

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	:10-Mar-2025 17:08:48 UTC
End Time	:10-Mar-2025 17:16:01 UTC
Run Time	:0:07:12
Test Result	:Failed
Reason	:One or more fingerprint mismatched. Please check detail report for more info

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	: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. Fingerprint Summary

Test Result	:Failed
No. of KPIs in Source	:9
No. of KPIs in Target	:9
No. of KPIs matched	:4
No. of KPIs mismatched	:5

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	:1001000
Target Row Count	:1001000

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', 900877), ('DRIVERS', 200539), ('MAIDEN', 800319), ('PASSPORT', 300807), ('SUFFIX', 800558)]  
List of Columns has null in target :[('DEATHDATE', 901273), ('DRIVERS', 200008), ('MAIDEN', 800625), ('PASSPORT', 299989), ('SUFFIX', 800405)]  
KPI Status :Failed  
Reason :Source or/and Target has nulls in some columns

5.4. KPI (4) - Row Level Null Check

No of null rows in source :0  
No of null rows in target :0  
KPI Status :Passed  
Reason :No rows are null

5.5. KPI (5) - Column Level empty Check

List of Columns has empty in source :[]  
List of Columns has empty in target :[]  
KPI Status :Passed  
Reason :Source or/and Target has no empty values in any columns

5.6. KPI (6) - Column Level Duplicate Check

List of Columns has duplicates in source:[('ADDRESS', 'BIRTHDATE', 'BIRTHPLACE', 'CITY', 'COUNTY', 'DEATHDATE', 'DRIVERS', 'ETHNICITY', 'FIRST', 'GENDER', 'HEALTHCARE\_COVERAGE', 'HEALTHCARE\_EXPENSES', 'LAST', 'LAT', 'LON', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'RACE', 'SSN', 'STATE', 'SUFFIX', 'ZIP', 'id']  
List of Columns has duplicates in target :[('ADDRESS', 'BIRTHDATE', 'BIRTHPLACE', 'CITY', 'COUNTY', 'DEATHDATE', 'DRIVERS', 'ETHNICITY', 'FIRST', 'GENDER', 'HEALTHCARE\_COVERAGE', 'HEALTHCARE\_EXPENSES', 'LAST', 'LAT', 'LON', 'MAIDEN', 'MARITAL', 'PASSPORT', 'PREFIX', 'RACE', 'SSN', 'STATE', 'SUFFIX', 'ZIP', 'id']  
KPI Status :Failed  
Reason :Source or/and Target has duplicate values in some columns  
Dup values match source & target :Not Matched

5.6.1 Source Duplicate - Column Level

S.No	Column Name	Duplicate Values
1	ADDRESS	ADXfCnxCEXocEiN

2	BIRTHDATE	1995-06-25 2000-07-03 2021-10-06 1946-02-25 1938-09-17 1961-11-01 1974-05-11 1951-11-16 1996-02-09 1938-03-12 2018-07-23 1939-04-28 1969-06-16 1930-04-07 1965-11-12 1966-11-30 2025-01-23 1950-07-29 1996-02-29 1983-06-16 1993-08-26 1946-07-15 1999-02-01 1970-11-04 1954-09-19 1993-12-19 1985-10-19 1935-10-22 1996-12-22 1972-05-02 2004-06-12 2013-02-17 1988-01-13 1959-01-22 1962-04-18 1933-09-13 1992-10-30 1993-09-26 1955-07-03 2002-06-24 1965-04-01 1990-11-30 1966-02-20 2015-03-09 1944-05-08 1949-08-28 1967-05-31 1936-11-28 1966-05-22 1944-05-25
3	BIRTHPLACE	dloniiUYFPs
4	CITY	LSWJdAweZl
5	COUNTY	oOOquzvTyq
6	DEATHDATE	2002-12-25 1998-05-16 2056-06-14 1985-10-19 1981-03-25 1996-10-17 2054-09-02 2051-02-17 2013-05-21 2025-10-21 2025-01-23 2026-05-09 1966-11-30 2053-08-22 2014-10-30 1982-07-02 1971-05-12 2038-10-03 2070-01-14 2000-12-26 1998-06-16 1996-12-22 2032-10-28 1976-10-06 2013-05-13 1993-09-26 2025-10-19 2073-11-15 1990-11-30 1961-06-12 2059-01-11 1995-12-01 1984-02-24 2011-10-03 2028-08-19 2048-05-16 1978-02-08 1972-05-02 2078-06-18 2036-09-07 2051-09-07 1993-08-26 2002-06-24 2006-05-21 2004-06-12 1999-02-01 1974-05-11 2010-10-04 1997-03-25 1999-08-14
7	DRIVERS	KRKBvUYNsj
8	ETHNICITY	Hispanic
9	FIRST	qwkUmh
10	GENDER	M

