

Golf Data Profiling Report

Full data quality assessment and pattern discovery
Generated February 11, 2026 · 2,141 shots across 46 sessions · 2025-01-04 to 2026-02-05

1. Executive Summary

This report profiles the golf_stats.db database containing launch monitor data from a Uneekor simulator. The dataset spans 13 months of practice sessions, rounds, and drills. Below are the key findings.

Category	Finding	Severity
Completeness	Core metrics (carry, speed, spin, angle) are 100% populated	Good
Completeness	session_type, lie_angle, and shot_tag are 100% null (never populated)	Info
Completeness	97 shots (4.5%) missing session_date — 2 sessions affected	Warning
Sentinels	3 sentinel values (99999+) found in carry, total, and club_speed	Warning
Sentinels	1 shot_type value of '99999' instead of a real category	Minor
Consistency	70 different club names for ~9 actual clubs — heavy fragmentation	Warning
Consistency	38 shots where carry > total distance (physically impossible)	Warning
Accuracy	383 shots with dynamic_loft = 0 (likely short-game or data gap)	Info
Accuracy	74 shots with carry < 5 yds (mishits or chip/putt data)	Info
Trends	Avg carry improved +18.5 yds from first half to second half of sessions	Good
Trends	Face angle control improved (avg face dropped 0.53 degrees)	Good
Trends	Club path consistency improved (avg path dropped 0.45 degrees)	Good

2. Column Completeness

Every column in the shots table was checked for null values, zero counts, and distinct value counts. The table is remarkably complete for a launch monitor dataset — all core measurement columns are 100% populated.

Column	Non-Null	Null	Null %	Distinct	Status
shot_id	2141	0	0.0%	2141	Complete
session_id	2141	0	0.0%	46	Complete
date_added	2141	0	0.0%	233	Complete
session_type	0	2141	100.0%	0	Sparse
club	2141	0	0.0%	70	Complete
carry	2141	0	0.0%	1410	Complete
total	2141	0	0.0%	1456	Complete
smash	2141	0	0.0%	87	Complete
club_path	2141	0	0.0%	894	Complete
face_angle	2141	0	0.0%	1151	Complete
ball_speed	2141	0	0.0%	1033	Complete
club_speed	2141	0	0.0%	694	Complete
side_spin	2141	0	0.0%	1373	Complete
back_spin	2141	0	0.0%	1906	Complete
launch_angle	2141	0	0.0%	1470	Complete
side_angle	2141	0	0.0%	812	Complete
dynamic_loft	2141	0	0.0%	1406	Complete
attack_angle	2141	0	0.0%	1076	Complete
impact_x	2141	0	0.0%	69	Complete
impact_y	2141	0	0.0%	61	Complete
side_distance	2141	0	0.0%	1493	Complete
descent_angle	2141	0	0.0%	1427	Complete
apex	2141	0	0.0%	1557	Complete
flight_time	2141	0	0.0%	352	Complete
shot_type	2141	0	0.0%	14	Complete
impact_img	155	1986	92.8%	155	Sparse

Column	Non-Null	Null	Null %	Distinct	Status
swing_img	155	1986	92.8%	155	Sparse
optix_x	2141	0	0.0%	70	Complete
optix_y	2141	0	0.0%	62	Complete
club_lie	2141	0	0.0%	950	Complete
lie_angle	0	2141	100.0%	0	Sparse
shot_tag	0	2141	100.0%	0	Sparse
session_date	2044	97	4.5%	33	95-99%
face_to_path	2020	121	5.7%	1209	80-95%
strike_distance	2020	121	5.7%	281	80-95%

3. Numeric Column Distributions

Percentile distributions for each numeric measurement column. Flags indicate potential data quality issues like sentinel values (99999), unexpected zeros, or extreme outliers.

