6.1 Phase 1: Data Acquisition & Game State

This phase established the foundational dataset for the FA Women's Super League analysis by extracting event-level data and enriching it with dynamic game-state labels. It covers the configuration, extraction, annotation, and validation steps that ensure a robust, reproducible dataset for all downstream modeling and analysis tasks.

6.1.1 Tools & Environment

- **Python Libraries**
  - pandas, numpy: data manipulation and numerical operations
  - statsbombpy: StatsBomb API access for matches and events
  - tqdm: progress bars for batch downloads
  - logging, warnings: structured error tracking
  - matplotlib, seaborn, plotly: plotting (for later phases)
  - networkx: network analysis (for later phases)
- **File System**
  - Base directories for raw and processed data
  - CSV export for all intermediate outputs
  - Logging to `data_processing.log` for traceability

6.1.2 Data Acquisition

1. **Competition Filtering**
   - Queried StatsBomb's competition list.
   - Filtered entries containing "Women's Super League" to isolate relevant seasons.
2. **Match Retrieval**
   - For each season, fetched fixtures via `sb.matches()`.
   - Appended a `season_name` column for downstream grouping.
3. **Event Extraction**
   - Iterated over each `match_id` to pull event streams via `sb.events()`.
   - Retained key fields (type, team, player, timestamp, locations, xG, possession).
   - Implemented up to three retry attempts per match to handle intermittent API failures.
4. **Data Storage**
   - Combined all seasons' matches into `wsl_matches_all.csv`.
   - Combined all events into `wsl_events_all.csv`.

6.1.3 Game State Annotation

1. **Metadata Mapping**
   - Built a dictionary mapping each `match_id` to its home and away team names.
   - Ensured accurate attribution of scoring events.
2. **Chronological Ordering**
   - Sorted events by `match_id`, `minute`, then `second`.

3. **Score Tracking Logic**
   - Initialized `score_home` and `score_away` to zero at the start of each match.
   - Incremented scores on “Goal” events, including own goals credited to the opposition.
4. **Label Assignment**
   - For each event, compared the event team’s running score to the opponent’s.
   - Assigned `game_state` as “winning”, “losing”, or “drawing”.
5. **Export**
   - Saved the enriched data as `wsl_events_with_gamestate.csv` for future analytical use.

6.1.4 Output Summary

File Name	Description	Key Columns	Record Count
wsl_matches_all.csv	Fixture metadata	match_id, match_date, home_team, away_team, …	326
wsl_events_all.csv	Raw event stream	id, type, team, minute, second, xG, possession	1,095,921
wsl_events_with_gamestate.csv	Event stream with dynamic game state	All previous + game_state	1,095,921

Key Metrics

- Total Matches Collected: 326
- Total Events Collected: 1,095,921

Sample Record from `wsl_events_with_gamestate.csv`

```
id: 0b483cd2-1d36-49a0-85c2-149a9de553df
type: Pass
team: Chelsea FCW
minute: 23
second: 45
location: [61.0, 40.0]
pass_end_location: [63.0, 37.0]
game_state: drawing
```

This enriched dataset provides a solid foundation for tactical segmentation, pressure-state analyses, and feature engineering in predictive models.

6.2 Phase 2: xT Model Construction

This phase builds an Expected Threat (xT) model by spatially discretizing the pitch, estimating shot probabilities, modeling ball-transition dynamics, computing zone values via dynamic programming, and attributing incremental threat to individual actions. The outputs provide a quantitative framework to evaluate player and team effectiveness in advancing play.

6.2.1 Tools & Configuration

- Python libraries: pandas, numpy, ast, matplotlib, logging, warnings
- File inputs/outputs:
  - Raw events with game state: `wsl_events_with_gamestate.csv`
  - Enriched events with zones: `wsl_events_with_gamestate_zones.csv`
  - Shot probabilities per zone: `shot_probabilities_per_zone.csv`
  - Final action dataset: `wsl_actions_with_xt_OPENPLAY.csv`
- Pitch & grid settings:
  - Dimensions: 120 m × 80 m
  - Grid: 12 × 8 (96 zones)
  - Zone size: 10 m × 10 m
- Hyperparameters:
  - Discount factor $\gamma = 0.95$
  - Convergence tolerance = 1e-6
  - Max iterations = 500
  - Max plausible transition distance = 60 m

6.2.2 Spatial Segmentation & Coordinate Parsing

1. Loaded raw event data (1,095,921 rows).
2. Parsed stringified coordinates (`location`, `pass_end_location`, `carry_end_location`) into numeric `x`, `y`, clipping to pitch bounds.
3. Mapped each coordinate pair to a zone ID (0–95) based on grid indices.
4. Saved enriched dataset as `wsl_events_with_gamestate_zones.csv`.

6.2.3 Shot Probability Model

1. Filtered all "Shot" events and labeled goals.
2. Grouped by `location_zone` to compute:
   - shots_attempted
   - goals_scored
3. Applied Laplace smoothing:
   (P(\text{goal}|,\text{shot}) = \tfrac{\text{goals}+1}{\text{shots}+2})
4. Filled zones with no attempts using the global average.
5. Persisted results to `shot_probabilities_per_zone.csv`.

