

# Data Quality Report: Comparison of exampleCSV\_source and exampleCSV\_target

No time restriction. All available data were analysed

(c) Universitätsklinikum Erlangen

15. Dezember 2021

## Inhaltsverzeichnis

|          |                                      |          |
|----------|--------------------------------------|----------|
| <b>1</b> | <b>Data Map</b>                      | <b>3</b> |
| 1.1      | Target Data System . . . . .         | 3        |
| 1.2      | Source Data System . . . . .         | 3        |
| <b>2</b> | <b>Completeness Checks</b>           | <b>4</b> |
| 2.1      | Validation . . . . .                 | 4        |
| 2.2      | Verification . . . . .               | 4        |
| <b>3</b> | <b>Conformance Checks</b>            | <b>5</b> |
| 3.1      | Value Conformance . . . . .          | 5        |
| <b>4</b> | <b>Detailed Descriptive Analysis</b> | <b>6</b> |
| 4.1      | Age in years . . . . .               | 6        |
| 4.2      | Amount of credit . . . . .           | 8        |
| 4.3      | Birthdate . . . . .                  | 10       |
| 4.4      | Credit worthy? . . . . .             | 12       |
| 4.5      | Current bank balance . . . . .       | 14       |
| 4.6      | Date of contact . . . . .            | 16       |
| 4.7      | Forename . . . . .                   | 18       |
| 4.8      | Income . . . . .                     | 20       |
| 4.9      | Job . . . . .                        | 22       |
| 4.10     | Name . . . . .                       | 24       |
| 4.11     | Person ID . . . . .                  | 26       |
| 4.12     | Sex . . . . .                        | 28       |

|          |  |           |
|----------|--|-----------|
| <b>5</b> | <b>Plausibility Checks</b>             | <b>30</b> |
| 5.1      | Atemporal Plausibility . . . . .       | 30        |
| 5.2      | Uniqueness Plausibility . . . . .      | 31        |
| <b>6</b> | <b>Appendix</b>                        | <b>32</b> |
| 6.1      | R-Package Version ‘DQAstats’ . . . . . | 32        |
| 6.2      | R-Package Version ‘DIZutils’ . . . . . | 32        |

# **1 Data Map**

## **1.1 Target Data System**

There is no data to display.

## **1.2 Source Data System**

There is no data to display.

## 2 Completeness Checks

### 2.1 Validation

Completeness checks (validation) evaluate the ETL (extract transform load) jobs. They compare for each variable the exact matching of the number of distinct values, the number of valid values (=n), and the number of missing values between the source data system and the target data system.

| Variable             | Check Distincts | Check Valids | Check Missings |
|----------------------|-----------------|--------------|----------------|
| Age in years         | passed          | passed       | passed         |
| Amount of credit     | passed          | passed       | passed         |
| Birthdate            | passed          | passed       | passed         |
| Credit worthy?       | passed          | passed       | passed         |
| Current bank balance | passed          | passed       | passed         |
| Date of contact      | passed          | passed       | passed         |
| Forename             | passed          | passed       | passed         |
| Income               | passed          | passed       | passed         |
| Job                  | passed          | passed       | passed         |
| Name                 | passed          | passed       | passed         |
| Person ID            | passed          | passed       | passed         |
| Sex                  | failed          | passed       | passed         |

### 2.2 Verification

| Variable             | Missings [%] (source) | Missings [%] (target) |
|----------------------|-----------------------|-----------------------|
| Age in years         | 0                     | 0                     |
| Amount of credit     | 56.52                 | 56.52                 |
| Birthdate            | 0                     | 0                     |
| Credit worthy?       | 0                     | 0                     |
| Current bank balance | 0                     | 0                     |
| Date of contact      | 0                     | 0                     |
| Forename             | 0                     | 0                     |
| Income               | 0                     | 0                     |
| Job                  | 0                     | 0                     |
| Name                 | 0                     | 0                     |
| Person ID            | 0                     | 0                     |
| Sex                  | 0                     | 0                     |

