Predicting MLS Goals For a Given Game

Team 8:
Moazzam Ali, Wesley Gao, Agha Yusuf Khan, Zone Li, & Zeyu 'Alan' Wang

Introduction

What is xG?

- xG (expected goals) calculates probabilities for individual shot actions.
- **Often misassociated with predicting team goals, it instead tracks performance of teams and players.**
- The primary aim is to minimize random effects associated with goal scoring when making such evaluations on players and teams.

Introduction

What about Goals?

- Our interest lies in determining the number of goals scored by the target team at the game.

- There are numerous potential applications for being able to make such predictions:

  - Ticket Sales for high event games
  - Fan-based pursuits (ie. sports betting)
  - Corporate Promotions

- With MLS playoffs currently underway, we viewed the 2024 season data as a good dataset to predict goals of playoff games.

Overview

01

Problem Statement & Data Description

02

Variable Selection Methodology

03

Regression Analysis & Goodness-of-Fit

04

Model Predictions & Recommendations

Background | Problem Statement

- *How can we accurately predict the number of goals scored by a team in a specific MLS game using 2024 season data?*

- *What predictors are most significant to predict the number of goals?*

Data Source:

- 2024 MLS Season Data from **FBRef**

- Focused on Goals For (“GF”) for each team; teams that played each other were constrained to separate rows (986)

- Variables included all tables from data source covering offensive, defensive, and goalkeeping statistics

Background | Problem Statement

Goals of the Project

- Predictions of playoff games in MLS
- Understanding of variables contributing to goals scored

Idea concentrated around a MLR model

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ \ldots + \hat{\beta}_k x_k$$

Shots on Target Variable

Goalkeeper Saves

Well…

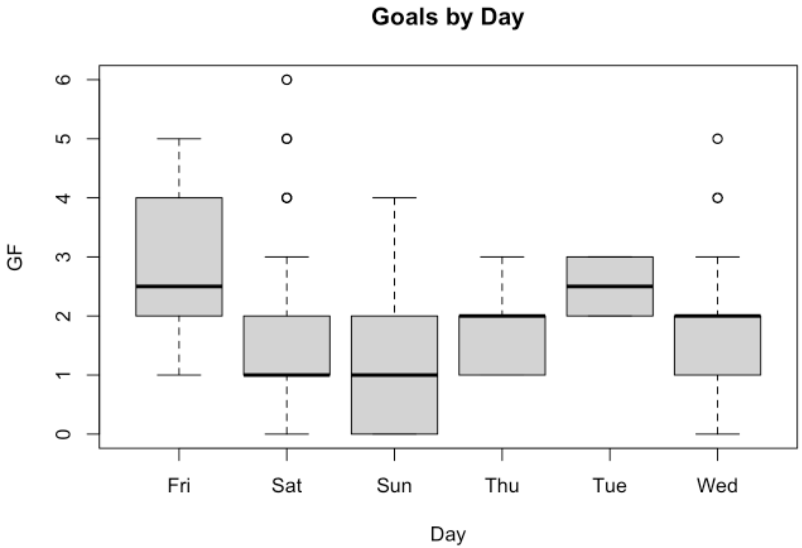
R^2 = ~95%

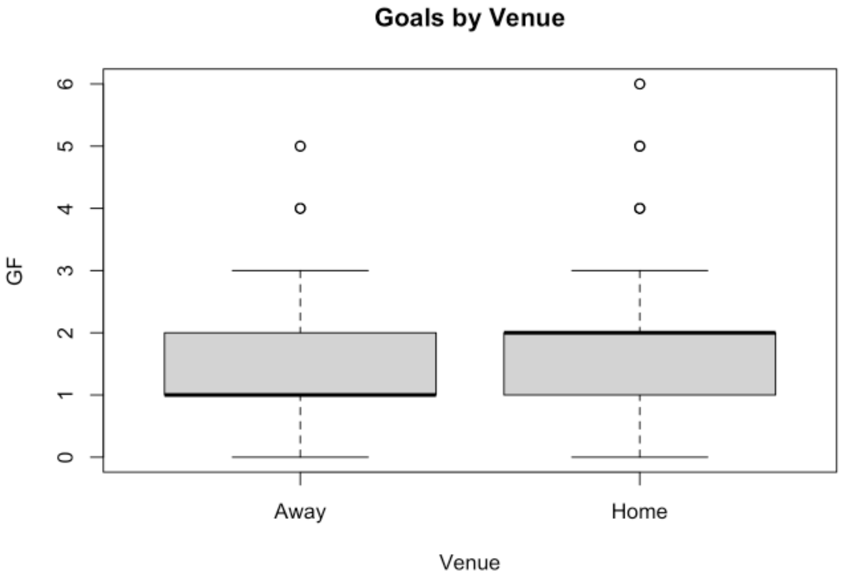
We're done right?