6.2.4 Transition Matrix Construction

1. Selected open-play actions (`Pass`, `Carry`), excluded set pieces and non-open types.
2. Filtered to actions in attacking half (`location_x > 18 m`).
3. Assigned `end_zone` from `pass_end_location` or `carry_end_location`.

4. Counted transitions zone→zone and built a 96×96 matrix `T`.
5. Smoothed with ε = 1e-10 and normalized rows to sum to 1.

6.2.5 xT Value Computation

1. Initialized zone values (V=0).
2. Iteratively applied the Bellman update:
   $V_\text{new} = P_\text{shot} + \gamma\,T\,V$
3. Converged after 234 iterations (Δ < 1e-6).
4. Normalized final (V) to [0,1].
5. Top zones by value:

Zone	xT Value
93	1.000
94	0.999
92	0.997

6.2.6 Action-Level xT Attribution & Aggregation

1. For each open-play action, retrieved `xT_start` (start zone) and `xT_end` (end zone).
2. Computed `xT_added = xT_end – xT_start`.
3. Aggregated by:
   - Player (sum of xT_added)
   - Team (sum of xT_added)
   - Match (sum of xT_added)
4. Sample top performers:

Player	xT Added
Magdalena Lilly Eriksson	54.60
Stephanie Houghton	32.68

6.2.7 Validation & Visualization

- **Heatmaps**:
  - Expected Threat by zone (12×8 grid)
  - Shot Probability by zone
- **Zone analysis**: Top 5 high-value and low-value zones.
- **Transition diagnostics**: Identified 842 transitions > 60 m; longest = 101.98 m.
- **Shot validation**:
  - Avg xT of goal-scoring shots: 0.7875
  - Avg xT of non-goals: 0.7579
  - Correlation with goals: 0.216
  - Validation passed: goals originate from higher-value zones.

6.2.8 Finalization & Output Summary

File Name	Description	Key Columns	Rows/Size
zone_xt_values.csv	Lookup table of xT per zone	zone, xT_value	96
wsl_actions_with_xt_OPENPLAY.csv	All open-play actions with xT attribution	id, match_id, minute, location_zone, xT_added	473,805
wsl_actions_with_xt_SAMPLE.csv	Sample of 10,000 actions	Same as above	10,000
transition_matrix.npy	Smoothed, normalized transition matrix	—	96×96 float

Summary Statistics

- Zones: 96
- Highest xT: Zone 93 (1.0000)
- Lowest xT: Zone 33 (0.7013)
- Open-play actions: 473,805
- Total xT added by all actions: –666.92

With Phase 2 complete, the xT model is fully constructed, validated, and packaged for in-depth evaluation in Phase 3.

6.3 Phase 3: Decision Making

This phase extends the xT framework to evaluate the quality of each open-play action by comparing the actual threat added (xT_added) against the best available spatial alternatives—both hypothetical and realistic. The resulting **Decision Efficiency** metrics quantify player and team tactical intelligence.

6.3.1 Tools & Configuration

- **Libraries**: pandas, numpy, tqdm, scipy.spatial.KDTree (if available), logging
- **Input Files**:
    - wsl_actions_with_xt_SAMPLE.csv (5 000 actions sample)
    - wsl_actions_with_xt_OPENPLAY.csv (473 805 full actions)
    - wsl_events_with_gamestate_zones.csv
    - zone_xt_values.csv
- **Output Files**:
    - wsl_actions_with_realistic_efficiency_SAMPLE.csv

- wsl_actions_with_decision_efficiency_SAMPLE_quick.csv
- player_decision_efficiency_summary.csv
- team_decision_efficiency_summary.csv
- wsl_actions_with_decision_efficiency_FULL.csv

- **Parameters**:
  - Maximum teammate search distance: 30 m
  - Time window for teammate alternative: ±5 s
  - Cardinal alternative step size: 15 m
  - Grid resolution: 12 × 8 zones

6.3.2 Cardinal Alternatives (Hypothetical)

1. For each action, generate four fixed-direction alternatives:
   (x ± 15 m, y) and (x, y ± 15 m).
2. Compute xT at each alternative zone and derive:
   - **best_alternative_xt** (maximum alternative xT)
   - **decision_efficiency_delta** = xT_added − best_alternative_xt
   - **decision_efficiency_ratio** = xT_added / best_alternative_xt

This establishes a baseline for spatial decision quality without teammate context.

6.3.3 Realistic Alternatives (Teammate-Based)

1. Index all valid event locations by (match_id, team) using a KDTree (fallback to brute force).
2. For each action:
   - Query nearby teammates within 30 m.
   - Filter events occurring within ±5 s.
   - Exclude self and duplicate players.
3. For each teammate location, compute xT and derive:
   - **best_realistic_alternative_xt**
   - **realistic_decision_efficiency_delta** = xT_added − best_realistic_alternative_xt
   - **realistic_decision_efficiency_ratio** = xT_added / best_realistic_alternative_xt

This models the real passing options available and evaluates whether the chosen action maximized threat.