Column	Min	P5	P25	P50	P75	P95	Max	Mean	Flags
carry	0.6	8.9	103.7	144.6	182.4	285.6	109359.9	195.2	sentinel, extreme max
total	0.4	11.1	106.9	150.4	192.0	307.2	109359.9	202.9	sentinel, extreme max
ball_speed	6.9	22.4	82.2	103.4	120.8	158.9	168.9	99.5	
club_speed	5.4	22.7	76.9	86.5	92.5	111.0	223691.8	185.2	sentinel, extreme max
smash	0.0	0.9	1.1	1.2	1.3	1.5	1.5	1.2	
launch_angle	-4.3	12.5	18.2	22.3	27.9	36.0	79.5	23.3	
back_spin	-989.0	1376	4330	6655	8667	10303	12404.0	6350.5	
side_spin	-2181.0	-594	-12	335	686	1324	3230.0	347.2	
side_distance	-62.6	-14.5	-3.3	0.2	5.6	25.1	99.9	2.0	
face_angle	-43.7	-6.7	-1.9	0.4	3.0	7.7	32.1	0.5	
club_path	-32.7	-6.7	-3.5	-1.7	-0.3	1.9	21.5	-2.0	extreme max
attack_angle	-15.0	-9.8	-5.8	-3.1	0.0	5.2	14.7	-2.9	
dynamic_loft	0.0	0.0	14.6	28.8	39.1	50.2	90.0	26.5	
apex	0.0	1.2	18.3	25.3	30.5	38.1	55.3	23.4	
descent_angle	0.0	24.3	38.1	44.0	47.5	51.0	76.0	41.8	
flight_time	0.0	1.0	4.6	5.5	6.2	7.0	8.1	5.1	

4. Data Quality Issues

4a. Club Name Fragmentation

The database contains 70 distinct club values for what are approximately 9 actual golf clubs. Many entries represent session names, practice drill types, or minor variations of the same club. Below are the most notable groupings where the same club appears under multiple names.

Actual Club	Name Variants Found	Total Shots
7 Iron	'7', '7 Iron', 'M 7', 'M 7 Iron'	124
8 Iron	'8', '8 Iron', 'Dst Compressor 8', 'Dst Compressor 8 Full'	108
6 Iron	'6', '6 Iron'	89
'PW', 'Wedge Pitching', 'Wedge Pitching Medium', 'Wedge Pitching Premium', 'Forward Impact Pw'		89
56 Degree	'56', 'M 56', 'Wedge 56 Premium'	51
9 Iron	'9', '9 Iron'	40
Warmup	10 different variants including 'Warmup', 'Wmup', 'Dst Warmup', etc.	426

Recommendation: Normalize club names in the database using the existing automation/naming_conventions.py module. This would consolidate ~70 entries into ~15-20 meaningful categories.

4b. Sentinel Values (99999)

The Uneekor system uses 99999 as a sentinel meaning 'no data.' Three instances were found that survived the clean_value() filter in golf_db.py:

Column	Value Found	Count	Context
carry	109,359.9	1 shot	Sgt Rd1 session
total	109,359.9	1 shot	Same shot as carry
club_speed	223,691.8	1 shot	Same shot
shot_type	'99999' (text)	1 shot	Should be a shot shape category

4c. Carry > Total Distance

38 shots have carry distance greater than total distance, which is physically impossible (total should always equal or exceed carry). The differences are small (0.2-1.8 yards) and concentrated in practice/round sessions. This appears to be a rounding or measurement artifact from the Uneekor system rather than a data import error.

4d. Missing Session Dates

97 shots (4.5%) across 2 sessions lack a session_date. These are: session 44266 (71 shots: Iron 1 + warmup 8 dst) and session 44187 (26 shots: club '9'). These sessions could be reclassified using the --manual or --from-listing commands.

4e. Unused Columns (100% Null)

Three columns are completely empty: session_type (intended for practice/round/fitting classification), lie_angle (string field, unused), and shot_tag (intended for warmup/practice/round/fitting tagging). The tag_catalog table defines 4 tags but none have been applied to shots.

5. Performance Patterns

5a. Improvement Over Time

Comparing the first 16 sessions to the last 17 sessions shows measurable improvement across all key metrics. Carry distance increased by 18.5 yards on average, while face angle and club path control both tightened.

Metric	First Half	Second Half	Change	Direction
Avg Carry Distance	142.4 yds	160.9 yds	+18.5 yds	Improvement
Avg Ball Speed	100.4 mph	107.8 mph	+7.4 mph	Improvement
Avg Smash Factor	1.190	1.230	+0.040	Improvement
Avg Face Angle	3.51 deg	2.98 deg	-0.53 deg	Improvement
Avg Club Path	2.73 deg	2.28 deg	-0.45 deg	Improvement

5b. What Drives Carry Distance?