Not so fast

Background | Data Description

Qualitative Variables – Day and Venue

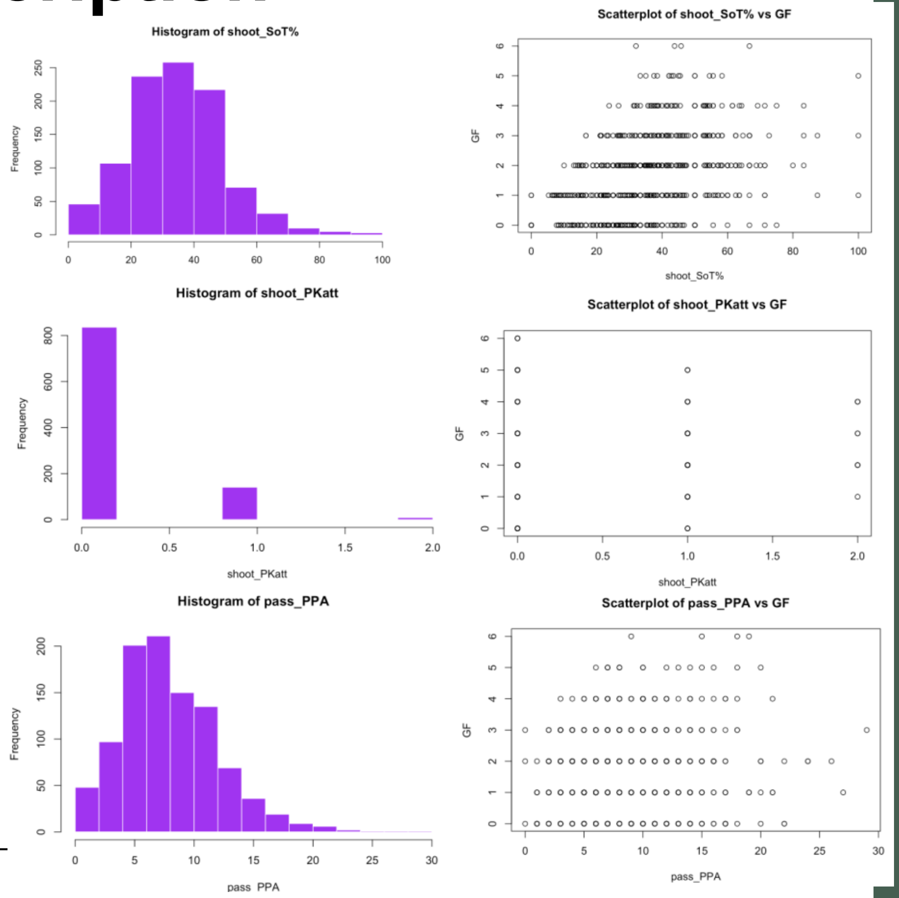




Background | Data Description

Quantitative Variables – Assorted Match Statistics for Target Team

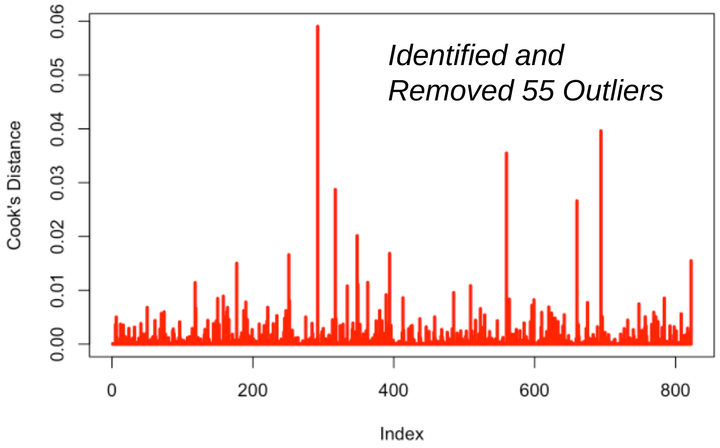
- Response: Goals for Target Team (GF)
- 161 variables covered shooting, passing, dribbling, defensive, goalkeeping, and penalties (challenge for variable selection)
- Some non-linear relationships identified across variables and response
- High Correlation amongst several variables indicating near similar or repeat variables
- Examples included Shots on Target%, Penalty Kick Attempts, Passes in Penalty Box [widely different distributions]



Variable Selection Methodology

Removal of Redundant and Irrelevant Variables → Bi-Directional Stepwise Regression → Outlier Removal, Mallow's CP & Lasso

- Exclusion of so-called derived statistics (e.g. Penalty Kicks that resulted into goals)
- Removed redundancies amongst variables selection (% versus counts) identified via correlation; manual removal resulted in 98 variables for stepwise
- Stepwise Regression Resulted in a model of 28 variables from which outlier removal and subsequent Lasso and Mallow's CP showcased final 23 variables



Selected Variables

23 Quantitative Variables

opp_keeper	opp_def	shoot	pass	passtype	poss	misc
Save%	Clr	SoT%	PPA	TB	Succ	2CrdY
Att.1	Mid 3rd	Sh	1/3		PrgDist	Won%
Att (GK)	Att 3rd	PKatt			CPA	PKcon
	Err				Mid 3rd	OG
	Sh					
	Blocks					

Multiple Linear Regression

MLR Model:

- Ran MLR using the 23 selected variables on the training data after removing the outliers

Statistical Inference:

- Testing for overall model significance ($\boldsymbol{\alpha = 0.05}$):
  - $H_0: \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_{23} = 0$
  - H_a: At least one of the coefficients is different from 0
  - Test Statistic: $F_{23,\,743} = 145.3$
  - p-value < 2.2e-16
  - Conclusion: Since p-value < α, we reject H_0, hence, the overall regression model is significant at 95% significance level

- **R-squared:** 81.81%; **Adjusted R-squared:** 81.25%
  - The MLR model explains 81.81% of variation in the Goals scored by the team.