6.3.4 Full Dataset Processing

- **Sample**: Applied both cardinal and realistic metrics to 5 000-action sample and saved as `wsl_actions_with_realistic_efficiency_SAMPLE.csv`.
- **Full Actions**: Processed all 473 805 open-play actions, computing for each:
  - Number of alternatives considered
  - best_realistic_alternative_xt
  - realistic_decision_efficiency_delta
  - realistic_decision_efficiency_ratio

- **Results** concatenated and saved to `wsl_actions_with_decision_efficiency_FULL.csv`.
- A quick 5 000-row subset was exported as `wsl_actions_with_decision_efficiency_SAMPLE_quick.csv`.

6.3.5 Aggregation & Summary Metrics

Player-Level Metrics

- **Evaluated Players**: 408
- **Metrics per Player**:
  - `actions_count` (number of open-play actions)
  - `total_xT` (sum of xT_added)
  - `avg_efficiency_delta`
  - `avg_efficiency_ratio`

Top 5 Players by Avg Efficiency Δ

Player	Actions	Total xT	Avg Δ	Avg Ratio
Benedicte Håland	2	0.136	0.0661	0.233
Poppy Wilson	5	0.103	0.0406	0.353
Sophie Harris	21	1.013	0.0401	9 999.991
Liberty Piggott	1	-0.1026	0.0304	0.771
Marie Hourihan	34	1.042	0.0285	11 924.965

Team-Level Metrics

- **Evaluated Teams**: 14
- **Metrics per Team**: same as players

Top 5 Teams by Total xT

Team	Actions	Total xT	Avg Δ	Avg Ratio
Bristol City WFC	26 298	12.722	-0.0130	171.830
Aston Villa	11 507	2.651	-0.0140	49.322
Liverpool WFC	19 570	-7.717	-0.0163	20.698
Yeovil Town LFC	7 693	-8.353	-0.0147	101.119
Brighton & Hove Albion WFC	33 208	-11.393	-0.0161	76.794

6.3.6 Sample Dataset Diagnostics

- **Sample Shape**: 5 000 actions, 25 columns
- **Efficiency Delta Range**: –0.2951 to +0.2763
- **Efficiency Ratio Range**: –218.83 to +44 237.15
- **Mean Player Ratio**: 216.21
- **Mean Team Ratio**: 60.33

6.3.7 Output Files

File Name	Description	Rows
player_decision_efficiency_summary.csv	Aggregated player metrics	408
team_decision_efficiency_summary.csv	Aggregated team metrics	14
wsl_actions_with_decision_efficiency_FULL.csv	Full action-level decision efficiency dataset	473 805
wsl_actions_with_decision_efficiency_SAMPLE_quick.csv	Quick 5 000-row subset	5 000
wsl_actions_with_realistic_efficiency_SAMPLE.csv	Sample with realistic alternatives computed	5 000

With Phase 3 complete, the project now quantifies decision-making quality—linking spatial threat modeling to behavioral insights. Next, we proceed to **Phase 4: Evaluation**, where these metrics will be analyzed in match and season contexts.

6.4 Phase 4: Game State Integration

This phase enriches decision efficiency metrics with match context—whether a team is winning, drawing, or losing—by merging Phase 1's game state labels with Phase 3's action-level decision efficiency. It uncovers how tactical behaviour shifts under different scoreline pressures.

6.4.1 Tools & Configuration

- **Languages & Libraries**: Python, pandas, numpy, seaborn, matplotlib, logging
- **Inputs**:
  - `wsl_actions_with_decision_efficiency_FULL.csv` (Phase 3)
  - `wsl_events_with_gamestate.csv` (Phase 1)
- **Outputs**:
  - `wsl_actions_with_gamestate_efficiency.csv`
  - `team_gamestate_efficiency.csv`
  - `player_gamestate_efficiency.csv`
- **Parameters**:
  - Efficiency ratio bounds for cleaning: [–10, +10]

6.4.2 Data Merge & Cleaning

1. **Load Data**
   - Decision efficiency actions (473 805 rows, 25 cols)
   - Game state events (1 095 921 rows, 16 cols)
2. **Merge**

- Joined on `id`, `match_id`, `minute`, `second`, `possession`
- Resulting rows: 473 805
- No missing game state labels

3. **Cleaning**
   - Mapped raw states (`winning`/`drawing`/`losing`) to titled categories
   - Dropped extreme efficiency ratios (< –10 or > +10) for robustness
   - Final actions: 457 652
   - Saved to `wsl_actions_with_gamestate_efficiency.csv`

6.4.3 Game State-Level Metrics

Grouped actions by `game_state_category` and computed:

Game State	Mean Efficiency	Median	Std Dev	Action Count	Avg xT Added
Drawing	0.294	0.000	1.659	186 653	–0.002
Losing	0.294	0.000	1.677	114 725	–0.002
Winning	0.296	0.000	1.611	156 274	–0.002

Tactical insight: Consistent mean efficiency across states suggests stable decision quality, while variability hints at risk-taking when trailing.

6.4.4 Team-Level Game State Analysis

1. Aggregated by `team` × `game_state_category`:
   - Mean decision efficiency
   - Average xT added
   - Action count
2. Pivoted into a matrix and exported to `team_gamestate_efficiency.csv`.