11	HEALTHCARE_COVERAGE	2332.0283 4295.82 9922.7 9508.8545 6039.391 9233.855 7202.031 8677.452 9151.201 7199.0083 4371.5713 1138.1304 9051.257 2841.8489 3402.4353 6302.516 5672.1807 6476.467 8373.373 9541.741 3218.9463 8618.104 9348.429 7311.551 8683.429 3838.0679 5035.4736 6766.944 8230.819 5874.9556 6271.314 4327.363 9508.24 8517.084 9869.938 5067.9414 9021.376 3419.4133 7391.938 2834.0647 8645.987 9755.84 5383.6924 4760.1143 5563.7793 2430.2488 6567.7104 5557.613 5695.6 3368.5684
12	HEALTHCARE_EXPENSES	89438.555 7914.237 78032.13 97387.57 73930.6 96596.73 24487.027 47904.16 73763.63 62658.613 35284.008 49049.93 88664.48 95825.68 29221.012 84458.336 72270.625 48773.918 80116.73 66578.84 85233.81 82314.805 78913.34 99287.016 11483.324 90369.445 45188.707 84168.03 80868.78 56432.773 4583.003 81258.48 96876.37 36522.727 37383.11 57153.926 88546.945 78378.195 78099.0 43357.234 82980.24 95811.75 99237.87 92759.016 58660.336 57088.465 86541.96 38048.46 83896.38 53739.72
13	LAST	URMXGQla

14	LAT	-36.174915 -38.83365 -72.512856 -82.30611 72.34066 -70.67419 -74.253975 75.150734 68.56535 -84.79017 82.08587 -38.47233 -70.503105 67.81425 -39.445072 -54.764626 65.030014 -67.09473 -71.631935 32.786507 -83.50698 73.061005 87.15616 -76.01226 -71.467224 -54.872097 71.89983 -75.44695 -86.84712 -35.380646 -84.11596 -18.64083 -42.292885 67.23778 55.741875 -11.575984 76.24993 70.88307 -15.46162 -74.22823 -38.34619 -59.220173 47.218693 -50.534386 -65.583336 -86.129456 -12.212008 56.63808 -51.08525 48.182552
15	LON	-64.37452 19.457335 7.548823 151.55872 -147.51355 -99.3252 -136.56503 175.46852 -158.66997 -144.25652 -149.07501 -129.04736 -136.37477 122.23118 -115.21688 114.82724 -82.54124 153.05023 95.38273 145.96896 -120.35809 134.65474 -116.42556 179.43033 65.06307 -122.30143 156.56236 -58.253273 167.27293 -171.44267 176.4207 80.46689 -148.84761 -143.86943 -84.23963 -83.072296 156.18481 151.21219 34.214268 77.20992 -40.000248 156.37428 -167.58911 143.07114 -130.55138 -150.38463 -84.37634 -121.73278 164.79257 -169.24231
16	MAIDEN	HOslOtnn
17	MARITAL	Widowed
18	PASSPORT	uSKsADpZo
19	PREFIX	Mr.
20	RACE	Other
21	SSN	167-89-2702
22	STATE	FL
23	SUFFIX	Sr.
24	ZIP	17395

