

Data Quality Assessment Report

Dataset: fact_txn

Generated: 2025-10-17 09:46:33

Executive Summary

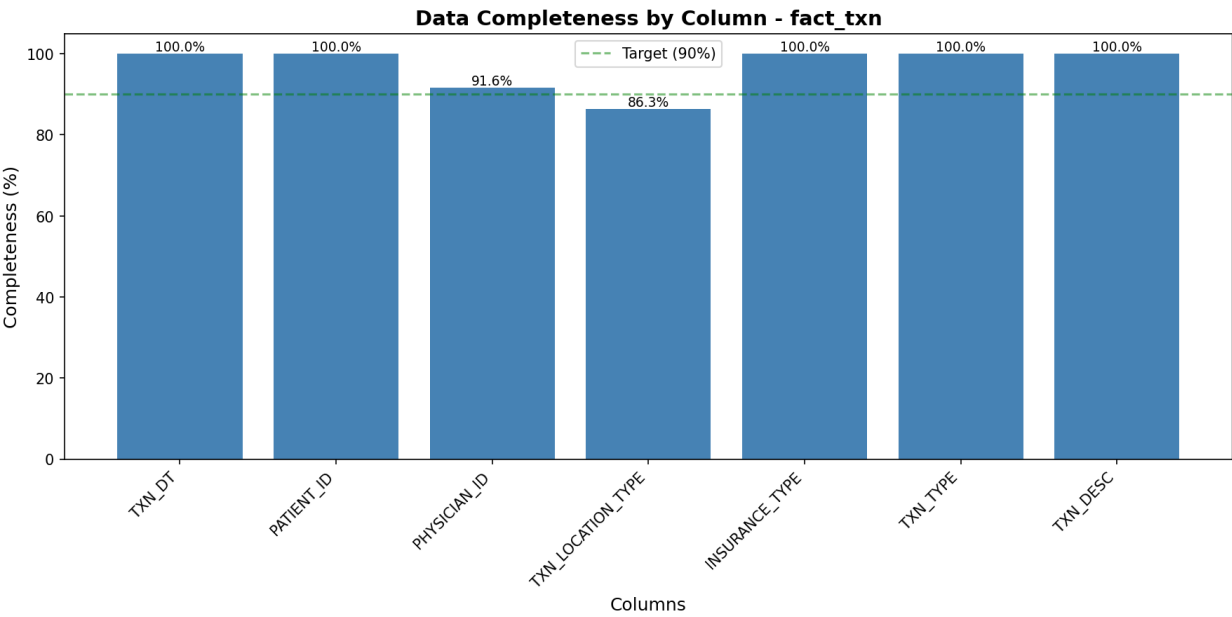
Metric	Value
Overall Quality Score	98.7%
Quality Grade	A (Excellent)
Total Records	115,274
Total Columns	7
Completeness Score	96.84%
Consistency Score	99.87%
Validity Score	100%

Data Quality Visualization



1. Data Completeness Analysis

Column	Total Records	Non-Null	Null Count	Null %	Completeness
TXN_DT	115,274	115,274	0	0.0%	100.0%
PATIENT_ID	115,274	115,274	0	0.0%	100.0%
PHYSICIAN_ID	115,274	105,555	9,719	8.43%	91.57%
TXN_LOCATION_TYPE	115,274	99,479	15,795	13.7%	86.3%
INSURANCE_TYPE	115,274	115,274	0	0.0%	100.0%
TXN_TYPE	115,274	115,274	0	0.0%	100.0%
TXN_DESC	115,274	115,274	0	0.0%	100.0%



2. Data Consistency Analysis

Duplicate Records: 152 (0.13%)

Format Issues Detected:

Column	Issue	Affected Rows
TXN_DESC	Mixed case	115274

3. Data Validity Analysis

Column	Data Type	Unique Values	Unique %
TXN_DT	object	2,030	1.76%
PATIENT_ID	int64	4,020	3.49%
PHYSICIAN_ID	float64	25,343	21.99%
TXN_LOCATION_TYPE	object	48	0.04%
INSURANCE_TYPE	object	5	0.0%
TXN_TYPE	object	4	0.0%
TXN_DESC	object	53	0.05%

Statistical Summary for Numeric Columns

Column	Min	Max	Mean	Median	Std Dev	Outliers
PATIENT_ID	1	4020	2396.16	2501.0	1054.0	0 (0.0%)
PHYSICIAN_ID	1.0	25343.0	14116.74	14998.0	7265.22	0 (0.0%)

5. Data Timeliness Analysis

Column	Earliest Date	Latest Date	Date Range (Days)
TXN_DT	2016-12-31 00:00:00	2022-07-22 00:00:00	2,029