Soccer Shots

Brendan Lafferty
08/05/2020

Table of Contents

01 The Question

What goes into making goal?

02 The Data

SCIENTIFIC DATA

wyscout

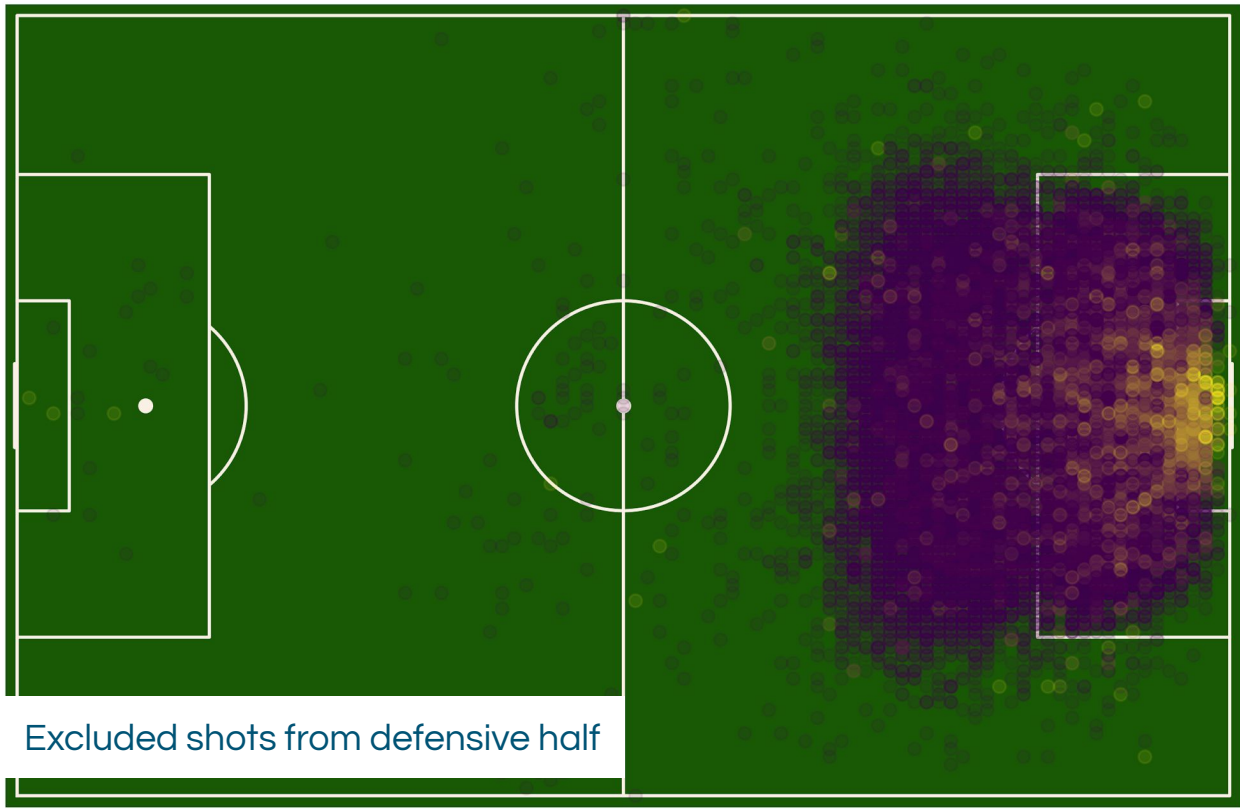
3,000,000+

Events meticulously recorded

200,000+

Shots Analysed

Shots Outcomes



Excluded shots from defensive half

Goals: Yellow

Misses: Purple

03 The Model

Linear Regression was chosen for its interpretability

Features with significance: ($p < .5$)

- Distance from the Goal
- Angular Size of the Goal
- Whether the ball was kicked vs controlled with the body or head
- Whether the ball was shot with the same side foot as the side of the field
- Whether there was a play from a dead ball in the last 30 seconds by the same team

Features with poor significance: ($p > .5$)

- The cumulative player advantage up to that point in the game
- Whether the player used their dominant foot

04 Takeaways: For Coaches

- The biggest predictors are distance and angular size of the goal
- Kicking the ball produced better

Look to create chances close to the goal that are near the ground and can be shot from the feet.

