

Durban July Day 2026 — Prediction Grading

Written 2026-07-05 · grades *durban_july_predictions.md* (frozen 2026-07-04 morning)

Grades *durban_july_predictions.md* against the finalised results in the refreshed `RacingDBDebug` (meeting 2026-07-04 Greyville, races 63407–63418, all `iRaceStatus=2`). Per that file's own protocol this grading lives in a separate dated note; the predictions file is untouched. This was a forecasting exercise, not a registered experiment — nothing here spends Tier B/holdout data or enters the experiment registry.

Scorecard

R	Predicted winner	Actual winner (final tote)	Win hit	Winner in pred top-5?	Top-5 set overlap
1	Sommerstern	Sommerstern (2.50)	✓	1st	3/5
2	Please Be True	Please Be True (2.40)	✓ exacta exact	1st	2/5
3	Imilenzeyokududuma (9th)	Magic Verse (8.50)	✗	not listed	2/5
4	Ahead Of The Facts	Ahead Of The Facts (2.20)	✓	1st	5/5
5	Alice B Toklas (2nd)	Palace Of Arvernia (9.20)	✗	not listed	3/5
6	Vibe Check	Vibe Check (3.20)	✓	1st	2/5
7	Note To Self	Note To Self (5.70)	✓ exacta exact	1st	3/5
8	Double Grand Slam (3rd)	Quickstepgal (5.30)	✗	predicted 2nd	4/5
9	Buffalo Storm Cody (2nd)	Jet Force (6.10)	✗	not listed	3/5
10	Talk To The Master	Talk To The Master (8.80)	✓	1st	3/5
11	Captain Selvie (8th)	Warm Reception (18.70)	✗	not listed	3/5
12	Clair De Lune (2nd)	Thought Control (14.60)	✗	not listed	2/5

Totals (the three pre-registered grading questions):

- **Top-1 hit rate: 6/12 (50%).** Raw ranking (model/market top) would have scored 5/12 — the single winner-level overlay move (R1 Sommerstern over co-fav United Front, who ran 9th) was correct. Blind favourite-backing \approx 4–5/12.
- **Top-5 coverage: 35/60 (58%).** Random-set baseline for these field sizes \approx 24/60 (40%). Winner was somewhere in the predicted five in 7/12; in the other 5 races the winner was absent from the five entirely.
- **Overlay vs raw ranking: won at winner level (+1), lost at place level (net –1 verifiable hit).** Detail below.

Extras: exact-order exactas in R2 and **R7 (the Durban July itself: Note To Self → Wish List)**; exact positions 10/60 (1st: 6, 2nd: 2, 3rd: 1, 4th: 0, 5th: 1 — the stated 4th/5th low-confidence caveat was justified: 1 exact hit in 24 slots below 3rd).

What went right

- **The two most-researched races were the two best.** Durban July: model-top pick won, predicted 1-2 exact, Viva's Liberte 4th, and the flagged "first alternate" Isivivane duly ran 5th. Gold Vase: all five predicted horses filled the actual first five (order shuffled), with the weight-swing promotion Master Redoute running 2nd.
- **Tipster-consensus overlays were the profitable subset:** R1 Sommerstern over the unendorsed market co-fav (won; United Front 9th), R10 Talk To The Master at 8.8 (won; the two wide-drawn co-favs marked out ran 9th and 11th), R9 Tenango promoted over the model's One Eye On Vegas (3rd vs 14th).
- Race-note pace reads were often literally correct — R6 (Vibe Check soft lead from the pole, won) and R8 (Quickstepgal controlled lead beating the odds-on fav — see failure #3).

Where it went wrong, and why

1. Horizon: three of the six missed winners were late steamers (the project's own validated signal)

Fixed-book odds movement of the missed-race winners, morning snapshot (~06:10 UTC) → final pre-off:

Winner	Morning	Final	Move
Palace Of Arvernia (R5)	12–15	6.5–8.1	halved — heavy steam
Jet Force (R9)	11–15	6.0–7.0	halved — heavy steam
Quickstepgal (R8)	6.0–6.5	4.5–4.8	steamed (fav drifted to 3rd)
Magic Verse (R3)	6.0–7.5	12–13.5	drifted, won anyway
Warm Reception (R11)	10–11	15–18.7	drifted, won at a big price
Thought Control (R12)	~16	21–34 fixed / 14.6 tote	drifted, won at a big price

The predictions' own caveat #1 ("the morning horizon strips out the live signals the edge lives on") was the operative failure mode: 60-min momentum — a confirmed signal in this project — was pointing straight at Palace Of Arvernia, Jet Force and Quickstepgal by the off, invisible at T-minus-hours. A live-card empirical echo of the EXP-0016 timing-decay finding. R11/R12 (drifting outsiders in low-grade F&M sprints) are best written off as variance no horizon would catch.