MLR MODEL OUTPUT

```
Call:
lm(formula = GF ~ `opp_keeper_Save%` + `shoot_SoT%` + shoot_Sh +
    shoot_PKatt + opp_def_Clr + `opp_def_Mid 3rd` + misc_2CrdY +
    opp_def_Sh + opp_keeper_Att.1 + `opp_def_Att 3rd` + pass_PPA +
    `opp_keeper_Att (GK)` + poss_Succ + opp_def_Err + passtype_TB +
    opp_def_Blocks + `pass_1/3` + poss_PrgDist + poss_CPA + `poss_Mid 3rd` +
    `misc_Won%` + misc_PKcon + misc_OG, data = train_data2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.46813 -0.30050 -0.00968  0.33847  1.37568

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.3887705  0.1529013   9.083  < 2e-16 ***
`opp_keeper_Save%`    -0.0284440  0.0007228 -39.351  < 2e-16 ***
`shoot_SoT%`           0.0307200  0.0015632  19.652  < 2e-16 ***
shoot_Sh               0.1220677  0.0071452  17.084  < 2e-16 ***
shoot_PKatt            0.8059395  0.0492253  16.372  < 2e-16 ***
opp_def_Clr           -0.0144333  0.0026984  -5.349 1.18e-07 ***
`opp_def_Mid 3rd`      0.0209673  0.0064819   3.235 0.001271 **
misc_2CrdY            -0.2587437  0.0877762  -2.948 0.003301 **
opp_def_Sh            -0.0327082  0.0148065  -2.209 0.027476 *
opp_keeper_Att.1      -0.0324386  0.0084630  -3.833 0.000137 ***
`opp_def_Att 3rd`     -0.0283306  0.0115157  -2.460 0.014113 *
pass_PPA               0.0174535  0.0060730   2.874 0.004169 **
`opp_keeper_Att (GK)` -0.0061628  0.0022326  -2.760 0.005915 **
poss_Succ              0.0128509  0.0057037   2.253 0.024543 *
opp_def_Err            0.0428843  0.0266268   1.611 0.107699
passtype_TB            0.0415468  0.0145914   2.847 0.004530 **
opp_def_Blocks        -0.0141128  0.0056458  -2.500 0.012644 *
`pass_1/3`            -0.0087073  0.0024465  -3.559 0.000396 ***
poss_PrgDist          -0.0003478  0.0001107  -3.142 0.001748 **
poss_CPA               0.0267541  0.0082482   3.244 0.001233 **
`poss_Mid 3rd`         0.0014573  0.0004452   3.273 0.001113 **
`misc_Won%`           -0.0026049  0.0014714  -1.770 0.077073 .
misc_PKcon             0.1082640  0.0462440   2.341 0.019488 *
misc_OG                0.1901937  0.0856876   2.220 0.026746 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4911 on 743 degrees of freedom
Multiple R-squared:  0.8181,	Adjusted R-squared:  0.8125
F-statistic: 145.3 on 23 and 743 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

Statistical Inference:

- Testing for individual coefficient significance ($\boldsymbol{\alpha = 0.05}$):
  - $H_0: \hat{\beta}_j = 0;\ H_a: \hat{\beta}_j \neq 0$ for j = 1, 2, …, 23
  - Test Statistic: t-value of $\hat{\beta}_j$
  - Using p-values for t-statistic at 95% significance level:
    - ➢ Intercept + 21 predictor coefficients significant
    - ➢ 2 predictor coefficients insignificant:

      opp_def_Err (Defender Errors)

      misc_won% (Aerial Duels Win%)

Notable Observations:

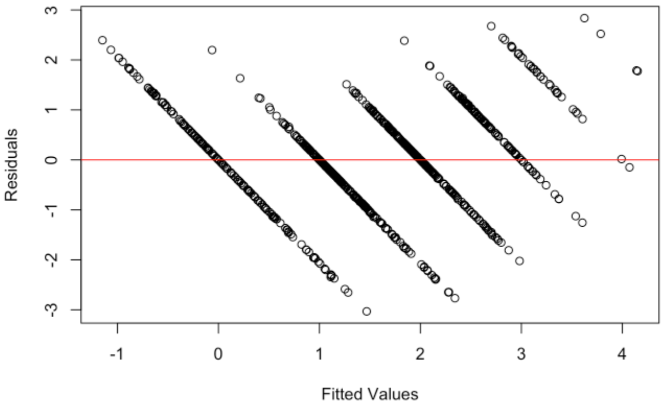
- Most of the offensive variable statistics have a positive coefficient
- Most of the opposition's defensive variable statistics have a negative coefficient
- misc_OG, misc_Pkcon have a positive coefficient *(Interpret Carefully!)*

MLR MODEL OUTPUT

