# Throw Ins: Gain the Upper Hand

**Rachel Yang & Anna Hartman**

## Table of Contents



## 1. Introduction

Throw-ins: the one time players are able to hold the ball. Our goal in this project is to analyze the success of throw-ins during soccer matches using throw-in location and other factors such as pass length, pass height, pass angle, and which player throws the ball in. There is a general lack of research on throw-in plays. While throw-ins are generally considered to be the least important set piece, we view them as a valuable, frequent occurrence – pretty much the only one that allows players to pause gameplay and pick up the ball.

## 2. Related Work

We had set out to find parts of the game of soccer that had not been focused on much in prior research or in general game strategy. To that end, we based some of our research on the paper"The undervalued set piece: Analysis of soccer throw-ins during the English Premier League 2018–2019 season" (Stone, Smith, and Barry, 2021) and aimed to expand upon their research.

## 3. Experiments and Results

### 3.1 Data

We mainly used the Statsbomb events data. The data included detailed features such as the types of passes, location of passes, length, height, and angle of passes, and more.

We define **success** a couple of different ways: One being **whether the team executing the throw-in successfully makes first contact with the ball**, another is **whether the team executing the throw-in retains possession for at least 7 seconds following the throw-in**, and last is **whether the throw-in leads to a shot opportunity**.

We define **failure** (for each success scenario) as an **<u>incomplete first contact by the throw-in team</u>**, when **<u>the throw-in team receives the ball but loses possession in less than 7 seconds</u>**, or if **<u>no shot chance is created</u>**.

Note that In our results, we combine the results of **possession** and **shot creation** into one section. We also want to note that we use the terms “incomplete” and “unsuccessful” interchangeably.

## 3.2 Methods

For our project, we used Jupyter Notebook and Python as our programming language of choice. We mainly used Pandas Dataframe to parse through and filter our data. We used statsmodel.api and scikit-learn to conduct our linear regressions. We then used matplotlib, mplsoccer, seaborn, and prettytable to visualize our data.

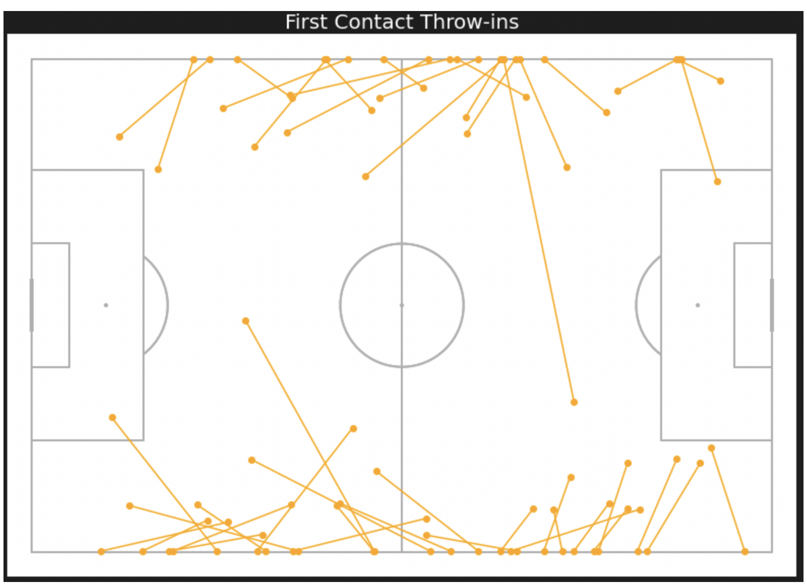
Before anything, we compiled all of our data into one csv file (full_data_sorted). This file includes all event data for all matches. We then converted this csv file into a pandas dataframe. From here, we filtered through our data to only get plays that were from throw-ins. At this point, our data is prepared for analysis.

