Quality Control Report of
test_dataset2.csv

This QC Report is created on **05/06/2020 21:07:16**

Version of Data Quality Control Tool: **2.0**

File path: tests/test_datasets/test_dataset2.csv

Total number of rows: **474**, Total number of columns: **26**

Is metadata JSON file with table schema provided? **No**

Data Cleaning performed? **No**

Number of rows with invalid values: **0**

Rows Data Completion Overall Statistics

Number of rows with only the ID column filled (the others are filled with nulls)	0 (0.0%)
Number of rows with nulls in ID column	0 (0.0%)
Number of rows with 100% (all 26) of the columns filled	0 (0.0%)
Number of rows with 75-99% (20 - 25) of the columns filled	368 (77.64%)
Number of rows with 50-74% (13 - 19) of the columns filled	69 (14.56%)
Number of rows with 25-49% (6 - 12) of the columns filled	0 (0.0%)
Number of rows with 0-24% (0 - 5) of the colums filled	37 (7.81%)

Rows Data Validation Overall Statistics

Number of rows with 100% (all 26) of the columns filled with valid data (nulls included)	474 (100.0%)
Number of rows with 75-99% (20 - 25) of the columns filled with valid data (nulls included)	0 (0.0%)
Number of rows with 50-74% (13 - 19) of the columns filled with valid data (nulls included)	0 (0.0%)
Number of rows with 25-49% (6 - 12) of the columns filled with valid data (nulls included)	0 (0.0%)
Number of rows with 0-24% (0 - 5) of the colums filled with valid data (nulls included)	0 (0.0%)

Report for column **subjectcode**

Type	text
Total number of rows	474
Number of rows with data	474
Completion percentage	100.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Count of unique values (for text variables)	474
Most frequent value	nG+MUN3T0AD001
Number of occurrences for most frequent value	1
5 most frequent values	['nG+MUN3T0AD001', 'nG+MUN3T0AD020', 'nG+FRA3T0AD008', 'nG+AMS1T5HC006', 'nG+FRA3T0AD013']
5 least frequent values	['nG+MAN3T0MCI009', 'nG+MIL3T0MCI009', 'nG+MAI1T5HC011', 'nG+AMS3T0MCI008', 'nG+MUN3T0HC001']

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **subjectageyears**

Type	integer
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Mode (most frequent value)	71
Number of occurrences for most frequent value	27
Minimum value	36
Maximum value	87
25% of records are below this value(limit value of the first quartile)	66
50% of records are below this value (median)	71
75% of records are below this value(limit value of the third quartile)	76

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **gender**

Type	nominal
Total number of rows	474
Number of rows with data	474
Completion percentage	100.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	F
Number of occurrences for most frequent value	247
List of category values	['F', 'M']
Number of categories	2

