

# testcase12\_parquet\_parquet\_mismatch\_manual

## null 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 :null  
Test Environment :SIT  
Start Time :12-Mar-2025 10:36:05 UTC  
End Time :12-Mar-2025 10:36:49 UTC  
Run Time :0:00:44  
Test Result :Passed  
Reason :Nulls are matching for each column

### 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 :Passed  
No of column has null in source :8  
No of column has null in target :8  
No of columns has null count match :8  
No of columns has null count mismatch :0

### 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

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

None

# testcase23\_parquet\_snowflake\_mismatch

## null comparison

### 1. Run Summary

Application Name	:etl_pipeline_goldlayer
Protocol Name	:DATFDemo
Protocol File Path	:datf_core//test/testprotocol/testselection_template.xlsx
Testcase Name	:testcase23_parquet_snowflake_mismatch
Testcase Type	:null
Test Environment	:SIT
Start Time	:12-Mar-2025 10:37:08 UTC
End Time	:12-Mar-2025 10:38:01 UTC
Run Time	:0:00:53
Test Result	:Failed
Reason	:Nulls are not matching for each column

### 2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:null
Source Connection Name	:RawS3Parquet
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Parquet
Source Format	:parquet
Source Name	:patients_source_parquet
Source Path	:test/data/source
Source Exclude Columns	:BIRTHDATE,DEATHDATE
Target Connection Name	:raw_snowflake_sql_connection
Target Connection Type	:snowflake
Target Connection Value	:raw_snowflake_sql_connection
Target Format	:table
Target Name	:SAMPLE_SRC_PATIENTS
Target Exclude Columns	:BIRTHDATE,DEATHDATE
Primary Keys	:id

### 3. Null Summary

Test Result	:Failed
No of column has null in source	:7
No of column has null in target	:7
No of columns has null count match	:1
No of columns has null count mismatch	:8

### 4. SQL Queries

#### 4.1 Source Query

```
SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP,id
FROM actualpatients
```

#### 4.2 Target Query

```
SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALTHCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP
FROM SAMPLE_SRC_PATIENTS
```

### 5. Columns having nulls at source and target

#### 5.1 Columns having nulls in source

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

1	DRIVERS	213
2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	ZIP	543

## 5.2 Columns having nulls in target

S.No	Column	Count
1	DRIVERS	213
2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	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	MAIDEN	840	LON	0
2	MARITAL	380	MAIDEN	840
3	PASSPORT	273	MARITAL	380
4	PREFIX	244	PASSPORT	273
5	RACE	0	PREFIX	244
6	SUFFIX	1159	STATE	0
7	ZIP	543	SUFFIX	1159
8	id	0	ZIP	543

# testcase24\_snowflake\_snowflake\_etljob

## null comparison

### 1. Run Summary

Application Name :etl\_pipeline\_goldlayer  
Protocol Name :DATFDemo  
Protocol File Path :datf\_core//test/testprotocol/testselection\_template.xlsx  
Testcase Name :testcase24\_snowflake\_snowflake\_etljob  
Testcase Type :null  
Test Environment :SIT  
Start Time :12-Mar-2025 10:38:26 UTC  
End Time :12-Mar-2025 10:39:17 UTC  
Run Time :0:00:50  
Test Result :Failed  
Reason :Nulls are not matching for each column

### 2. Configuration Details

Compare Type :likeobjectcompare  
testquerygenerationmode :Auto  
Testcase Type :null  
Source Connection Name :raw\_snowflake\_sql\_connection  
Source Connection Type :snowflake  
Source Connection Value :raw\_snowflake\_sql\_connection  
Source Format :table  
Source Name :SAMPLE\_SRC\_PATIENTS  
Source Exclude Columns :BIRTHDATE,DEATHDATE  
Target Connection Name :raw\_snowflake\_sql\_connection  
Target Connection Type :snowflake  
Target Connection Value :raw\_snowflake\_sql\_connection  
Target Format :table  
Target Name :SAMPLE\_TGT\_PATIENTS  
Target Exclude Columns :BIRTHDATE,DEATHDATE  
Primary Keys :ID