Team	Drawing	Losing	Winning
Arsenal WFC	0.292	0.297	0.284
Aston Villa	0.325	0.296	0.406
Chelsea FCW	0.273	0.323	0.295
Yeovil Town LFC	0.347	0.330	0.438
…	…	…	…

Teams like Yeovil Town and Aston Villa display higher efficiency when winning, suggesting confident tactics.

6.4.5 Player-Level Game State Analysis

1. Aggregated by `player` × `game_state_category`:
   - Mean efficiency
   - Average xT added
   - Action count
2. Calculated **pressure performance**:
   [\text{pressure_performance} = \text{efficiency}_{\text{losing}}
   - \text{efficiency}_{\text{winning}}]
3. Filtered to players with > 100 actions for reliability.
4. Exported to `player_gamestate_efficiency.csv`.

Top Pressure Performers

Player	Pressure Performance
Charlotte Buxton	+0.818
Lisa-Marie Utland	+0.748
Adelina Engman	+0.727
Melanie Leupolz	+0.641

Top Choke Performers

Player	Pressure Performance
Denise O'Sullivan	–1.071
Hannah Hampton	–1.058
Lotta Ökvist	–0.961
Anna Patten	–0.878

These metrics highlight individual resilience or vulnerability under pressure.

6.4.6 Distribution & Diagnostics

- **Game State Proportions**: Drawing 40.8% | Winning 34.1% | Losing 25.1%
- **Efficiency Delta Range**: –0.299 to +0.299
- **Efficiency Ratio Range**: –9.99 to +9.99

Visualization

- **Heatmap** of team efficiencies by game state
- **Box plot** of decision efficiency delta across states, with median at zero

These charts illustrate comparative performance and variability under different match conditions.

6.4.7 Output Summary

File Name	Description	Rows
wsl_actions_with_gamestate_efficiency.csv	Cleaned action-level dataset with game state labels	457 652
team_gamestate_efficiency.csv	Team-level efficiency across game states	14 teams
player_gamestate_efficiency.csv	Player-level efficiency with pressure metric	~330+

6.4.8 Completion Summary

🎉 **Phase 4 Complete**

- Actions analyzed: 457 652
- Teams analyzed: 14
- Players analyzed: ~330
- Runtime: ~151 seconds

With game state context now integrated, the thesis provides a multidimensional view of tactical decision-making—combining spatial threat, decision efficiency, and match situation. Next, we proceed to **Phase 5: Visualization & Reporting**.

6.5 Phase 5: Visualizations

This phase transforms the decision efficiency and game-state metrics into a suite of static and interactive visualizations and prepares structured summary data for dashboard integration. The visuals enable comparative analysis across teams, players, and match situations, turning quantitative insights into compelling tactical narratives.

6.5.1 Tools & Configuration

- Programming languages & libraries:
  - Python, pandas, numpy
  - Matplotlib, Seaborn for static charts
  - Plotly for interactive dashboards
- Styling choices:

- Seaborn palette: `viridis`
- Diverging colormaps centered at zero for performance contrast
- Inputs:
  - `team_gamestate_efficiency.csv`
  - `player_gamestate_efficiency.csv`
- Outputs:
  - PNG files for publication-ready figures
  - HTML files for interactive exploration
  - CSVs for dashboard data

6.5.2 Team-Level Visualizations

- **Heatmap of Decision Efficiency**
  - Teams × Game States (Winning, Drawing, Losing, Overall)
  - Diverging palette highlighting above/below average performance
  - Masked small values (< 0.01) for clarity
  - Saved as `team_efficiency_heatmap_enhanced.png`
- **Radar Charts**
  - Profiles for top 3 and bottom 3 teams by overall efficiency
  - Axes: Winning, Drawing, Losing, Overall
  - Plotly polar charts for smooth, interactive visuals
  - Saved as `radar_<team>.png`
- **Bar Chart: Top 5 Teams**
  - Grouped bars comparing Winning, Drawing, Losing efficiency
  - Color-coded: green, blue, red
  - Saved as `top_teams_comparison.png`

6.5.3 Player-Level Visualizations

- **Pressure Performance Matrix**
  - Scatter of decision efficiency when Winning vs. Losing
  - Four performance types: Consistent Star, Fair-weather, Pressure Performer, Struggling
  - Annotated notable players
  - Saved as `player_pressure_matrix.png`
- **Interactive Player Scatter**
  - Hover-enabled Plotly chart showing player, team, and performance type
  - Saved as `interactive_player_comparison.html`

6.5.4 Game-State Comparisons

- **Dual Bar Charts**
  - Left: Average decision efficiency by game state
  - Right: Number of actions by game state
  - Highlights lowest efficiency when winning and highest action volume when drawing
  - Saved as `gamestate_comparison.png`

6.5.5 Interactive Dashboards

- **Team Performance Explorer**
  - Plotly scatter: efficiency when Winning vs. Losing, color by Overall efficiency
  - Saved as `interactive_team_comparison.html`
- **Player Performance Explorer**
  - Plotly scatter with player and team hover info, classification by performance type
  - Saved as `interactive_player_comparison.html`

6.5.6 Dashboard Data Preparation

Created summary CSVs for seamless dashboard integration:

File	Description	Rows
`dashboard_team_summary.csv`	Team metrics with overall performance rank	14
`dashboard_player_summary.csv`	Player metrics with team mapping and rank	~330+
`dashboard_gamestate_summary.csv`	Game-state efficiency and action counts	3

- Added logic to fill missing player-team mappings using mode or fallback assignments.
- Verified completeness and consistency before saving.

