

Care Gap Detection Dataset

CGDD v1.0 — Dataset Fact Sheet

1,000,000

Records

149

Fields

12

Domains

97.4

Cert Score

Source Intelligence

Source Type	Pct	Primary Sources
CMS_SOURCE	28.0%	CMS Medicare Claims, Enrollment, HEDIS Reporting
MEDICAID_SOURCE	12.0%	CMS Medicaid Claims, Managed Care Enrollment
HEDIS_SOURCE	20.0%	NCQA HEDIS Technical Specifications, Measure Reporting
STAR_SOURCE	12.0%	CMS Star Ratings, Measure Weights, Plan Performance
CLINICAL_SOURCE	10.0%	CDC Guidelines, USPSTF Recommendations, ADA, ACC/AHA
DERIVED_INTELLIGENCE	12.0%	Risk scores, gap closure models, predictive analytics
SYNTHETIC_ENRICHMENT	6.0%	Edge case enrichment, data completeness

Distribution Tiers

Tier	File	Records	Contents
Sample	Sample.zip	500	Sample CSV, Data Dictionary, Overview, Fact Sheet, Glossary, Release N
Developer	Developer.zip	25,000	Dataset CSV (25K), Data Dictionary, Metadata, Schema, all PDFs, License
Professional	Professional.zip	250,000	Dataset CSV+Parquet (250K), Source Lineage, Coverage Report, all PDFs
Enterprise/OEM	Enterprise_OEM.zip	1,000,000	Full dataset, all 8 registries, Quality Assessment, Certification Report