

Final Report: Comparing LLM Responses vs. Python Script Results

Objective

The purpose of this project was to use a Large Language Model (LLM) to answer analytical questions about an NBA player dataset and then compare those answers with results calculated directly from the data using a Python script. This process helps evaluate the accuracy and reasoning capabilities of LLMs when provided with structured data.

Dataset

The dataset used contains NBA player statistics for a given season, including fields such as player name, team, age, games played, minutes played, total points, free throw percentage, and awards.

Questions Asked

The following questions were asked of both the LLM and the Python script:

1. Who has the most points per match?
2. Who has the most awards?
3. Who has the greatest free throw percentage?
4. What is the average age of each team?
5. Which player has played the most minutes in total?

Python Script Results

1. Most Points Per Match:

Joel Embiid - 34.69 points per match

2. Most Awards:

Shai Gilgeous-Alexander - 5 awards

Awards: MVP-2, DPOY-7, CPOY-3, AS, NBA1

3. Greatest Free Throw Percentage:

David Roddy - 100.00%

4. Average Age by Team:

Team

2TM 27.79

3TM 29.75

ATL 25.90

BOS 26.88

BRK 26.13

CHI 27.73

CHO 25.79

CLE 26.92

DAL 27.00

DEN 27.17

DET 25.73

GSW 27.85

HOU 25.08

IND 26.06

LAC 29.29

LAL 26.92

MEM 25.05

MIA 28.43

MIL 30.31

MIN 27.08

NOP 25.75

NYK 26.65

OKC 25.47

ORL 24.54

PHI 28.35

PHO 27.53

POR 24.07

SAC 27.00

SAS 23.31

TOR 25.78

UTA 24.75

WAS 26.00

Name: Age, dtype: float64

5. Most Minutes Played:

DeMar DeRozan - 2989 minutes

LLM Responses

Here are the answers to your questions based on the provided NBA player statistics:

- **Who has the most points per match?**
 - **Joel Embiid** scored the most points per match, averaging approximately 34.69 points per game.
- **Who has the most awards?**
 - **Shai Gilgeous-Alexander** received the most awards, with a total of 5 different awards listed.
- **Who has the greatest free throw percentage?**
 - **David Roddy** had the highest free throw percentage, achieving a perfect 1.000 (100%).
- **What is the average age of each team?**
 - Here are the average ages (top 10 examples for readability):

Team	Avg. Age
MIL	30.31
3TM	29.75
LAC	29.29
MIA	28.43
PHI	28.35
GSW	27.85
2TM	27.79

Team	Avg. Age
CHI	27.73
PHO	27.53
DEN	27.17

- (Other teams also calculated; let me know if you want the full breakdown.)
- **Which player has played the most minutes in total?**
 - **DeMar DeRozan** played the most total minutes at 2,989.

Comparison and Observations

- **Accuracy:** The Python script provides the “ground truth,” directly calculated from the dataset. In most cases, LLMs like Perplexity are accurate with direct, fact-based queries if given structured data. However, subtle data nuances or inconsistencies can lead to errors in LLM answers if the prompt isn’t clear or the data is too ambiguous for the LLM to process correctly.
- **Prompt Engineering:** Clearly specifying how statistics are to be calculated (e.g., “points per match” means total points divided by games played) is crucial when querying an LLM.
- **Edge Cases:** The script identifies perfect FT% (100%) and handles cases like missing data or players with zero games. LLMs sometimes need explicit instructions to do this.
- **Subjectivity:** For more subjective or open-ended questions, the LLM may require extra context or definitions.
- **Consistency:** Python scripts produce consistent results, whereas LLM responses may vary depending on prompt phrasing.

Lessons Learned

- **Validation is essential:** Always verify LLM-generated answers against calculations from the raw data.
- **Prompt specificity matters:** The clarity and precision of the prompt directly impact the accuracy of LLM responses.
- **LLMs are good for exploration:** LLMs can quickly generate hypotheses or summaries, but should not be relied on for final numbers without verification.

Conclusion

LLMs like Perplexity can answer many direct data questions accurately if provided with structured data and clear instructions. However, for research and reporting, validating LLM answers with scripting (like Python and pandas) is essential for accuracy and reliability.