```
Call:
lm(formula = GF ~ `opp_keeper_Save%` + `shoot_SoT%` + shoot_Sh +
    shoot_PKatt + opp_def_Clr + `opp_def_Mid 3rd` + misc_2CrdY +
    opp_def_Sh + opp_keeper_Att.1 + `opp_def_Att 3rd` + pass_PPA +
    `opp_keeper_Att (GK)` + poss_Succ + opp_def_Err + passtype_TB +
    opp_def_Blocks + `pass_1/3` + poss_PrgDist + poss_CPA + `poss_Mid 3rd` +
    `misc_Won%` + misc_PKcon + misc_OG, data = train_data2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.46813 -0.30050 -0.00968  0.33847  1.37568

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.3887705  0.1529013   9.083  < 2e-16 ***
`opp_keeper_Save%`    -0.0284440  0.0007228 -39.351  < 2e-16 ***
`shoot_SoT%`           0.0307200  0.0015632  19.652  < 2e-16 ***
shoot_Sh               0.1220677  0.0071452  17.084  < 2e-16 ***
shoot_PKatt            0.8059395  0.0492253  16.372  < 2e-16 ***
opp_def_Clr           -0.0144333  0.0026984  -5.349 1.18e-07 ***
`opp_def_Mid 3rd`      0.0209673  0.0064819   3.235 0.001271 **
misc_2CrdY            -0.2587437  0.0877762  -2.948 0.003301 **
opp_def_Sh            -0.0327082  0.0148065  -2.209 0.027476 *
opp_keeper_Att.1      -0.0324386  0.0084630  -3.833 0.000137 ***
`opp_def_Att 3rd`     -0.0283306  0.0115157  -2.460 0.014113 *
pass_PPA               0.0174535  0.0060730   2.874 0.004169 **
`opp_keeper_Att (GK)` -0.0061628  0.0022326  -2.760 0.005915 **
poss_Succ              0.0128509  0.0057037   2.253 0.024543 *
opp_def_Err            0.0428843  0.0266268   1.611 0.107699
passtype_TB            0.0415468  0.0145914   2.847 0.004530 **
opp_def_Blocks        -0.0141128  0.0056458  -2.500 0.012644 *
`pass_1/3`            -0.0087073  0.0024465  -3.559 0.000396 ***
poss_PrgDist          -0.0003478  0.0001107  -3.142 0.001748 **
poss_CPA               0.0267541  0.0082482   3.244 0.001233 **
`poss_Mid 3rd`         0.0014573  0.0004452   3.273 0.001113 **
`misc_Won%`           -0.0026049  0.0014714  -1.770 0.077073 .
misc_PKcon             0.1082640  0.0462440   2.341 0.019488 *
misc_OG                0.1901937  0.0856876   2.220 0.026746 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4911 on 743 degrees of freedom
Multiple R-squared:  0.8181,	Adjusted R-squared:  0.8125
F-statistic: 145.3 on 23 and 743 DF,  p-value: < 2.2e-16
```

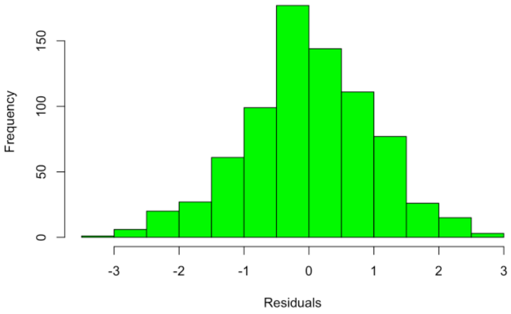
MLR Goodness of Fit

Fitted Values vs Residuals

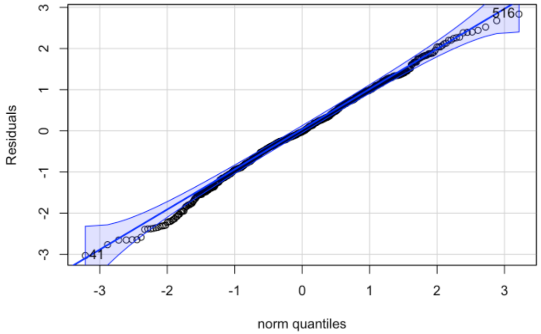


Constant Variance & Independence Assumption
(Violated – Non-constant Variance)

Histogram of Residuals



Normality Plot (QQ Plot)



Normality Assumption
(Satisfied – Approximately Normal)

- Linearity Assumption: Most of the predictors exhibit a linear relationship with residuals

- VIF Analysis:
  - VIF Threshold: Max(10, 1/(1 – 81.81%)) = Max(10, 5.49) = 10
  - Max VIF among all variables = 3.81
  - Conclusion: Since VIF values for all variables are below the VIF threshold the model does not exhibit multicollinearity

Poisson Regression

Poisson Regression Model:

- Ran Poisson Regression model using the 23 selected variables on the training data after removing the outliers
- Number of Shots is used to account for Exposure

Statistical Inference:

- Testing for overall model significance ($\boldsymbol{\alpha = 0.05}$):
  - $H_0: \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_{22} = 0$
  - H_a: At least one of the coefficients is different from 0
  - Test Statistic: Null Deviance – Residual Deviance = 638.14, Degrees of Freedom = 766-744 = 22
  - p-value = 0
  - Conclusion: Since p-value < α, we reject H_0, hence, the overall regression model is significant at 95% significance level

POISSON REGRESSION MODEL OUTPUT

