

An Immovable Object Meets an Unstoppable Force:

Does Defense Produce Success in International Football?

Kyle Tan (500564538)

Supervised by Tamer Abdou

2/xxx/2023

Table of Contents

[Revised Abstract 4](#_Toc127609849)

[Foreword: 4](#_Toc127609850)

[Abstract: 4](#_Toc127609851)

[Literature Review 5](#_Toc127609852)

[Defining Research Questions: 5](#_Toc127609853)

[Methodology 7](#_Toc127609854)

[Defining Defense 7](#_Toc127609855)

[Data Collection + Cleaning: 7](#_Toc127609856)

[First Iteration: 9](#_Toc127609857)

[Second Iteration 14](#_Toc127609858)

[Third Iteration 16](#_Toc127609859)

[Index 23](#_Toc127609860)

[Variable Details 23](#_Toc127609861)

[Bibliography 25](#_Toc127609862)

# Revised Abstract

## Foreword:

This project a live work in progress, as such, things change. New information may come to light, old assumptions may be put into question. As more research, the initial exploratory analysis, and a more critical look at the information provided is completed, so changes some of the details previously discussed in the abstract. Below is a revised project abstract, with respect to exploratory analysis, overall report structure, and critical reading on existing works.

## Abstract:

“Attack wins you games, defense wins you titles”, these are the words of one of the most recognized football managers, Sir Alex Ferguson (Smith, 2017). This theory is an often-said rhetoric that spans all sorts of sports (Davis & Suryawanshi, 2023). However, with the scope of this project, this theory will be analyzed through the scope of international football (soccer). It needs to be recognized that the psychology, challenges, and difficulties in each different competition varies quite a bit. League football, for example, takes place over months, with different levels of support, control, time, and accountabilities than cup football. Time has passed since Sir Alex said the famous quote. Since then there has been major developments and changes in the dogma of both club and international football, from the rise of possession seen in Vincent Del Bosques’ successful Spain side (MARCA, 2012) to the rise in popularity and success in positional play seen in Pep Guardiola’s Manchester City (Breaking the Lines, 2022) all of which leads away from a traditionally “defensive” mindset. Despite the success and popularity of these developments, we saw in the most recent world cup, teams with fantastic defense records like Argentina and Morocco saw success and trophies (FOX Sports, 2023). The question this project aims to answer is: does the popular rhetoric still apply to today’s international cup football competitions such as the World Cup? Using data provided from FBREF.com, this project will aim to use predictive analytics techniques to find pattern and correlations in the matches of two of most recent champions (USA of Women’s World Cup 2019, and England of Women’s Euro 2022). Techniques proposed here are mainly to detect any sort of patterns and trends within the champions’ journey to victory. As defensive statistics for soccer is an incredibly complex and understudied area (Winterburn, 2017), this project will also have a further aim to develop a set of reliable way to quantify defensive statistics of a team.

# Literature Review

## Defining Research Questions:

This project aims to answer one simple question: Does defense win titles in the context of international cup football? With more research done into previously completed projects and works, the question needs be further fine tuned and expanded.

Analytics within the world of football is relatively young and underdeveloped as compared to the other world of sports like Basketball, Baseball, and even American football (Anderson & Scally, 2013). One big question that was generated through the research process is: What defines defense? Other big leagues and sports like the NBA (NBA, 2023) and the NFL (NFL, 2023) have well defined, standardized, and an accepted set of defensive stats that help tell the story of a team’s defensive performance. Perhaps this is due to the way the games are played, perhaps this is due to a lack of research. While analysis for defensive stats and simply stats such as xGA (expected goals against) are developed, they still provide a bit of a simple approach towards looking at the complexity that is the defensive system in a football team. There has been attempts to develop a sound view to look at defensive by previous experts (Winterburn, 2017) (Brownell, 2013), they all agree on one fact: there is much works to be done, and there’s much more than meets the eye.

One of the attempts to answer this question, and an inspiration for this project was done by popular football analytics YouTube channel Tifo Football. In their attempt to tackle the question they looked at a model that involved analyzing historical performances of the English Premier League teams with respect to goals scored and conceded (TIFO Football, 2020). They also aimed to look at the effects of winning with and without the key defenders. Overall, the stats they used involved looking at goals. The analysis done by Winterbrown and Brownell focused on looking at the stats of tackles, interceptions, and clearances. Perhaps these are the good stats to use to find a link between defense and success. Further steps will be taken to develop a robust set of stats to tell the story of defense. This may produce some challenge as defense is a bit of a subjective topic within football.

TIFO had concluded through their historical evidence and third party evidence that within league football, there isn’t a strong correlation that says defense indeed does win titles, with 64% of titles won by the higher scorers, 41% won by the least conceded, and 70% won by the team with the best difference (TIFO Football, 2020). If anything, a balanced of developing both attack and defensive is favorable.

Winning and losing, and the experience of winning titles is not unique to international cup football. Although the games differ, the psychology, business operations, and the wheel that crank out champions for other games have things in common. Other research done into the link between defense and success cannot be ignored. Especially since much more academic work has been done.

In their research, Robst et al, looked at the link between defense and their success in the world of American football (Robst, VanGilder, Berri, & Vance, 2011). With their work, they aimed to help football teams answer what’s more investing into? Defensive players or the offensive, as there is an existing salary cap that exists in the NFL that lead teams to have this question. With their model, they used the well-defined pre-existing defensive records of the NFL to develop a matrix that showed there is a difference between playoff success for teams that more offensively inclined vs. the ones that are defensively inclined. Although the difference is incredibly minute, it favors the offense. Overall, they found very little evidence teams benefit from focusing on one or the other. The differences are marginal. With respect to American football, focusing on both is important.