04 Takeaways: For Players

- Using your dominant foot is not a significant factor in the model

Have confidence in taking shots with the foot that is available

A Word of Caution

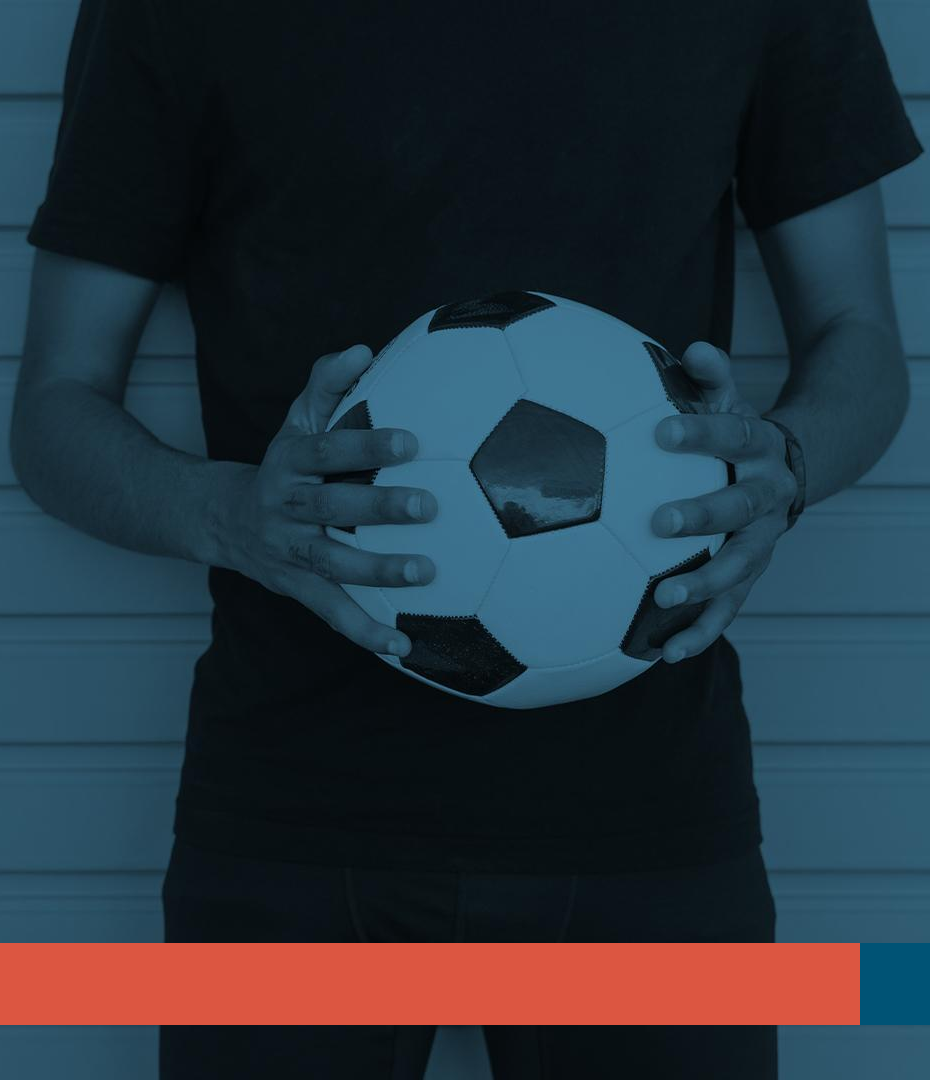
Don't use this model to say:

- Players should try to kick the ball as opposed to heading the ball as kicking is more successful.
- The model predicted Wondo should have scored in 2014 against Belgium; why did he let us down!