```
Call:
glm(formula = GF ~ `opp_keeper_Save%` + `shoot_SoT%` + offset(log(shoot_Sh)) +
    shoot_PKatt + opp_def_Clr + `opp_def_Mid 3rd` + misc_2CrdY +
    opp_def_Sh + opp_keeper_Att.1 + `opp_def_Att 3rd` + pass_PPA +
    `opp_keeper_Att (GK)` + poss_Succ + opp_def_Err + passtype_TB +
    opp_def_Blocks + `pass_1/3` + poss_PrgDist + poss_CPA + `poss_Mid 3rd` +
    `misc_Won%` + misc_PKcon + misc_OG, family = poisson, data = train_data2)

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)           -1.6368004  0.2605210  -6.283 3.33e-10 ***
`opp_keeper_Save%`    -0.0240194  0.0013353 -17.988  < 2e-16 ***
`shoot_SoT%`           0.0285941  0.0026258  10.890  < 2e-16 ***
shoot_PKatt            0.6046352  0.0730198   8.280  < 2e-16 ***
opp_def_Clr           -0.0085644  0.0046945  -1.824   0.0681 .
`opp_def_Mid 3rd`      0.0036169  0.0109517   0.330   0.7412
misc_2CrdY            -0.1746966  0.1728558  -1.011   0.3122
opp_def_Sh             0.0256918  0.0217789   1.180   0.2381
opp_keeper_Att.1      -0.0052389  0.0122877  -0.426   0.6699
`opp_def_Att 3rd`     -0.0076037  0.0194306  -0.391   0.6956
pass_PPA               0.0119283  0.0098011   1.217   0.2236
`opp_keeper_Att (GK)` -0.0008443  0.0037045  -0.228   0.8197
poss_Succ              0.0086183  0.0093192   0.925   0.3551
opp_def_Err            0.0006628  0.0409606   0.016   0.9871
passtype_TB            0.0121589  0.0225918   0.538   0.5904
opp_def_Blocks        -0.0136678  0.0096241  -1.420   0.1556
`pass_1/3`            -0.0025772  0.0041765  -0.617   0.5372
poss_PrgDist          -0.0001635  0.0001867  -0.876   0.3812
poss_CPA               0.0096154  0.0131264   0.733   0.4638
`poss_Mid 3rd`         0.0002516  0.0007303   0.344   0.7305
`misc_Won%`            0.0002661  0.0024288   0.110   0.9128
misc_PKcon             0.0126055  0.0715827   0.176   0.8602
misc_OG                0.1343878  0.1345177   0.999   0.3178
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 841.77  on 766  degrees of freedom
Residual deviance: 203.63  on 744  degrees of freedom
AIC: 1761.9

Number of Fisher Scoring iterations: 4
```

```
# Checking for model significance
1-pchisq((841.77-203.63),(766-744))
```

```
## [1] 0
```

Poisson Regression

Statistical Inference:

- Testing for individual coefficient significance ($\boldsymbol{\alpha = 0.05}$):
  - $H_0: \hat{\beta}_{\mathrm{j}} = 0;\ H_a: \hat{\beta}_j \neq 0$ for j = 1, 2, …, 22
  - Test Statistic: z-value of $\hat{\beta}_j$ (Wald test)
  - Using p-values for z-statistic at 95% significance level:
    - ➢ Intercept + 3 predictor coefficients significant:

      opp_keeper_Save% (Keeper’s Save%)

      shoot_SoT% (Shot’s on Target)

      Shoot_Pkatt (Penalty Kicks Attempted)
    - ➢ 19 predictor coefficients insignificant:

- Testing for Subsets of Coefficients significance ($\boldsymbol{\alpha = 0.05}$):
  - Reduced Model: First 5 terms ($\bar{\beta}_0, \bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4$)
  - $H_0: \hat{\alpha}_{\mathrm{i}} = 0;\ H_a: \hat{\alpha}_{\mathrm{i}} \neq 0$ for i = 1, 2, …, 18
  - Wald Test: X2 = 379.5, df = 5; p-value = 0
  - Conclusion: Since p-value > α, we fail to reject H_0, hence, the other variables do not have significant explanatory power at 95% significance level

POISSON REGRESSION MODEL OUTPUT

```
Call:
glm(formula = GF ~ `opp_keeper_Save%` + `shoot_SoT%` + offset(log(shoot_Sh)) +
    shoot_PKatt + opp_def_Clr + `opp_def_Mid 3rd` + misc_2CrdY +
    opp_def_Sh + opp_keeper_Att.1 + `opp_def_Att 3rd` + pass_PPA +
    `opp_keeper_Att (GK)` + poss_Succ + opp_def_Err + passtype_TB +
    opp_def_Blocks + `pass_1/3` + poss_PrgDist + poss_CPA + `poss_Mid 3rd` +
    `misc_Won%` + misc_PKcon + misc_OG, family = poisson, data = train_data2)

