

SET TIMER

INTRO

As well as soon becoming a fully-fledged data nerd, I'm also a self-confessed football nerd. But even if you're not a football nerd like me, you almost definitely know Lionel Messi.

SPACE

LIONEL MESSI

When Lionel Messi recently signalled his intention to leave Barcelona, many football fans found themselves in a state of shock and sorrow. It didn't just feel like the end of AN era, but the end of THE era. The threat of his departure, and the approaching final act of his career, is what prompted me to begin the enormous work of analysing his legacy.

SPACE

HONOURS/RECORDS

Not only is Messi widely regarded as the greatest player of all-time, some of the Barcelona teams he's been a part of could easily be considered some of the best club sides of all-time – as can be illustrated by the number of trophies won during the Messi era.

Messi has been playing for the Barcelona since 2004/05 season, but for the purpose of this analysis I only looked at data from the 2005/06 season onwards. I also discounted data for last season (2019/20), as the season was obviously heavily impacted by COVID-19.

TABLEAU MAP

INFLUENTIAL TEAMMATES & WELCOMING OPPONENTS

First off, I wanted to look at whether the team-mates Messi played with, or the opposition he faced, had much of an influence on his output. I couldn't find much of a correlation between any player or opposition, and that's possibly because Messi has been so consistent throughout his career, scoring with anyone and against anyone...

Well, almost everyone...of the 41 clubs Messi has played against in his La Liga career, there are only 3 that he hasn't scored against:

(SE) Real Murcia (currently in 3rd tier, only played once in 2007/08, 5-3 win, 2 assists)

(S) Xerez (currently in 4th tier, in La Liga 2009/10, was sub both games, 1 assist)

(S) Cadiz (only played once in 2005/06, 1 assist, Cadiz now back in La Liga)

BACK TO GOOGLE SLIDES

SPACE

GOAL CLASSIFICATION

Using supervised machine learning, I looked at whether it was possible to predict whether a shot would result in a goal or not.

I looked at all shots taken by Barcelona players in the Messi era (7655 shots in total), basing the model on the following 6 variables...

- Season
- Play Pattern – regular play, from corner/free kick/throw in/counter/goal kick/keeper/kick-off
- Player
- Shot type – open play, corner, free kick, penalty, kick-off
- Body Part – right foot, left foot, head, other
- Shot technique – backheel, diving header, half volley, lob, normal, overhead kick, volley

TABLEAU GOAL VISUALISATION

I achieved an accuracy score of 0.87 using the logistic regression model – we can predict whether a shot will result in a goal or not with 87% certainty.

BACK TO GOOGLE SLIDES

SPACE

MESSI'S OUTPUT BY SEASON 1

There are obviously a number of ways in which output can be measured for a footballer, but I'll focus on the simplest measures – goals and assists (the final pass leading up to a goal).

SPACE - GRAPH

Here we see Messi's goals (bars) and assists (line) throughout his career. To make things even easier, I combined these into what's known as 'goal involvements'.

SPACE - GRAPH

I could've calculated a simple linear regression, but this would've produced an unrealistic, constantly rising line.

There appears to be a clear peak at the 2011/12 season (2012 = 91 goals) so I decided to look at plotting two separate lines of regression – one of the upward trend towards 2011/12, and one for the downward trend from 2011/12 onwards.

SPACE

MESSI'S OUTPUT BY SEASON 2

GRAPH

The upward trend line fits nicely and has an r-squared value of 0.985.

SPACE - GRAPH

But the downward trend was harder to plot a straight line for, with an r-squared value of 0.253.

The dips appear to be linked with injuries (79 days/12 days in 2013/14, 62 days/13 games in 2015/16), so looked instead at 'returns per game' to improve accuracy, but this didn't work – in fact, the r-squared for both plots decreased slightly. Therefore I decided to look at quadratic regression instead.

SPACE

MESSI'S OUTPUT BY SEASON 3

GRAPH

Overall this was a much better fit, with an adjusted r-squared value of 0.743.

SPACE

PREDICTING MESSI'S DECLINE

Using our quadratic linear regression model, I wanted to predict Messi's output as he approaches retirement.

Changed the x-axis from 'seasons' to 'age' (Messi's birthday is 24 June 1987, which is conveniently between seasons, so I went with his age at the start of the season).

SPACE - GREEN LINE

The green line shows the predicted goal involvements for an additional 5 years, with the model giving us the following predicted numbers...

SPACE - PREDICTED RESULTS

Messi's actual goal involvements last season in 2019/20 was 44 – so it's clear that Messi isn't declining as quick as the model suggests.

Further proof that Messi is no mere footballer, more an artist that has transcended his sport for a whole generation and could maybe continue to do so for a few years longer.

SPACE

QUESTIONS?

Compilation - <https://youtu.be/PSanJ5swYBM?t=348>

Athletic - <https://www.youtube.com/watch?v=YBaXSEkzAEE>

Betis - <https://www.youtube.com/watch?v=zq0Az3HTmLU>

<https://giphy.com/gifs/soccer-goat-messi10-JQ4yLSltNc5jGfSPDZ>

<https://giphy.com/gifs/fcbarcelona-barcelona-xUA7aT4k7JLXH71zG0>

INTRO MUSIC - <https://www.youtube.com/watch?v=Y1fiOJDXA-E&feature=youtu.be&t=187>

Google Drive data - <https://drive.google.com/drive/u/0/folders/1g76yuSsOAeZOpyonhNRyGPxocogRZkE3>