Brian Delgado
12/1/25
FINC306 and FINC403
ProjectReport

Goal Scoring Careers of Four Futbol Players

1. Introduction

Professional soccer clubs invest millions of dollars in transfers and salaries for attacking players. These decisions depend heavily on expectations about future performance, including how many goals a player is likely to score, how long they will remain at their peak, and how quickly they will decline. Because of this, it is useful to study a player's "career curve" in the same way we might study the life cycle of a financial asset.

In this project, I analyze the goal-scoring careers of four elite forwards: Lionel Messi, Cristiano Ronaldo, Neymar, and Robert Lewandowski. For each player, I utilize historical goal data to construct a time series of goals per year, fit a quadratic curve to approximate their career trajectory, and employ derivatives to analyze how their performance evolves over time. The main questions are:

- When does each player reach peak goal-scoring form?
- How quickly does their performance improve before the peak and decline afterward?
- How do the four careers compare to each other in terms of timing and duration of peak performance?

The results are summarized with an R script, an Excel file, and a Tableau dashboard. Together, these tools connect the math to a real sports context and provide a visual way to compare all four players.

2. Data and Methods

The data for this project comes from Kaggle goal-by-goal logs for each player. Each CSV file contains one row per goal with information such as date, tournament, club, opponent, match result, and the minute of the goal. I used R to read four separate files (messi.csv, ronaldo.csv, neymar.csv, and lewandowski.csv) and combine them into a single data set.

Because the goal logs are at the individual-goal level, the first step was to aggregate the data into goals per calendar year for each player. In R, I converted the date strings into real dates, extracted the year, and then counted how many goals each player scored in each year. I also created a time variable t that measures the number of years since the player's first recorded goal. For example, t = 0 corresponds to the first season in the data set, t = 1 to the next season, and so on.

To model the career curve, I fit a quadratic function of the form

$G(t) = a * t^2 + b * t + c$

where G(t) is the number of goals scored in year t, and a, b, and c are constants estimated from the data. I fit a separate quadratic model for each player using least-squares regression in R. The first derivative of this function is

$G'(t) = 2 * a * t + b$

and represents the approximate rate of change in goals per year at time t. When G'(t) is positive and large, the player is improving quickly. When G'(t) is close to zero, performance is relatively stable. When G'(t) is negative, performance is declining.

Using the fitted model for each player, I computed:

- Fitted goals G(t) for each year t
- The slope G'(t) for each year t
- The year with the maximum fitted goals, which I interpret as the estimated peak scoring season

In addition to the fitted curve, I also computed a discrete year-to-year change in goals using

$DeltaGoals(t) = G(t + 1) – G(t).$

Positive values of DeltaGoals(t) indicate improvement compared to the previous year, while negative values indicate a decline.

All of these steps are implemented in the R script players_code.R. The script reads the four CSV files, aggregates goals by year, creates the time variable t, fits the quadratic models, and exports the combined data to a file called players_goals_per_year.csv. This file is then used in Excel and

Tableau to create graphs and dashboards. The R script, the Excel file, and the Tableau packaged workbook are all linked from the project website.

3. Results

The quadratic models fitted to goals per year produced reasonable “career curves” for all four forwards. For each player, the fitted function G(t) captures the typical pattern of rapid growth in the early seasons, followed by a peak period with very high goal totals, and then a gradual decline. After converting the time variable t back into calendar years, I obtained an estimate of the peak scoring year and the corresponding predicted number of goals for each player.

The model suggests that Lionel Messi reached his maximum annual scoring output around 2015, with an estimated peak of about 51.4 goals in that season. Cristiano Ronaldo also peaked in 2015, with a very similar estimated maximum of about 48.5 goals. In the fitted curves for both players, the slope G'(t) is strongly positive during their early and mid-career years, indicating rapid improvement, and then becomes negative after 2015, indicating declining annual goal totals as they age and move between clubs and leagues.

For Neymar, the model identifies an earlier and lower peak. His fitted curve reaches a maximum around 2013, with an estimated 23.1 goals. After this point, the derivative G'(t) becomes smaller and eventually negative, consistent with a combination of injuries, role changes, and fewer games in some seasons. Compared with Messi and Ronaldo, Neymar’s peak is both earlier in time and lower in terms of total goals per year.

Robert Lewandowski shows a different pattern. His estimated peak occurs later in his career, around 2019, with approximately 41.7 goals. His career curve remains relatively high and flat for several years before 2019, suggesting a long period of sustained elite performance rather than a single sharp spike. In other words, Lewandowski appears to have a longer “plateau” of high goal scoring, while Messi and Ronaldo have slightly sharper peaks around 2015 at very high levels.

Overall, the fitted models and their derivatives show that all four players experience strong early-career growth, but their peak timing and peak height differ:

- Messi: peak around 2015, about 51.4 goals
- Ronaldo: peak around 2015, about 48.5 goals
- Neymar: peak around 2013, about 23.1 goals
- Lewandowski: peak around 2019, about 41.7 goals

These differences are visible in both the raw goals-per-year data and in the smoothed career curves. The Tableau dashboard created for this project displays:

- A line chart of goals per year by player, with time on the horizontal axis and goals on the vertical axis, and player as the color.

- A histogram of DeltaGoals (year-to-year change in goals) to show how often each player has big jumps or drops from one season to the next.

An interactive Player filter on the dashboard allows the user to focus on one player at a time or compare multiple players.

4. Conclusion

This project used goal-by-goal data for Lionel Messi, Cristiano Ronaldo, Neymar, and Robert Lewandowski to build yearly scoring profiles and fit quadratic models to their careers. By treating time as years since the player's first professional goal and analyzing the derivative of the fitted curve, I identified approximate peak scoring years and periods of improvement and decline for each player. The results indicate that Messi and Ronaldo reached similar, extremely high peaks in 2015, Neymar's best year occurred earlier in 2013 with a more modest maximum, and Lewandowski's peak arrived later in 2019 with a strong but slightly lower goal total than Messi and Ronaldo.

From a financial and decision-making perspective, these career curves illustrate why clubs and agents care about the timing of peak performance. A player who peaks later and maintains a long stretch of high output, like Lewandowski, may justify longer contracts and higher wages over an extended period. A player who peaks earlier and then faces injuries or role changes, as in Neymar's case, may be viewed as a riskier investment. Messi and Ronaldo's overlapping peak around 2015 shows how rare it is for two players to sustain such extreme levels of scoring for many consecutive seasons.

The analysis also has limitations. The models do not adjust for minutes played, strength of the league, quality of teammates, expected goals (xG), or non-scoring contributions such as assists and defensive work. The quadratic model is relatively simple and may not capture all the nuances of a player's career. However, despite these limitations, the project demonstrates how basic calculus tools—such as curve fitting and derivatives—can be applied to real sports data to quantify aging curves and support more informed discussions about player value and performance over time.

5. References

Kaggle: Goal-scoring data sets for Lionel Messi, Cristiano Ronaldo, Neymar, and Robert Lewandowski.
https://www.transfermarkt.com/