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)          -1.6368004  0.2605210  -6.283 3.33e-10 ***
`opp_keeper_Save%`   -0.0240194  0.0013353 -17.988  < 2e-16 ***
`shoot_SoT%`          0.0285941  0.0026258  10.890  < 2e-16 ***
shoot_PKatt           0.6046352  0.0730198   8.280  < 2e-16 ***
opp_def_Clr          -0.0085644  0.0046945  -1.824   0.0681 .  
`opp_def_Mid 3rd`     0.0036169  0.0109517   0.330   0.7412    
misc_2CrdY           -0.1746966  0.1728558  -1.011   0.3122    
opp_def_Sh            0.0256918  0.0217789   1.180   0.2381    
opp_keeper_Att.1     -0.0052389  0.0122877  -0.426   0.6699    
`opp_def_Att 3rd`    -0.0076037  0.0194306  -0.391   0.6956    
pass_PPA              0.0119283  0.0098011   1.217   0.2236    
`opp_keeper_Att (GK)` -0.0008443  0.0037045  -0.228   0.8197    
poss_Succ             0.0086183  0.0093192   0.925   0.3551    
opp_def_Err           0.0006628  0.0409606   0.016   0.9871    
passtype_TB           0.0121589  0.0225918   0.538   0.5904    
opp_def_Blocks       -0.0136678  0.0096241  -1.420   0.1556    
`pass_1/3`           -0.0025772  0.0041765  -0.617   0.5372    
poss_PrgDist         -0.0001635  0.0001867  -0.876   0.3812    
poss_CPA              0.0096154  0.0131264   0.733   0.4638    
`poss_Mid 3rd`        0.0002516  0.0007303   0.344   0.7305    
`misc_Won%`           0.0002661  0.0024288   0.110   0.9128    
misc_PKcon            0.0126055  0.0715827   0.176   0.8602    
misc_OG               0.1343878  0.1345177   0.999   0.3178    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 841.77  on 766  degrees of freedom
Residual deviance: 203.63  on 744  degrees of freedom
AIC: 1761.9

Number of Fisher Scoring iterations: 4
```

```
Wald test:
----------

Chi-squared test:
X2 = 9.2, df = 18, P(> X2) = 0.95
```

Poisson Reduced Model

```
Call:
glm(formula = GF ~ `opp_keeper_Save%` + `shoot_SoT%` + offset(log(shoot_Sh)) +
    shoot_PKatt + opp_def_Clr, family = "poisson", data = train_data2)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -1.679151   0.121366 -13.835   <2e-16 ***
`opp_keeper_Save%` -0.024038   0.001209 -19.885   <2e-16 ***
`shoot_SoT%`        0.028396   0.001991  14.260   <2e-16 ***
shoot_PKatt         0.604502   0.070088   8.625   <2e-16 ***
opp_def_Clr        -0.007106   0.004026  -1.765   0.0775 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 841.77  on 766  degrees of freedom
Residual deviance: 212.81  on 762  degrees of freedom
AIC: 1735.1

Number of Fisher Scoring iterations: 4
```

```
# GOF Test
with(reduced_poisson, cbind(res.deviance = deviance, df = df.residual,
 p = pchisq(deviance, df.residual, lower.tail = FALSE)))
```