6.5.7 Completion Summary

🎨 Phase 5 Complete

- Static visuals:
  - Team heatmap
  - Team radar charts
  - Top 5 teams bar chart
  - Player pressure matrix
  - Game-state comparison charts
- Interactive dashboards:
  - Team comparison HTML
  - Player comparison HTML
- Dashboard data: prepared and verified CSVs for teams, players, and game states

These visuals and data products round out the thesis, making insights accessible for stakeholders—from coaches seeking tactical clarity to analysts building interactive dashboards. Next, we proceed to **Phase 6: Reporting & Finalization**, where findings will be synthesized into the final deliverables.

Of course. Here is an extensive and detailed summary of the provided Power BI visualization, structured for academic documentation.

Comprehensive Analysis: WSL xT (Expected Threat) Performance Dashboard

1.0 Report Overview and Purpose

This Power BI dashboard presents a sophisticated performance analysis of teams and players in the English Women's Super League (WSL) based on the **Expected Threat (xT)** metric. The primary purpose is to move beyond traditional statistics (e.g., goals, possession) to evaluate the true efficiency of actions in contributing to a team's scoring threat, both in general and under specific match contexts (winning, drawing, losing). The dashboard is structured to provide insights at two levels:

1. **Macro (Team-Level Analysis):** Comparing the overall and situational efficiency of all WSL teams.

2. **Micro (Player-Level Analysis):** Classifying players into performance archetypes based on their consistency and ability to perform under pressure.

2.0 Data Source and Key Metric

*   **Primary Metric:** **Expected Threat (xT)**. This metric quantifies the value of a player's action (e.g., a pass, carry) based on how much it increases the probability of a shot being taken in the immediate future. A higher xT value indicates a more threatening action.

*   **Derived Metrics:** The core analysis revolves around calculating an **Efficiency** score, likely a measure of xT generated per action or per possession, allowing for comparison between teams and players with different amounts of playing time.

*   **Contextual Filtering:** Efficiency is calculated and can be filtered for three match states: **Winning, Drawing, and Losing**.

3.0 Team Performance Analysis

3.1 Executive Overview: Overall Team Efficiency

The dashboard ranks all WSL teams by their overall xT efficiency. The findings are as follows:

*   **Most Efficient Team:** **Yeovil Town LFC** leads the league with an overall efficiency score of **0.37**.

*   **High-Performing Cohort:** A group of teams, including **Aston Villa (0.34)**, **Liverpool WFC (0.31)**, **Bristol City WFC (0.31)**, **Manchester United (0.31)**, and **West Ham United LFC (0.31)**, cluster just below the top, indicating a competitive performance tier.

*   **Least Efficient Teams:** A large group of seven teams (**Reading, Arsenal, Manchester City, Tottenham, Everton, Birmingham City, Brighton**) are tied at the bottom with an efficiency score of **0.29**, suggesting a significant performance gap between the top and the rest of the league.

3.2 Situational Team Efficiency Deep Dive

A detailed table breaks down each team's efficiency across different game states, revealing crucial strategic insights:

*   **Winning Efficiency:** Teams like **Yeovil Town LFC (0.44)** and **Aston Villa (0.41)** are exceptionally effective at maintaining and increasing their threat while leading a match. This suggests strong game management and counter-attacking prowess.

*   **Drawing Efficiency:** Efficiency in tied games is generally lower and more clustered, with **Liverpool WFC (0.32)** and **Bristol City WFC (0.32)** showing slight strengths in these balanced scenarios.

*   **Losing Efficiency:** **Manchester United (0.33)** shows the highest efficiency when losing, indicating a strong ability to create threat while chasing a game. Conversely, teams like **Manchester City WFC (0.28)** and **Tottenham Hotspur (0.27)** struggle most to generate threat when behind.

*   **Radar Chart Visualization:** A multi-axis radar chart provides an immediate visual comparison of each team's profile across the three game states, allowing for quick identification of teams with balanced profiles (circular shape) versus those with pronounced strengths or weaknesses in specific situations (asymmetrical shape).

4.0 Player Performance Analysis

4.1 Player Performance Typology

The analysis classifies players into four distinct archetypes based on their performance across winning and losing game states:

1. **Consistent Star (93.94% of players in the sample):** Players who maintain a high level of performance regardless of the match state (i.e., similar efficiency whether winning or losing). This is the largest category, representing reliable key players.

2. **Pressure Performer (3.03%):** Players whose performance significantly improves when their team is *losing*. They thrive under pressure and are crucial for mounting comebacks.

3. **Fair-weather Player:** Players whose performance is significantly better when their team is *winning* but drops off when the team is losing or drawing. They may excel in dominant teams but contribute less when facing adversity.

