

Ronaldo's Return: A Red Herring For The Red Devils?



Presented by:

Samarth Sajeesh, Aadi Krishnan, Alec Zhang

Eric Lu, Fisher Zhao

Background



2020/2021 Premier League

Club	Points
Manchester City	86
Manchester Utd	74
Liverpool	69
Chelsea	67



2021/2022 Premier League

Club	Points
Manchester City	93
Liverpool	92
Chelsea	74
Tottenham	71
Arsenal	69
Manchester Utd	58

2021/2022 PL Player Stats

Player	Goals
1  Salah	23
2  Son	23
3  Ronaldo	18
4  Kane	17

2020/2021 Serie A Player Stats*

Player	Goals
1  Ronaldo	29
2  Lukaku	24
3  Muriel	22
4  Vlahovic	21

*Covid season





Null Hypothesis

Manchester United's changes in performance during the 2021-2022 period were not affected by Cristiano Ronaldo's involvement.



Research Hypothesis

Cristiano Ronaldo's presence made a statistically significant impact on Manchester United's performances.

Goals and Understanding

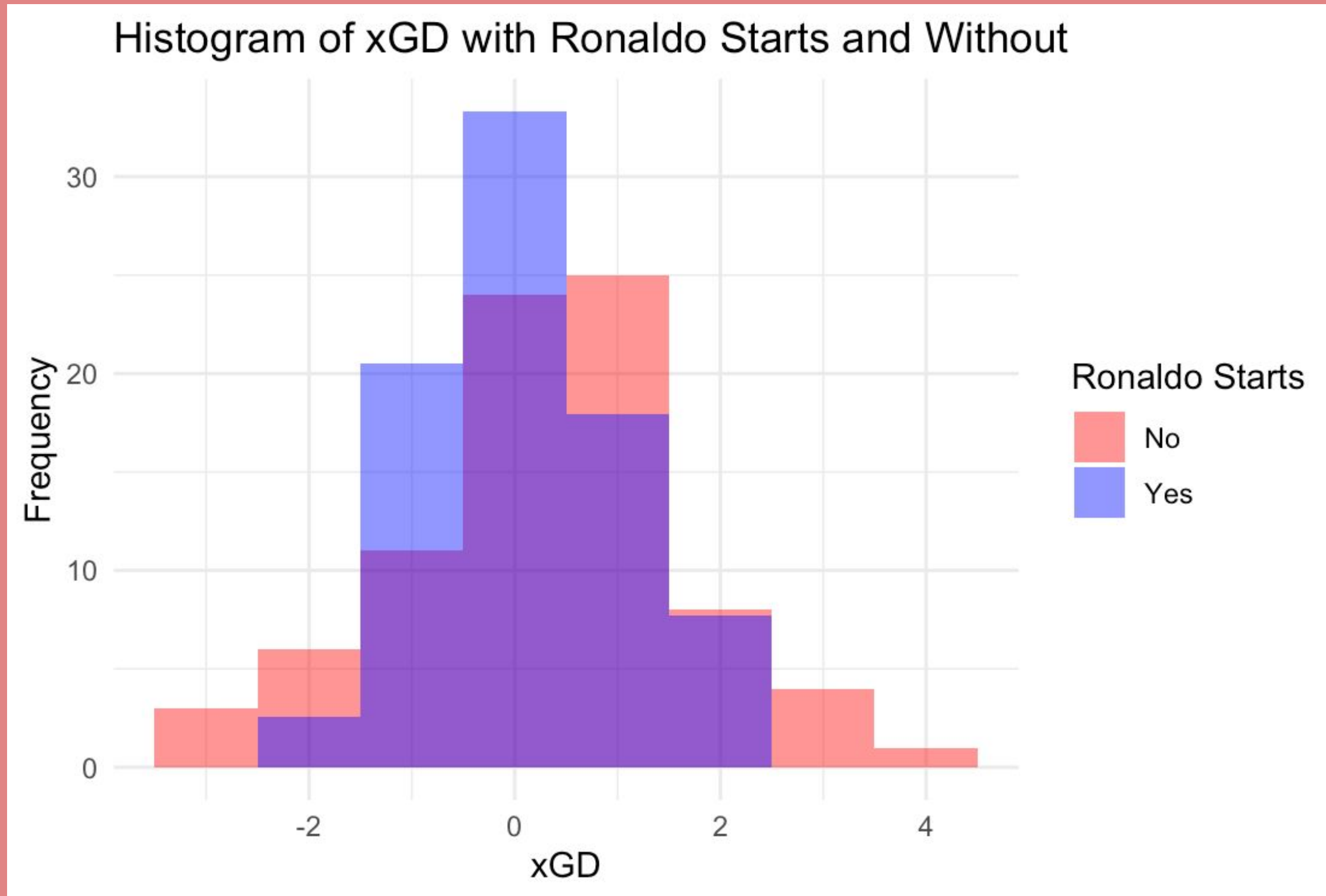
1. Identify Unique Factors that Ronaldo contributed to
2. Create models to show Ronaldo's impact
3. Work through any Confounding Factors