05 Extras

- Future Features:
  - Number of passes that lead up to the shot
  - The direction the ball came from in relation to where the shot occured
  - Whether the chance resulted from a turnover in the attacking half

Thank You!



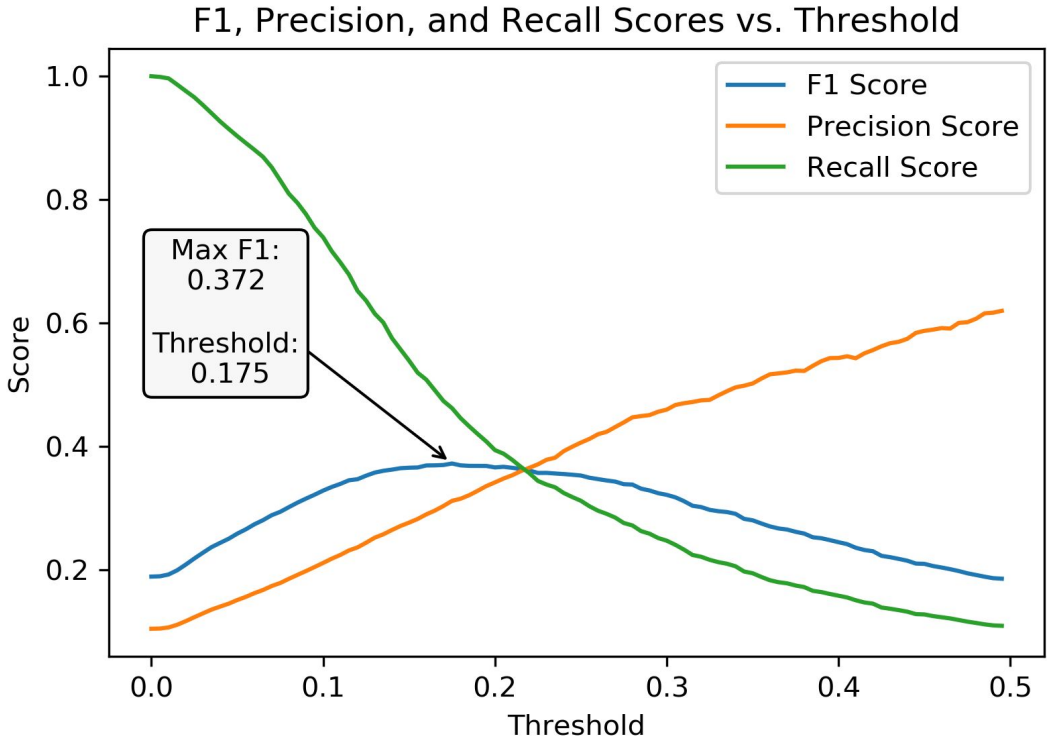
Appendices

Model Coefficients

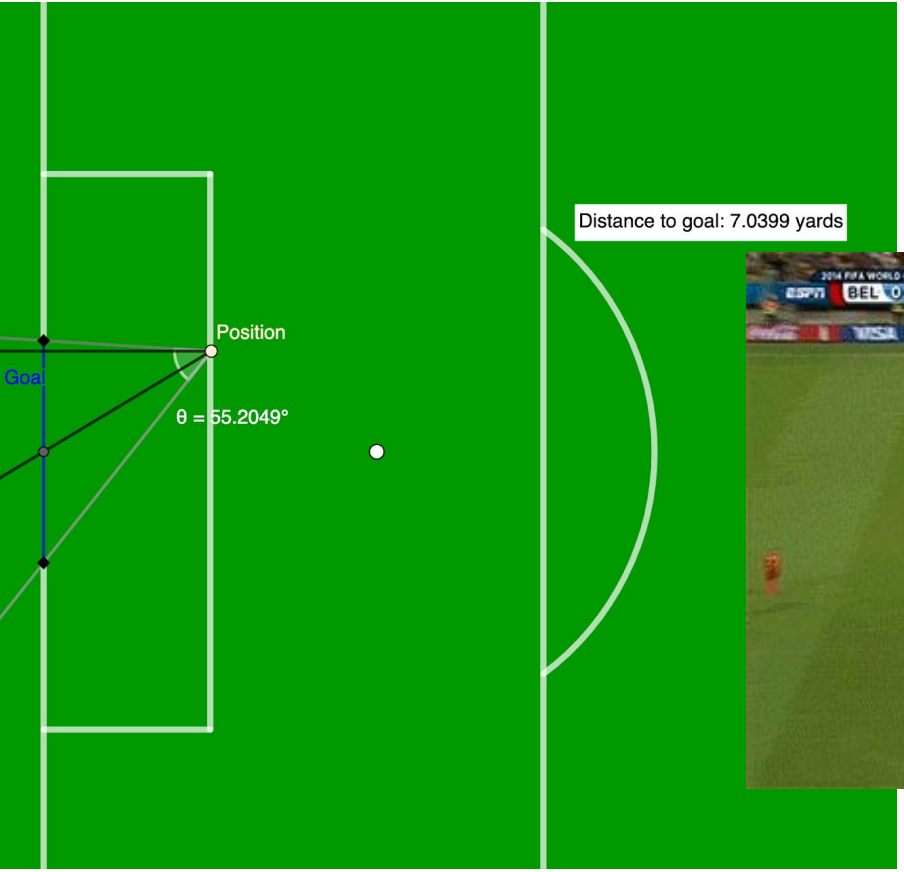
	coef	std err	P>|z|	[0.025	0.975]
Intercept	-3.4147	0.025	0.000	-3.463	-3.366
dist_std*	-0.9403	0.018	0.000	-0.975	-0.906
ang_std*	0.3645	0.011	0.000	0.343	0.386
kicked	1.0043	0.023	0.000	0.960	1.049
side_of_field_matching_foot	0.0160	0.010	0.122	-0.004	0.036
corner	-0.0728	0.026	0.005	-0.123	-0.022
free_kick	-0.0501	0.031	0.103	-0.110	0.010
free_kick_cross	-0.0241	0.037	0.516	-0.097	0.049
free_kick_shot	-0.0930	0.092	0.310	-0.272	0.086
goal_kick	0.0496	0.043	0.248	-0.035	0.134
penalty	0.1793	0.185	0.333	-0.183	0.542
throw_in	-0.0563	0.024	0.019	-0.103	-0.009