Pearson correlations between each metric and carry distance reveal what most strongly predicts longer shots. Ball speed dominates ($r = +0.98$), followed by club speed and smash factor. Higher launch angles correlate with shorter carries, reflecting the difference between full-swing clubs and wedges/short game.

Metric	Correlation (r)	Interpretation
Ball Speed	+0.984	Strong positive
Club Speed	+0.905	Strong positive
Smash Factor	+0.855	Strong positive
Launch Angle	-0.762	Strong negative (higher loft = shorter clubs)
Apex Height	+0.730	Strong positive
Dynamic Loft	-0.626	Moderate negative
Attack Angle	+0.591	Moderate positive
Side Distance	+0.557	Moderate (longer clubs miss more laterally)
Club Path	-0.421	Moderate negative
Back Spin	-0.329	Weak negative
Face Angle	-0.286	Weak negative

5c. Club Performance Profiles

Each actual club shows a distinct performance signature. The table below compares key metrics across all clubs in the bag. Notable: the 1 Iron and 6 Iron produce nearly identical carry distances (~193 yds) but through very different profiles — the 6 Iron is more efficient (higher smash, tighter dispersion).

Club	Shots	Carry	Total	Speed	Smash	Launch	Spin	Side	Face
Driver	53	285.0	308.1	159.8	1.438	13.6	2007	19.1	2.62
1 Iron	49	192.5	204.1	119.6	1.272	21.6	5281	18.0	4.06
6 Iron	80	193.0	203.7	124.4	1.324	16.6	5886	9.2	3.04
7 Iron	64	186.0	197.7	121.3	1.308	17.4	5390	9.2	3.15
8 Iron	63	155.5	162.4	109.9	1.27	19.9	7338	7.4	2.25
9 Iron	14	155.0	159.6	108.5	1.202	20.6	9271	4.2	2.88
PW	17	126.5	131.4	92.8	1.193	23.3	8448	4.7	2.52
GW	8	132.6	139.0	95.4	1.128	29.5	7305	3.3	1.26

5d. Shot Shape Distribution

Overall, 65% of shots are classified as 'straight,' 15% as 'hookslice' (starts left, curves right), and 13% as 'drawfade.' The Driver shows the most shape variety — only 40% straight vs 82% for the 7 Iron — consistent with the Driver being harder to control.

5e. Session Fatigue Pattern

An analysis of shot position within each session (first 10 shots vs. last 10, for sessions with 25+ shots) reveals a significant performance drop toward the end. Avg carry drops from 185 yds to 116 yds, and ball speed from 118 mph to 87 mph. This likely reflects session structure (warmup with full clubs first, then short game drills later) rather than pure fatigue — but could also indicate diminishing returns from long sessions.

Position	Avg Carry	Avg Speed	Avg Smash	Shots
First 10 shots	185.4 yds	117.7 mph	1.277	330
Middle shots	147.3 yds	101.4 mph	1.199	1,262
Last 10 shots	116.0 yds	87.2 mph	1.139	330

5f. Practice Volume

Month	Sessions	Shots
2025-01	10	538
2025-02	2	141

Month	Sessions	Shots
2025-12	6	401
2026-01	14	940
2026-02	1	24

January 2026 was the most active month with 14 sessions and 940 shots. There is a 9-month gap (March-November 2025) with no recorded data.

6. Recommendations

Action	Details	Priority
Clear obsolete values in carry, total, and club_speed. These should be nulled or removed. One shot_type of '99999' should be reclassified.		High
Normalize club names using the normalize_club_names module. This would make filtering and analysis much more reliable.		High
Backfill missing last_shot_date for 97 shots across 2 sessions by running 'python automation_runner.py reclassify-dates --manual' to fix these.		Medium
Apply shot tags to the tag_catalog table. Tagging shots as warmup/practice/round/fitting would enable better filtering.		Medium
Investigate carry vs total distance. While the difference is minimal (< 2 yds), this violates physics. Consider clamping total = max(carry, total).		Low
Populate 100% of shot_type. This would add a useful dimension for analysis.		Low