Data Quality Analysis Detailed Report

testcase1_parquet_parquet_mismatch

1. Configuration Details

1.1. Source Configuration Details

connectionname	:
connectiontype	:
path	:
format	:
name	:
excludecolumns	:
filter	:
testquerygenerationmode	:Auto
delimiter	:
test_case_name	:testcase1_parquet_parquet_mismatch
autoscripttype	:source
autoscriptpath	:
comparetype	:s2tcompare
filename	:

1.2. Target Configuration Details

connectionname	:
connectiontype	:
path	:
format	:
name	:
excludecolumns	:
filter	:
testquerygenerationmode	:Auto
delimiter	:
test_case_name	:testcase1_parquet_parquet_mismatch
autoscripttype	:target
autoscriptpath	:
comparetype	:s2tcompare
filename	:

2. Testcase Details

2.1. Testcase_COUNTY_Length_Equal_Validation

Test Status	:Failed
Column Name	:COUNTY
DQ Validation	:Length
Expected Length	:10
Element Count	:1000
Unexpected Count	:1000
Sample Mismatches	:['Hampden County', 'Middlesex County', 'Hampden County', 'Middlesex County', 'Suffolk County', 'Hampden County', 'Plymouth County', 'Franklin County', 'Bristol County', 'Norfolk County', 'Hampden County', 'Norfolk County', 'Norfolk County', 'Suffolk County', 'Norfolk County', 'Norfolk County', 'Middlesex County', 'Plymouth County', 'Middlesex County', 'Middlesex County']

2.2. Testcase_CITY_Length_Between_Validation

Test Status	:Failed
Column Name	:CITY
DQ Validation	:Length
Expected Length Range	:10-23
Element Count	:1000
Unexpected Count	:740
Sample Mismatches	:['Chicopee', 'Chicopee', 'Lowell', 'Boston', 'Pembroke', 'Colrain', 'Walpole', 'Medfield', 'Needham', 'Boston', 'Quincy', 'Waltham', 'Norwell', 'Malden', 'Ashland', 'Peabody', 'Boston', 'Malden', 'Peabody', 'Abington']

2.3. Testcase_DRIVERS_Null_Validation

Test Status	:Failed
Column Name	:DRIVERS
DQ Validation	:Null
Expected Value	:'null'
Element Count	:1000
Unexpected Count	:814
Sample Mismatches	:['S99984236', 'S99962402', 'S99972682', 'S99974448', 'S99915787', 'S99954048', 'S99978036', 'S99913545', 'S99993444', 'S99957470', 'S99927965', 'S99983425', 'S99917383', 'S99916076', 'S99927174', 'S99979946', 'S99937445', 'S99924952', 'S99933543', 'S99962675']

2.4. Testcase_ETHNICITY_NotBeNull_Validation

Test Status	:Passed
Column Name	:ETHNICITY
DQ Validation	:NotBeNull
Expected Value	:'not null'
Element Count	:1000
Unexpected Count	:0
Sample Mismatches	:[]

2.5. Testcase_id_Unique_Validation

Test Status	:Passed
Column Name	:id
DQ Validation	:Unique
Expected Value	:to be unique
Element Count	:1000
Unexpected Count	:0
Sample Duplicate Values	:[]

2.6. Testcase_GENDER_DistinctSet_Validation

Test Status	:Failed
Column Name	:GENDER
DQ Validation	:DistinctSet
Expected Distinct Value	:['M','F']
Actual Distinct Value	:['F', 'M']

2.7. Testcase_ColumnCount_Validation

Test Status	:Failed
DQ Validation	:ColumnCount
Expected Column Count	:56
Actual Column Count	:26

2.8. Testcase_HEALTHCARE_COVERAGE_Sum_Between_Validation

Test Status	:Failed
Column Name	:HEALTHCARE_COVERAGE
DQ Validation	:Sum
Expected Sum Range	:9000-10002
Actual Sum	:12911883.180000005