Another similar to attempt to look at the link between success and defense was a thesis completed by Thomas Burkett (Burkett, 2021). Within his research he developed QQ plots to analyze the link using, once again, the well developed set of stats available for the NBA. With is work, it is also concluded that a mix of defense and offense is the key.

With respect to all that’s been researched before. This project and its goals hasn’t been exactly replicated by any experts in the past. This project aims to fill in the gap that exists for the defensive analysis of international cup football. Most work done in the past has been relative to other sports, or league football. This work is worth doing in the light of what’s already been done as it helps shed light on the minute differences in cup football. It also aims to help develop a further understanding in defensive statistical analysis for football overall.

# Methodology

## Defining Defense

Defense in football is a fairly subjective topic. How a team defends and keeps a clean sheet varies greatly depending on their systems, players, and plans.

For the scope of this project, we define defense as the style of play where the focus is for the players to stay back in their own 3rd, get compact, hold their lines with little to no emphasis of ball retention, possessional play. Simply put, we define defense as teams who ‘park the bus’, a style of play described as a team that sits behind the ball in an effort to block the goal, moving the defensive pressure back closer to the goal, and allowing opposition to have possession (Young, 2015).

## Data Collection + Cleaning:

The main strategy and idea of this project is to, given the limited resources the researcher has, collect data from sources online, clean it, and find patterns and correlations between performance stats. An extra aim is here to create a “robust measure for rating defense”.

All stats and measures are measured and compared between the sampled teams to help paint a picture of how each team performs.

Data used in this project is collected from FBREF.com.

The initial step of this project is collection and the organization of raw, open-source data. The data is essentially turned from just a simple collection of related, useful numbers into a meaningful and relevant set of statistics that can be used to the purpose of this project’s specific analysis.

First, knowledge was gained in regards to the topic of the project – the Women’s World Cup 2019. As there are 24 participating teams, only a sample of the teams are chosen. The scope of this project will aim to look at only the relatively successful teams, in this case we are looking at the top 4 teams, and the runner ups (FBREF, 2023). The teams are listed below in order:

1. USA
2. Netherlands
3. Sweden
4. England
5. Germany
6. France
7. Italy
8. Norway

The website is chosen as it provides a good starting point to create a robust set of defensive statistics. Their database includes a set of fairly detailed “defensive actions” for all players and teams. The first step in data cleaning was to download all the separate documents containing the respective stats for each of the 8 teams mentioned above.

Once this is collected, an aggregate was collected. Meaning, instead of looking at individual player stats, we will be looking at the team stats as well as the averaged team stats. This project does not aim to look at the individual contributions, but rather the overall team performance. Once the sum and averages for each team is collected, they are combined into one data frame to hold the relevant team data.

For the sake of future comparisons, two data frames were create: one data frame is created that holds purely just the team stats, and one is created that holds teams stats along with average stats.

For each of the respective nations, we are also looking at both defensive performance for and against, as this helps paint a picture of how each teams performs against their opponents. It will help show relatively to their opposition is a team performing offensively, defensively, and how by how much.

An initial EDA report is created on this aggregated and combined “country defensive stats” data frame (with no averages included).

The EDA shows that near all of the stats participating is under the “real number” type as such, this project will involve working with a large number of numerical data. There are no missing data, as such this doesn’t need to be dealt with.

However, one item to note is that most of the stats are highly correlated to each other. This may produce problems later, and is noted.

Outlier detection was done on the data frame as well. First, the non-numerical column containing the country names (‘Players’) was removed to create a purely numerical data frame. Then this outlier detection/removal was done a basis of Z-score.

For the scope of this project, we will define an outlier as a value that’s beyond 3 z-scores (99.7%) of the mean. The filter applied rounds the Z-score and is also able to tell me stats that are near outlier status as well.

The filter on was done in Python and simply told me of which teams have an outlier. Further inspection was done in R to tell exactly which stat of which team is the outlier and what the respective z-scores are.

Based on the criterion above, with this iteration/dataframe the only outlier is the ‘error’ stat for England with a Z-score of 3.05. This stat will be removed altogether in the future steps. But this outstanding stat will be kept in mind for discussion.

## First Iteration:

Once the data is cleaned up and organized into something relevant and meaningful to the scope of this project, we will try to make meaning of the datasets.

The first comparison is to understand and check the summary of each of the column stats. This was provided on the EDA, and organized into the chart below:

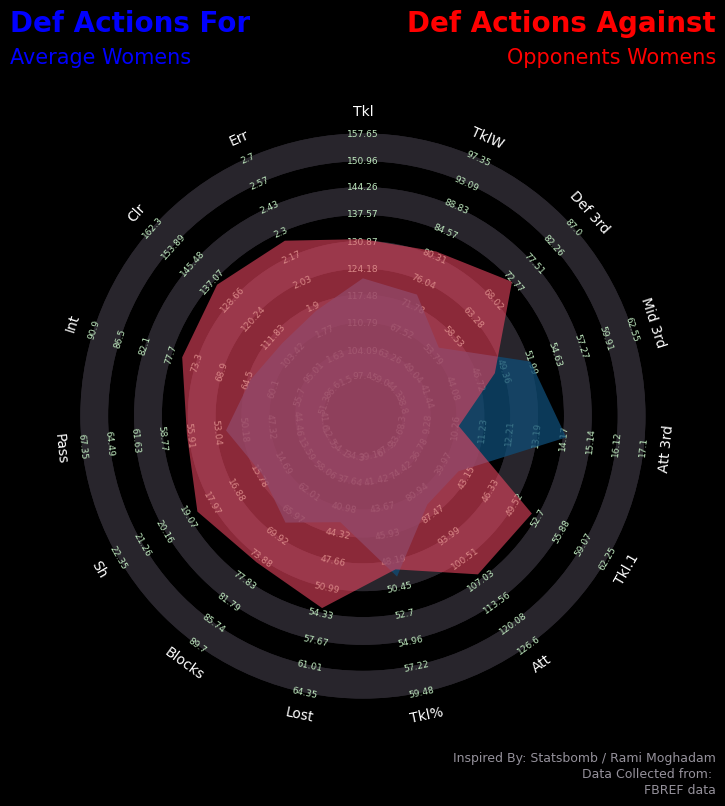
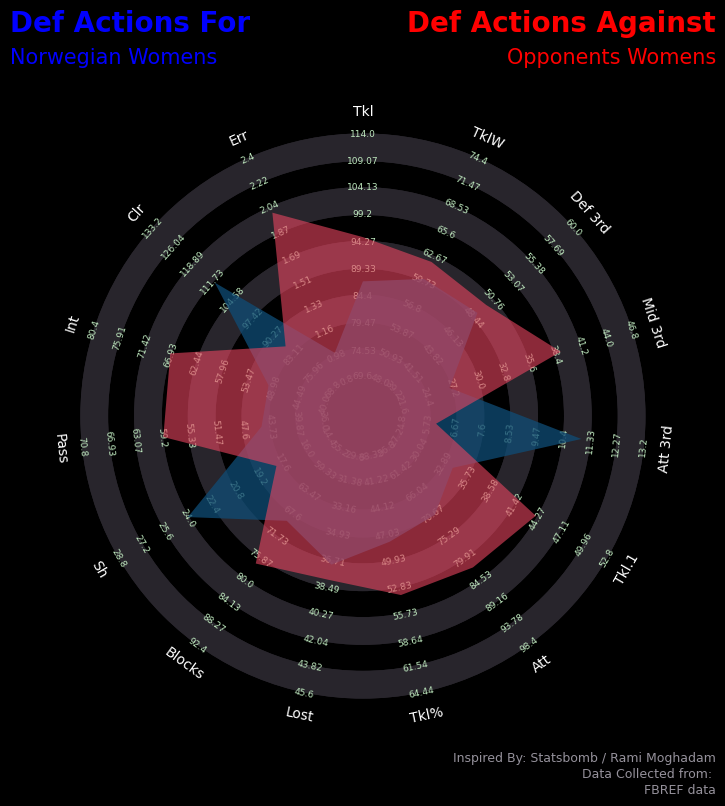
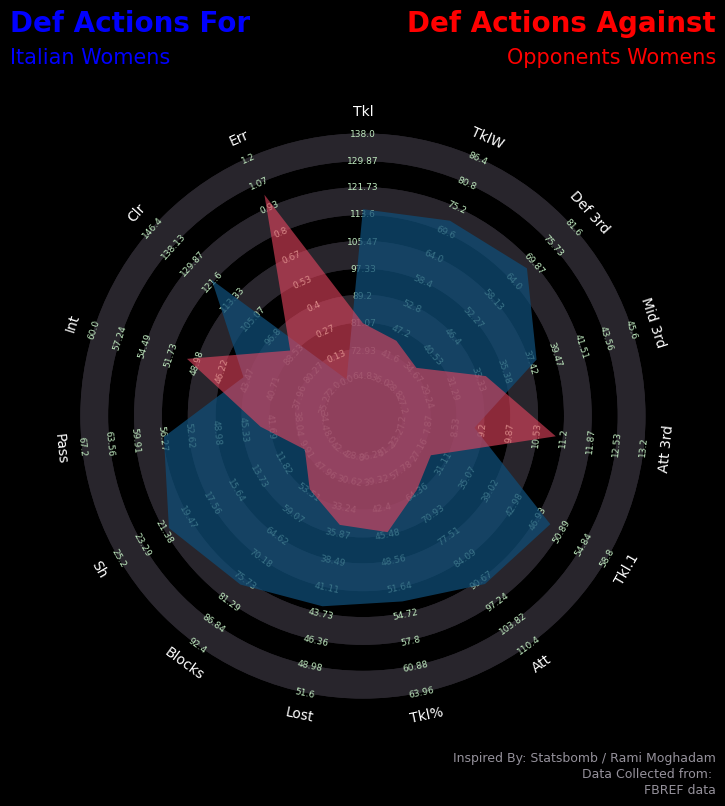
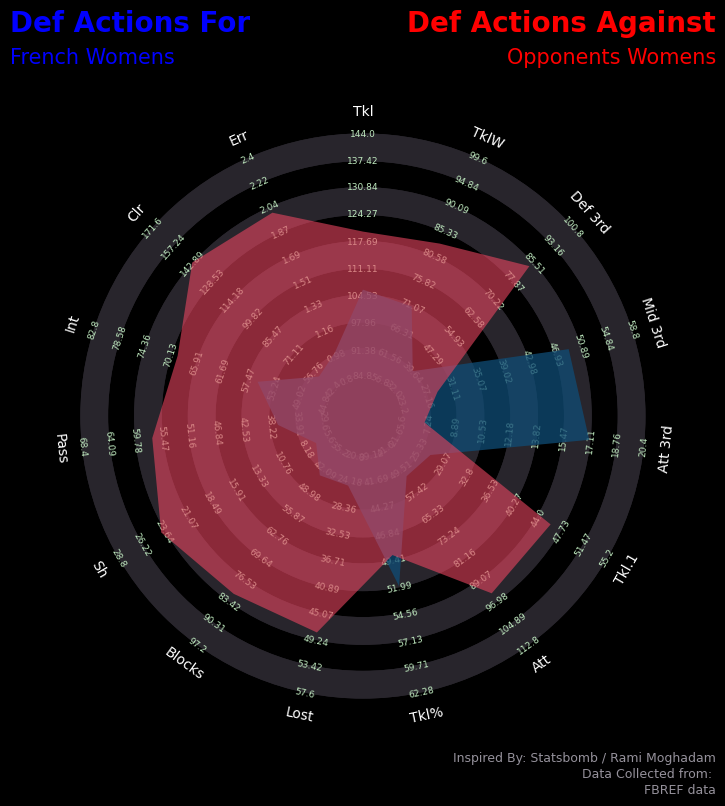
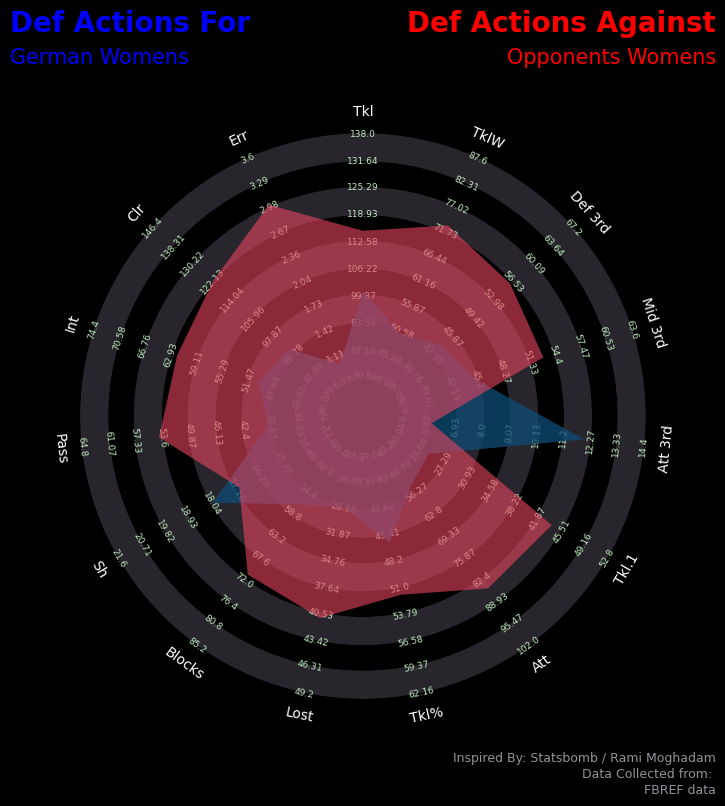
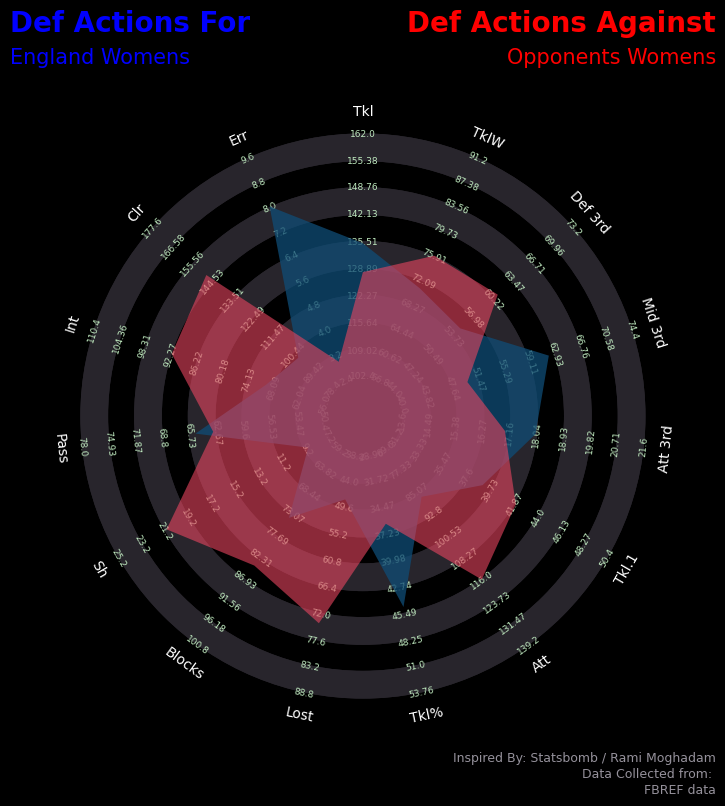
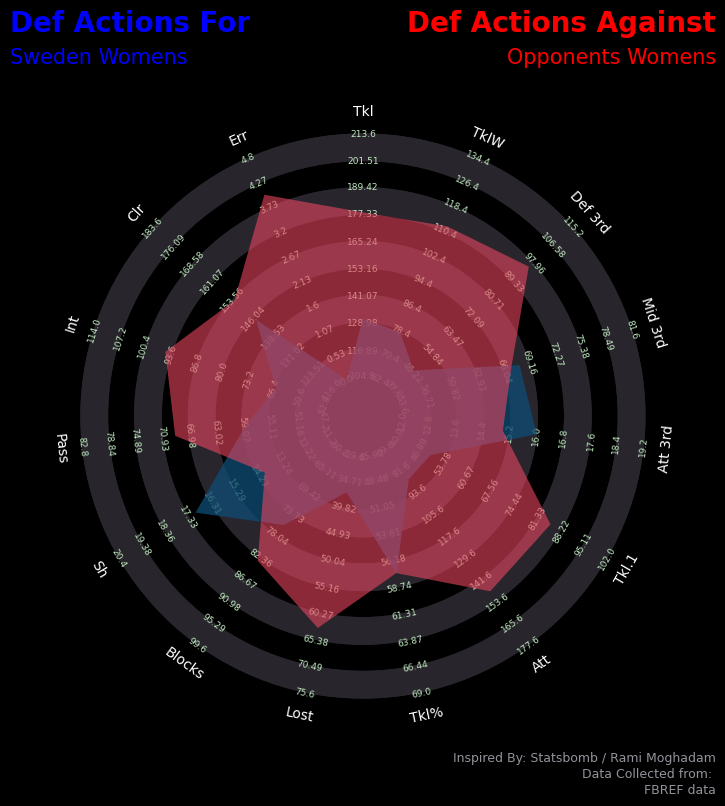
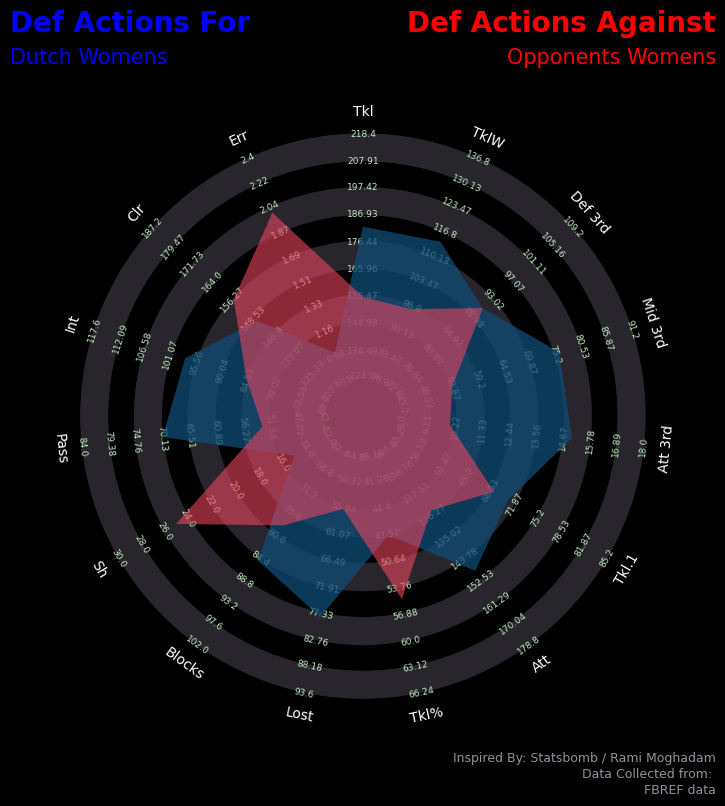
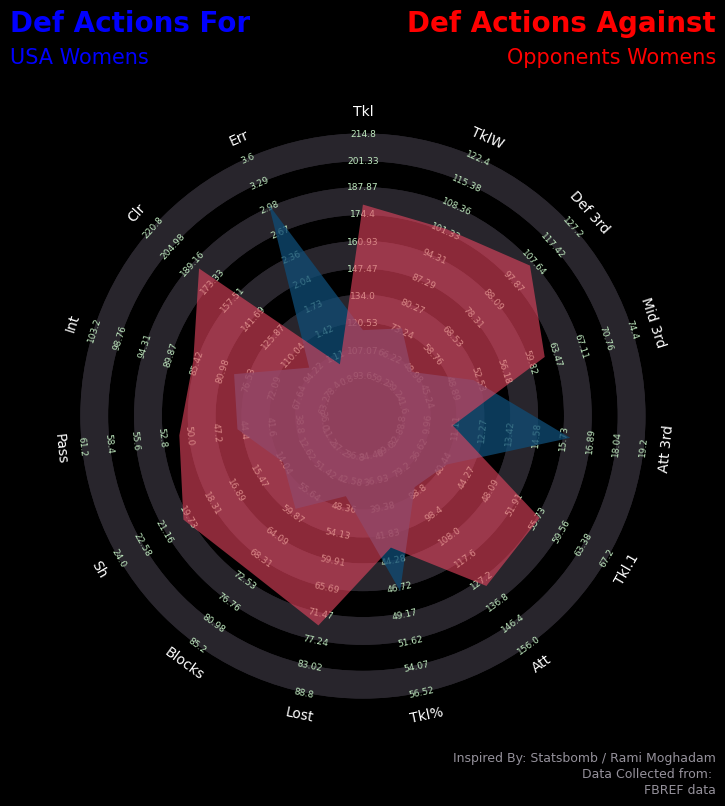
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stat | Stat Description | Mean | Min | Max |
| Age | Age | 27.09 | 25.5 | 28.8 |