```
##      res.deviance  df p
## [1,]     212.8094 762 1
```

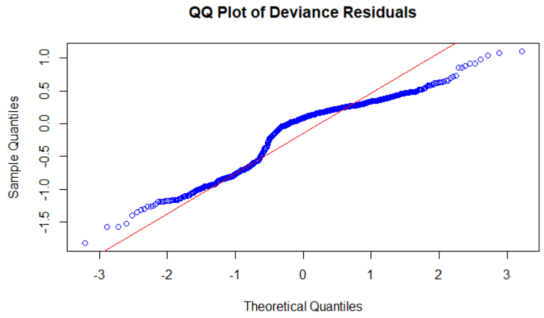
Reduced Model:

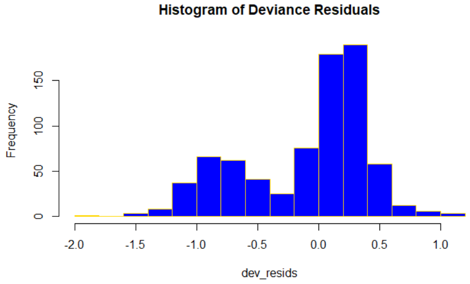
- First five variables from larger Poisson model
- Number of Shots is used to account for Exposure

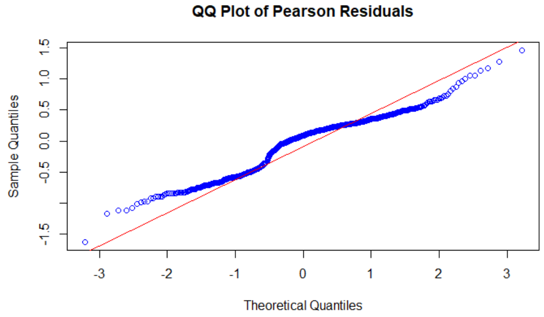
Hypothesis Testing Procedure:

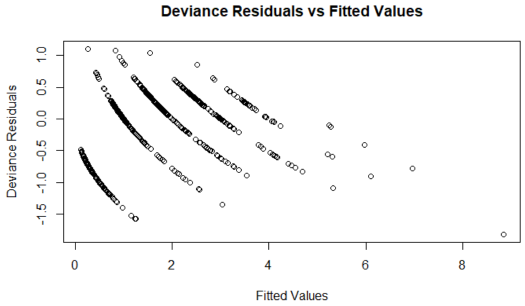
- Testing for Goodness of Fit of the model ($\boldsymbol{\alpha = 0.05}$):
  - H_0: the Poisson model fits the data
  - H_a: the Poisson model does not fit the data
  - Test Statistic: Residual Deviance = 212.81, Degrees of Freedom = 762
  - p-value = 1
  - Conclusion: Since p-value > α, we fail to reject H_0, hence, the reduced Poisson model fits the data at 95% significance level

Poisson Goodness of Fit – Visual Analysis









Prediction Accuracy

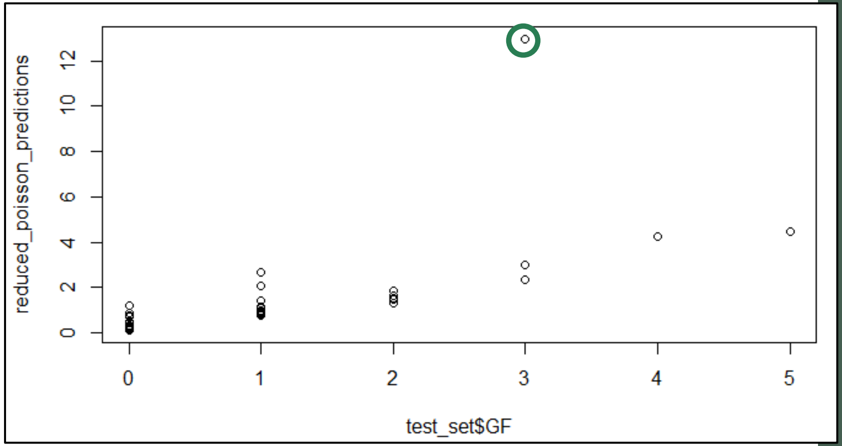
Model Prediction Accuracies	Full Validation Set for 1 Iteration			Validation for 100 iterations					
	MSPE	MAE	PM	MSPE		MAE		PM	
				Mean	Median	Mean	Median	Mean	Median
Poisson Model	0.405	0.427	0.241	0.743	0.462	0.497	0.494	0.493	0.295
Reduced Poisson	0.339	0.402	0.201	0.560	0.417	0.462	0.463	0.369	0.270
MLR Model	0.427	0.492	0.254	0.406	0.401	0.476	0.476	0.267	0.261

- Reduced Poisson Model is most accurate over the full validation set while the MLR is more accurate over the sampled 100 iterations.
- MLR model has inaccuracies associated with negative goals and the violation of the MLR Constant Variance Assumption from Goodness-of-Fit Analysis.
- While further optimization may be needed, the Reduced Poisson model may be the better model in the long run due to its handling of the discrete, bounded response.

Recommendations

- Poisson Regression Models have higher variability in model performance depending on the chosen data suggesting overfitting.
- Poisson Reduced Model may yield more accurate and preferred results due to a simpler model
- MLR may not perform well on a bounded, discrete range.
- Feature engineering could enable deeper understanding of goal scoring determinants.
- Training models tailored onto individual team season performances could yield differing and potentially better results.

Reduced Poisson Model predictions on Round One playoff data



Atlanta United Round 1 (v Inter Miami CF)

Results:	Model Predictions:	xG-xGA:
L (1 – 2)	L (1 – 2)	(1 – 3.3)
W (2 – 1)	W (2 – 1)	(1.4 – 1.5)
W (3 – 2)	W (3 – 1)	(1.8 – 2.7)

Thank You

Appendix

Selected Variables

Table of Chosen Predictors

Name	Simple Name	Short Description	Type of Predictor
opp_keeper_Save%	Keeper's Save %	Number of Shots on Target saved by the opponent's keeper	Quantitative
opp_keeper_Att.1	Goal Kick's Attempted	Number of Goal Kicks attempted by the opponent's keeper	Quantitative
opp_keeper_Att (GK)	Passes Attempted by Keeper	Number of Passes attempted by the opponent's keeper	Quantitative
