testcase12_parquet_parquet_mismatch_manual duplicate comparison

1. Run Summary

Application Name	:etl_pipeline_goldlayer
Protocol Name	:DATFDemo
Protocol File Path	:datf_core//test/testprotocol/testselection_template.xlsx
Testcase Name	:testcase12_parquet_parquet_mismatch_manual
Testcase Type	:duplicate
Test Environment	:SIT
Start Time	:11-Mar-2025 14:15:28 UTC
End Time	:11-Mar-2025 14:15:41 UTC
Run Time	:0:00:12
Test Result	:Failed
Reason	:Duplicates

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Manual
Testcase Type	:duplicate
Source Connection Name	:RawS3Bucket
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Bucket
Source Format	:parquet
Source Name	:patient_data_source_parquet
Source Path	:test/data/source
Target Connection Name	:CuratedS3Bucket
Target Connection Type	:aws-s3
Target Connection Value	:CuratedS3Bucket
Target Format	:parquet
Target Name	:patient_data_target_parquet
Target Path	:test/data/target
Primary Keys	:id

3. Duplicate Summary

Test Result	:Failed
No. of rows in Source	:1,001,000
No. of distinct rows in Source	:1,000,000
No. of duplicate rows in Source	:1,000
No. of rows in Target	:1,001,000
No. of distinct rows in Target	:1,000,000
No. of duplicate rows in Target	:1,000

4. SQL Queries

4.1 Source Query

SELECT Count(DISTINCT id) FROM rawpatients

4.2 Target Query

SELECT Count(DISTINCT id) FROM curatedpatients