## 3 Conformance Checks

### 3.1 Value Conformance

Value conformance checks (verification) compare for each variable the values of each data system to predefined constraints. Those constraints can be defined for each variable and data system individually in the metadata repository (MDR).

| Variable             | Check Source Data | Check Target Data |
|----------------------|-------------------|-------------------|
| Age in years         | passed            | failed            |
| Amount of credit     | passed            | passed            |
| Birthdate            | passed            | passed            |
| Credit worthy?       | passed            | passed            |
| Current bank balance | passed            | passed            |
| Date of contact      | passed            | passed            |
| Income               | passed            | failed            |
| Sex                  | passed            | failed            |
| pl.atemporal.item01  | passed            | failed            |

## 4 Detailed Descriptive Analysis

### 4.1 Age in years

The age of the person at the time of contact.

#### 4.1.1 Representation in source data system

- Variable: AGE
- Table: dqa\_example\_data\_01.csv

##### Overview:

- Variable name: dqa\_age\_years
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 21

##### Results:

|          |       |
|----------|-------|
| Mean     | 63    |
| Minimum  | 22    |
| Median   | 64    |
| Maximum  | 94    |
| SD       | 22.56 |
| Negativ  | 0     |
| Zero     | 0     |
| Positive | 23    |
| OutLo    | 0     |
| OutHi    | 0     |
| Skewness | -0.46 |
| Kurtosis | -1.15 |
| Variance | 509   |
| Range    | 72    |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min | max | unit |
|-----|-----|------|
| 0   | 110 | a    |

#### 4.1.2 Representation in target data system

- Variable: AGE
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_age\_years
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 21

##### Results:

|          |         |
|----------|---------|
| Mean     | 67.35   |
| Minimum  | 22      |
| Median   | 64      |
| Maximum  | 175     |
| SD       | 32.45   |
| Negativ  | 0       |
| Zero     | 0       |
| Positive | 23      |
| OutLo    | 0       |
| OutHi    | 1       |
| Skewness | 1.25    |
| Kurtosis | 2.83    |
| Variance | 1052.87 |
| Range    | 153     |

##### Value conformance:

- Conformance check: failed
- Constraining values/rules:

| min | max | unit |
|-----|-----|------|
| 0   | 110 | a    |

- Extrem values are not conform with constraints.

## 4.2 Amount of credit

That's the amount of credit the person has used

### 4.2.1 Representation in source data system

- Variable: CREDIT-AMOUNT
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_credit\_amount
- Variable type: integer
  - n: 23
  - Valid values: 10
  - Missing values: 13
  - Distinct values: 10

#### Results:

|          |              |
|----------|--------------|
| Mean     | 39220        |
| Minimum  | 12200        |
| Median   | 33350        |
| Maximum  | 72800        |
| SD       | 21447.19     |
| Negativ  | 0            |
| Zero     | 0            |
| Positive | 10           |
| OutLo    | 0            |
| OutHi    | 0            |
| Skewness | 0.39         |
| Kurtosis | -1.56        |
| Variance | 459981777.78 |
| Range    | 60600        |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min | max | unit  |
|-----|-----|-------|
| 0   | Inf | money |



#### 4.2.2 Representation in target data system

- Variable: CREDIT-AMOUNT
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_credit\_amount
- Variable type: integer
  - n: 23
  - Valid values: 10
  - Missing values: 13
  - Distinct values: 10

##### Results:

|          |              |
|----------|--------------|
| Mean     | 39220        |
| Minimum  | 12200        |
| Median   | 33350        |
| Maximum  | 72800        |
| SD       | 21447.19     |
| Negative | 0            |
| Zero     | 0            |
| Positive | 10           |
| OutLo    | 0            |
| OutHi    | 0            |
| Skewness | 0.39         |
| Kurtosis | -1.56        |
| Variance | 459981777.78 |
| Range    | 60600        |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min | max | unit  |
|-----|-----|-------|
| 0   | Inf | money |