2. Coverage: two winners were never discussed at all, one of them a 6/1 morning chance

Magic Verse was ~6.0 in the morning market — a top-3/4 market chance in R3 — and does not appear once in the predictions, even as a danger. Jet Force (11–15 morning) likewise. The qualitative layer only ever discussed the ~5–7 horses the tipster sheets covered, and the tipster sources copy each other's shortlists — so the research inherited their blind spots. (Whether Magic Verse was the model's unlisted 4th in R3 can't be verified without a rescore; the write-up only named model ranks 1, 2, 3 and 5.)

3. Draw reasoning was the worst-performing overlay — asserted, not measured, and partly stale

- The day's 12 winners came from gates 10, 3, 3, 9, 13, 1, 11, 6, 4, 4, 7, 9 — median 6.5 in ~13-runner fields. **No low-draw bias materialised**, yet "low-draw-favouring track" was a premise in at least four races. Draw-driven promotions almost all flopped: Nightshine (11th), December Dawn (11th), Royal Tiara (9th), Callmegetrix (9th), Serengetti Sun (9th), Poursomesugaronme (8th). The lone successes: Town Crier (3rd) and marking out R10's wide co-favourites.
- **Stale-gate trap (R12)**: after four scratchings the field was re-drawn 1–11. Wedding Vow's "poor draw 13" became gate 8/11 and Izibulo's "14" became 5/11 — both were demoted for wide gates they no longer had, and finished 5th and 3rd, while the horses promoted over them (December Dawn, Poursomesugaronme) ran 11th and 8th. The file's scratch rule promoted picks but never revisited draw logic. (R6's Nightshine showed the file *could* track gate inheritance when the scratch was known by the morning; R12's were later.) Same category: R10's promoted reserve Landoftherisingsun (rail draw) ran 2nd, unanalysed.

4. Overriding the model at place level went 1-for-3

Verifiable model-race overrides: Tenango over model #5 → 3rd ✓ (+1 hit); Capsaicin over model-5th World Of Our Own → 11th vs 3rd ✗ (-1); December Dawn over model-3rd Wedding Vow → 11th vs 5th ✗ (-1); Field Marshal over model-3rd Tenpenny → both unplaced (neutral). The model's "boring" lower-ranked picks outran their narrative replacements twice out of three.

5. Narrative right, ranking not acted on (R8)

The race note described the winning scenario exactly ("if Quickstepgal gets an easy controlled lead she can make the champion work hard") — and still left the odds-on favourite on top. When your own pace analysis contradicts the market rank, either move the pick or explicitly lower confidence.

What to do better next time

1. **Re-score at the validated horizon.** The full day's odds are now in the DB — an alerter-style T-10 replay (`--simulate 2026-07-04 / day-scorer at T-10`) would quantify how much of the miss was horizon: would steam have promoted Palace Of Arvernia, Jet Force, Quickstepgal? If predicting a future card, publish the morning list but add a T-10 "late-money watchlist" update pass.
2. **Force full-field coverage.** Rank every runner and require a sentence on any horse above ~8% market probability. A rule that mechanical would have caught Magic Verse and Jet Force.
3. **Measure draw bias before using it.** One SQL/Parquet check of Greyville win-rate by gate tercile per surface/distance instead of repeating tipster copy. And re-check gates after every scratching — the DB shows fields are re-drawn, so scratch handling must revise draw-based adjustments, not just promote picks.
4. **Restrict overlays to the classes that showed value.** Tipster-consensus divergence and pace/weight narratives earned their moves; ad-hoc price-"value" and draw promotions lost. Never displace a model pick without a signal class that has evidence behind it — the same discipline the project applies to retired angles.
5. **Freeze the counterfactual.** The file pre-registered its grading questions (good) but didn't record the raw consensus top-5 for the market-consensus races, so the overlay-vs-ranking comparison is only fully scorable in the 5 model races. Next time, freeze both lists side by side.

Honest bottom line

50% winners and 58% top-5 coverage is a solid public-forecast scorecard — headlined by an exact Durban July exacta — but it is *not* evidence of edge: the six hits paid 2.2–8.8 (mostly market leaders), all six misses paid 5.3+, and the one systematic recoverable failure (late steam) is precisely the already-known T-10 signal. The exercise's real value is that it reproduced, on a single live card, the project's central finding: the information that decides these races arrives in the last hour.