Task 07: Stakeholder Report

Title & Purpose

Evaluating LLM-Driven Sports Analytics for Player Performance Decisions.

Purpose: To assess the reliability of LLM-generated insights on basketball performance data and recommend decisions for coaching staff based on validated analytics.

Executive Summary

This report evaluates the effectiveness of large language model (LLM) outputs in analyzing structured sports data, specifically from the Syracuse Men's Basketball 2024–2025 season. The LLM responses were benchmarked against Python-based descriptive statistics and custom metrics.

Key Findings:
• Chris Bell is the top-performing bench player and a strong candidate for a starting role (Low Risk).
• A custom impact metric (PTS + AST + REB) confirmed consistency with LLM's ranking.
• All numeric prompts were validated; minor variance observed in subjective consistency metrics.

Confidence: High for numeric claims; Moderate for interpretation tasks.
Recommendation: Promote Chris Bell to starter; monitor for consistency (Low Risk).

Background & Decision Question

The coaching staff seeks recommendations on player promotion and bench utilization for the upcoming season. This report helps validate whether AI-generated player analysis can support such decisions.

Decision Question: Can we reliably use LLM-generated summaries of structured sports data to inform staffing decisions?

Data & Methods

Dataset: Syracuse Men's Basketball 2024–2025 season stats (32 rows, 16 players). Metrics include Points, Assists, Rebounds, Minutes Played, FG%, and Game Participation (GP-GS).

Tools: Python (pandas, statistics, matplotlib), ChatGPT-4.

Methods: Descriptive analysis (mean, std, min, max), player grouping, impact score metric (PTS + AST + REB), visualizations, LLM prompt-response capture.

Findings

• Chris Bell emerged as the top bench player (9.3 PPG, 13 starts out of 33 games).
• LLM correctly identified the top scorer and high-impact players using the custom formula.
• Visualization (bar chart) of top scorers confirmed AI's rankings.

Uncertainty Quantification: Numeric rankings are robust. Consistency evaluations carry subjectivity. No major outliers or missing values after cleaning.

Recommendations (Risk-Tiered)

• Promote Chris Bell to a starting role based on consistent scoring (Low Risk).
• Investigate duplicate entry for Kyle Cuffe Jr. before formal evaluation (Medium Risk).
• Refrain from personnel changes based solely on subjective LLM reasoning without coach review (High Risk).

Ethical / Legal Concerns

• AI-generated outputs must not substitute coaching expertise.
• Privacy is maintained; dataset contains no PII.
• Duplicated or inconsistent player data must be reconciled to avoid misinterpretation.

Next Steps

• Present findings to coaching staff.
• Develop automated dashboard for season tracking.
• Investigate fairness/bias: evaluate representation by position or playing time.
• Submit GitHub repo + form as required.

Appendices

• Python scripts: python_phase2.ipynb
• LLM prompts: llm_august15_prompts_with_chart.md
• README: Task_05_Descriptive_Stats
• Raw outputs and figures
• Dataset lineage and cleaning logs

LLM Labeling: All AI content is marked and validated. Model: GPT-4. Prompt File: llm_august15_prompts_with_chart.md.