

HubSpot Data Quality Audit

This is a sample of the report every client receives. The portal, the numbers, and the findings below are fictional and exist to show the format, depth, and level of detail. Real reports follow exactly this structure.

Prepared for	Sample Co (fictional B2B SaaS)
Portal	HubSpot Professional
Records audited	48,213 contacts · 9,870 companies · 2,315 deals
Access	Read-only private-app token
Delivered	3 business days after access

SUMMARY

Three scores, five things to fix first.

CONTACTS

61/100

Duplicates and missing lifecycle stage drive most of the loss.

COMPANIES

72/100

Domain hygiene is decent. Association coverage is not.

DEALS

54/100

Pipeline dates and amounts cannot be trusted for forecasting.

FIX FIRST, IN THIS ORDER

1. ☐ **Close or re-date 137 open deals with close dates in the past.** Until this is done, every forecast report is overstating the current quarter.
2. ☐ **Populate lifecycle stage on 9,016 contacts.** 18.7% of the database is invisible to every lifecycle-based workflow and report.
3. ☐ **Merge 2,025 duplicate contacts by normalized email.** Sequences and scoring are splitting across records.
4. ☐ **Associate 2,314 companies that have no contacts.** Most are import residue from 2024 and can be archived rather than fixed.
5. ☐ **Fill amount on 261 open deals.** 11.3% of the open pipeline carries no value.

LEAVE ALONE FOR NOW

- ☐ Free-mail company domains (3.2%). Mostly legitimate sole traders. Not worth the cleanup effort this quarter.
- ☐ Custom-property sprawl. 41 properties with under 1% fill. Annoying, not harmful. Archive when convenient.

CONTACTS · 48,213 RECORDS · SCORE 61

Findings

CHECK	RESULT	RECORDS	SEVERITY	HOW TO REPRODUCE
Duplicates by normalized email	4.2%	2,025	HIGH	Export email, lowercase, strip plus-tags and dots on gmail. Group and count >1.
Lifecycle stage blank	18.7%	9,016	HIGH	Filter: Lifecycle stage is unknown.
Contact owner blank	12.4%	5,978	MED	Filter: Contact owner is unknown.
No company association	9.8%	4,725	MED	Filter: Associated company count = 0.
Hard-bounced or invalid email	3.1%	1,495	MED	Filter: Email hard bounce reason is known, plus regex fail on export.
Role-based addresses (info@, sales@)	1.6%	771	LOW	Regex on local part against role list.
Original source blank or "Offline" with no drill-down	22.9%	11,041	MED	Filter: Original source is unknown OR drill-down 1 is unknown.
No activity in 18 months, still Lead or above	31.5%	15,187	LOW	Filter: Last activity date more than 18 months ago AND lifecycle stage ≥ Lead.
Duplicates by name + company	1.1%	530	LOW	Export first, last, company. Normalize and group.

WHAT THIS MEANS

Nearly one in five contacts sits outside the lifecycle model, so any report segmented by stage undercounts by that much. The duplicate rate is moderate for a portal this age, but it is concentrated in the last 12 months of form submissions, which points at a form setting rather than legacy imports.

LIKELY ROOT CAUSES

- Form "create new contact for new email" left on for a gated-content form.
- A 2025 list import without a lifecycle stage mapping.
- Round-robin owner assignment excludes one team.

COMPANIES · 9,870 RECORDS · SCORE 72

Findings

CHECK	RESULT	RECORDS	SEVERITY	HOW TO REPRODUCE
No associated contacts	23.4%	2,314	MED	Filter: Associated contact count = 0.
Domain blank or malformed	9.1%	898	MED	Filter: Company domain is unknown, plus regex fail on export.
Duplicates by domain	2.7%	266	MED	Export domain, strip www, group and count >1.
Industry blank	41.2%	4,066	LOW	Filter: Industry is unknown.
Free-mail domain (gmail, outlook...)	3.2%	316	LOW	Domain in free-mail list.
Company owner blank	6.9%	681	LOW	Filter: Company owner is unknown.

DEALS · 2,315 RECORDS · SCORE 54

CHECK	RESULT	RECORDS	SEVERITY	HOW TO REPRODUCE
Open stage, close date in the past	5.9%	137	HIGH	Filter: Deal stage is not Closed Won / Lost AND Close date before today.
Amount blank in open pipeline	11.3%	261	HIGH	Filter: Open deals AND Amount is unknown.
Stuck in stage over 90 days	17.8%	412	MED	Filter: Open deals AND Time in current stage > 90 days.
No associated company	8.4%	194	MED	Filter: Associated company count = 0.
Closed lost, no lost reason	38.6%	894	MED	Filter: Closed Lost AND Closed lost reason is unknown.
Deals in a pipeline unused since 2024	2.2%	51	LOW	Group by pipeline; check last create date.

CROSS-OBJECT

Where the objects disagree with each other.

CHECK	RESULT	RECORDS	SEVERITY	HOW TO REPRODUCE
Contact is Customer, no closed-won deal on its company	6.3%	3,038	HIGH	Filter contacts: Lifecycle = Customer AND associated company has 0 Closed Won deals.
Deal owner differs from company owner	27.1%	627	MED	Export both owners; compare.
Lead status set on contacts below MQL	4.0%	1,929	LOW	Filter: Lead status is known AND Lifecycle stage < MQL.
Custom properties with under 1% fill	41 props	—	LOW	Property settings; sort by "used in records".

METHOD

- Read-only access only. No records were created, edited, merged, or deleted.
- Every count comes from a saved filter or a full export, taken on the audit date. Filters are named in the "How to reproduce" column so your team can re-run them.
- Duplicate checks normalize case, whitespace, and common email variants before grouping.
- Scores weight each check by the number of standard reports and workflows it distorts.

WHAT HAPPENS NEXT

- The video walkthrough covers the "fix first" list on page 2 in order, with the filters open on screen.
- This report is yours. Hand it to your team, your agency, or your HubSpot partner. Nothing in it requires us.
- Access to the portal can be revoked today.