

## testcase12\_parquet\_parquet\_mismatch\_manual fingerprint 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	:fingerprint
Test Environment	:SIT
Start Time	:07-Mar-2025 11:51:40 UTC
End Time	:07-Mar-2025 11:52:54 UTC
Run Time	:0:01:13
Test Result	:Failed
Reason	:Few columns has nulls in src & tgt. columns has duplicates in src & tgt.

### 2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Manual
Testcase Type	:fingerprint
Source Connection Name	:RawS3Bucket
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Bucket
Source Format	:parquet
Source Name	:patients_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	:patients_source_parquet
Target Path	:test/data/source
Primary Keys	:id

### 3. Fingerprint Summary

Test Result	:Failed
No. of KPIs in Source	:5
No. of KPIs in Target	:5
No. of KPIs matched	:3
No. of KPIs mismatched	:2

### 4. SQL Queries

#### 4.1 Source Query

```
SELECT  
ADDRESS,BIRTHDATE,BIRTHPLACE,CITY,COUNTY,DEATHDATE,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTH  
CARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,  
STATE,SUFFIX,ZIP,id FROM rawpatients
```

#### 4.2 Target Query

```
SELECT  
ADDRESS,BIRTHDATE,BIRTHPLACE,CITY,COUNTY,DEATHDATE,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTH  
CARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,  
STATE,SUFFIX,ZIP,id FROM curatedpatients
```

### 5. Fingerprint Comparison Report

#### 5.1. KPI (1) - Row Count

Source Row Count	:1171
------------------	-------

Target Row Count	:1171
KPI Status	:Passed
Reason	:Source and Target row count is matching

## 5.2. KPI (2) - Column Count

Source Column Count	:25
Target Column Count	:25
KPI Status	:Passed
Reason	:Source column count and target column count is matching

## 5.3. KPI (3) - Column Level Null Check

List of Columns has null in source	:['DEATHDATE', 'DRIVERS', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'SUFFIX', 'ZIP']
List of Columns has null in target	:['DEATHDATE', 'DRIVERS', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'SUFFIX', 'ZIP']
KPI Status	:Failed
Reason	:Source or/and Target has nulls in some columns

## 5.4. KPI (4) - Column Level Duplicate Check

List of Columns has duplicates in source:	['BIRTHDATE', 'BIRTHPLACE', 'CITY', 'COUNTY', 'DEATHDATE', 'DRIVERS', 'ETHNICITY', 'FIRST', 'GENDER', 'HEALTHCARE_COVERAGE', 'LAST', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'RACE', 'SSN', 'STATE', 'SUFFIX', 'ZIP']
List of Columns has duplicates in target	:['BIRTHDATE', 'BIRTHPLACE', 'CITY', 'COUNTY', 'DEATHDATE', 'DRIVERS', 'ETHNICITY', 'FIRST', 'GENDER', 'HEALTHCARE_COVERAGE', 'LAST', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'RACE', 'SSN', 'STATE', 'SUFFIX', 'ZIP']
KPI Status	:Failed
Reason	:Source or/and Target has duplicate values in some columns

## 5.5. KPI (5) - Check Sum Row Level

KPI Status	:Passed
Reason	:Data is matching between source and target

### 5.5.1 Addition Rows in Source

None

### 5.5.2 Addition Rows in Target

None