# testcase12_parquet_parquet_mismatch_manual

## null comparison

## 1. Run Summary

| | |
|---|---|
| Application Name | :etl_pipeline_goldlayer |
| Protocol Name | :DATFDemo |
| Protocol File Path | :/Workspace/Shared/QE_ATF_Latest/datf_core/test/testprotocol/traversedtestprotocol.xlsx |
| Testcase Name | :testcase12_parquet_parquet_mismatch_manual |
| Testcase Type | :null |
| Test Environment | :SIT |
| Start Time | :16-Jan-2025 10:08:46 UTC |
| End Time | :16-Jan-2025 10:09:21 UTC |
| Run Time | :0:00:34 |
| Test Result | :Failed |
| Reason | :Null count is not matching between source and target for each column. Please look into section 5.3 for more information |

## 2. Configuration Details

| | |
|---|---|
| Compare Type | :likeobjectcompare |
| testquerygenerationmode | :Manual |
| Testcase Type | :null |
| 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. Null Summary

| | |
|---|---|
| Test Result | :Failed |
| No of column has null in source | :9 |
| No of column has null in target | :8 |
| No of columns has null count match | :8 |
| No of Columns has null count mismatch | :1 |

## 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. Columns having nulls at source and target

### 5.1 Columns having nulls in source

| S.No | Column | Count |
|---|---|---|

| | | |
|---|---|---|
| 1 | DEATHDATE | 1000 |
| 2 | DRIVERS | 213 |
| 3 | MAIDEN | 840 |
| 4 | MARITAL | 380 |
| 5 | PASSPORT | 273 |
| 6 | PREFIX | 244 |
| 7 | SUFFIX | 1159 |
| 8 | ZIP | 543 |
| 9 | Column1 | 1171 |

## 5.2 Columns having nulls in target

| S.No | Column | Count |
|---|---|---|
| 1 | DEATHDATE | 1000 |
| 2 | DRIVERS | 213 |
| 3 | MAIDEN | 840 |
| 4 | MARITAL | 380 |
| 5 | PASSPORT | 273 |
| 6 | PREFIX | 244 |
| 7 | SUFFIX | 1159 |
| 8 | ZIP | 543 |

## 5.3 Columns having null count mismatch between source and target

| S.No | Source Column Name | Src Null Count | Target Column Name | Tgt Null Count |
|---|---|---|---|---|
| 1 | Column1 | 1171 | Column1 | 0 |

## 1. Run Summary

| | |
|---|---|
| Application Name | :etl_pipeline_goldlayer |
| Protocol Name | :DATFDemo |
| Protocol File Path | :/Workspace/Shared/QE_ATF_Latest/datf_core/test/testprotocol/traversedtestprotocol.xlsx |
| Testcase Name | :testcase15_parquet_parquet_match_likeobject |
| Testcase Type | :null |
| Test Environment | :SIT |
| Start Time | :16-Jan-2025 10:09:28 UTC |
| End Time | :16-Jan-2025 10:10:00 UTC |
| Run Time | :0:00:32 |
| Test Result | :Failed |
| Reason | :Null count is not matching between source and target for each column. Please look into section 5.3 for more information |

## 2. Configuration Details

| | |
|---|---|
| Compare Type | :likeobjectcompare |
| testquerygenerationmode | :Auto |
| Testcase Type | :null |
| 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. Null Summary

| | |
|---|---|
| Test Result | :Failed |
| No of column has null in source | :9 |
| No of column has null in target | :8 |
| No of columns has null count match | :8 |
| No of Columns has null count mismatch | :1 |

## 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 actualpatients

### 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 expectedpatients

## 5. Columns having nulls at source and target

### 5.1 Columns having nulls in source

| S.No | Column | Count |
|---|---|---|

| | | |
|---|---|---|
| 1 | DEATHDATE | 1000 |
| 2 | DRIVERS | 213 |
| 3 | MAIDEN | 840 |
| 4 | MARITAL | 380 |
| 5 | PASSPORT | 273 |
| 6 | PREFIX | 244 |
| 7 | SUFFIX | 1159 |
| 8 | ZIP | 543 |
| 9 | Column1 | 1171 |

## 5.2 Columns having nulls in target

| S.No | Column | Count |
|---|---|---|
| 1 | DEATHDATE | 1000 |
| 2 | DRIVERS | 213 |
| 3 | MAIDEN | 840 |
| 4 | MARITAL | 380 |
| 5 | PASSPORT | 273 |
| 6 | PREFIX | 244 |
| 7 | SUFFIX | 1159 |
| 8 | ZIP | 543 |

## 5.3 Columns having null count mismatch between source and target

| S.No | Source Column Name | Src Null Count | Target Column Name | Tgt Null Count |
|---|---|---|---|---|
| 1 | Column1 | 1171 | Column1 | 0 |