| 90s | Total Mins played/90. Measure of games played. | Categorial as per EDA | Irrelevant stat anyways. | Will be removed in the future. |
| Tkl | # of Tackles | 126.56 | 81 | 182 |
| TklW | # of Tackles Won to get possession. | 77.44 | 45 | 114 |
| Def.3rd | # of Tackles Won in defensive 3rd | 63.94 | 36 | 106 |
| Mid.3rd | # of Tackles won in the middle 3rd | 50.31 | 27 | 76 |
| Att.3rd | # of Tackles won in the final 3rd | 12.31 | 6 | 18 |
| Tkl1 | # of dribblers tackled successfully | 46.94 | 25 | 85 |
| Att | # of dribblers tackled attempts | 95.19 | 52 | 149 |
| Tkl% | Tkl1/Att.  A measure of success % in tackling dribblers | 49.26 | 36.2 | 57.5 |
| Lost | # of drillbers tackled unsuccessfully | 48.25 | 25 | 78 |
| Blocks | # of times ball stopped by someone standing in its path | 71.19 | 44 | 85 |
| Sh | Shots Blocked | 17.13 | 7 | 25 |
| Pass | Passes Blocked | 54.06 | 37 | 70 |
| Int | Interceptions | 69.94 | 44 | 98 |
| Tkl+Int | Players tackles + interceptions | 196.5 | 131 | 280 |
| Clr | Clearances | 121.75 | 53 | 184 |
| Err | Errors | 2.06 | 0 | 8 |