## 3.3 Analysis & Results

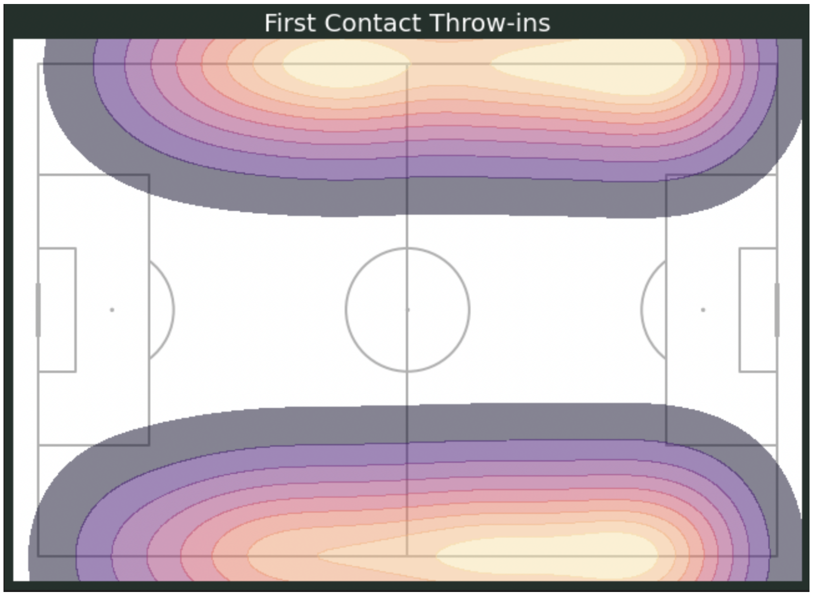
### 3.3.A Success Defined as First-Contact by Throw in Team

We begin by exploring our data for **<u>success defined by the throw-in team making first contact</u>**. This allowed us to familiarize ourselves with relevant variables and establish a strong foundation for the same analysis across *all matches* in the dataset.

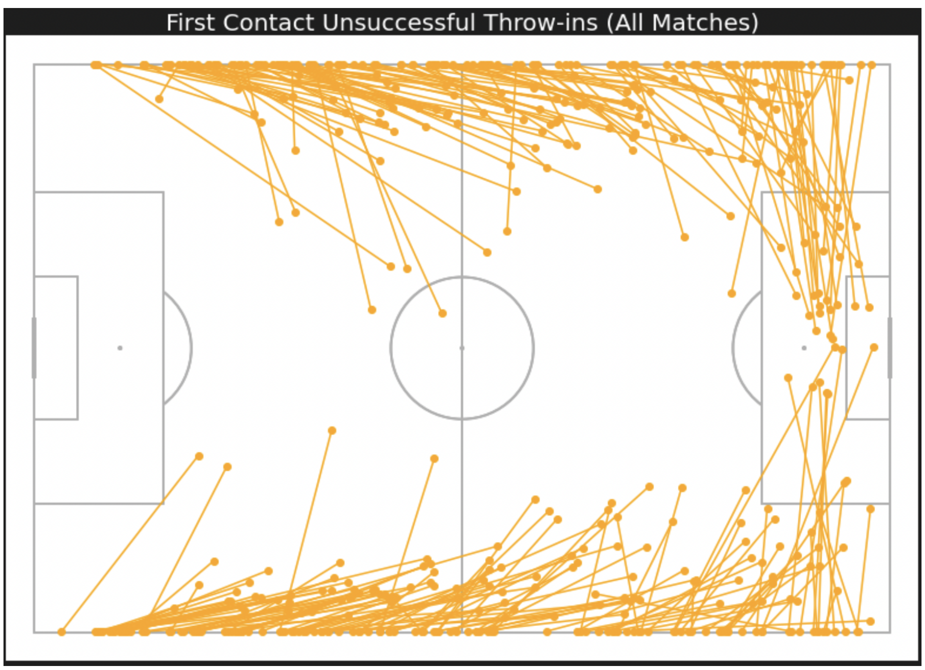
The diagram below shows starting, ending, length, and direction of throw-ins from *one match*. It shows that throw-ins are clustered around the middles of each sideline.



The following heat map confirms this trend among *all matches*, showing that the majority of throw-ins occur near the middles of each sideline.




Below is a diagram of the unsuccessful throw-ins across *all matches*. We will investigate a series of variables that might predict our results. Let's begin!