25	id	27 63 303 389 474 754 964 51 65 379 418 541 558 192 234 243 440 14 279 860 346 415 608 636 846 446 598 941 957 385 736 982 17 135 292 738 809 932 59 77 542 453 463 493 939 15 165 188 878 180
----	----	--

### 5.6.2 Target Duplicate - Column Level

S.No	Column Name	Duplicate Values
1	ADDRESS	BMdCtYUwnZPBSLO
2	BIRTHDATE	2018-07-23 1951-03-26 1970-08-18 2019-05-23 1963-06-14 1999-04-23 1978-02-08 1969-06-16 1975-10-24 1970-05-29 1999-02-01 1981-03-25 1945-10-21 1967-05-31 1928-02-10 1960-10-16 1944-06-20 1982-05-17 1996-02-09 1993-08-26 2001-07-13 1995-12-01 2000-12-26 2020-11-19 1962-06-06 1974-03-26 2002-05-26 1993-04-09 1942-12-28 1968-06-15 1998-05-16 1951-11-16 2018-08-21 1932-04-20 1979-09-24 1951-05-31 1959-01-22 2023-11-30 2014-10-30 1987-05-29 1999-08-14 1976-10-06 2002-06-24 1983-06-16 1946-09-26 1992-10-30 1968-06-22 1971-05-12 2002-12-25 2004-12-29
3	BIRTHPLACE	iMfoNgqGOg
4	CITY	vyZdOaPDSC
5	COUNTY	VKAAvLZtXp

6	DEATHDATE	2053-08-22 2014-09-26 2056-06-14 2095-12-18 2064-11-16 1970-11-04 2062-03-21 1997-12-11 2036-03-06 2053-08-07 1984-07-16 1954-09-19 2003-11-26 2039-09-17 1982-05-17 1970-08-18 2002-06-24 2079-07-19 1974-03-26 1993-09-26 2021-10-06 2070-01-14 2019-05-23 2059-03-27 2010-10-04 2042-11-22 2096-08-29 1992-10-30 2067-09-23 2072-02-09 2025-10-19 2075-08-13 1963-06-14 2094-11-21 1949-09-28 2062-05-19 1992-07-20 2027-04-26 2051-01-06 1992-07-10 2046-08-04 1980-06-01 1982-07-02 1969-01-04 2040-01-25 1939-10-26 1976-09-16 1998-06-16 2113-06-27 2034-12-13
7	DRIVERS	eraEkPDvck
8	ETHNICITY	Hispanic
9	FIRST	ndHrDn
10	GENDER	M
11	HEALTHCARE_COVERAGE	2474.0137 3567.183 6317.8984 4480.7305 8614.98 2561.5264 5836.6597 5634.6875 9403.682 6579.8086 326.58804 5382.731 4395.5317 8103.9736 3937.457 9732.6045 3871.5342 9414.272 8124.8477 9128.082 7763.2266 4814.4116 6690.552 2427.2183 6295.9824 2983.1692 7667.961 9722.081 5799.533 9485.78 6125.6865 5531.631 9357.568 9487.409 7938.6123 3452.9011 9819.801 5619.5195 6717.0986 5357.607 5562.4595 7565.6084 8789.097 8128.0254 3533.517 6574.1963 9485.636 6700.2944 5320.6626 5386.64