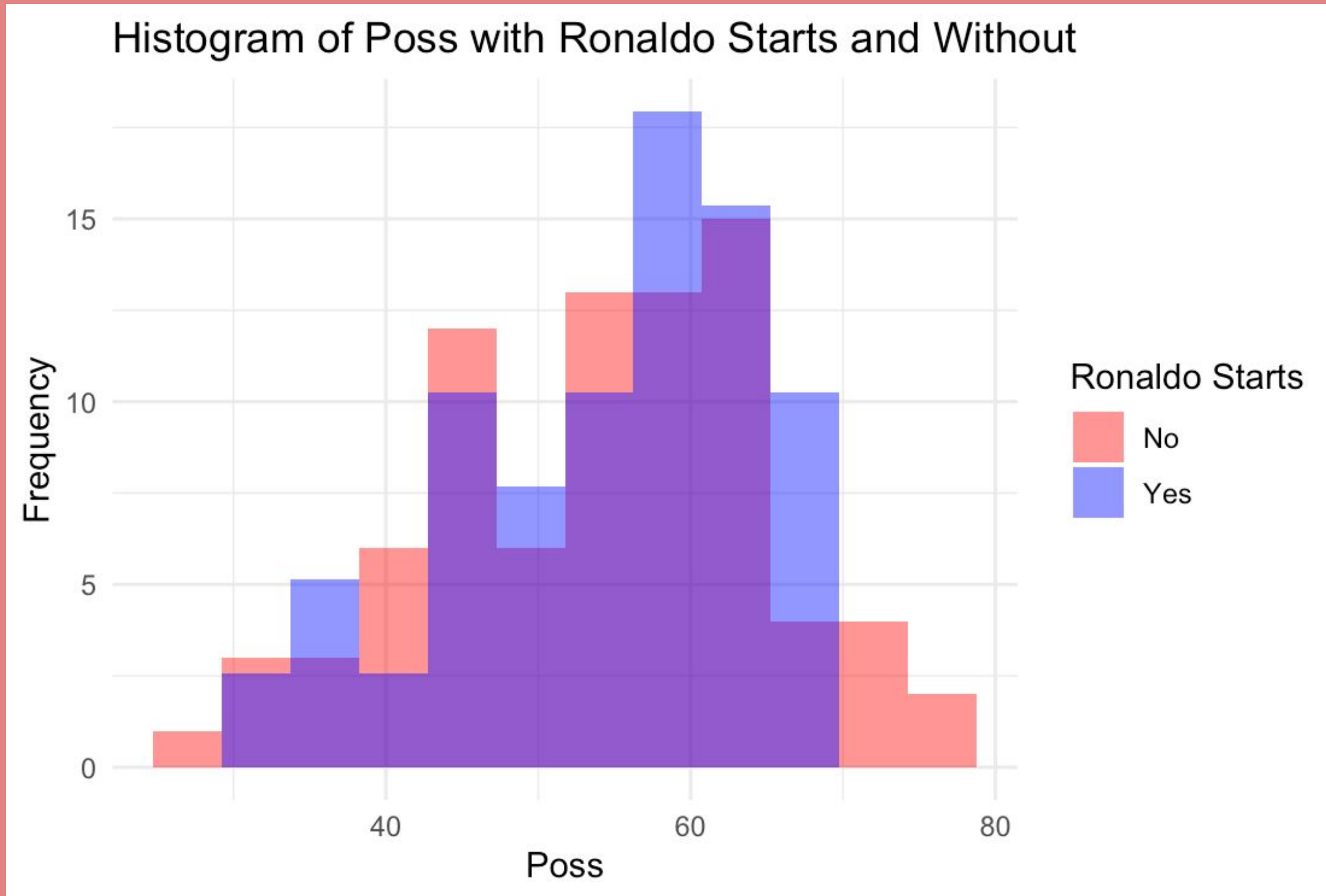
Glossary

- 1. “Expected” - The likelihood of an action in a game**
- 2. Expected Goal(xG): The likelihood of a shot converting to a goal based on distance, angle, etc.**
- 3. Goal Difference(GD):
Goals for the home team-Goals for the away team**