Out of 1948 throw-ins, 256 throw-ins (shown above) were incomplete. This represents a 13.14% failure rate of throw-in first contact.

| Total Throw-ins | Sucessful First Contact | Unsucessful First Contact |
| --- | --- | --- |
| 1948 | 1692 | 256 |

When looking at throw-in **lengths**, we defined a long throw-in as one that was longer than 18.9 yards, the average length of a throw-in. We found that there were a total of 855 long throw-ins and 1093 short throw-ins. 21.75% of long throw-ins were incomplete compared to only 6.4% incomplete short throw-ins.

| Throw-in Length | Total | Failure Rate |
| --- | --- | --- |
| Long Throw-ins | 855 | 21.75% |
| Short Throw-ins | 1093 | 6.4% |

As for **height**, there were a total of 513 high throw-ins and 1435 low throw-ins. However, 37.82% of high throw-ins were unsuccessful, and only 4.32% of low throw-ins were incomplete. This leads us to conclude that long, high throw-ins are riskier than short, low throw-ins.

| Throw-in Height | Total | Failure Rate |
| --- | --- | --- |
| High Throw-ins | 513 | 37.82% |
| Low Throw-ins | 1435 | 4.32% |

We now move on to the **angle** of throw-ins. The angle of the throw-in pass is in radians with 0 pointing towards the attacking team's goal, positive values indicating a clockwise angle, and negative values indicating a counterclockwise angle. We defined a lateral throw-in as one between -1 and 1 radian, a backward throw-in as less than -1 radians, and a forward throw-in as greater than 1. The amount of lateral, backward, and forward throw-ins were similar, with lateral being the most common. However, lateral throw-ins also had the highest failure rate of 26.34% compared to backward and forward throw-ins with 5.71% and 4.75% respectively.

| Throw-in Angle | Total | Failure Rate |
|---|---|---|
| Lateral Throw-ins | 729 | 26.34% |
| Backward Throw-ins | 630 | 5.71% |
| Forward Throw-ins | 589 | 4.75% |

Lastly, we looked at the **location** of throw-ins. We split the location into 2 basic areas: penalty and attacking. The penalty area is from 0 to 18 yards and 102 to 120 yards. The attacking area is from 18 to 102 yards. We discovered that although less balls are thrown into the penalty area, the failure rate is higher at 17.55%.

| Throw-in Location | Total | Failure Rate |
|---|---|---|
| Penalty | 359 | 17.55% |
| Attacking | 1589 | 12.15% |

At this point, let's **tabulate results to recap all our variables: length, height, angle, and location.**

We are now able to see that short, low throw-ins tend to be more successful than long, high throw-ins. Additionally, backward and forward throw-ins are more successful compared to lateral throw-ins. Lastly, throwing the ball in the attacking area leads to a higher chance of a successful throw-in.

| Throw-in Feature | Total | Complete | Incomplete | Sucess Rate |
|---|---|---|---|---|
| Long Throw-ins | 855 | 669 | 186 | 78.25% |
| Short Throw-ins | 1093 | 1023 | 70 | 93.6% |
| High Throw-ins | 513 | 319 | 194 | 62.18% |
| Low Throw-ins | 1435 | 1373 | 62 | 95.68% |
| Lateral Throw-ins | 729 | 537 | 192 | 73.66% |
| Backward Throw-ins | 630 | 594 | 36 | 94.29% |
| Forward Throw-ins | 589 | 561 | 28 | 95.25% |
| Penalty (Location) Throw-ins | 359 | 296 | 63 | 82.45% |
| Attacking (Location) Throw-ins | 1589 | 1396 | 193 | 87.85% |

Now, to model the relationship(s) between each **independent variable** and **success (as defined by first contact made by the throw-in team)**, we conduct linear regressions.

We expect that for every one radian increase in **angle**, there is a 0.0014 decrease in the probability of throw-in success. Based on our r-squared value of 4.8253772105821824e-05, throw-in angle explains an extremely small proportion of the variation in throw-in success. Because the p-value is greater than 0.05, our data is not statistically significant.
This is not the best result, so we attempt to do linear regression on throw-in length!

