📊 GPT vs Claude Comparison
(IPL 2022 Dataset, Advanced Questions Q6–Q8)

This section compares outputs from **ChatGPT** and **Claude** against **Python validation** for advanced research questions.

⚠️ Q6: Overall Wins by Defending vs Chasing

Question: *Did teams win more by defending or by chasing in IPL 2022?*

- **ChatGPT Answer:**
  - Chasing wins: **40**
  - Defending wins: **34**
  - *Conclusion: Chasing slightly more successful.*
- **Claude Answer (Narrative):**
  - Chasing wins: **41 (55.4%)**
  - Defending wins: **33 (44.6%)**
  - *Conclusion: Chasing more successful.*
- **Python Validation:**
  - Chasing wins: **37**
  - Defending wins: **37**
  - *Conclusion: Perfectly balanced (50–50).*

Analysis:
Both GPT and Claude leaned toward a "chasing advantage," consistent with IPL narrative bias, but factually wrong. Shows a shared tendency to generate *plausible cricket storylines* instead of strict calculations.

⚠️ Q7: Batting-Friendly & Chase-Friendly Venues

Question: *Which venues were most batting-friendly (highest average 1st inns) and most chase-friendly (highest % of wins by wickets), considering venues with ≥5 matches?*

- **ChatGPT Answer:**
  - Batting-friendly: **Brabourne Stadium (~176 avg)** ✅
  - Chase-friendly: **DY Patil Stadium (~62%)** ❌
- **Claude Answer:**
  - Batting-friendly: **DY Patil Stadium (~175 avg)** ❌
  - Chase-friendly: **Brabourne Stadium (~60%)** ❌
- **Python Validation:**
  - Batting-friendly: **Brabourne Stadium (avg 177.2, n=16)** ✅
  - Chase-friendly: **Wankhede Stadium (13/21 = 61.9%)** ✅

Analysis:
ChatGPT partially correct (batting-friendly), Claude failed on both. Both models confused chase-friendly venue, likely because percentages were close — instead of calculating, they **guessed**.

⚠️ Q8: Close-Game Performance

Question: *Who handled "close games" best (≤10 runs or ≤3 wickets; teams with ≥3 such games)?*

- **ChatGPT Answer:**
  - Rajasthan Royals — **67%** ❌
  - Gujarat Titans — **60%** (approx) ✅
  - RCB — **50%** ❌
- **Claude Answer:**
  - Rajasthan Royals — **67% (4/6)** ❌
  - Gujarat Titans — **60% (3/5)** ❌
  - RCB — **50% (3/6)** ❌
- **Python Validation:**
  - Gujarat — **2/3 (66.7%)** ✅
  - Lucknow — **2/3 (66.7%)** ✅
  - Mumbai — **1/3 (33.3%)**
  - Kolkata — **0/4 (0%)**

Analysis:

Both GPT and Claude **hallucinated stats for Rajasthan/RCB** and ignored Lucknow. This shows a clear **bias toward famous/high-profile teams** and a failure to reflect true dataset values.

✅ Key Takeaways

- **Shared bias:** Both GPT and Claude favored *plausible cricket narratives* (chasing advantage, famous teams).
- **Hallucination:** Both models invented statistics not supported by the dataset.
- **Narrative vs Data Gap:** Claude sometimes contradicted its own outputs (narrative vs structured).
- **Validation Importance:** Python scripts caught systematic errors that would otherwise go unnoticed.
- **Research Value:** Comparing multiple LLMs shows these are *not isolated errors* — they highlight fundamental weaknesses in LLM reasoning over structured data.

📑 Summary Table

Question	ChatGPT	Claude	Python (Ground Truth for both)	Notes
Q6 Wins (Def vs Chase)	40–34 (chasing edge) ❌	41–33 (chasing edge) ❌	37–37 (balanced) ✅	Both biased toward "chasing advantage"
Q7 Venues	Brabourne / DY Patil ❌ (half correct)	DY Patil / Brabourne ❌	Brabourne / Wankhede ✅	GPT partly right, Claude wrong both
Q8 Close Games	Rajasthan, Gujarat, RCB ❌	Rajasthan, Gujarat, RCB ❌	Gujarat, Lucknow, Mumbai, Kolkata ✅	Both hallucinated, ignored Lucknow

---

👉 This comparison strengthens the study:

LLMs cannot yet be trusted for precise statistical reporting without external validation.