Quality Control Report

This QC Report is created on **31/03/2021 12:56:40**

Version of Data Quality Control Tool: **3.0**

File path: /home/kostis/HBP/DataQualityControlTool/tests/test_datasets/test_dataset.csv

Total number of rows: **20**, Total number of columns: **4**

Is metadata JSON file with table schema provided? **Yes**

Data Cleaning performed? **No**

Number of rows with invalid values: **9**

Number of columns containing invalid values:

Rows Data Completeness Overall Statistics

Number of rows with 100% (all 4) of the columns filled	14 (70.0%)
Number of rows with 75-99% (3 - 3) of the columns filled	4 (20.0%)
Number of rows with 50-74% (2 - 2) of the columns filled	1 (5.0%)
Number of rows with 25-49% (1 - 1) of the columns filled	1 (5.0%)
Number of rows with 0-24% (0 - 0) of the colums filled	0 (0.0%)

Rows Data Validation Overall Statistics

Number of rows with 100% (all 4) of the columns filled with valid data (nulls included)	11 (55.0%)
Number of rows with 75-99% (3 - 3) of the columns filled with valid data (nulls included)	6 (30.0%)
Number of rows with 50-74% (2 - 2) of the columns filled with valid data (nulls included)	2 (10.0%)
Number of rows with 25-49% (1 - 1) of the columns filled with valid data (nulls included)	1 (5.0%)
Number of rows with 0-24% (0 - 0) of the colums filled with valid data (nulls included)	0 (0.0%)

Report for column **Patient_id**

Type	integer
Total number of rows	20
Number of rows with data	19
Completion percentage	95.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Mode (most frequent value)	1
Number of occurrences for most frequent value	1
Minimum value	1
Maximum value	20
25% of records are below this value(limit value of the first quartile)	5
50% of records are below this value (median)	10
75% of records are below this value(limit value of the third quartile)	14

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

Report for column **Diagnosis Categories**

Type	nominal
Total number of rows	20
Number of rows with data	9
Completion percentage	45.0%
Number of rows with constraint violations	7
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	Category1
Number of occurrences for most frequent value	5
List of category values	['Another3', 'Category1', 'Category2']
Number of categories	3

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Catgry2	Category2
CATEGORR2	Category2
Anoter1	Another3
CATEGOR1	Category1
CategGory1	Category1

Values that will be replaced with null

not value, vecw,

Report for column **Gender_num**

Type	nominal
Total number of rows	20
Number of rows with data	16
Completion percentage	80.0%
Number of rows with constraint violations	2
Number of rows with datatype violations	0
Data Cleansing applied?	No

Statistics Table

Most frequent value	2
Number of occurrences for most frequent value	9
List of category values	[1, 2]
Number of categories	2

Suggested corrections for datatype violations

Suggested corrections for constraint violations

Values that will be replaced with null

3,

Report for column **Date_visit**

Type	date
Total number of rows	20
Number of rows with data	14
Completion percentage	70.0%
Number of rows with constraint violations	0
Number of rows with datatype violations	4
Data Cleansing applied?	No

Statistics Table

Mode (most frequent value)	1995-03-01
Number of occurrences for most frequent value	1
Minimum value	1985-12-12
Maximum value	2020-05-31

Suggested corrections for datatype violations

12-06-2009	12/06/2009
16062020	16/06/2020

Suggested corrections for constraint violations

Values that will be replaced with null

32/04/2015, not a date,