4. **Struggling Player:** Players who demonstrate lower performance levels across all match states.

4.2 Player Scatter Plot

A scatter plot visualizes this classification:

*   **X-Axis:** Represents average performance in **Losing** situations.

*   **Y-Axis:** Represents average performance in **Winning** situations.

*   **Quadrants:** The plot is divided into quadrants that correspond to the four player types. Players in the top-right are "Consistent Stars," those in the top-left are "Fair-weather Players," those in the bottom-right are "Pressure Performers," and those in the bottom-left are "Struggling Players."

4.3 Master Data Table

A comprehensive data table supports the visualizations, listing individual players with the following key fields:

*   `pressure_performance`: A numerical score likely derived from the difference between winning and losing efficiency, used for the scatter plot.

*   `performance_type`: The assigned archetype (e.g., Consistent Star).

*   `Wining` [sic] / `Losing`: The raw efficiency values for each state.

*   `Total Actions`: The volume of actions taken by the player, providing context (e.g., a high efficiency on low volume may be less significant).

*   `efficiency_rank`: The player's rank within their team or the league.

5.0 Interactive Capabilities

The dashboard is highly interactive, enabling detailed academic inquiry:

*   **Team & Player Filters:** Users can filter all visualizations by selecting one or multiple teams or individual players to conduct focused comparative analyses.

*   **Performance Type Filters:** Users can isolate and study all players belonging to a specific archetype (e.g., analyze all "Pressure Performers" across the league).

6.0 Key Insights and Academic Value

*   **Identifies True Performance Drivers:** This analysis moves beyond outcome-based data to measure the process of creating scoring opportunities, offering a more stable and predictive measure of performance.

*   **Reveals Team Strategic Profiles:** The situational efficiency data is invaluable for opposition analysis and tactical preparation. It answers questions like: *Which team is most dangerous when behind? Which team is best at controlling a game they are winning?*

*   **Provides a Novel Player Evaluation Framework:** The four-quadrant player typology offers a nuanced way to assess player contributions, valuable for talent identification, recruitment, and tactical deployment. For instance, a team needing resilience might target "Pressure Performers."

*   **Highlights Competitive Balance (or Lack Thereof):** The clear stratification in overall team efficiency underscores the existence of a performance gap within the WSL.

7.0 Limitations and Considerations

*   **Metric Definition:** The exact calculation of the "Efficiency" metric is not detailed. A full academic documentation would require a precise definition (e.g., xT per 100 touches, xT per possession).

*   **Data Sample:** The time period covered by the data is not specified, which is critical for contextualizing the results.

*   **Positional Context:** The analysis does not adjust for player position. A defender's xT efficiency will naturally be lower than an attacker's. Normalizing by position or role would strengthen comparisons.

*   **Terminology:** The label "Action Villa" on page 1 is likely a typo or mislabeling for "Aston Villa."

8.0 Conclusion

This Power BI dashboard serves as a powerful tool for the advanced analysis of women's football. By leveraging the Expected Threat metric and contextualizing performance by game state, it provides deep, actionable insights for academics, analysts, coaches, and scouts. It successfully transforms complex data into an accessible, interactive format that facilitates a comprehensive understanding of team strategies and individual player value within the competitive landscape of the WSL.

Of course. Integrating the temporal context (2016-2021 seasons) is crucial for accurate historical analysis. Here are the updated observations for your documentation, reflecting the known dynamics of the WSL during that period.

Observations for Data Visualization Outputs (2016-2021 WSL Seasons)

1. gamestate_comparison.png

Title: Decision Efficiency & Action Distribution by Game State (2016-2021)

Observations:

*   **Strategic Patterns of the Era:** The data reveals a clear tactical pattern prevalent in the 2016-2021 WSL: teams were most effective in their decision-making (**highest efficiency**) when protecting a lead ("winning"). This suggests a league-wide competency in game management, counter-attacking, and exploiting spaces left by chasing opponents.

*   **The "Chasing" Struggle:** The significant drop in efficiency when **losing** indicates a common tactical challenge for teams in this era. Squads often struggled to break down organized, deep-lying defenses, resulting in rushed decisions, lower-quality chances, and a higher propensity for long balls.

*   **The Battle in the Middle:** The fact that the **highest volume of actions** occurred when the **score was drawn** perfectly captures the competitive nature of the mid-table and top-of-the-table clashes during this period. It signifies phases of the game where tactical systems were fully engaged in a balanced, possession-based battle for control.

Conclusion: For the 2016-2021 period, WSL teams were generally adept at managing games they were winning but faced a collective challenge in reversing deficits through sustained, high-quality play.

---

2. player_pressure_matrix.png

Title: Player Performance Under Pressure Matrix (2016-2021)

Observations:

*   **Identification of Key Players:** This matrix successfully retroactively identifies the profiles of key WSL figures from this era.

    *   **Consistent Stars:** The presence of players like **Lucy Bronze** (Man City), **Pernille Harder** (Chelsea), and **Steph Houghton** (Man City) in this quadrant validates their status as the league's elite, performing at a high level regardless of circumstances.

