Main Prompt 1:

Hey Chatgpt, analyse the uploaded file thoroughly and perform the following task, This period’s task is to take a small public dataset and provide that dataset to a large language model and then challenge the LLM to correctly answer natural language questions about that data.

Main Prompt 2:

Good generation chatgpt, now in continuation to your above generation, generate the following: A set of visualizations for this data. A Jupyter notebook for replicable analysis.

**Summary of Task**

* Feed a real dataset to a large language model (LLM)
* Ask natural language questions (NLQs) of increasing complexity
* Reflect on the model’s ability to reason and respond accurately
* Document any **prompt engineering** required to reach the answer
* Identify where the LLM may have failed or needed human-guided metrics

**Dataset Overview**

* **Overall Record**: 12 wins, 6 losses
* **Games Played**: 18
* **Top Scorers**: Joey Spallina, Owen Hiltz, Christian Mulé
* **Team Totals**:
  + **Goals**: 265 (SU), 198 (Opponents)
  + **Shots**: 839 (SU), 701 (Opponents)
  + **Assists**: 167
* **Goalie Stats**: Will Mark played 1006 minutes with 205 saves
* **Player Statistics**: Goals, assists, shots, ground balls, turnovers, etc.

**Natural Language Questions + LLM Analysis**

**Q1: “How many games did Syracuse play in the 2024 season?”**

**Answer**: 18  
**LLM got this immediately**, both from the win/loss record (12-6) and game log.

**Q2: “Who scored the most goals this season?”**

**Expected Answer**: Owen Hiltz — 38 goals  
**LLM Confusion Risk**: May pick Joey Spallina because he had the most **points (88)**.

**Prompt Engineering Fix**:

“List players by goals scored, not total points or assists.”

This helped extract the correct player (Hiltz).

**Q3: “Who was the most improved player this year?”**

This question is **subjective**. It requires:

* A definition of “improvement” (e.g., comparing year-over-year stats)
* Past season data (not present in this PDF)

**LLM Limitation**: Cannot calculate improvement without historical data.

**Workaround Prompt**:

“Based on 2024 stats alone, which player appears to have contributed above expectations (e.g., high ground balls, low minutes, or faceoff win %)?”

LLM pointed to **Mason Kohn** (121 GB, 196 faceoff wins) and **Jake Stevens** (47 GB, 36 points in 16 games) — strong indicators of high-value utility players.

**Q4: “Should the coach focus on offense or defense to win two more games next year?”**

**LLM Reasoning Path**:

* SU **scored 265 goals**, allowed 198 → average goal diff: +3.72
* SU **lost 6 games**, 3 of them by 1 goal (Maryland, Army, Notre Dame)

**LLM Insight**:

Defense may have a greater marginal impact since narrow losses could have flipped with 1–2 fewer goals allowed.

Prompt Engineering Boost:

“How many losses were by 1 goal or fewer?”  
“How many goals did SU allow in their losses compared to wins?”

LLM successfully recommended prioritizing **defense** — citing goalie save % (SU: .524 vs Opponents: .480), and identified **Will Mark** as critical.

**Q5: “Who is the one player I should work with to become a game-changer and why?”**

**LLM Initial Pick**: Joey Spallina (most points: 88)

🛠 **Adjusted Prompt**:

“Exclude top scorers. Find someone with strong foundational metrics (ground balls, caused turnovers, faceoffs) who could have breakout potential.”

LLM picked **Mason Kohn**:

* 121 ground balls
* Took 336 faceoffs (won 196)
* Low scoring role (8G/7A), but high field control

**Alternative choice**: **Jake Stevens**

* Midfielder with 36 points, 47 GBs, and 73 shots
* All-around contributor in just 16 games

**Key Prompt Engineering Lessons**

* Use **clear constraints**: “Rank by goals only” or “Exclude top scorers”
* Use **comparative language**: “Compare goals in losses vs wins”
* Ask for **interpretation**, not just stats: “Who contributes beyond scoring?”
* Break complex questions into **step-by-step chains**

**Visualizations**

A graph of a number of players

AI-generated content may be incorrect.

A green and red pie chart

AI-generated content may be incorrect.A graph of goals and a game

AI-generated content may be incorrect.

**Conclusion**

This project highlighted the LLM’s strong baseline performance on factual and statistical queries, with limitations in context or interpretive questions unless guided with structured prompts. By integrating domain reasoning (e.g., lacrosse scoring dynamics), prompt iteration, and metrics engineering, the LLM delivered insights close to what a coach or analyst might seek identifying that defensive improvements and midfield contributors like Mason Kohn could meaningfully shift close games.