shoot_SoT%	Shots on Target	Number of Shots on Target by the team	Quantitative
shoot_Sh	Shots	Number of Shots by the team	Quantitative
shoot_PKatt	Penalty Kicks Attempted	Penalty kicks attempted by the team	Quantitative
pass_PPA	Passes into Penalty Area	Number of completed passes into the 18-yard box	Quantitative
pass_1/3	Passes into the Final Third	Number of completed passes that enter into the 1/3 of the pitch closest to the goal	Quantitative
passtype_TB	Through Balls Completed	Number of completed passes sent b/w back defenders of the opposition into open space	Quantitative

Selected Variables

Name	Simple Name	Short Description	Type of Predictor
poss_Succ	Successful Take-Ons	Number of defenders taken-on successfully by dribbling past them	Quantitative
poss_PrgDist	Progressive Carrying Distance (in Yards)	Distance player moved the ball towards the opponent’s goals	Quantitative
poss_CPA	Carries into Penalty Area	Number of carries into the penalty area by the team	Quantitative
poss_Mid 3rd	Touches in Middle Third	Number of touches by the team in the middle third of the pitch	Quantitative
misc_2CrdY	Second Yellow Card	Number of second yellow cards picked up by the team	Quantitative
misc_Won%	Aerial Duels Win%	Number of aerial duels won by the team as percentage of total number of aerial duels	Quantitative
misc_PKcon	Penalty Kicks Conceded	Number of penalty kicks conceded by the team	Quantitative
misc_OG	Own Goals	Number of own goals by the team	Quantitative

Selected Variables

Name	Simple Name	Short Description	Type of Predictor
opp_def_Clr	Defender’s Clearances	Number of clearances by the opposition defenders	Quantitative
opp_def_Mid 3rd	Tackles in Middle Third	Number of tackles by the opposition in the middle third of the pitch	Quantitative
opp_def_Att 3rd	Tackles in Attacking Third	Number of tackles by the opposition in the attacking third of the pitch	Quantitative
opp_def_Err	Defender Errors	Number of opposition defender’s mistakes leading to a shot	Quantitative
opp_def_Sh	Shots Blocked	Number of shots blocked by the opposition	Quantitative
opp_def_Blocks	Blocks	Number of times ball is blocked by the opposition by standing in the ball’s path	Quantitative

MLS Playoffs Round One – GF v xG

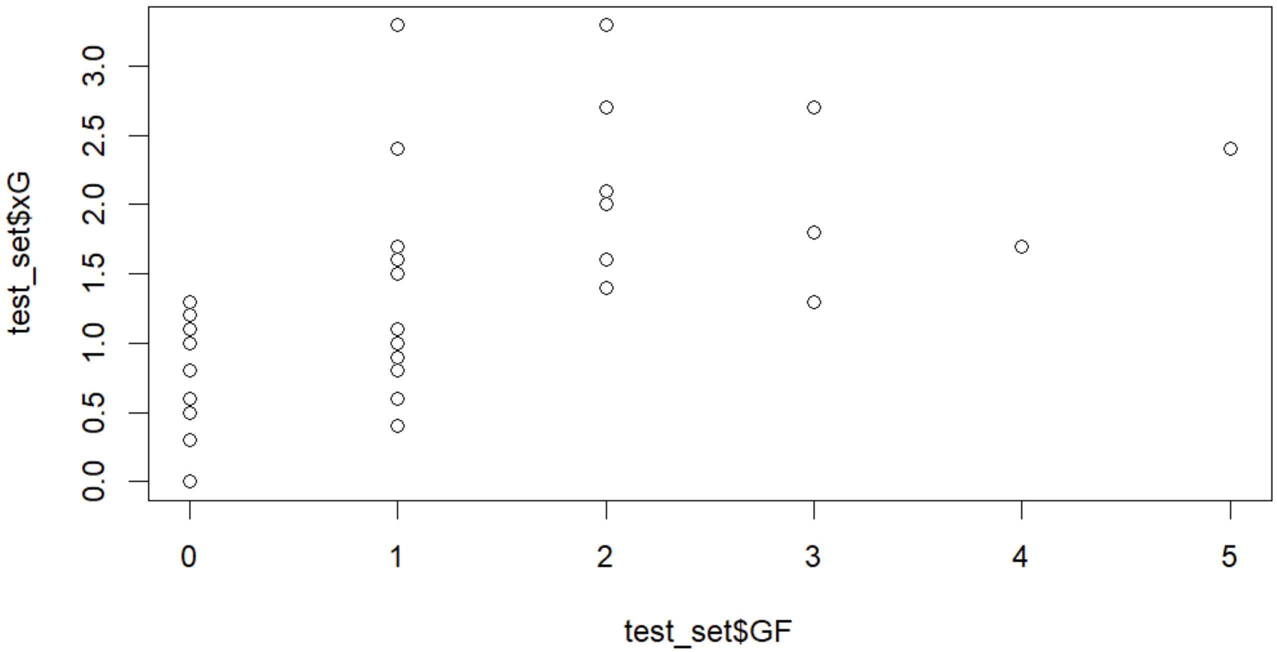


Photo Credits

1. "Football Goal GIF by Major League Soccer," Jul. 25, 2021. Available: https://giphy.com/gifs/mls-peru-sounders-ruidiaz-va2KuYtHbuL24i3Ygq

2. B. Bastello, "Step Inside: Mercedes-Benz Stadium – Home of the Atlanta Falcons," *Ticketmaster Blog*, Nov. 20, 2024. Available: https://blog.ticketmaster.com/step-inside-mercedes-benz-stadium-atlanta-ga/

3. ColumbusCrew.com, "Supporters Promotions | Columbus Crew," *ColumbusCrew.com*. Available: https://www.columbuscrew.com/supporters/promotions

4. "FBRef." Available: https://fbref.com/en/squads/1ebc1a5b/2024/matchlogs/c22/passing_types/Atlanta-United-Match-Logs-Major-League-Soccer