2.9. Testcase_ColumnOrder_Validation

Test Status	:Failed
DQ Validation	:ColumnOrder
Expected Column Order	:['id', 'BIRTHDATE', 'DEATHDATE', 'SSN', 'DRIVERS', 'PASSPORT', 'PREFIX', 'FIRST', 'LAST', 'SUFFIX', 'MAIDEN', 'MARITAL', 'RACE', 'ETHNICITY', 'GENDER', 'BIRTHPLACE', 'ADDRESS', 'CITY', 'STATE', 'COUNTY', 'ZIP', 'LAT', 'LON', 'HEALTHCARE_EXPENSES', 'DName', 'HEALTHCARE_COVERAGE']
Actual Column Order	:['Unnamed: 0', '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']
Column Order Mismatch	:[{'Expected Column Position': 0, 'Expected': 'id', 'Found': 'Unnamed: 0'}, {'Expected Column Position': 1, 'Expected': 'BIRTHDATE', 'Found': 'ADDRESS'}, {'Expected Column Position': 2, 'Expected': 'DEATHDATE', 'Found': 'BIRTHDATE'}, {'Expected Column Position': 3, 'Expected': 'SSN', 'Found': 'BIRTHPLACE'}, {'Expected Column Position': 4, 'Expected': 'DRIVERS', 'Found': 'CITY'}, {'Expected Column Position': 5, 'Expected': 'PASSPORT', 'Found': 'COUNTY'}, {'Expected Column Position': 6, 'Expected': 'PREFIX', 'Found': 'DEATHDATE'}, {'Expected Column Position': 7, 'Expected': 'FIRST', 'Found': 'DRIVERS'}, {'Expected Column Position': 8, 'Expected': 'LAST', 'Found': 'ETHNICITY'}, {'Expected Column Position': 9, 'Expected': 'SUFFIX', 'Found': 'FIRST'}, {'Expected Column Position': 10, 'Expected': 'MAIDEN', 'Found': 'GENDER'}, {'Expected Column Position': 11, 'Expected': 'MARITAL', 'Found': 'HEALTHCARE_COVERAGE'}, {'Expected Column Position': 12, 'Expected': 'RACE', 'Found': 'HEALTHCARE_EXPENSES'}, {'Expected Column Position': 13, 'Expected': 'ETHNICITY', 'Found': 'LAST'}, {'Expected Column Position': 14, 'Expected': 'GENDER', 'Found': 'LAT'}, {'Expected Column Position': 15, 'Expected': 'BIRTHPLACE', 'Found': 'LON'}, {'Expected Column Position': 16, 'Expected': 'ADDRESS', 'Found': 'MAIDEN'}, {'Expected Column Position': 17, 'Expected': 'CITY', 'Found': 'MARITAL'}, {'Expected Column Position': 18, 'Expected': 'STATE', 'Found': 'PASSPORT'}, {'Expected Column Position': 19, 'Expected': 'COUNTY', 'Found': 'PREFIX'}, {'Expected Column Position': 20, 'Expected': 'ZIP', 'Found': 'RACE'}, {'Expected Column Position': 21, 'Expected': 'LAT', 'Found': 'SSN'}, {'Expected Column Position': 22, 'Expected': 'LON', 'Found': 'STATE'}, {'Expected Column Position': 23, 'Expected': 'HEALTHCARE_EXPENSES', 'Found': 'SUFFIX'}, {'Expected Column Position': 24, 'Expected': 'DName', 'Found': 'ZIP'}, {'Expected Column Position': 25, 'Expected': 'HEALTHCARE_COVERAGE', 'Found': 'id'}]

2.10. Testcase_BIRTHDATE_Regexp_Validation

Test Status	:Failed
column Name	:BIRTHDATE
DQ Validation	:Regexp
Expected Pattern	:'^\d{4}-\d{2}-\d{2}$'
Element Count	:1000
Unexpected Count	:1000
Sample Mismatches	:[datetime.datetime(1989, 5, 25, 0, 0), datetime.datetime(1983, 11, 14, 0, 0), datetime.datetime(1992, 6, 2, 0, 0), datetime.datetime(1978, 5, 27, 0, 0), datetime.datetime(1996, 10, 18, 0, 0), datetime.datetime(2017, 7, 27, 0, 0), datetime.datetime(2003, 12, 13, 0, 0), datetime.datetime(2019, 5, 15, 0, 0), datetime.datetime(1970, 5, 16, 0, 0), datetime.datetime(2016, 7, 4, 0, 0), datetime.datetime(2004, 12, 19, 0, 0), datetime.datetime(1991, 7, 3, 0, 0), datetime.datetime(1989, 6, 7, 0, 0), datetime.datetime(1982, 9, 1, 0, 0), datetime.datetime(1958, 7, 1, 0, 0), datetime.datetime(1957, 2, 25, 0, 0), datetime.datetime(1959, 5, 26, 0, 0), datetime.datetime(1954, 4, 4, 0, 0), datetime.datetime(1958, 4, 11, 0, 0), datetime.datetime(1983, 9, 2, 0, 0)]

2.11. Testcase_DEATHDATE_Regexplist_Validation

Test Status	:Failed
column Name	:DEATHDATE
DQ Validation	:Regexplist
Expected Pattern	:['^\d{4}-\d{2}-\d{2}$','^\d{4}/\d{2}/\d{2}$']
Element Count	:1000
Unexpected Count	:139
Sample Mismatches	:[datetime.datetime(2012, 11, 21, 0, 0), datetime.datetime(1991, 4, 30, 0, 0), datetime.datetime(2017, 11, 24, 0, 0), datetime.datetime(2006, 1, 8, 0, 0), datetime.datetime(1989, 1, 17, 0, 0), datetime.datetime(1987, 3, 29, 0, 0), datetime.datetime(1967, 7, 4, 0, 0), datetime.datetime(1986, 2, 15, 0, 0), datetime.datetime(2016, 10, 16, 0, 0), datetime.datetime(2012, 1, 12, 0, 0), datetime.datetime(2007, 12, 11, 0, 0), datetime.datetime(2015, 8, 31, 0, 0), datetime.datetime(1927, 7, 8, 0, 0), datetime.datetime(2010, 3, 28, 0, 0), datetime.datetime(2019, 5, 15, 0, 0), datetime.datetime(1991, 2, 17, 0, 0), datetime.datetime(2016, 2, 2, 0, 0), datetime.datetime(1968, 1, 23, 0, 0), datetime.datetime(2019, 3, 23, 0, 0), datetime.datetime(1989, 11, 25, 0, 0)]