Task 06 – Descriptive Statistics & LLM Comparison Part2

Dataset

- **Name:** 2023–24 NBA Player Statistics
- **Source:** Downloaded from Basketball Reference
- **Format:** Excel(xlsx) file
- **Size:** 32 columns × ~500 rows (one per player)

Key Columns Used

Player, G, PTS, FG, FGA, AST, TRB, TOV, Trp-Dbl, Awards

Project Goals

1. Compute descriptive statistics using Python (pandas) to get **ground truth** results.
2. Ask the same questions to **Perplexity AI** (LLM).
3. Compare the answers for **accuracy, reasoning, and alignment**.
4. Identify where the LLM succeeded or failed.

Questions Tested

We tested **10 questions**, progressing from **simple factual retrieval** to **complex analytical reasoning**:

1) 1)Who scored the most total points?
2) Which player had the highest FG%?
3) Which player had the most triple-doubles?
4) Who had the highest points per game?
5) Who had the best assist-to-turnover ratio?
6) Who had the highest combined (PTS+REB+AST) per game?

7) Who had the highest contribution index?
8) Who is most likely to get a triple-double in a future game?
9) If top 5 scorers are removed, who is the new top scorer?
10) As a coach, which player would you build a team around?

Python Ground Truth Results

1) Luka Dončić (2,370)
2) Daniel Gafford (77.96%)
3) Domantas Sabonis (26)
4) Joel Embiid (34.69 PPG)
5) Shake Milton (∞)
6) Luka Dončić (52.9)
7) Luka Dončić (52.9)
8) Luka Dončić (52.9)
9) Anthony Edwards (2,049)
10) Luka Dončić

Perplexity LLM Results:

1 Luka Dončić (2,370)

2 Daniel Gafford (72.5%)

3 Domantas Sabonis (26)

4 Luka Dončić (33.9 PPG)

5 Tyrese Haliburton (4.73)

6 Luka Dončić (53.09)

7 Luka Dončić (53.09)

8 Domantas Sabonis (triple-double every ~3 games)

9 Anthony Edwards (2,049)

10 Luka Dončić

Comparison & Accuracy Table

Question	Match?	Notes
1	Yes	Exact match
2	No	LLM underreported FG% (likely rounding or using subset)
3	Yes	Exact match
4	No	LLM missed Embiid, possibly applied games threshold
5	No	Ignored ∞ ratio, gave realistic alternative
6	Yes	Minor rounding difference
7	Yes	Minor rounding difference

Question	Match?	Notes
8	No	Chose Sabonis due to triple-double rate, not metric definition
9	Yes	Exact match
10	Yes	Same player and reasoning

Findings

- **Accuracy Rate:** 7/10 correct or mostly correct
- **Strengths:** Excellent at direct retrieval from provided table; good at summarizing and explaining results.
- **Weaknesses:**
    - Sometimes it applies its **own thresholds** (e.g., excluding low-games players).
    - Handles edge cases differently (e.g., ignores infinite values).
    - May switch **metric definition** if not reinforced (Q8).
- **Key Insight:** Clearly defining metrics in the prompt increases LLM accuracy. Without explicit instructions, the LLM may default to a *plausible* but different interpretation.