

# Data Cleaning Software Evaluation

## Questions to Ask Vendors

Use this checklist during vendor demos and RFP processes to ensure you're evaluating what matters.

---

### 1. Duplicate Detection

- What matching algorithms do you use (exact, fuzzy, semantic)?
- Can I adjust sensitivity thresholds for different use cases?
- How do you handle company name variations (Inc., LLC, Corp.)?
- What is your false positive rate for duplicate detection?
- Can I preview matches before merging?

### 2. Format Standardization

- How many phone number formats do you support?
- Do you support international formats (E.164)?
- Do you detect and correct common email typos?
- How do you handle professional credentials (MD, PhD, CPA)?
- What address components can you standardize?

### 3. Missing Value Handling

- What methods do you use for imputation?
- How do you calculate confidence scores for predictions?
- Can I set thresholds for automatic vs. manual review?
- What data relationships does the system learn from?
- Can I disable imputation for specific fields?

### 4. Anomaly Detection

- What detection algorithms do you use (Z-score, IQR, Isolation Forest)?
- Can I adjust sensitivity by field type?
- How are flagged anomalies presented for review?
- Can I teach the system what is normal for my data?
- Do you detect logical impossibilities (future birthdates, negative values)?

## 5. Security & Governance

- What security certifications do you hold (SOC 2, ISO 27001)?
- Where is data processed and stored?
- Can you provide a SOC 2 Type II report?
- How long are audit logs retained?
- Do you support SSO and role-based access controls?
- Is data encrypted in transit and at rest?

## 6. Integrations

- Which platforms do you integrate with natively?
- Is the integration read-only or bidirectional?
- How do you handle conflicts when pushing data back?
- What is the sync frequency?
- Do you provide API access for custom integrations?