From looking at the table above, there are some stats I do not want to include in future analysis. The Tkl+Int is a bit redundant as it’s simply a combination of two other simple stats, they don’t significantly contribute to the question posed by this project. Other stats such as 90s and age is removed as it’s just irrelevant to the purpose of this project.

Once this is completed, comparison diagrams are created to generate an easy visualization of each team’s performance relative to their respective oppositions. Each diagram is generated with respect to each team’s performance vs how other teams performed against them.

For the sake of easy comparison, two more “teams” are added to the rows: “average” and “averageOpp” to show the average defensive actions of the teams playing and how their opponents done against them, respectively.

The set of comparison diagrams are included below:



From looking at the comparison relative to how they perform vs their opponents, right away the overwhelming pattern is that for most of the teams, the opposition performs much more defensive actions than their respective partners.

This is especially evident in the average chart with the only the Mid.3rd and Att.3rd stats being performed at a greater rate. Both of which are poor indicators of defensive performance.

Italy is the only nation to nation to statistically perform more “defensively”, although you can also include the Netherlands/Dutch on that list.

Italy’s performances are the only one where the defensive actions are greater than that of their oppositions. With the Def.3rd, TklW, Blocks, Tkl1, and Clr being the very standout indicators.

The very key tell is the Def.3rd stat. Italy has the highest relative number of actions completed in the defensive 3rd compared to their opponents (68 : 36). Similarly, it has the lowest relative comparison for Att.3rd. This tells me most of the defensive actions completed by Italy is actually completed in front of their goal. This is a team that doesn’t press the opponent and invites the pressure into their own half. A team that fits our definition of defensive play.

Similarly, the Netherlands completed the highest number of actions in the defensive 3rd at 91 (compared to Italy’s 68). Although their relative defensive 3rd is much higher than that of Italy (91:91). This indicates the Netherlands plays a slightly more hybrid game. The final 3rd stats for the Dutch is much higher both relatively, and objectively. This would indicate the Dutch plays a game that utilizes their defensive third more. Their very high Att.3rd stat relative to their opposition indicates they take the game to the opponents and try to press. Traditional defending is secondary.

|  |  |  |  |
| --- | --- | --- | --- |
| Player | Def 3rd | Mid 3rd | Att 3rd |
| Netherlands | 91 | 76 | 15 |
| NetherlandsOpp | 91 | 54 | 10 |
| Italy | 68 | 38 | 9 |
| ItalyOpp | 36 | 34 | 11 |
| Average | 73.75 | 55.38 | 52.13 |
| AverageOpp | 81.13 | 72.5 | 48.5 |

Based on these patterns displayed above, for the most part, a highly defensive style doesn’t seem to fit with the styles that the successful teams play. The general trend for the successful teams at Women’s World Cup 2019 is not to play a traditionally defensive style.

## Second Iteration

After the initial iteration of the data frame, there are some changes make in order to help better tell the story of each team’s performance.

Specifically, more stats are added to the data frame to help paint a more vivid picture. The further aim of this iteration is to help find which stats are linked to success.

The stats of possession and where the possession occurs is added to help further look at where the general action happens for each team. This helps us know where the game is played for each team.

The stats of goals for and against, as well as expected goals and against is also included to help show relative success of each team’s defensive systems. These are the simplest indictor of success and goals scored and goals conceded are direct indicators of if a team will win or not.

Finally, set piece statistics are also included in this new iteration, particularly the number of passes from “dead ball” situation (free kicks, corners, kick offs, throw ins, goal kicks). This is also an indicator of defensive focus, as part of defensive play focuses on slowing down the game and essentially not allowing the ball to be played, with certain teams such as Jose Bordalas’ Getafe seeing as little as 42 mins of the ball being in play while the other 48 was dead balls and time wasting. These stats will perhaps further help tell the story of a team’s defensive performance.

The error stat is simply removed as it contained an outlier and doesn’t contribute much to the overall story of defensive performance.

Once these stats are added, the previous outlier detection filter is used again. This time, because error is removed, the only “outlier” detected is USA’s goals sitting at a Z-score of 2.97. This isn’t exactly 3 as mentioned before, therefore not removed, but will be kept in consideration for future discussions.

Once outlier detection is completed and understood, a new set of measures is created to help measure and tell the relative performances between teams and their opponents.

This measure was created by using the following formula:

This measure will simply standardize all the numbers and create a measure of relative performance. For example is a Tkl stat shows 1.054 this would indicate the home team does 5.4% better than their opponents.

Note: from here on, the stats will be discussed using the relative measures unless mentioned.

With this new dataframe created, a new EDA is done on this. Once EDA is completed, a new row of averages is inserted to the dataframe to help make comparisons.

To help visualize the general trends of each team. The following line graph is created to compare the overall actions of each team with respect to the average. Two graphics are created, one with the goals and xG included, and one with out. This is due to the reason that the USA team’s goal and xG statistic is almost an outlier with Z-scores of 2.97 and 2.45 respectively, and thus skews the y-axis of the graphs and makes it less easy to interpret. To make the graph more meaningful this outstanding stat was removed for better visualization.

Including USA Goals and xG

USA Goals and xG removed.

With the revised graph, the average line is colored red and made to stand out to show a comparative trend between the best teams of the tournament.