12	HEALTHCARE_EXPENSES	95013.45 32222.83 14466.469  34128.61 68588.27 86192.71  43714.348 67536.836 78089.6 7 70958.86 20530.947 91902. 664 93382.33 31533.332 9619 1.85 37963.203 56353.266 71 495.914 85414.4 72177.73 34 515.457 93050.445 87345.52  38369.344 58941.316 57926.1 8 35116.414 39037.902 82156 .34 62426.754 66294.9 73145. 164 61130.734 80452.81 2065 0.674 74520.55 67912.57 841 57.85 90960.914 77923.65 70 142.14 36712.934 52988.1 56 655.695 74939.22 60608.13 4 5349.41 59086.805 81457.78  18020.09
13	LAST	mhsJsvKd
14	LAT	-74.70412 71.5954 75.41035 5 4.721607 -69.82209 50.99160 4 -64.1288 83.44563 69.37636  76.55224 -88.53318 83.00242 6 70.787575 -50.59904 13.430 673 4.138988 -81.17896 -58.8 8424 -44.67755 -72.79852 -60 .56479 84.3939 -74.32498 -81 .77681 44.899036 80.785736  87.14705 -58.47235 36.30455 4 -80.74235 -78.51176 -74.30 6145 -86.039276 -84.8921 -83 .23455 66.41307 75.58711 66. 41118 64.63207 79.96896 -85. 57225 46.666386 43.743122 7 6.15521 -73.44445 35.27558 5 8.044094 77.10943 -72.73924  53.767693



15	LON	113.871796 -117.64139 -102.414375 102.79169 178.24393 178.47488 -151.10045 149.3523 -108.656586 144.615 64.94117 -89.81952 123.70778 -171.18729 102.672905 152.1052 -100.27339 136.1528 -119.4149 -120.32284 161.45494 158.1419 -132.11752 -179.60558 -177.31209 -166.7786 -157.25455 -164.11705 -75.76287 -152.75513 -65.796585 174.82425 135.51749 -176.4452 -173.55716 152.25089 114.19776 -140.15263 -173.94994 48.710796 162.43169 -160.36923 133.39085 -126.63217 -147.73172 179.96666 178.59157 -34.197285 141.39594 -28.148561
16	MAIDEN	dhuqqoeC
17	MARITAL	Divorced
18	PASSPORT	uFHdivQyy
19	PREFIX	Ms.
20	RACE	Asian
21	SSN	266-70-6856
22	STATE	NY
23	SUFFIX	III
24	ZIP	88546
25	id	27 63 303 389 474 754 964 51 65 379 418 541 558 192 234 243 440 14 279 860 346 415 608 636 846 446 598 941 957 385 736 982 17 135 292 738 809 932 59 77 542 453 463 493 939 15 165 188 878 180

5.7. KPI (7) - Row Level Duplicate Check

No of duplicate rows in source :1000  
No of duplicate rows in target :1000  
KPI Status :Failed  
Reason :One or more rows are duplicated

5.8 Primary Key Duplicate Check

Source Row Count :1001000  
Target Row Count :1001000  
KPI Status :Passed  
Reason :Source and Target row count is matching

5.8.1 Primary Key Duplicate Check - Source

S.No	id	count
1	27	2
2	63	2
3	303	2

4	389	2
5	474	2
6	754	2
7	964	2
8	51	2

### 5.8.2 Primary Key Duplicate Check - Target

S.No	id	count
1	27	2
2	63	2
3	303	2
4	389	2
5	474	2
6	754	2
7	964	2
8	51	2

### 5.9. KPI (9) - CheckSum Row Level

KPI Status:Failed

Reason:Data is not matching between source and target

#### 5.9.1 Addition Rows in Source

S.No	hashval
1	0220751618574bf3bfcf495b633aa9cb
2	029d99347cdae3900473637cdfa37d8b
3	0c72e1ecff51fe83051d9a796faa1064
4	0ea71e926316da934cc6819d694aca93
5	128fa3c130b9bdcf2202cf13184c41d4
6	158479ce1377e81e11d472ad03054532
7	1701d81614415691269ad7fad5078f71
8	1dd798b65bcd9f8e24c3f761a255e773

#### 5.9.2 Addition Rows in Target

S.No	hashval
1	0eb9776ede72734c20e921c051097b39
2	14eeac6c522417bc9aae184fad623bf3
3	17d888c5d0075c2282eee0a389f45e48
4	1f74d2756a97f6b7cd14a9d75debffaf
5	1fa8fc575bf1e75ee81120815f6c2dc4
6	2ad15535510d79645e9d89332a1a6b1c
7	2c014e21bc959a29a66c2466602d7f1f
8	316d98f1083d0349edc4c11338cc066b