How was Man United's xGD affected by Ronaldo



How did Ronaldo affect United's possession



Pythag Model - The Luck Factor

Predicted points =
$$\frac{GF^{1.2}}{GF^{1.2} + GA^{1.2}}$$

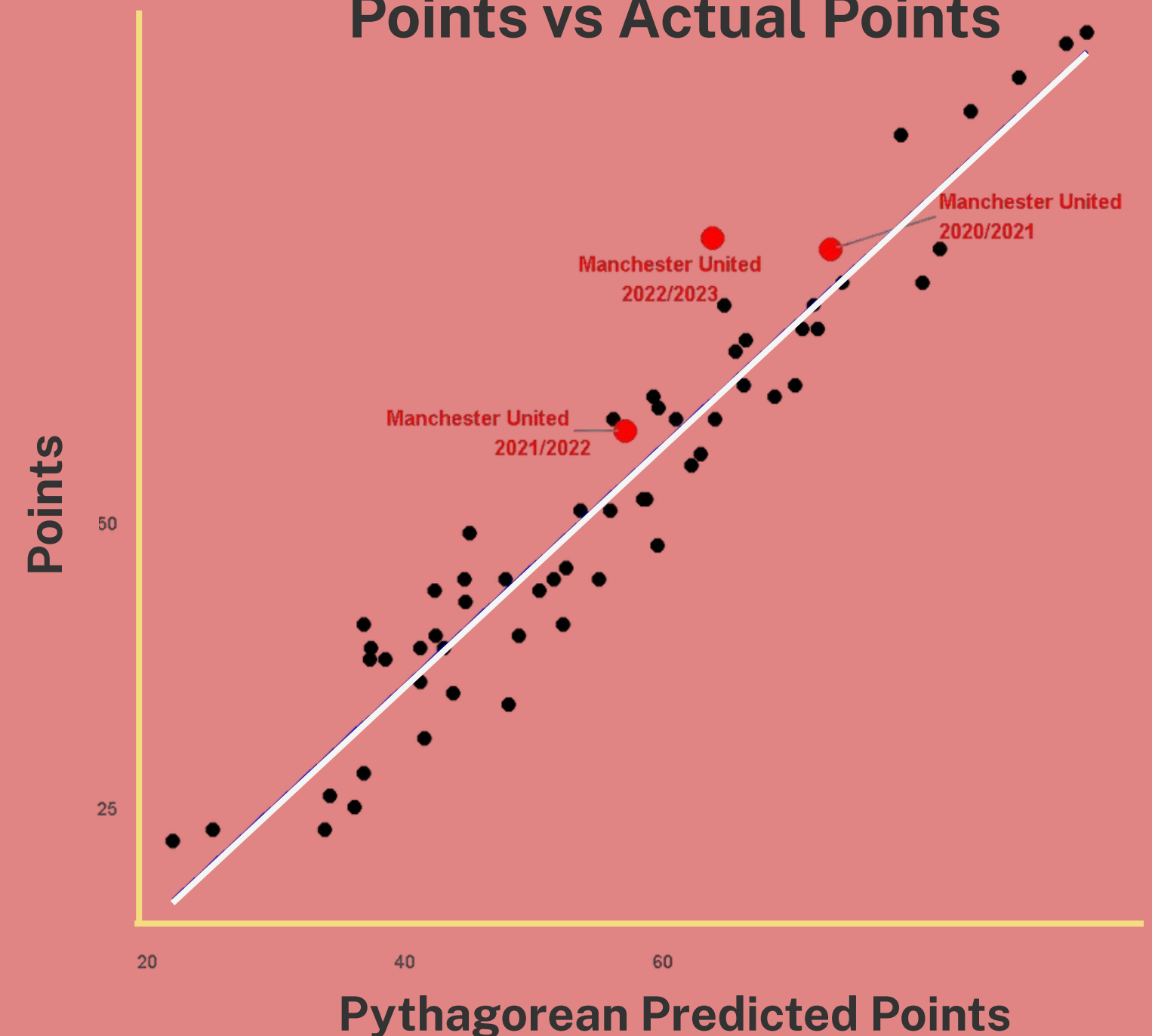
Flaws

- Variation increases a lot with fewer games
- Stats bomb RMSE is 2.74 due to outliers in data

