Baroody Data Solutions
DATA PROFILING CHECKLIST

A structured, general purpose checklist for assessing data quality before analysis, reporting or dashboarding

DATA STRUCTURE CHECKS

1. Row Level & Structure Checks

- ○ Record total number of rows
- ○ Segment records into key status groups (i.e. active vs. inactive, open vs. closed)
- ○ Validate adherence to expected column data types
- ○ Calculate % of NULL and/or Blank fields

2. Primary Key & Uniqueness

- ○ Check for uniqueness (no duplicate primary keys)
- ○ Check for nulls

3. Foreign Key & Relationship Integrity

- ○ Identify reference or look-up fields
- ○ Find % of records with valid references
- ○ Identify orphaned records
- Confirm relationships match business expectations

DATA CONSISTENCY CHECKS

4. Date Fields

- ○ Review earliest and latest values
- ○ Check for future dated records
- ○ Validate logical ordering of related fields

5. Numeric Fields

- ○ Review max, min & average values
- ○ Identify unexpected zero or negative values
- ○ Flag extreme outliers

6. Text Fields

- ○ Find % or records with text populated
- ○ Review max, min & average text length
- ○ Identify empty strings vs. NULL values

BUSINESS LOGIC CHECKS

7. Logical Checks

- ○ Identify potential logical duplicates
- ○ Find invalid overlapping time periods
- ○ Compare related fields for consistency

8. Reasonableness Checks

- ○ Flag unusually long or short date ranges
- ○ Validate key ratios or per-unit metrics
- ○ Look for repeated phrases or patterns
- ○ Flag combinations that appear unlikely or inconsistent
- ○ Distinguish true errors from valid edge-cases

DATA READINESS SUMMARY

Overall Readiness

- ☐ High – data is reliable for most reporting
- ☐ Medium – usable with noted limitations
- ☐ Low – significant issues must be addressed

Key Findings

- ☐ No major structural issues identified
- ☐ Some missing or inconsistent fields
- ☐ Business rules partially violated
- Incomplete data for certain time periods

notes

_____
_____
_____
_____

Known Risks & Limitations

- ☐ Missing values in critical fields
- ☐ Inconsistent definitions and fields
- ☐ Potential duplicates or overlaps
- Historical gaps or migration issues

notes

_____
_____
_____
_____