Step 1. Stakeholder & Decision Context

Purpose: Define who the analysis serves and what decisions depend on it.

Methods: The primary stakeholders are coaching staff, sports analysts, and Syracuse athletic leadership. Their decision revolves around using LLM-generated narratives and statistical analyses to inform coaching interventions, training adjustments, and player evaluations. Risk levels vary from low (operational tweaks) to high (personnel decisions with scholarship or career implications).

Findings: The stakeholder context was explicitly documented, ensuring that all subsequent steps tied directly to decision relevance.

Implications: Anchoring the workflow in stakeholder needs ensures recommendations are interpretable, actionable, and risk-tiered.

Step 2. Data Provenance & Scope

Purpose: Ensure data integrity and transparency.

Methods: Data came from curated Syracuse Women's Lacrosse 2025 game logs and player statistics, collected from official athletic records. Privacy concerns were minimal since the dataset contains performance metrics, not sensitive personal data. Limitations include a single-season scope and lack of contextual variables (e.g., defensive quality, fatigue tracking).

Findings: Data lineage was documented, file sources logged, and known gaps flagged.

Implications: Stakeholders can trust that results are reproducible and free from hidden biases introduced by data provenance.

Step 3. Re-Create / Validate Descriptive Results

Purpose: Verify that descriptive statistics and visuals underlying LLM narratives are accurate.

Methods: Using Python (pandas, matplotlib), key descriptive outputs were regenerated:

- Top goal scorers (bar chart)
- Win/loss season trend (line plot)
- Goals vs. shots correlation (scatterplot)
- Points per game (≥5 games) (bar chart)

Random seeds were fixed for reproducibility, and outputs were saved to a results/ folder.

Findings: Visualizations confirmed LLM descriptions (e.g., Emma Ward as a consistent scoring leader, declining accuracy patterns after mid-game).

Implications: This step validated that LLM summaries matched real statistical outputs, reducing risk of fabricated or exaggerated claims.

Step 4. LLM Prompt & Transcript Capture

Purpose: Maintain transparency of AI involvement.

Methods: All LLM prompts and raw outputs were archived in a prompts/ folder. An annotated version was created noting edits: removal of redundancy, correction of statistical claims, and clarification of terminology.

Findings: Documentation showed where human judgment corrected AI text, e.g., fixing unsupported claims about shot efficiency.

Implications: This ensures accountability stakeholders know which insights are machine-generated and which are human-verified.

Step 5. Quantify Uncertainty

Purpose: Provide confidence intervals and robustness to support statistical claims.

Methods: Bootstrap resampling (10,000 iterations, seed=42) estimated confidence intervals for player-level points-per-game. Distributions for Emma Ward and Emma Muchnick were visualized.

Findings:

- Emma Ward's mean PPG = 3.2 (95% CI: 3.1-3.8).
- Confidence intervals excluded zero effect, supporting claims of consistent performance.

Implications: Coaches can act with moderate-to-high confidence when interpreting player contributions.

Step 6. Sanity Checks & Domain Validation

Purpose: Detect data quality issues and validate claims against domain knowledge.

Methods:

- Missingness analysis (no significant gaps).
- Outlier detection on goals per game (flagged one extreme).
- T-tests on early vs. late game shooting accuracy (not statistically significant, $p > 0.05$).

Findings: Data quality was strong; outliers were noted but did not distort team-level insights.

Implications: Confidence in dataset integrity, though stakeholders should avoid overinterpreting individual anomalies.

Step 7. Bias & Fairness Checks

Purpose: Ensure analyses do not introduce inequitable conclusions.

Methods: Players were grouped by position and playing time to check for disparities. Overrepresentation was noted among starters, with bench players under-sampled in statistical records.

Findings: LLM narratives tended to highlight starters disproportionately, though their statistical dominance partly justified this.

Implications: Coaches should be aware of under-representation when making decisions about second-string athletes.

Step 8. Robustness & Sensitivity

Purpose: Test whether findings hold under perturbations.

Methods:

- Removed top-1 and top-2 shooters, re-ran leaderboards.
- Normalized stats by possessions instead of games.
- Tested CI-width sensitivity under different seeds.

Findings: Rankings remained stable ($\rho = 0.97$ Spearman correlation). CI widths varied slightly but did not alter conclusions.

Implications: Recommendations are robust to reasonable data perturbations, increasing stakeholder trust.

Step 9. Recommendation Tiers

Purpose: Deliver actionable next steps at different risk levels.

Findings & Tiers:

- **Operational (Low Risk):** Provide fatigue-monitoring and coaching to high-shooting players (Ward).
- **Investigatory (Medium Risk):** Collect multi-season and defensive-context data for richer evaluation.
- **High-Stakes (High Risk):** Personnel changes should require human review, ethical oversight, and multi-source validation.

Implications: Stakeholders have a clear, risk-aligned decision framework**.**

Conclusion

This workflow documents a rigorous, transparent process for validating LLM-assisted analytics in sports. Each step links raw data, statistical validation, and stakeholder actionability, ensuring that recommendations are trustworthy, reproducible, and ethically grounded.