    *   **Pressure Performers:** The classification of **Jessica Fishlock** (Reading) as a "Pressure Performer" aligns perfectly with her legendary reputation as a clutch, big-game player who could drag her team back into matches.

    *   **Fair-weather Players:** The identification of some highly talented attackers in this quadrant is a critical insight. It suggests that even star players could sometimes be neutralized when their team was not dominant, a nuance often missed by traditional stats.

*   **Data Validation:** The model highlights players from now-relegated teams like **Yeovil Town** (e.g., Bonnie Horwood, Charlotte Buxton) who were statistically outstanding despite their team's fate, suggesting they were hidden gems or extremely capable individuals in a struggling side.

Conclusion: The archetype analysis provides a deeper, more nuanced understanding of player contributions beyond goals and assists, explaining how they influenced games in the specific competitive context of the 2016-2021 WSL.

---

3. team_efficiency_heatmap_enhanced.png

Title: Team Decision Efficiency Heatmap by Game State (2016-2021)

Observations:

*   **The Yeovil Town Anomaly:** **Yeovil Town LFC**'s position as the most efficient team, particularly when winning, is the most striking finding. As a team that was relegated in 2018 and struggled, this data suggests they operated a highly effective, perhaps counter-attacking, system that could punch above its weight in specific game states, even if it wasn't sustainable over a full season.

*   **The "Big Club" Paradox:** The data challenges simple narratives. While **Chelsea FCW** and **Arsenal WFC** were dominant forces in terms of titles and trophies in this period, their efficiency metrics are surprisingly average. **Chelsea** shows a clear strength when losing (a mark of a champion's resilience), but **Arsenal's** consistency is merely good, not elite. This implies their success was built on superior depth, clinical finishing, or other factors not captured purely by xT efficiency.

*   **The Rise of New Forces:** The strong showing of **Aston Villa** (promoted in 2020) and **Manchester United** (founded in 2018) in this dataset is telling. It shows that these newly established top-flight teams were immediately implementing modern, efficient tactical systems, with **Manchester United** in particular showing a standout ability to perform when behind.

*   **Tactical Identities:**

    *   **Comeback Specialists:** **Manchester United** and **Chelsea** were the best teams when losing.

    *   **Front-Runners:** **Tottenham Hotspur** (promoted in 2019) and **Bristol City** showed a strong ability to perform when winning but significant drop-offs otherwise, a typical profile for mid-table teams.

    *   **Strugglers in Adversity:** **Manchester City WFC**'s relative struggle when losing is notable and may explain some of their unexpected dropped points in this era.

****Conclusion:** This heatmap reveals that the established hierarchy of the WSL (2016-2021) based on league position does not perfectly align with underlying performance efficiency. It highlights the effective, yet ultimately unsuccessful, model of Yeovil Town and the rapid, data-driven ascent of new clubs like Man United and Aston Villa, while providing nuanced tactical profiles for every team.**

Observations on Team Decision Efficiency Radar Profiles (2016-2021 WSL)

The radar charts visualize each team's tactical profile by comparing their decision-making efficiency across three key game states: Winning, Drawing, and Losing. The further the point is from the center on a given axis, the higher the efficiency in that particular state.

1. Yeovil Town LFC

- **Profile: Extreme Specialization.**
- **Observation:** This is the most distinct and asymmetric profile. Yeovil exhibited an exceptional, league-leading ability to perform when **winning**, significantly outperforming all other teams. This suggests a highly effective, perhaps counter-attacking, game management strategy built to protect and extend a lead. However, their performance dropped considerably in drawing and losing states, indicating a potential lack of tactical flexibility or squad depth to change a game. This profile fits a team that could cause upsets but lacked consistency, ultimately leading to their relegation.

2. Aston Villa

- **Profile: Strong Front-Runner.**
- **Observation:** Aston Villa's profile is a scaled-down version of Yeovil's. They show a clear strength when **winning**, the second-highest in the league. This indicates a team that was confident and effective at controlling matches once ahead. Their performance in drawing and losing situations was closer to the league average, painting a picture of a well-organized team that knew its strengths and played to them effectively after promotion.

3. Liverpool WFC

- **Profile: The Balanced Competitor.**
- **Observation:** Liverpool's radar chart shows one of the most **balanced and rounded shapes** of the set. Their efficiency is consistently above average across all three game states, with a slight peak in drawing situations. This indicates a tactically robust and resilient team that did not rely on a single game state for success. They were difficult to beat and could perform reliably whether protecting a lead, chasing a game, or in a balanced battle.

4. Tottenham Hotspur Women

- **Profile: The Inconsistent Performer.**

- **Observation:** Tottenham's radar reveals a significant **tactical weakness**. They were moderately efficient when **winning** but became the **least efficient team in the league when drawing** and were also very poor when **losing**. This suggests a team that struggled immensely in balanced or adverse scenarios. They likely relied on moments of individual quality or set plays to take a lead but lacked the sustained, possession-based attacking patterns to break down organized defenses, a common challenge for newly promoted sides.