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           pass_outcome   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                   0.09391
Date:                Mon, 01 Aug 2022   Prob (F-statistic):              0.759
Time:                        01:09:32   Log-Likelihood:                -746.32
No. Observations:                1948   AIC:                             1497.
Df Residuals:                    1946   BIC:                             1508.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1478      0.008     18.364      0.000       0.132       0.164
pass_angle    -0.0014      0.005     -0.306      0.759      -0.010       0.008
==============================================================================
Omnibus:                      666.079   Durbin-Watson:                   1.935
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1582.630
Skew:                           1.984   Prob(JB):                         0.00
Kurtosis:                       4.937   Cond. No.                         1.78
==============================================================================
```

We expect that for every one yard increase in throw-in **length**, there is a 0.0113 increase in the probability of throw-in success. Based on our r-squared value of 0.0867, throw-in length explains a small proportion of the variation in throw-in success. However, it is greater than throw-in angle. The p-value is less than 0.05, so our data is statistically significant!
This is better, so now let's try throw-in height!

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           pass_outcome   R-squared:                       0.087
Model:                            OLS   Adj. R-squared:                  0.086
Method:                 Least Squares   F-statistic:                     184.8
Date:                Mon, 01 Aug 2022   Prob (F-statistic):           2.83e-40
Time:                        01:09:33   Log-Likelihood:                -658.01
No. Observations:                1948   AIC:                             1320.
Df Residuals:                    1946   BIC:                             1331.
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          -0.0662      0.018     -3.776      0.000      -0.101      -0.032
pass_length     0.0113      0.001     13.594      0.000       0.010       0.013
==============================================================================
Omnibus:                      572.688   Durbin-Watson:                   1.925
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1227.108
Skew:                           1.738   Prob(JB):                    3.44e-267
Kurtosis:                       4.741   Cond. No.                         47.9
==============================================================================
```

We expect that for every one yard increase in <u>**height**</u>, there is a 0.3471 decrease in the probability of throw-in success. Based on our r-squared value of 0.185, throw-in height explains a modest proportion of the variation in throw-in success, greater than both throw-in angle and throw-in length. The p-value is less than 0.05, so our data is statistically significant! Now, we move on to the throw-in location!

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           pass_outcome   R-squared:                       0.185
Model:                            OLS   Adj. R-squared:                  0.185
Method:                 Least Squares   F-statistic:                     443.1
Date:                Mon, 01 Aug 2022   Prob (F-statistic):           8.56e-89
Time:                        01:09:34   Log-Likelihood:                -546.55
No. Observations:                1948   AIC:                             1097.
Df Residuals:                    1946   BIC:                             1108.
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.4035      0.014     28.515      0.000       0.376       0.431
pass_height    -0.3471      0.016    -21.050      0.000      -0.379      -0.315
==============================================================================
Omnibus:                      468.788   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              928.662
Skew:                           1.431   Prob(JB):                    2.21e-202
Kurtosis:                       4.803   Cond. No.                         3.67
==============================================================================
```

We expect that for every one yard along the side of the field—**<u>location</u>**—there is a 0.0013 decrease in the probability of throw-in success. Based on our r-squared value of 0.011, throw-in location explains a rather small proportion of the variation in throw-in success, despite still being greater than throw-in angle. **Nonetheless, individual linear regressions ignore the fact that all four variables—not just one alone—may be predictive of throw-in outcome.**

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           pass_outcome   R-squared:                       0.011
Model:                            OLS   Adj. R-squared:                  0.010
Method:                 Least Squares   F-statistic:                     21.64
Date:                Mon, 01 Aug 2022   Prob (F-statistic):           3.51e-06
Time:                        01:09:35   Log-Likelihood:                -735.60
No. Observations:                1948   AIC:                             1475.
Df Residuals:                    1946   BIC:                             1486.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.2320      0.020     11.726      0.000       0.193       0.271
x_loc         -0.0013      0.000     -4.652      0.000      -0.002      -0.001
==============================================================================
Omnibus:                      654.812   Durbin-Watson:                   1.928
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1536.169
Skew:                           1.955   Prob(JB):                         0.00
Kurtosis:                       4.908   Cond. No.                         178.
==============================================================================
```

