2023 Women's World Cup:
A Tactical Analysis of Individual Risks on Team Performance

Christopher Spann
christopher.andrew.spann@gmail.com
https://github.com/cas0105

Abstract

This analysis aims to uncover the intricate relationship between on-field risks taken by individual players and the overall success of their teams in the World Cup. Specifically, we focus on analyzing dribbling and interception events during the group stage matches of the 2023 Women's World Cup. Our objective is to identify patterns and assess the impact of these statistics on crucial metrics such as goal differential and progression to later stages of the tournament. The subsequent narrative will walk through the comprehensive methodologies employed to extract meaningful insights.

Data Collection and Preprocessing

The foundation of this analysis relies on data from the open-source repository by Statsbomb, accessed and processed using the Python programming language in a Jupyter notebook. The statsbombpy package was utilized to retrieve data, focusing on matches with a competition ID of 72 and a season ID of 107 corresponding to the 2023 Women's World Cup.

A dedicated dataframe was created to isolate group stage matches. The matches dataframe was further utilized to construct another dataframe summarizing team statistics in those group stage matches, including total goals, goal differential, and dummy variables indicating advancement to later tournament stages.

There were a total of 1,626 dribble events and 1,013 total interception events during the 2023 Women's World Cup. A total of 32 teams participated in the world cup. Individual stats by team were calculated using groupby functions and the resulting statistics were analyzed to create insights related to the effect of dribbling and interceptions on team performance.

Dribbling and Interception Analysis

For the analysis of dribbling and interception statistics, we applied the following steps:

1. Data Selection:
    - Creation of separate dataframes for dribble and interception events, filtering based on event type.

    - Restriction of datasets to events occurring during the group stage using match IDs.
2. Statistical Grouping:
    - Aggregation of statistics related to dribbling and interceptions, grouped by team name in new dataframes.
3. Metrics Examined for Dribble Analysis:
    - Total dribbles attempted, completed, and success rate.
    - Breakdown of dribbles by pitch thirds, success rates in each third, and more.
4. Metrics Examined for Interception Analysis:
    - Total interceptions attempted, successful interceptions, and success rates.
    - Location-based breakdown of successful interceptions, average successful interceptions per game, and more.

Visualization and Interpretation

To provide a holistic view of the insights obtained, we employed various visualization techniques:

1. Bar Charts: Python's plot function was utilized to create bar charts depicting dribble and interception success rates by team during the group stage.

2. Top Teams and Bottom Teams: Utilized Python's groupby function to identify top and bottom teams based on relevant statistics.

3. Correlation Analysis: Evaluated correlations between variables using a correlation heat map and specific correlation plots.

4. Pitch Graphics: Employed Python to plot dribble and interception events for top and bottom teams on a pitch graphic. The analysis also included the process of visualizing average dribble and interception locations for each team on a pitch graphic.

Results and Presentation

Insights derived from the analysis were presented in a compelling PowerPoint presentation. Key focuses included:

1. Team Evaluation: Identification of top teams based on dribbling and interception success rates.

2. Statistical Analysis: Comparative analysis of teams excelling in these statistics versus those that did not.

3. Impact on Team Performance: Visualization of how dribbling and interceptions impact goal differential and tournament progression.

Dribble Analysis
Team
Total Dribbles Attempted
Total Dribbles Completed
Dribble Success Rate
Dribbles completed – attacking third
Dribbles completed – neutral third
Dribbles completed – defending third
Dribbles attempted – attacking third
Dribbles attempted – neutral third
Dribbles attempted – defending third
Dribble success rate – attacking third
Dribble success rate – neutral third
Dribble success rate – defending third
Total Goals Scored
Total Goals Against
Goal Differential
Advanced to R16 Dummy
Advanced to QF Dummy
Advanced to SF Dummy
Advanced to Final Dummy
Average X Location for Dribbles Attempted
Average Y Location for Dribbles Attempted

Interception Analysis
Team
Total Interceptions Attempted
Success Out Interceptions
Won Interceptions
Total Successful Interceptions
Total Successful Interceptions In play
Successful Interceptions – attacking third
Successful Interceptions – neutral third
Successful Interceptions – defending third
Attempted Interceptions – attacking third
Attempted Interceptions – neutral third

Attempted Interceptions – defending third
Interception Success Rate – attacking third
Interception Success Rate – neutral third
Interception Success Rate – defending third
Interception Success Rate
Average Successful Interceptions per Game
Total Goals Scored
Total Goals Against
Goal Differential
Advanced to R16 Dummy
Advanced to QF Dummy
Advanced to SF Dummy
Advanced to Final Dummy
Average X Location for Interceptions Attempted
Average Y Location for Interceptions Attempted

Conclusion

The analysis concludes that dribble success rate, average dribble location, and average interception location are strongly correlated with goal differential. Particularly, the average location of interception attempts emerged as a significant indicator of teams likely to advance in the tournament. These findings offer valuable insights into the nuanced dynamics of player risks and team success in the 2023 Women's World Cup.

The results of the analysis can offer teams insight as to how they can tactically prepare for tournaments with a group stage. For example, interception attempts higher up on the pitch are more likely to result in a higher goal differential and advancement to later rounds; therefore, pressing in the attacking and neutral thirds should be encouraged. In addition, teams should emphasize good decision-making for players attempting dribbles. Having a high dribble success rate, especially in the neutral third of the pitch, is correlated with a better goal differential.

References

Link to Statsbomb Github - https://github.com/statsbomb