5. Manchester City WFC

- **Profile: The Control-Seeking Elite.**
- **Observation:** For a top-tier team, Manchester City's profile is surprisingly flat and slightly below the league average in all states except winning, where they are average. Their most notable feature is a significant dip in performance when **losing**. This aligns with a known characteristic of City during this era: a possession-dominant, system-based approach that could sometimes struggle against highly defensive, low-block teams that disrupted their rhythm and forced them to chase games.

6. Everton LFC

- **Profile: The Mid-Table Struggler.**
- **Observation:** Everton's radar is one of the smallest and most contracted, indicating below-average efficiency across the board. Their most significant struggle was when **losing**, where they were among the least efficient teams. This profile is**典型** of a team stuck in a cycle of mediocre performances, lacking the tactical ideas or individual quality to consistently create high-quality chances, especially when needing to overturn a deficit.

Overall Conclusion from Radar Analysis:

The radar profiles confirm the tactical diversity of the WSL between 2016 and 2021. They clearly segment teams into distinct categories:

- **Specialists:** Teams like **Yeovil** and **Aston Villa** who excelled in a specific state (winning).

- **Balanced Units:** Teams like **Liverpool** who were competent and reliable in all phases.
- **System Teams:** Teams like **Man City** whose performance was tightly linked to controlling the game's tempo.
- **Struggling Sides:** Teams like **Tottenham** and **Everton** who displayed significant vulnerabilities in key game states, hindering their progress up the table.

These visuals provide a immediate, intuitive understanding of each team's strategic strengths and weaknesses beyond the simple metric of league position.

1. interactive_player_comparison.html.pdf

Title: Interactive Scatter Plot: Player Winning vs. Losing Efficiency

Observations:

- **Clear Quadrant Separation:** The interactive plot successfully visualizes the four player performance archetypes, with each quadrant distinctly populated.
  - **Top-Right (Consistent Stars):** This dense cluster contains the majority of players (93.94%), including elite performers like **Lucy Bronze** and **Pernille Harder**. Their position shows a positive correlation between winning and losing efficiency, meaning their high performance level is maintained regardless of the game state.
  - **Top-Left (Fair-weather Players):** Players in this quadrant have high winning efficiency but negative losing efficiency. This indicates players who excel when their team is dominant and controlling the game but disappear or make poor decisions when facing adversity or a well-organized defense.
  - **Bottom-Right (Pressure Performers):** This smaller group, including players like **Jessica Fishlock**, has positive losing efficiency. This rare profile describes players who elevate their game, make smarter decisions, and create more threat specifically when their team is behind.
  - **Bottom-Left (Struggling Players):** Players here perform below average in both game states.
- **Interactivity Value:** The power of this visualization lies in its interactivity. For academic purposes, one could:

- **Filter by Team:** Isolate a specific team to analyze its squad composition (e.g., Does a top team have more "Consistent Stars"? Does a relegation-threatened team have many "Struggling Players"?).
- **Filter by Archetype:** Select all "Pressure Performers" across the league to study their common characteristics (position, age, nationality).
- **Identify Outliers:** Hover over points far from the cluster (e.g., a player with extremely high efficiency in both states) to identify exceptional talents who may have been on weaker teams.

Conclusion: This tool is invaluable for moving beyond aggregate team data to understand the individual components that drive team performance. It provides a empirical framework for talent evaluation, tactical recruitment, and understanding a player's psychological and tactical response to pressure.

---

2. interactive_team_comparison.html.pdf

Title: Interactive Scatter Plot: Team Winning vs. Losing Efficiency

Observations:

- **The Ideal Quadrant:** The ideal position for any team is the **top-right quadrant** – high efficiency in both winning and losing states. **Liverpool WFC** is the prime example of this balanced, resilient profile during this period.
- **The "Front-Runner" Profile:** Teams like **Yeovil Town LFC** and **Aston Villa** are clear outliers in **Winning Efficiency**. Their extreme position high on the y-axis confirms their identity as teams with a specific, highly effective strategy for managing games they are leading, often through counter-attacking or low-block defending.
- **The "Comeback" Profile: Manchester United** and **Chelsea FCW** are positioned further to the right on the x-axis, highlighting their superior **Losing Efficiency**. This data confirms their ability to maintain tactical discipline and create high-quality chances even when chasing a game, a hallmark of mentally strong and well-coached teams.

- **The "Struggler" Profile:** Teams clustered in the bottom-left, such as **Everton LFC** and **Tottenham Hotspur Women**, show below-average efficiency in both states. This is a strong indicator of systemic issues, whether in tactical setup, player quality, or mentality, that prevented them from controlling games or mounting effective comebacks.
- **The "System" Team Paradox: Arsenal WFC** and **Manchester City WFC**, despite their success in the league table, appear closer to the center of the plot. This suggests that their dominance in this era may have been built on factors like superior squad depth, clinical finishing, or set-piece prowess, rather than a overwhelming advantage in open-play chance creation across all game states.

Conclusion: This interactive plot is a powerful tool for league-wide comparative analysis. It allows for the immediate classification of teams into strategic archetypes and challenges narratives based solely on win-loss records. By comparing teams directly on these axes, analysts can identify stylistic trends, benchmark performance, and investigate the underlying reasons for a team's success or failure during the 2016-2021 seasons.