**To that end, we must perform multiple regression analysis, incorporating all of our variables to predict throw-in outcome**. Upon conducting the regression, we find that pass height and pass angle are two variables whose relationship with throw-in success is statistically significant (p-value ≤ 0.05). We expect that for each 1 yard increase in pass height, there is a 0.2971 decrease in the probability of throw-in success. And, for each 1 yard increase in pass length, we predict a 0.0054 increase in the probability of throw-in success. Pass angle and pass location were not statistically significant variables in predicting throw-in success. Overall, our model's r-squared is 0.202, which we believe to be quite high for real-world data.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           pass_outcome   R-squared:                       0.202
Model:                            OLS   Adj. R-squared:                  0.201
Method:                 Least Squares   F-statistic:                     123.2
Date:                Mon, 01 Aug 2022   Prob (F-statistic):           7.94e-94
Time:                        01:13:16   Log-Likelihood:                -526.15
No. Observations:                1948   AIC:                             1062.
Df Residuals:                    1943   BIC:                             1090.
Df Model:                           4
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.2855      0.030      9.573      0.000       0.227       0.344
pass_height    -0.2971      0.018    -16.350      0.000      -0.333      -0.261
pass_angle      0.0027      0.004      0.669      0.504      -0.005       0.011
pass_length     0.0054      0.001      6.250      0.000       0.004       0.007
x_loc          -0.0003      0.000     -1.210      0.226      -0.001       0.000
==============================================================================
Omnibus:                      452.091   Durbin-Watson:                   1.932
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              881.015
Skew:                           1.388   Prob(JB):                    4.90e-192
Kurtosis:                       4.774   Cond. No.                         330.
==============================================================================
```

We realize that linear regressions may not be the best model for our data. We are now searching for new models that may be a better fit. We also want to research more variables that may influence the success rate of a throw-in. The data shows that the coefficient constant is 0.202 which indicates that there are other variables that have potential to influence the success of a throw-in!

### 3.3.B First-Contact Success by Player Making Throw-in

We conducted further analysis on throw-ins per team and player. We found the player that throws the **most balls per team and analyzed their throw-in styles**. For example, take a look at Benjamin Pavard from France. In one competition season, he threw in 24 total balls with 4 incomplete throw-ins for a success rate of 83.33%. We have also included a couple of other players.

| Player | Team | Total Throw-ins | Total Incomplete Throw-ins | Success Rate | Failure Rate |
|---|---|---|---|---|---|