r^2 : 0.92

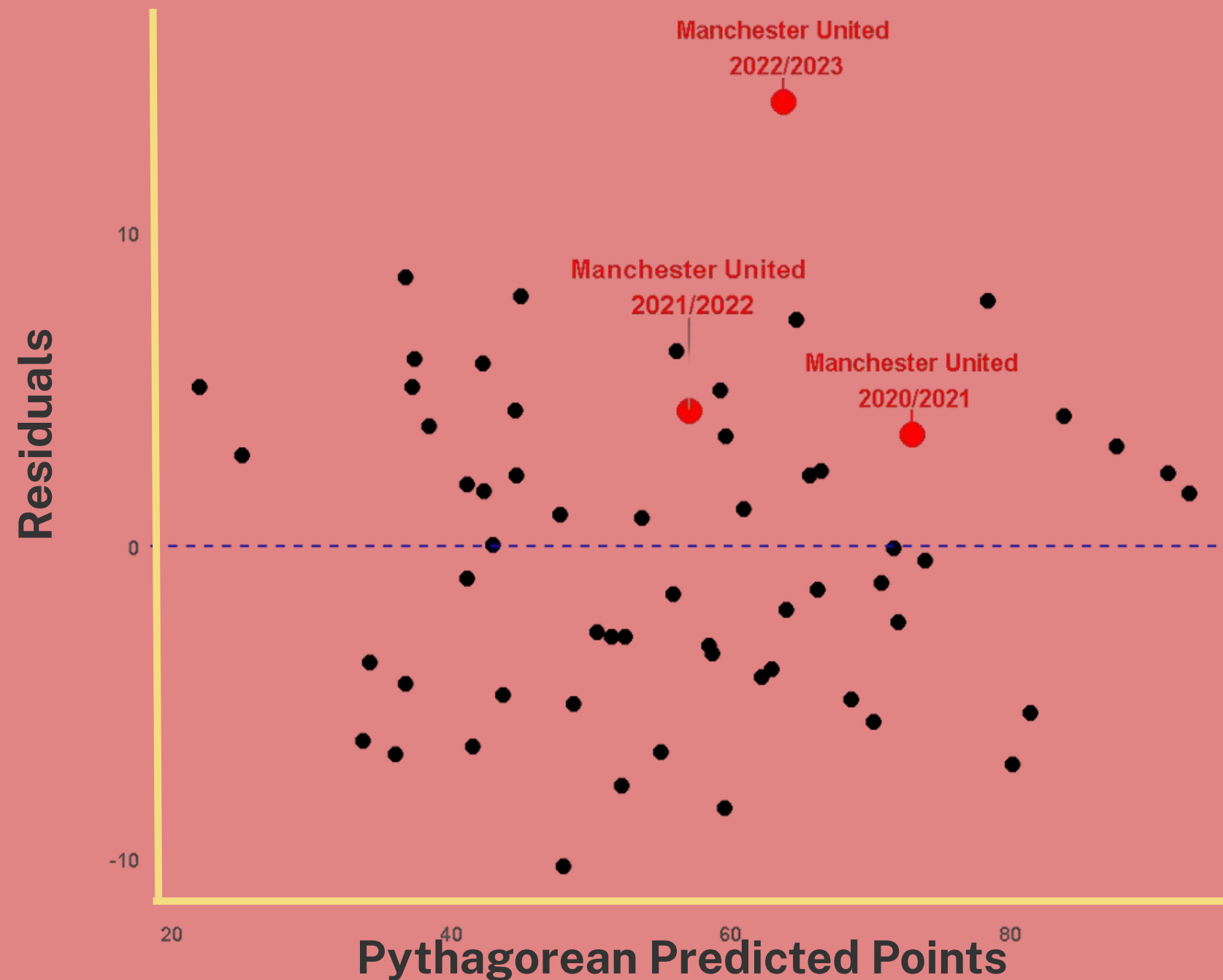
RMSE: 3.47pts

Pythag Predicted Points vs Actual Points



Residuals - The Luck Factor

Residuals for
Pythagorean Values



$$Z\text{-Score} = \frac{X - \mu}{\sigma}$$

- X is residual ($Y_i - Y_p$)

20/21: 0.7128868

21/22: 0.8598177

22/23: 2.8208009

The Ronaldo Coefficient[®]

Parametric
=

$$RC \text{ (Ronaldo)} + \alpha \text{ (H/A)} + \beta \text{ (TPI)}$$

What we are controlling for:

Team strength

Home/Away*

Parametrics tracking Ronaldo's impact:

Possession

xGD

Ronaldo Dummy variable

1 - Starting

0 - Not Starting

Confounding Factors

Team Strength

Home-Field Advantage

Relegated Teams



Controlling for Team Strength

$$\text{TPI} = \frac{\text{GF}^{1.2}}{\text{GF}^{1.2} + \text{GA}^{1.2}}$$

**As per the team's form
in the last 10 games**

For Gameweeks in which less than 10 games were played, previous seasons data was used