## 4.3 Birthdate

The date of birth written as dd.mm.yyyy

### 4.3.1 Representation in source data system

- Variable: BIRTHDATE
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_birthdate
- Variable type: datetime
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

#### Results:

|         |            |
|---------|------------|
| Min.    | 1921-02-19 |
| 1st Qu. | 1932-09-17 |
| Median  | 1951-07-03 |
| Mean    | 1950-09-25 |
| 3rd Qu. | 1965-05-10 |
| Max.    | 1990-05-26 |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules: ''

### 4.3.2 Representation in target data system

- Variable: BIRTHDATE
- Table: dqa\_example\_data\_02.csv

#### Overview:

- Variable name: dqa\_birthdate
- Variable type: datetime
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

#### Results:

|         |            |
|---------|------------|
| Min.    | 1921-02-19 |
| 1st Qu. | 1932-09-17 |
| Median  | 1951-07-03 |
| Mean    | 1950-09-25 |
| 3rd Qu. | 1965-05-10 |
| Max.    | 1990-05-26 |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules: ''

## 4.4 Credit worthy?

Indicates whether the person is creditworthy at the time of the contact

### 4.4.1 Representation in source data system

- Variable: CREDIT-WORTHY
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_credit\_worthy
- Variable type: enumerated
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 2

#### Results:

| dqa_credit_worthy | Freq | % Valid |
|-------------------|------|---------|
| no                | 13   | 56.522  |
| yes               | 10   | 43.478  |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules: ' yes, no '

#### 4.4.2 Representation in target data system

- Variable: CREDIT-WORTHY
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_credit\_worthy
- Variable type: enumerated
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 2

##### Results:

| dqa_credit_worthy | Freq | % Valid |
|-------------------|------|---------|
| no                | 13   | 56.522  |
| yes               | 10   | 43.478  |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules: ' yes, no '

## 4.5 Current bank balance

The bank-balance at the time of contact

### 4.5.1 Representation in source data system

- Variable: BANK-BALANCE
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_bank\_balance
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 22

#### Results:

|          |               |
|----------|---------------|
| Mean     | 35152.17      |
| Minimum  | -34200        |
| Median   | 18800         |
| Maximum  | 124100        |
| SD       | 39516.63      |
| Negativ  | 2             |
| Zero     | 0             |
| Positive | 21            |
| OutLo    | 0             |
| OutHi    | 0             |
| Skewness | 0.62          |
| Kurtosis | -0.4          |
| Variance | 1561564426.88 |
| Range    | 158300        |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min  | max | unit  |
|------|-----|-------|
| -Inf | Inf | money |

#### 4.5.2 Representation in target data system

- Variable: BANK-BALANCE
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_bank\_balance
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 22

##### Results:

|          |               |
|----------|---------------|
| Mean     | 26395.65      |
| Minimum  | -64200        |
| Median   | 12800         |
| Maximum  | 124100        |
| SD       | 46097.8       |
| Negative | 4             |
| Zero     | 0             |
| Positive | 19            |
| OutLo    | 0             |
| OutHi    | 0             |
| Skewness | 0.33          |
| Kurtosis | -0.37         |
| Variance | 2125006798.42 |
| Range    | 188300        |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min  | max | unit  |
|------|-----|-------|
| -Inf | Inf | money |

## 4.6 Date of contact

Date of contact

### 4.6.1 Representation in source data system

- Variable: CONTACT-DATE
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_contact\_date
- Variable type: datetime
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 23

#### Results:

|         |            |
|---------|------------|
| Min.    | 2011-10-12 |
| 1st Qu. | 2012-08-11 |
| Median  | 2013-10-02 |
| Mean    | 2013-10-28 |
| 3rd Qu. | 2014-12-21 |
| Max.    | 2015-12-20 |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules: ''



#### 4.6.2 Representation in target data system

- Variable: CONTACT-DATE
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_contact\_date
- Variable type: datetime
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 23

##### Results:

|         |            |
|---------|------------|
| Min.    | 2011-10-12 |
| 1st Qu. | 2012-08-11 |
| Median  | 2013-10-02 |
| Mean    | 2013-10-28 |
| 3rd Qu. | 2014-12-21 |
| Max.    | 2015-12-20 |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules: ''

## 4.7 Forename

The Forename of the person.