### 3. Null Summary

Test Result :Failed  
No of column has null in source :7  
No of column has null in target :9  
No of columns has null count match :3  
No of columns has null count mismatch :6

### 4. SQL Queries

#### 4.1 Source Query

```
SELECT  
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT  
HCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP  
FROM SAMPLE_SRC_PATIENTS
```

#### 4.2 Target Query

```
SELECT  
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT  
HCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP  
FROM SAMPLE_TGT_PATIENTS
```

### 5. Columns having nulls at source and target

#### 5.1 Columns having nulls in source

S.No	Column	Count
1	DRIVERS	213

2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	ZIP	543

## 5.2 Columns having nulls in target

S.No	Column	Count
1	DRIVERS	213
2	ETHNICITY	6
3	MAIDEN	839
4	MARITAL	381
5	PASSPORT	273
6	PREFIX	244
7	RACE	5
8	SUFFIX	1157
9	ZIP	541

## 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	ETHNICITY	0	ETHNICITY	6
2	MAIDEN	840	MAIDEN	839
3	MARITAL	380	MARITAL	381
4	RACE	0	RACE	5
5	SUFFIX	1159	SUFFIX	1157
6	ZIP	543	ZIP	541

# testcase31\_snowflake\_parquet\_validation

## null comparison

### 1. Run Summary

Application Name :etl\_pipeline\_goldlayer  
Protocol Name :DATFDemo  
Protocol File Path :datf\_core//test/testprotocol/testselection\_template.xlsx  
Testcase Name :testcase31\_snowflake\_parquet\_validation  
Testcase Type :null  
Test Environment :SIT  
Start Time :12-Mar-2025 10:39:40 UTC  
End Time :12-Mar-2025 10:40:02 UTC  
Run Time :0:00:21  
Test Result :Failed  
Reason :Nulls are not matching for each column

### 2. Configuration Details

Compare Type :likeobjectcompare  
testquerygenerationmode :Auto  
Testcase Type :null  
Source Connection Name :raw\_snowflake\_sql\_connection  
Source Connection Type :snowflake  
Source Connection Value :raw\_snowflake\_sql\_connection  
Source Format :table  
Source Name :rtPCR\_diagnostic\_lab\_testing  
Target Connection Type :aws-s3  
Target Format :parquet  
Target Name :patients\_target\_parquet\_mismatch  
Target Path :test/data/target  
S2T Path :test/s2t/s2t\_31\_snowflake\_parquet\_val.xlsx  
Primary Keys :id,state

### 3. Null Summary

Test Result :Failed  
No of column has null in source :0  
No of column has null in target :8  
No of columns has null count match :0  
No of columns has null count mismatch :2

### 4. SQL Queries

#### 4.1 Source Query

```
SELECT
DATE,FEMA_REGION,NEW_RESULTS_REPORTED,OVERALL_OUTCOME,STATE,STATE_FIPS,STATE_NAME,TOTAL_RESULTS_REPORTED FROM rtPCR_diagnostic_lab_testing
```

#### 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 rtPCR_target
```

### 5. Columns having nulls at source and target

#### 5.1 Columns having nulls in source

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

#### 5.2 Columns having nulls in target

<b>S.No</b>	<b>Column</b>	<b>Count</b>
1	DEATHDATE	998
2	DRIVERS	213
3	MAIDEN	838
4	MARITAL	380
5	PASSPORT	273
6	PREFIX	244
7	SUFFIX	1157
8	ZIP	543

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

<b>S.No</b>	<b>Source Column Name</b>	<b>Src Null Count</b>	<b>Target Column Name</b>	<b>Tgt Null Count</b>
1	STATE_FIPS	0	DEATHDATE	998
2	STATE_NAME	0	DRIVERS	213