| Benjamin Pavard | France | 24 | 4 | 83.33 | 16.67 |

| Player | Team | Total Throw-ins | Total Incomplete Throw-ins | Success Rate | Failure Rate |
|---|---|---|---|---|---|
| Mehmet Zeki Çelik | Turkey | 18 | 0 | 100.0 | 0.0 |

| Player | Team | Total Throw-ins | Total Incomplete Throw-ins | Success Rate | Failure Rate |
|---|---|---|---|---|---|
| Stephen O"Donnell | Scotland | 27 | 11 | 59.26 | 40.74 |

When looking specifically at each players' throw-ins, we can observe **characteristics like length, height, and angle.** Similarly, we were able to find these data points for each player on the team.

| Kamil Jóźwiak | | |
|---|---|---|
| Throw-in Length | Throw-in Height | Throw-in Angle |
| ('Long Pass', 7)<br>('Short Pass', 12) | ('Low Pass', 16)<br>('High Pass', 3) | ('lateral', 5)<br>('backwards', 14)<br>('forwards', 0) |

| Giovanni Di Lorenzo | | |
|---|---|---|
| Throw-in Length | Throw-in Height | Throw-in Angle |
| ('Long Pass', 19)<br>('Short Pass', 37) | ('Low Pass', 43)<br>('High Pass', 13) | ('lateral', 19)<br>('backwards', 36)<br>('forwards', 1) |

| Mário Figueira Fernandes | | |
|---|---|---|
| Throw-in Length | Throw-in Height | Throw-in Angle |
| ('Long Pass', 12)<br>('Short Pass', 6) | ('Low Pass', 10)<br>('High Pass', 8) | ('lateral', 9)<br>('backwards', 9)<br>('forwards', 0) |

These are great insights for teams and their opponents. Teams can use this data to see if their throw-ins tend to advance the ball forward, for instance, or assess if more drills are needed to control high throw-in passes. In addition, opponents can analyze the tendencies of certain players, and set up their defense accordingly. If a player with a low success percentage is determined, opponents can exploit this weakness for defensive gain.

### 3.3.C Success Defined as 7 Seconds of Possession and Shot Creation

In this section, we analyze success defined as retaining possession for at least seven seconds based. Rather than taking into account our prior four variables, we focus **only on location as it is a proxy for the degree of defensive pressure that might force a turnover of the ball.**

Separately, we **will also analyze success defined as shot creation since, like possession, it involves subsequent action after first contact is made.**

We found that teams throwing the ball were about equally likely to retain possession in both the penalty areas (67.59%) and middle areas (65.04%) of the field.

| Throw-in Location | Successes | Total | Success Rate |
|---|---|---|---|
| Penalty | 1047 | 1549 | 67.59% |
| Attacking | 227 | 349 | 65.04% |

Before completing a linear regression, we suspected that it might not model our data well. This is because a soccer field has two goals, so we would expect rates of successful possession to be lower at both respective ends, not a display of linear increase from one end to the other. Once we performed the linear regression, our hypothesis was confirmed. The r-squared value was 0.0147 which means the model explains a very, very small proportion of the observed variation in success rate.

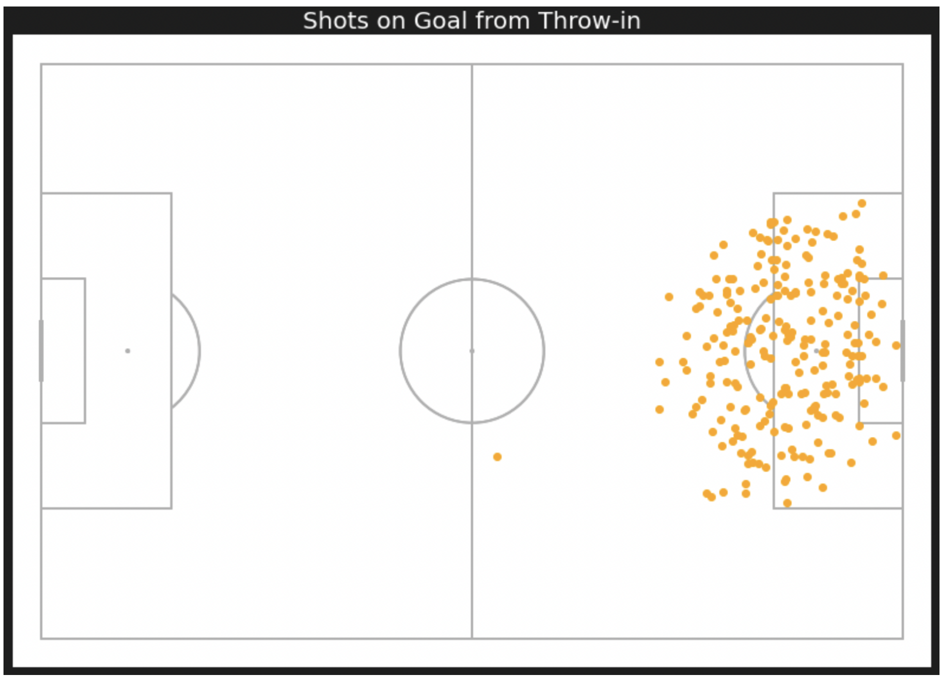
```
r_sq = model.score(x, y)
print(f"coefficient of determination: {r_sq}")
print(f"intercept: {model.intercept_}")
print(f"slope: {model.coef_}")
```

