date, df$birthday, units = "days")) / 365

# players < 23
young_players <- df %>% filter(age < 23)


trend_df <- young_players %>%
  group_by(player_api_id) %>%
  filter(!is.na(overall_rating)) %>%
  arrange(date) %>%
  summarise(
    player_name = first(player_name),
    age_latest = max(age),
    rating_start = first(overall_rating),
    rating_end = last(overall_rating),
    rating_trend = rating_end - rating_start,
    recent_rating = last(overall_rating),
    n_ratings = n()
  ) %>%
  filter(n_ratings >= 3) %>%
  arrange(desc(rating_trend))

top_young <- trend_df %>% top_n(50, wt = rating_trend)

# Rating trend vs Current rating
ggplot(top_young, aes(x = rating_trend, y = recent_rating, label = player_name)) +
  geom_point(color = "blue", size = 3) +
  geom_text(size = 3, hjust = -0.1, vjust = 0.5) +
  labs(
    title = "Up and Coming Players (Age < 23)",
    x = "Improvement in the Overall Rating",
    y = "Most Recent Rating"
  ) +
  theme_minimal()
```




In this second plot, we can see that many players with th age 21-23 cluster around 75-80 rating which show maturity of the players. Some of the younger players between 18-20 show potential and high improvement and also a strong development prospective. To get the most promising players to book, the focus should be on the dark blue larger bubbles in the upper middle which indicates good growth potential

```
# dataframe
bubble_df <- young_players %>%
  filter(!is.na(overall_rating)) %>%
  arrange(player_api_id, date) %>%
  group_by(player_api_id) %>%
  summarise(
    player_name = first(player_name),
    age_latest = max(age),
    rating_start = first(overall_rating),
    rating_end = last(overall_rating),
    potential = last(potential),
    recent_rating = last(overall_rating),
    rating_trend = rating_end - rating_start,
    n = n()
  ) %>%
  filter(n >= 3, recent_rating > 65) %>%
  mutate(headroom = potential - recent_rating)

# Ploting the bubble chart
ggplot(bubble_df, aes(x = age_latest, y = recent_rating,
                      size = rating_trend, color = headroom)) +
```

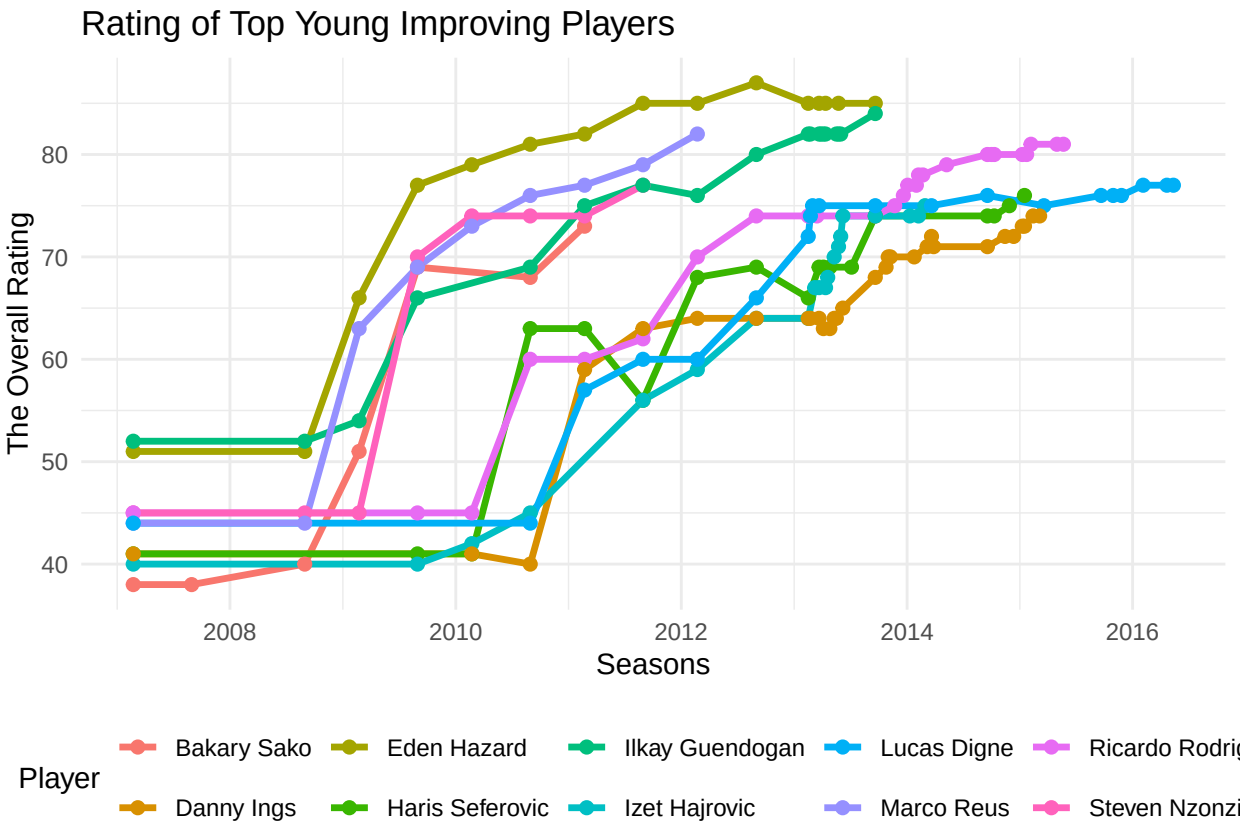
```
  geom_point(alpha = 1.5) +
  scale_color_gradient(low = "green", high = "blue") +
  scale_size(range = c(2, 10)) +
  labs(
    title = "Age vs Rating with Growth Potential",
    x = "Age",
    y = "Current Rating",
    size = "Improvement",
    color = "Potential Gap"
  ) +
  theme_minimal()
```


In this next plot, it shows players like Marco Reus and Eden Hazard indicate the highest and fastest growth reaching above 85. All the other players shown in this plot have a significantly high improvement in their ratings from 2008-2016. Most the other players between 2012 and 2014 peaked which show an early career breakthroughs The players like Izet Hajrovic and Danny Ings show gradual improvement but still remain below 80. From this trends, we can confirm that the value as the high potential bookings are based on consistent upward development

```
# Filtering young players vs ratings
young_players <- df %>%
  filter(age < 23, !is.na(overall_rating))

# top improvers
top_improv <- young_players %>%
  arrange(player_api_id, date) %>%
```

```
  group_by(player_api_id) %>%
  summarise(
    player_name = first(player_name),
    rating_start = first(overall_rating),
    rating_end = last(overall_rating),
    growth = rating_end - rating_start,
    count = n()
  ) %>%
  filter(count >= 4, growth > 5) %>%
  slice_max(growth, n = 10)


selected_players <- young_players %>%
  filter(player_api_id %in% top_improv$player_api_id)

# slope chart
ggplot(selected_players, aes(x = date, y = overall_rating, group = player_name, color = player_name)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(
    title = "Rating of Top Young Improving Players",
    x = "Seasons",
    y = "The Overall Rating",
    color = "Player"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```

```
## 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.
```


Rating of Top Young Improving Players
The Overall Rating
80
70
60
50
40
2008
2010
2012
2014
2016
Seasons
Player
Bakary Sako
Eden Hazard
Ilkay Guendogan
Lucas Digne
Ricardo Rodri
Danny Ings
Haris Seferovic
Izet Hajrovic
Marco Reus
Steven Nzonzi