### 4.7.1 Representation in source data system

- Variable: FORENAME
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_forename
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

#### Results:

| dqa_forename | Freq | % Valid |
|--------------|------|---------|
| Geraldine    | 3    | 13.043  |
| Zenaida      | 3    | 13.043  |
| Williams     | 2    | 8.696   |
| Wayne        | 2    | 8.696   |
| Dorothy      | 2    | 8.696   |
| Lawrence     | 1    | 4.348   |
| Janet        | 1    | 4.348   |
| Martin       | 1    | 4.348   |
| Georgina     | 1    | 4.348   |
| Elliott      | 1    | 4.348   |
| Gilberto     | 1    | 4.348   |
| Annie        | 1    | 4.348   |
| Karen        | 1    | 4.348   |
| John         | 1    | 4.348   |
| Susan        | 1    | 4.348   |
| Elijah       | 1    | 4.348   |

#### 4.7.2 Representation in target data system

- Variable: FORENAME
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_forename
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

##### Results:

| dqa_forename | Freq | % Valid |
|--------------|------|---------|
| Geraldine    | 3    | 13.043  |
| Zenaida      | 3    | 13.043  |
| Williams     | 2    | 8.696   |
| Wayne        | 2    | 8.696   |
| Dorothy      | 2    | 8.696   |
| Lawrence     | 1    | 4.348   |
| Janet        | 1    | 4.348   |
| Martin       | 1    | 4.348   |
| Georgina     | 1    | 4.348   |
| Elliott      | 1    | 4.348   |
| Gilberto     | 1    | 4.348   |
| Annie        | 1    | 4.348   |
| Karen        | 1    | 4.348   |
| John         | 1    | 4.348   |
| Susan        | 1    | 4.348   |
| Elijah       | 1    | 4.348   |

## 4.8 Income

The income of the person at the time of contact

### 4.8.1 Representation in source data system

- Variable: INCOME
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_income
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 23

#### Results:

|          |               |
|----------|---------------|
| Mean     | 68826.09      |
| Minimum  | 3000          |
| Median   | 59000         |
| Maximum  | 145000        |
| SD       | 46841.76      |
| Negativ  | 0             |
| Zero     | 0             |
| Positive | 23            |
| OutLo    | 0             |
| OutHi    | 0             |
| Skewness | 0.23          |
| Kurtosis | -1.48         |
| Variance | 2194150197.63 |
| Range    | 142000        |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules:

| min | max | unit  |
|-----|-----|-------|
| 0   | Inf | money |

#### 4.8.2 Representation in target data system

- Variable: INCOME
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_income
- Variable type: integer
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 23

##### Results:

|          |               |
|----------|---------------|
| Mean     | 68391.3       |
| Minimum  | -5000         |
| Median   | 59000         |
| Maximum  | 145000        |
| SD       | 47502.86      |
| Negativ  | 1             |
| Zero     | 0             |
| Positive | 22            |
| OutLo    | 0             |
| OutHi    | 0             |
| Skewness | 0.19          |
| Kurtosis | -1.44         |
| Variance | 2256521739.13 |
| Range    | 150000        |

##### Value conformance:

- Conformance check: failed
- Constraining values/rules:

| min | max | unit  |
|-----|-----|-------|
| 0   | Inf | money |

- Extrem values are not conform with constraints.