This graphic helps show a few key trends between the best teams:

* Most defensive actions are done in the attacking 3rd – most teams are not defending in their own half using a traditional method.
* There’s only 1 standout team that has most of their defensive actions in the defensive 3rd – traditional defending is not the main paradigm of the successful teams.
* The main trend is to play very positively, in the sense that, the game is played in the opposition half. Most touches and actions are completed in the attacking 3rd then the middle 3rd as shown by the average trend line.

## Third Iteration

How does these trend help answer the question: does defence win titles? At this moment, it doesn’t provide a definitive answer. It does help us understand what successful teams do with respect to this recent tournament. And for the most part, it does not seem like the traditional defensive systems are the ones that are most popular.

With this iteration, I plan to create another comparative index to compare the link between defensive actions and “success” at the tournament. This will help tell me if defensive actions are indeed linked to success in tournament football, or if a separate set of indicators are more indicative.

To define success, I will use the relative xG stat added in the second iteration. As this stat shows a good ratio of goals expected to be scored by the team and goals expected to be scored against them.

To define defense, an average between the relative defensive actions in each 3rd, tackles attempted, blocks, interception, and clearances (Def.3rd, Mid.3rd, Att.3rd, Tkl., blocks, int, clr) was taken to show on average which teams do more defensive actions. This indicator will be known as “Defence Actions Total”.

The following was generated using these indicators – sorted ascending by xG:

|  |  |  |
| --- | --- | --- |
| Player | xG | DefenceActionTotal |
| Norway | 0.594059406 | 1.030206 |
| Netherlands | 1.163043478 | 1.160198 |
| Italy | 1.253968254 | 1.31468 |
| Sweden | 1.65 | 0.816789 |
| Germany | 2.285714286 | 0.938046 |
| Avg | 2.289504787 | 0.998734 |
| England | 2.732142857 | 0.919003 |
| France | 3.827586207 | 0.986687 |
| USA | 4.80952381 | 0.824266 |

With the data displayed above, I want to see if there is indeed significant change to xG when Defence Actions Total is also changed. Thus telling me if my set of defensive indicators affect the outcome of a game for better or for worse, or if any.

Because of the fact that Defence Action Totals is a combined average of multiple indicators I will analyze this using two methods:

* Simple linear regression between xG and Defence Action Total
* Multiple linear regression between xG and all stats under Defence Action Total (Def.3rd, Mid.3rd, Att.3rd, Tkl., blocks, int, clr).

Simple Linear Regression:

First, I looked at the correlation and the general trend between the two stats, alongside the correlation.



Chart

Description automatically generated

Trend of Defence Actions Total as relative xG increases.

Both of these two analyses show a fairly negative trend.

Next the linear regression model was created and analyzed.:

A picture containing text

Description automatically generated

Linear Regression Model

Text, letter

Description automatically generated

Summary of the linear regression model.

Based on the coefficients details shown in the summary, this also agrees with the negative relationship between xG and defence total stats. The fact that there are no stars next to defence action total leads me to believe this isn’t a significant variable. Indicating not the strongest relationship. Similarly, the fairly low value of the adjusted R-squared also tells me this regression model doesn’t explain much of the variability in the outcome.

Multiple Linear Regression:

With this model the aim is to look more in depth into the defence action total stat and see which one of the defensive actions (if any) help contribute to a higher relative xG.

Correlation is summarized in the table below:

|  |  |
| --- | --- |
| Input Indicator | Correlation to xG |
| Def 3rd | -0.5838682 |
| Mid 3rd | 0.2074494 |
| Att 3rd | 0.2797702 |
| Tkl. | 0.4177418 |
| Blocks | -0.5766966 |
| Int | -0.01566973 |
| Clr | -0.8735534 |

Trend line where:

Chart, line chart

Description automatically generatedBlack = Def 3rd.

Red = Mid 3rd.

Green = Att 3rd.

Blue = Tkl.

Purple = Block

Orange = Int

Yellow = Blocks

Both of these initial looks at the data shows varying degrees of different correlations. Most of which seem to be a medium to fairly strong negative correlation. The line graph especially shows how “frantic” the relations are. It’s difficult to fit a line of best.

The following are generated form the multiple linear regression analysis:



Graphical user interface, text

Description automatically generated

Looking at this more detailed model we can see a p-value of 4.297e-10 which is tells me this is fairly significant, at least one of the variables significantly related to xG.

Looking at the coefficients, the Att 3rd, Tkl. Int, and Blocks variables are positively growing alongside xG, meaning these are the stats where the more you do the better your relative xG is likely to be. The number of stars for each variable can also tell me these variables are all individually significant to the changes in relative xG.

Looking at the various t-values of each variable, it seems they’re all fairly small. However, Int is the one to stand out as it has the highest value, telling me this variable is most associated with changes in relative xG.

The very standout stat here is the R - Squared, with a perfect score of 1, indicating all of the variability in the outcome is explained by the model. This is possibly the result of the high correlation between input variables as detected by the initial EDA.

Further Analysis due to Unexpected Issues:

Since a R - Squared of 1 is produced, further regression analysis was done, particularly doing variable selection using forward selection.

This was original done with just the defense total action stats included in the multiple regression model above, but also produced a perfect model.

Table

Description automatically generated

Based on this, all the variables were used in forward selection to see if a better set of indicators can be created. The ‘player’ indicator was removed as it’s non-numerical.

Graphical user interface, text

Description automatically generated

This once again produced a perfect model with R - Squared of 1. However this analysis did help provide the robust measure of indicators to help see what relates to a team having a good relative xG.

Based on the forward selection these are (in order):

* AttPen – touches in the opposition penalty area.
* Goals – goals scored
* Tkl. - % of opposition dribblers tackled
* Blocks - # times the ball is blocked by standing in its way.
* Int - # of interceptions.
* Att.3rdT - # of touches in the attacking third/opposition side.
* Pass- # of passes blocked.