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **_3rdventricle**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	1.677
Standard deviation	0.517
Minimum value	0.488
Maximum value	3.769
25% of records are below this value(limit value of the first quartile)	1.262
50% of records are below this value (median)	1.664
75% of records are below this value(limit value of the third quartile)	1.994
Outlier upper bound	3.228
Outlier lower bound	0.125
Total number of outliers (outside 3 std.dev)	4
Rows with outliers	[(13, Decimal('3.4425')), (17, Decimal('3.7693')), (321, Decimal('3.3332')), (425, Decimal('3.2846'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **_4thventricle**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	2.018
Standard deviation	0.538
Minimum value	0.252
Maximum value	6.115
25% of records are below this value(limit value of the first quartile)	1.671
50% of records are below this value (median)	1.928
75% of records are below this value(limit value of the third quartile)	2.271
Outlier upper bound	3.631
Outlier lower bound	0.406
Total number of outliers (outside 3 std.dev)	6
Rows with outliers	[(13, Decimal('3.8623')), (186, Decimal('6.1146')), (310, Decimal('3.7549')), (313, Decimal('4.2231')), (321, Decimal('3.8497')), (419, Decimal('0.25159000000000004'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **rightaccumbensarea**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	0.355
Standard deviation	0.064
Minimum value	0.0
Maximum value	0.683
25% of records are below this value(limit value of the first quartile)	0.323
50% of records are below this value (median)	0.358
75% of records are below this value(limit value of the third quartile)	0.39
Outlier upper bound	0.546
Outlier lower bound	0.164
Total number of outliers (outside 3 std.dev)	8
Rows with outliers	[(69, Decimal('0.00009')), (104, Decimal('0.00066')), (186, Decimal('0.0226')), (251, Decimal('0.14255')), (323, Decimal('0.68265')), (419, Decimal('0.03898')), (436, Decimal('0.14032')), (439, Decimal('0.08978'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **leftaccumbensarea**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	0.391
Standard deviation	0.068
Minimum value	0.0
Maximum value	0.662
25% of records are below this value(limit value of the first quartile)	0.354
50% of records are below this value (median)	0.394
75% of records are below this value(limit value of the third quartile)	0.43
Outlier upper bound	0.594
Outlier lower bound	0.187
Total number of outliers (outside 3 std.dev)	7
Rows with outliers	[(69, Decimal('0.0003')), (104, Decimal('0.00296')), (186, Decimal('0.02263')), (323, Decimal('0.66196')), (419, Decimal('0.13301')), (436, Decimal('0.16158')), (439, Decimal('0.095'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **rightamygdala**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	0.797
Standard deviation	0.118
Minimum value	0.182
Maximum value	1.253
25% of records are below this value(limit value of the first quartile)	0.72
50% of records are below this value (median)	0.802
75% of records are below this value(limit value of the third quartile)	0.871
Outlier upper bound	1.151
Outlier lower bound	0.443
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(186, Decimal('0.18236')), (323, Decimal('1.2526'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **leftamygdala**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	0.793
Standard deviation	0.122
Minimum value	0.186
Maximum value	1.314
25% of records are below this value(limit value of the first quartile)	0.71
50% of records are below this value (median)	0.798
75% of records are below this value(limit value of the third quartile)	0.878
Outlier upper bound	1.158
Outlier lower bound	0.429
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(186, Decimal('0.18575999999999998')), (323, Decimal('1.3139'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **brainstem**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	17.426
Standard deviation	2.21
Minimum value	2.083
Maximum value	27.231
25% of records are below this value(limit value of the first quartile)	16.058
50% of records are below this value (median)	17.444
75% of records are below this value(limit value of the third quartile)	18.746
Outlier upper bound	24.057
Outlier lower bound	10.796
Total number of outliers (outside 3 std.dev)	4
Rows with outliers	[(33, Decimal('10.4775')), (186, Decimal('2.0826')), (330, Decimal ('27.2308')), (419, Decimal('9.0519'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **rightstgsuperiortemporalgyrus**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	7.079
Standard deviation	0.871
Minimum value	3.441
Maximum value	9.974
25% of records are below this value(limit value of the first quartile)	6.534
50% of records are below this value (median)	7.076
75% of records are below this value(limit value of the third quartile)	7.592
Outlier upper bound	9.691
Outlier lower bound	4.467
Total number of outliers (outside 3 std.dev)	3
Rows with outliers	[(104, Decimal('3.5')), (186, Decimal('3.4409')), (332, Decimal('9.9739'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **leftstgsuperiortemporalgyrus**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	6.97
Standard deviation	0.839
Minimum value	3.472
Maximum value	9.339
25% of records are below this value(limit value of the first quartile)	6.435
50% of records are below this value (median)	7.013
75% of records are below this value(limit value of the third quartile)	7.467
Outlier upper bound	9.486
Outlier lower bound	4.454
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(104, Decimal('3.8933')), (186, Decimal('3.4718'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **righttmptemporalpole**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	7.438
Standard deviation	0.92
Minimum value	2.871
Maximum value	10.768
25% of records are below this value (limit value of the first quartile)	6.831
50% of records are below this value (median)	7.336
75% of records are below this value (limit value of the third quartile)	8.023
Outlier upper bound	10.197
Outlier lower bound	4.679
Total number of outliers (outside 3 std.dev)	4
Rows with outliers	[(186, Decimal('2.8712')), (251, Decimal('4.2566')), (273, Decimal('4.6739999999999995')), (323, Decimal('10.7681'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **lefttmptemporalpole**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	7.235
Standard deviation	0.928
Minimum value	2.561
Maximum value	10.493
25% of records are below this value(limit value of the first quartile)	6.604
50% of records are below this value (median)	7.166
75% of records are below this value(limit value of the third quartile)	7.809
Outlier upper bound	10.019
Outlier lower bound	4.45
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(186, Decimal('2.5612')), (323, Decimal('10.4928'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column
righttrifgtriangularpartoftheinferiorfrontalgyrus