## 4.9 Job

The job of the person at the time of contact

### 4.9.1 Representation in source data system

- Variable: JOB
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_job
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 15

#### Results:

| dqa_job      | Freq | % Valid |
|--------------|------|---------|
| Bank manager | 3    | 13.043  |
| Magician     | 3    | 13.043  |
| Student      | 2    | 8.696   |
| Pilot        | 2    | 8.696   |
| Lawyer       | 2    | 8.696   |
| Singer       | 2    | 8.696   |
| Photographer | 1    | 4.348   |
| Farmer       | 1    | 4.348   |
| Professor    | 1    | 4.348   |
| Engineer     | 1    | 4.348   |
| Researcher   | 1    | 4.348   |
| Chemist      | 1    | 4.348   |
| Gardener     | 1    | 4.348   |
| Psychologist | 1    | 4.348   |
| Comedian     | 1    | 4.348   |

#### 4.9.2 Representation in target data system

- Variable: JOB
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_job
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 15

##### Results:

| dqa_job      | Freq | % Valid |
|--------------|------|---------|
| Bank manager | 3    | 13.043  |
| Magician     | 3    | 13.043  |
| Student      | 2    | 8.696   |
| Pilot        | 2    | 8.696   |
| Lawyer       | 2    | 8.696   |
| Singer       | 2    | 8.696   |
| Photographer | 1    | 4.348   |
| Farmer       | 1    | 4.348   |
| Professor    | 1    | 4.348   |
| Engineer     | 1    | 4.348   |
| Researcher   | 1    | 4.348   |
| Chemist      | 1    | 4.348   |
| Gardener     | 1    | 4.348   |
| Psychologist | 1    | 4.348   |
| Comedian     | 1    | 4.348   |

## 4.10 Name

The Surname of the person.

### 4.10.1 Representation in source data system

- Variable: NAME
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_name
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

#### Results:

| dqa_name  | Freq | % Valid |
|-----------|------|---------|
| Jackson   | 3    | 13.043  |
| Staggs    | 3    | 13.043  |
| Rodriguez | 2    | 8.696   |
| Burdett   | 2    | 8.696   |
| Simpson   | 2    | 8.696   |
| Daniels   | 1    | 4.348   |
| Dardar    | 1    | 4.348   |
| Jones     | 1    | 4.348   |
| Cook      | 1    | 4.348   |
| Eatmon    | 1    | 4.348   |
| Kenney    | 1    | 4.348   |
| Stock     | 1    | 4.348   |
| Shuck     | 1    | 4.348   |
| Malloy    | 1    | 4.348   |
| Kirkland  | 1    | 4.348   |
| Sutton    | 1    | 4.348   |



#### 4.10.2 Representation in target data system

- Variable: NAME
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_name
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

##### Results:

| dqa_name  | Freq | % Valid |
|-----------|------|---------|
| Jackson   | 3    | 13.043  |
| Staggs    | 3    | 13.043  |
| Rodriguez | 2    | 8.696   |
| Burdett   | 2    | 8.696   |
| Simpson   | 2    | 8.696   |
| Daniels   | 1    | 4.348   |
| Dardar    | 1    | 4.348   |
| Jones     | 1    | 4.348   |
| Cook      | 1    | 4.348   |
| Eatmon    | 1    | 4.348   |
| Kenney    | 1    | 4.348   |
| Stock     | 1    | 4.348   |
| Shuck     | 1    | 4.348   |
| Malloy    | 1    | 4.348   |
| Kirkland  | 1    | 4.348   |
| Sutton    | 1    | 4.348   |

## 4.11 Person ID

Each person has its own person-id. It stays the same over the whole live of the person and does not change.

### 4.11.1 Representation in source data system

- Variable: PERSON\_ID
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_person\_id
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

#### Results:

| dqa_person_id | Freq | % Valid |
|---------------|------|---------|
| 1             | 3    | 13.043  |
| 7             | 3    | 13.043  |
| 5             | 2    | 8.696   |
| 11            | 2    | 8.696   |
| 15            | 2    | 8.696   |
| 2             | 1    | 4.348   |
| 3             | 1    | 4.348   |
| 4             | 1    | 4.348   |
| 6             | 1    | 4.348   |
| 8             | 1    | 4.348   |
| 9             | 1    | 4.348   |
| 10            | 1    | 4.348   |
| 12            | 1    | 4.348   |
| 13            | 1    | 4.348   |
| 14            | 1    | 4.348   |
| 16            | 1    | 4.348   |