From these indicators only Tkl. Blocks, Int, Pass are somewhat defense related.

Based on the linear regression, multiple regression, and forward selection that was done, there are a few very standout conclusion and patterns that can be detected:

* There is a negative correlation between defensive action totals and relative xG.
  + This says the more defensive actions done, the less likely you are to win a game.
* Certain defensive actions still do contribute to the success of a match:
  + Tackles, interceptions, and passes blocked are all fairly positive contributors according to forward selection.
  + The standout fact is that these actions are directly related to this project’s definition of defensive play. As most positive, attacking teams also aim to tackle, intercept and block passes.
* The main indicator of a successful is not in defensive but rather where the game takes place.
  + Teams that play in the opponent’s half (high levels of Mid.3rd, Mid3rdT, Att.3rd, AttPen are all highly related to the performances of the successful teams and their relative xG)

# Discussion

To be expanded as project goes along.

# Index

## Variable Details

|  |  |
| --- | --- |
| Variable Name | Description |
| Player | Name of nation/opposition |
| Age | Age |
| 90s | Minutes played by 90 – a measure of how many games someone played. |
| Tkl | # of players tackled total |
| TklW | # of tackles won |
| Def.3rd | Defensive actions in the team’s defensive third. |
| Mid.3rd | Defensive actions in the team’s middle third. |
| Att.3rd | Defensive actions in the team’s final third. |
| Tkl.1 | # of opposition dribblers successfully tackled |
| Att | # of dribbles attempted to tackle |
| Tkl. or Tkl% | Tkl.1 / Att – a measure of the success rate of tackling dribblers |
| Lost | # of opposition dribblers unsuccessfully tackled. |
| Blocks | # of times the ball is blocked by standing in its way |
| Sh | # of shots blocked |
| Pass | # of passes blocked |
| Int | # of interceptions |
| Tkl + Int | Tkl + Int |
| Clr | # of clearances completed. Clearances being an action where the player kicks the ball away or out for safety. |
| Err | # of mistakes leading to an opposition shot. |
| DefPen | Touches taken in the team’s own penalty area. |
| Def3rdT | Touches taken in the team’s own defensive third. |
| Mid3rdT | Touches taken in the team’s middle 3rd. |
| Att3rdT | Touches taken in the team’s final 3rd. |
| AttPen | Touches taken in the opposition’s penalty box. |
| Deadball | # of passes from free kicks, corners, kickoffs, throw ins, and goal kicks. |
| AeriaDuels | % of aerial duels won |
| Goals | Goals for |
| xG | Expected goals for. A measure for how good a team should be performing.  A probability that a shot will result in a goal based on the characteristics of that shot and the events leading up to is.  As provided by Opta.  (FBREF, 2023) |

# Bibliography

Anderson, C., & Scally, D. (2013). *The Numbers Game: Why Everything You Know About Soccer Is Wrong.* New York: Penguin Books.

Breaking the Lines. (2022, July 28). *What is Juego de Posición?* Retrieved from Breaking the Lines: https://breakingthelines.com/tactical-analysis/what-is-juego-de-posicion/

Brownell, P. (2013, July 31). *Which Stats Are Most Important for Measuring Defenders?* Retrieved from Bleacher Report: https://bleacherreport.com/articles/1722602-which-stats-are-most-important-for-measuring-defenders

Burkett, T. (2021). *Does Defense Actually Win Championships? Using Statistics to Examine One of the Greatest Stereotypes in Sports?* Columbia: University of South Carolina.

Davis , A., & Suryawanshi, N. (2023, 1 14). *Northwestern Sports Analytics Group*. Retrieved from Does Defense Win Championships?: https://sites.northwestern.edu/nusportsanalytics/2020/12/15/does-defense-win-championships/

FBREF. (2023, February 11). *2019 FIFA Women's World Cup Stats*. Retrieved from FBREF: https://fbref.com/en/comps/106/2019/2019-Womens-World-Cup-Stats

FOX Sports. (2023). *FIFA World Cup 2022 Defensive Stats*. Retrieved from Fox Sports: https://www.foxsports.com/soccer/morocco-men-team-summary-stats?category=defensive&sort=t\_int&season=2022&sortOrder=desc&groupId=12

MARCA. (2012, 9 12). *Spain beats its possession record for the Del Bosque era*. Retrieved from MARCA: https://www.marca.com/2012/09/12/en/football/national\_teams/1347468397.html

NBA. (2023, February 4). *NBA Advanced Stats*. Retrieved from NBA: https://www.nba.com/stats/teams/defense

NFL. (2023, February 4). *2022 Team Stats - Defense*. Retrieved from NFL: https://www.nfl.com/stats/team-stats/defense/passing/2022/reg/all

Robst, J., VanGilder, J., Berri, D., & Vance, C. (2011). Defense Wins Championships? The Answer from the Gridiron. *International Journal of Sports Finance*, 72-84.

Smith, A. (2017, 8 30). *Does attack or defence win titles?* Retrieved from Premier League: https://www.premierleague.com/news/464218

TIFO Football. (2020, July 2). Does Defence Actually Win Titles? London, England.

Winterburn, S. (2017, January 7th). *What can we learn from defensive statistics?* Retrieved from Football3665: https://www.football365.com/news/what-can-we-learn-from-defensive-statistics

Young, J. (2015, September 28). *WHEN SHOULD A TEAM PARK THE BUS? (ABRIDGED)*. Retrieved from American Soccer Analysis: https://www.americansocceranalysis.com/home/tag/Parking+the+Bus#:~:text=Although%20he%20would%20later%20say,effort%20to%20block%20the%20goal.