Type	numerical
Total number of rows	474
Number of rows with data	433
Completion percentage	91.35%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	3.45
Standard deviation	0.428
Minimum value	1.148
Maximum value	4.98
25% of records are below this value(limit value of the first quartile)	3.172
50% of records are below this value (median)	3.431
75% of records are below this value(limit value of the third quartile)	3.699
Outlier upper bound	4.732
Outlier lower bound	2.167
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(186, Decimal('1.1485')), (330, Decimal ('4.9804'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **lefttrifgtriangularpartoftheinferiorfrontalgyrus**

Type	numerical
Total number of rows	474
Number of rows with data	434
Completion percentage	91.56%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	3.681
Standard deviation	0.478
Minimum value	0.0
Maximum value	4.995
25% of records are below this value(limit value of the first quartile)	3.389
50% of records are below this value (median)	3.684
75% of records are below this value(limit value of the third quartile)	3.985
Outlier upper bound	5.115
Outlier lower bound	2.247
Total number of outliers (outside 3 std.dev)	2
Rows with outliers	[(186, Decimal('1.4140000000000001')), (419, Decimal ('0.00001'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **rightttgtransversetemporalgyrus**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	1.451
Standard deviation	0.2
Minimum value	0.116
Maximum value	2.288
25% of records are below this value(limit value of the first quartile)	1.325
50% of records are below this value (median)	1.428
75% of records are below this value(limit value of the third quartile)	1.566
Outlier upper bound	2.05
Outlier lower bound	0.852
Total number of outliers (outside 3 std.dev)	3
Rows with outliers	[(104, Decimal('0.1162')), (234, Decimal('2.2885')), (419, Decimal('0.78203'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **leftttgtransversetemporalgyrus**

Type	numerical
Total number of rows	474
Number of rows with data	437
Completion percentage	92.19%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

mean	1.551
Standard deviation	0.235
Minimum value	0.311
Maximum value	2.522
25% of records are below this value(limit value of the first quartile)	1.387
50% of records are below this value (median)	1.537
75% of records are below this value(limit value of the third quartile)	1.694
Outlier upper bound	2.255
Outlier lower bound	0.847
Total number of outliers (outside 3 std.dev)	4
Rows with outliers	[(64, Decimal('2.3404')), (104, Decimal('0.31108')), (234, Decimal('2.5225')), (238, Decimal('2.259'))]

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **minimentalstate**

Type	integer
Total number of rows	474
Number of rows with data	359
Completion percentage	75.74%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Mode (most frequent value)	29
Number of occurrences for most frequent value	61
Minimum value	4
Maximum value	30
25% of records are below this value(limit value of the first quartile)	23
50% of records are below this value (median)	27
75% of records are below this value(limit value of the third quartile)	29

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **agegroup**

Type	nominal
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	70-79y
Number of occurrences for most frequent value	170
List of category values	['+80y', '-50y', '50-59y', '60-69y', '70-79y']
Number of categories	5

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **edsdcategory**

Type	nominal
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	CN
Number of occurrences for most frequent value	151
List of category values	['AD', 'CN', 'MCI']
Number of categories	3

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **alzheimerbroadcategory**

Type	nominal
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	CN
Number of occurrences for most frequent value	151
List of category values	['AD', 'CN', 'MCI']
Number of categories	3

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **parkinsonbroadcategory**

Type	nominal
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	Other
Number of occurrences for most frequent value	217
List of category values	['CN', 'Other']
Number of categories	2

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **neurodegenerativescategories**

Type	nominal
Total number of rows	474
Number of rows with data	368
Completion percentage	77.64%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	CN
Number of occurrences for most frequent value	151
List of category values	['AD', 'CN', 'MCI']
Number of categories	3

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **dataset**

Type	nominal
Total number of rows	474
Number of rows with data	474
Completion percentage	100.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	edsd
Number of occurrences for most frequent value	474
List of category values	['edsd']
Number of categories	1

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **empty**

Type	text
Total number of rows	474
Number of rows with data	0
Completion percentage	0.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null