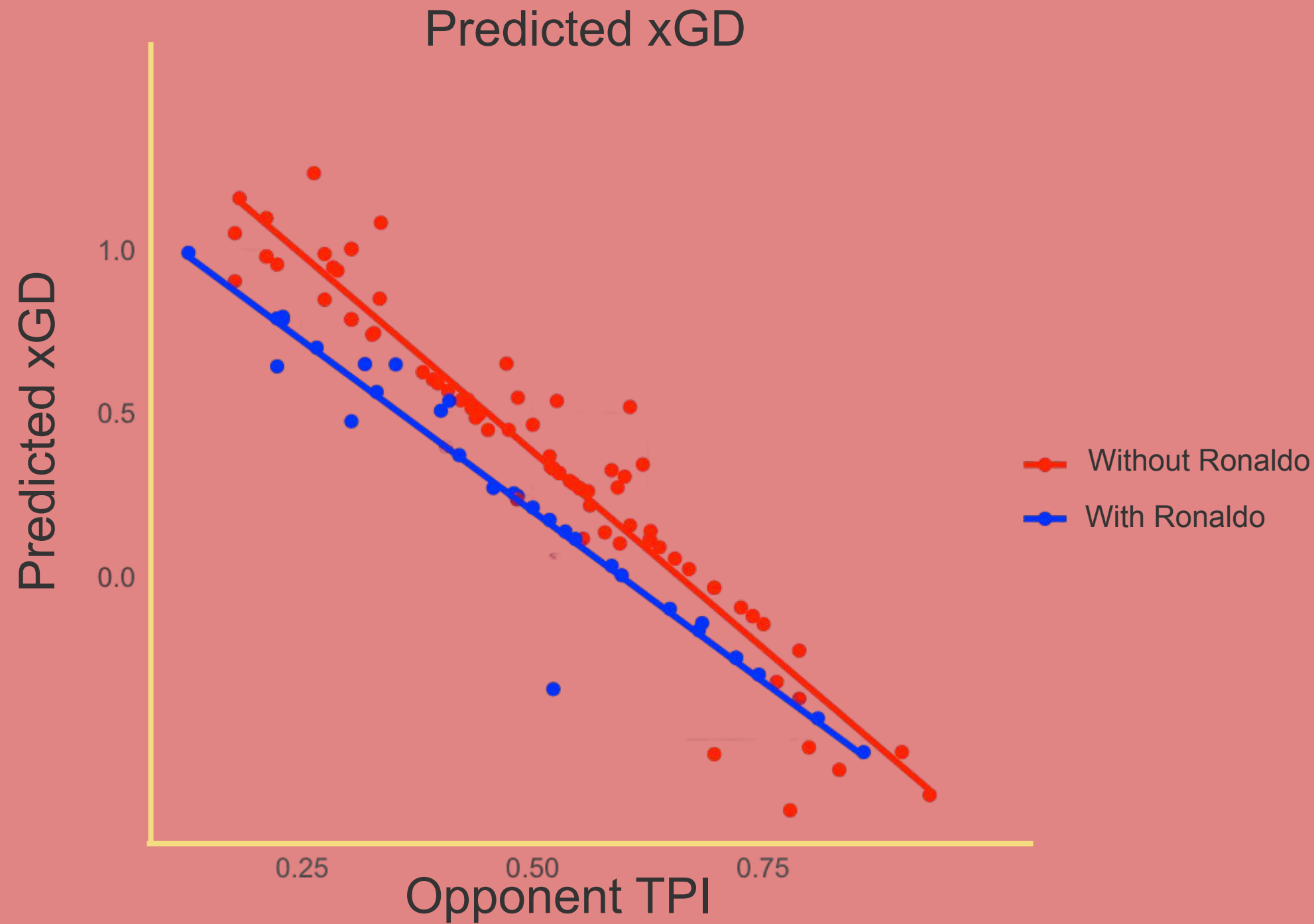
Controlling for Home field advantage

$$\text{HFA} = \frac{\text{Home}(\text{GF}-\text{GA}) - \text{Away}(\text{GF}-\text{GA})}{38}$$

(2017-19 Seasons)