```
coefficient of determination: 0.014712869011338503
intercept: 0.5429441047435098
slope: [0.00195444]
```

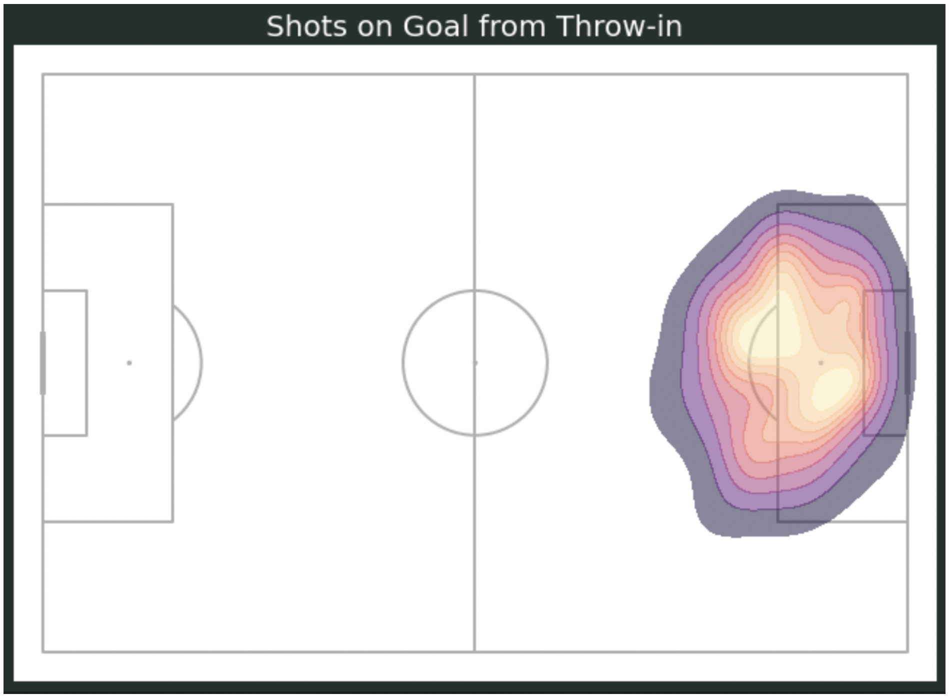
```
y_pred = model.predict(x)
print(f"predicted response:\n{y_pred}")
```

Now, we move on to analyzing shots on goal from throw-ins. We found that there were 1289 shots on goal total across *all matches*. 230 of these shots were from a throw-in play. This means 17.8% of shots on goal were from throw in plays. This suggests that not only can throw-ins lead to scoring opportunities, they have before.

Here we diagram shots on goal from throw-ins (assume this is for one of the teams).




This is a heat map representation of the previous diagram. The strategy for teams throwing the ball should be to create a play where the ball will end up near areas where shots tend to be clustered around (the light yellow region).




## 4. Practical Applications

There are several immediate uses for our work. Defensively, teams can strategize when clearing the ball out of play – they may be able to control what section of the sideline they kick it out at, allowing them to direct the throw-in to areas of lower success for the throw-in team. Offensively, our work demonstrates some of the strategies for throwing the ball in successfully when it comes to variables like angle and distance, as well as giving data on which players are best at performing throw-ins. Moreover, teams can adjust their throw-in set pieces, leading to more scoring opportunities. Coaching staff can also analyze the throw-ins of players on other teams who often throw the ball in and use our data to predict their tendencies, such as throwing in the ball low or high, in order to best defend against their throw-in. Our data can also help determine which players are best at throw-ins and others who may need some practice.

## 5. Future Work

To extend our project, we would like to continue exploring models that more accurately represent our data; we realized that linear regressions might not provide the best fit. We also want to explore other markers of success for the throw-in team, such as different throw-in strategies that result in shots on goal. Moreover, much of our analysis focused on the player throwing in the ball as well as the recipient, but it would also be interesting to analyze the position of all surrounding players since they are likely to be involved in any subsequent passing. These insights could allow the opposing team to now strategically set up their defense to defend against these surrounding attackers. Lastly, we would create visualizations to accompany these new pieces of information.

## References

"The undervalued set piece: Analysis of soccer throw-ins during the English Premier League 2018–2019 season" (Stone, Smith, and Barry, 2021)

## Acknowledgements

Thank you to Women in Sports Data for providing this opportunity for us to research this topic. We also want to thank our mentor, LJ Rader, for helping us throughout our process. Lastly, we want to thank Statsbomb for the ample amount of data they provided us with to do our research!