Analysis of NBA Player Performance Using Python and ChatGPT (2022–23)

Updated submission aligned to the course Task 5 requirements. This version loads the dataset from Nba\_data.csv by default and includes a reproducible notebook, visuals, and saved outputs for the reporting form.

# Overview

Goal: Answer six descriptive questions using Python and contrast with ChatGPT’s narrative strengths. Data columns include Player, PTS, REB, AST, STL, BLK, FG%, 3P%, FT%. The notebook: Task\_05\_Complete\_NBA\_LLManalysis\_NbaData.ipynb.

# Dataset

Primary file name: Nba\_data.csv (root).

Percentages (FG%, 3P%, FT%) are normalized to 0–1 if provided in 0–100 form.

# Methods

• Data cleaning (type coercion, percent normalization)

• Deterministic queries for Q1, Q2, Q5, Q6

• Z-score composite for 'most complete' and 'best shooter'

• Visuals: top-10 scorers bar chart, assists histogram, PTS vs AST scatter

• Saved outputs: q3\_composite\_table.csv and summary\_answers.json

# Questions & Answers

## Q1. Who scored the most points per game?

Python

top\_scorer = df.loc[df['PTS'].idxmax()]

Output example: Joel Embiid (33.1 PPG)

ChatGPT

Given a table of PPG, LLM returns Joel Embiid (33.1).

Match: Yes

## Q2. Who is the best all-around player (PTS + REB + AST)?

Python

df['ALL\_AROUND'] = df[['PTS','REB','AST']].sum(axis=1, skipna=True)

best\_all = df.loc[df['ALL\_AROUND'].idxmax()]

Output example: Luka Dončić (≈49 combined).

ChatGPT

Returns Luka Dončić with 32.4 + 8.6 + 8.0 ≈ 49.

Match: Yes

## Q3. Who is the “most complete” player?

Approach

Use a transparent composite: mean of per-feature z-scores across available stats (PTS, REB, AST, STL, BLK, FG%, 3P%, FT%).

ChatGPT

Gives a narrative rationale (e.g., Jokic/Giannis) but depends on definition. No single 'ground truth'.

## Q4. Who is the best shooter?

Python

Leaders by stat: max FG%, max 3P%, max FT%; plus a composite across the three as a tie-breaker.

ChatGPT

Nuanced: Curry for 3P, Durant for FG and FT, or composite for overall efficiency.

Match: Partial (definition-dependent).

## Q5. How many players averaged ≥ 25 PPG?

Python

int((df['PTS'] >= 25).sum())

Example result on sample: 15

ChatGPT

Correct when table is parsed accurately.

## Q6. How many players averaged ≥ 8 AST?

Python

int((df['AST'] >= 8).sum())

Example result on sample: 5

ChatGPT

Can miss boundary values depending on prompt (e.g., ≥ vs >).

# Visuals

1) Top-10 scorers bar chart

2) Assists histogram

3) PTS vs AST scatter

# Validation & Limitations

• Deterministic questions have definitive code answers

• 'Most complete' and 'best shooter' depend on criteria; composite makes this explicit

• Small subset included; ready to swap in larger dataset

# Files Included

• Task\_05\_Complete\_NBA\_LLManalysis\_NbaData.ipynb

• Nba\_data.csv

• q3\_composite\_table.csv

• summary\_answers.json

• docs/AI\_Street\_Interview\_Script.md

• docs/Reporting\_Form\_Answers.md

• README\_Nba\_data.md