xGD Conclusion



T-value : -0.660
P-value: 0.51082

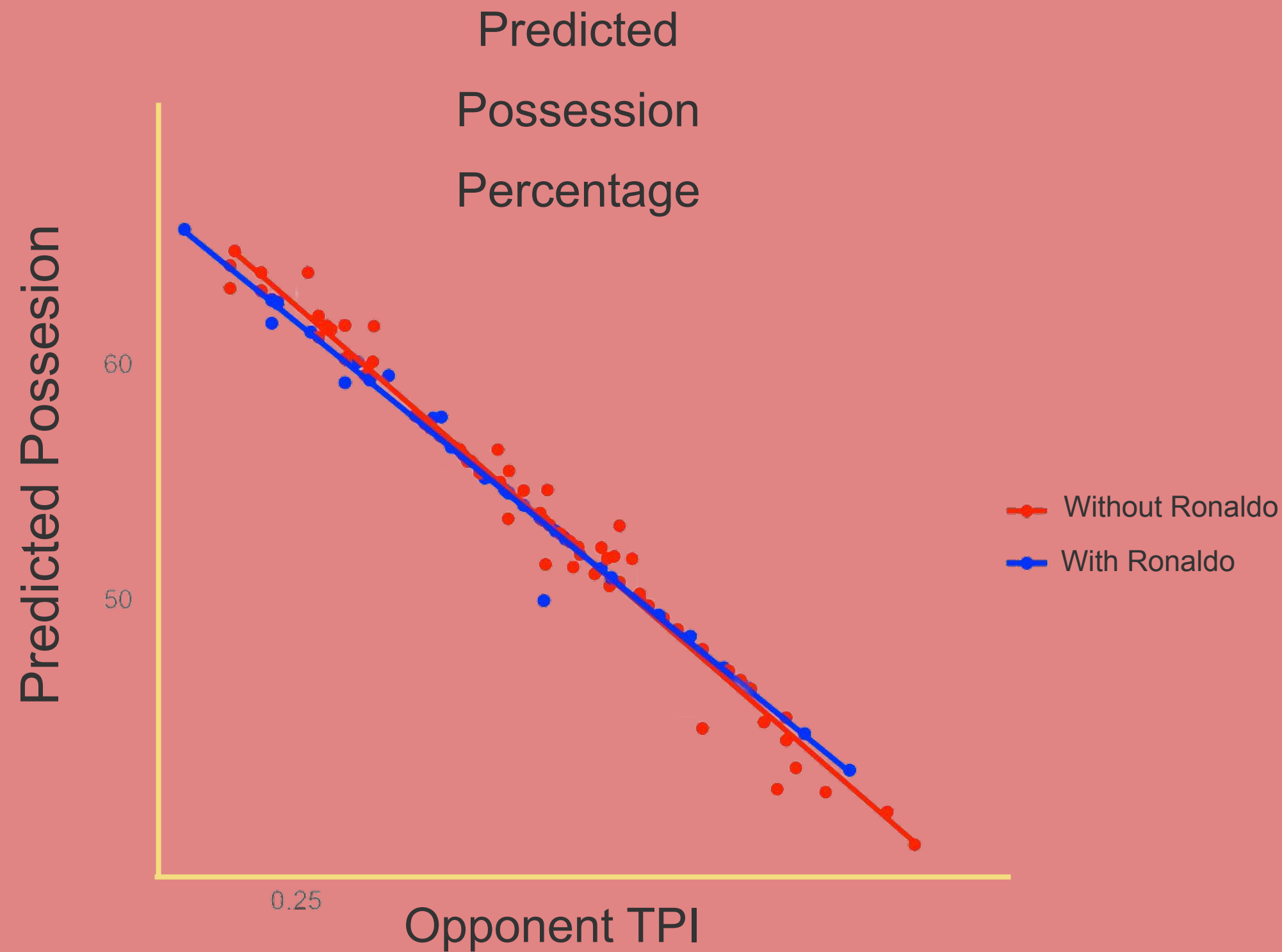
Equation MLR

xGD =

$$\begin{aligned} & -0.165(\text{RonaldoStarts}) - \\ & 2.088(\text{Opponent_TPI}) - \\ & 1.117(\text{HA}) + 1.802 \end{aligned}$$

Possession Conclusion

T-value : -0.025
P-value: 0.980



Equation MLR

$$\text{Possession} = -0.046(\text{RonaldoStarts}) - 31.282(\text{Opponent_TPI}) - 7.33(\text{HA}) + 72.15783$$

Next Steps

MU Strength

Strength of MU team
based on value of squad

Injuries

Important players injured
or number of injuries for
that match

Manager Change

4 Managers from
2020-2022, 2 interim, 2
permanent

Sources

[Pinnacle - Home Field Advantage Formula Reference](#)

[WorldFootballR](#)

[Statsbomb - Pythagoren Theorem](#)

[Cristiano Ronaldo Images](#)

[FBref](#)

[transfermarkt](#)

Questions?



Thank you