*Normalized centered with a Standard Scaler

Model Performance

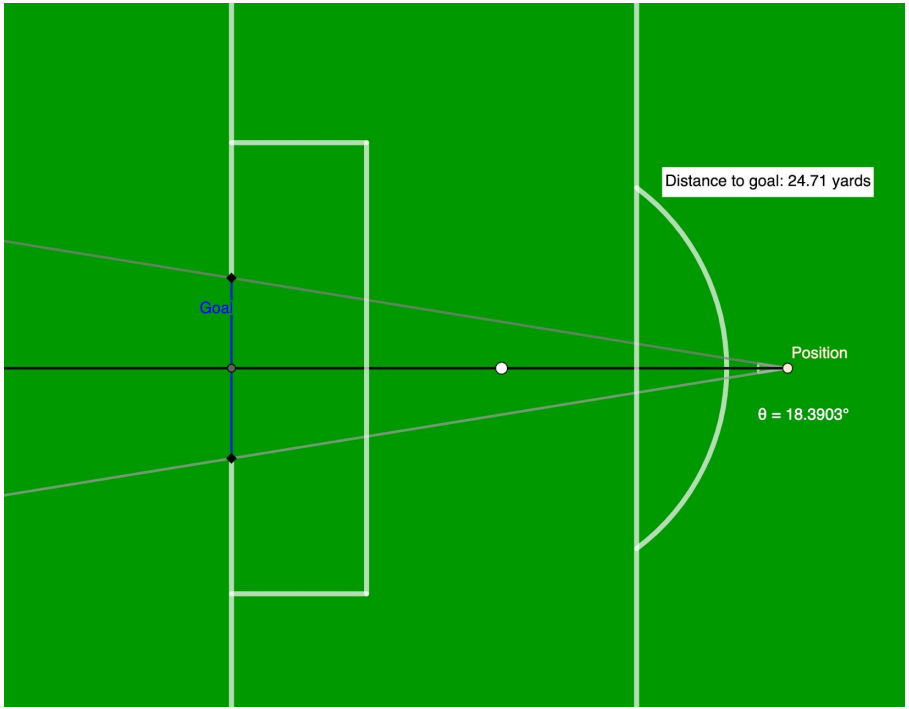


Wondo's Miss



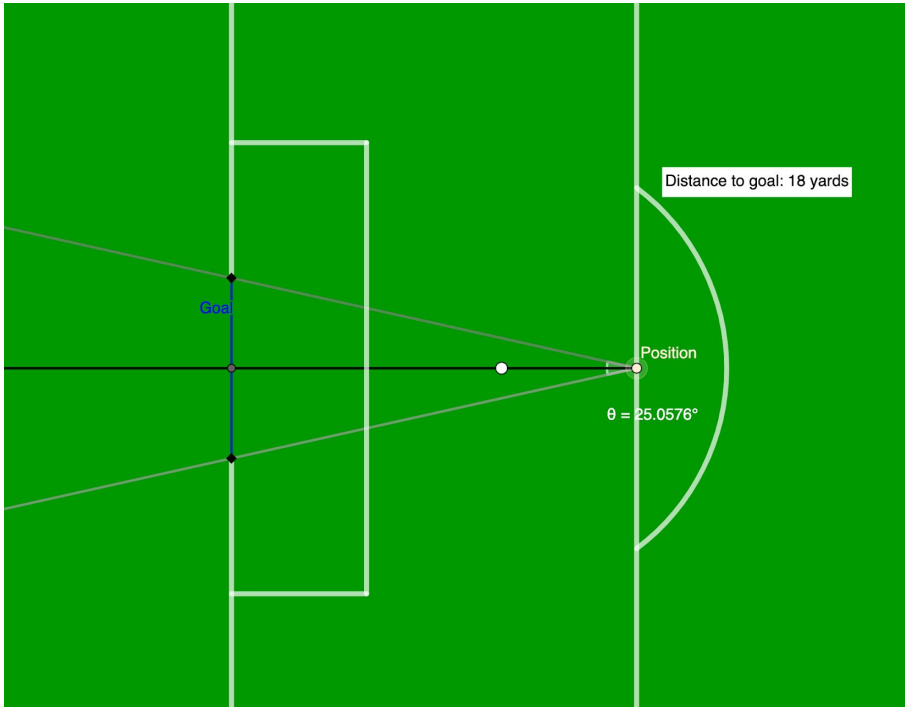
- 62% Chance
  - Throw in less than 30s ago
  - Foot matching Side
  - Kick

Model Predictions:



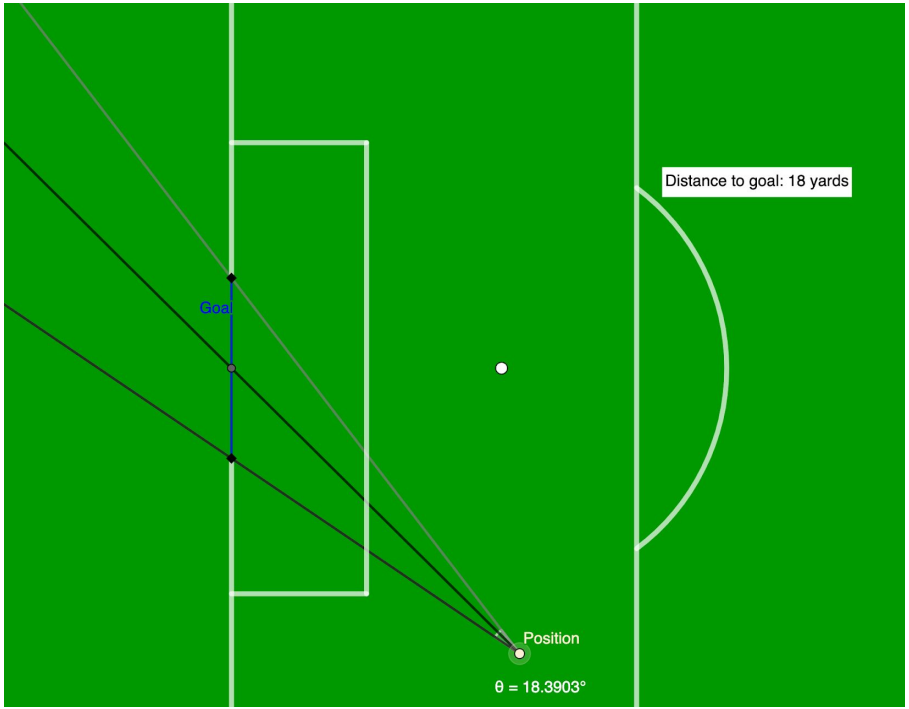
- 19% Chance
  - No other factors

Model Predictions:



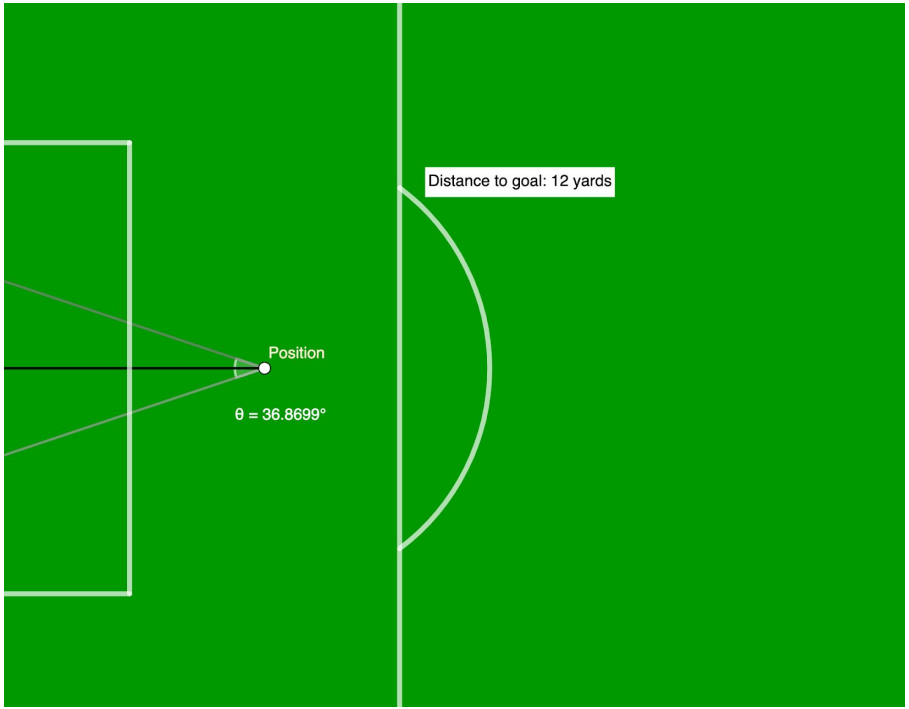
- 23 % chance
  - No other factors

Model Predictions:



- 20% chance
  - No other factors

Model Predictions:



- 47% chance
  - No other factors

Other model performances

Although many out performed the logistic regression, due to the lower of interpretability the none of these were chosen

Model	F1 Performance on Validation
Random Forest	.484
Adaboost	.163
Extra Trees	.477
Logistic Model	.368