#### 4.11.2 Representation in target data system

- Variable: PERSON\_ID
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_person\_id
- Variable type: string
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 16

##### Results:

| dqa_person_id | Freq | % Valid |
|---------------|------|---------|
| 1             | 3    | 13.043  |
| 7             | 3    | 13.043  |
| 5             | 2    | 8.696   |
| 11            | 2    | 8.696   |
| 15            | 2    | 8.696   |
| 2             | 1    | 4.348   |
| 3             | 1    | 4.348   |
| 4             | 1    | 4.348   |
| 6             | 1    | 4.348   |
| 8             | 1    | 4.348   |
| 9             | 1    | 4.348   |
| 10            | 1    | 4.348   |
| 12            | 1    | 4.348   |
| 13            | 1    | 4.348   |
| 14            | 1    | 4.348   |
| 16            | 1    | 4.348   |

## 4.12 Sex

The sex of the person in one letter: m, f or x for unknown.

### 4.12.1 Representation in source data system

- Variable: SEX
- Table: dqa\_example\_data\_01.csv

#### Overview:

- Variable name: dqa\_sex
- Variable type: enumerated
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 2

#### Results:

| dqa_sex | Freq | % Valid |
|---------|------|---------|
| f       | 13   | 56.522  |
| m       | 10   | 43.478  |

#### Value conformance:

- Conformance check: passed
- Constraining values/rules: ' m, f, x '

#### 4.12.2 Representation in target data system

- Variable: SEX
- Table: dqa\_example\_data\_02.csv

##### Overview:

- Variable name: dqa\_sex
- Variable type: enumerated
  - n: 23
  - Valid values: 23
  - Missing values: 0
  - Distinct values: 3

##### Results:

| dqa_sex | Freq | % Valid |
|---------|------|---------|
| f       | 12   | 52.174  |
| m       | 10   | 43.478  |
| abc     | 1    | 4.348   |

##### Value conformance:

- Conformance check: failed
- Constraining values/rules: ' m, f, x '
- Levels that are not conform with the value set:  
abc

## 5 Plausibility Checks

### 5.1 Atemporal Plausibility

#### 5.1.1 Pl.atemporal.Item01

Persons with a negative bank balance cannot be credit worthy

##### 5.1.1.1 Representation in source data system

- Variable 1: dqa\_credit\_worthy
- Variable 2: dqa\_bank\_balance
- Filter criterion variable 2 (regex): ^(-)
- Join criterion: dqa\_person\_id

##### Overview:

No data available for reporting

##### Results:

| dqa_credit_worthy | Freq | % Valid |
|-------------------|------|---------|
| no                | 2    | 100     |

##### Value conformance:

- Conformance check: passed
- Constraining values/rules: ' no '

##### 5.1.1.2 Representation in target data system

- Variable 1: dqa\_credit\_worthy
- Variable 2: dqa\_bank\_balance
- Filter criterion variable 2 (regex): ^(-)
- Join criterion: dqa\_person\_id

##### Overview:

No data available for reporting

##### Results:

| dqa_credit_worthy | Freq | % Valid |
|-------------------|------|---------|
| no                | 2    | 50      |
| yes               | 2    | 50      |

##### Value conformance:

- Conformance check: failed
- Constraining values/rules: ' no '
- Levels that are not conform with the value set:  
yes

## **5.2 Uniqueness Plausibility**

### **5.2.1 dqa\_\_name**

The last name of a person must be identical in all entries for one person ID.

#### **5.2.1.1 Representation in source data system**

- Plausibility check: passed
- Message: No duplicate occurrences of dqa\_\_person\_\_id found in association with dqa\_\_name.

#### **5.2.1.2 Representation in target data system**

- Plausibility check: passed
- Message: No duplicate occurrences of dqa\_\_person\_\_id found in association with dqa\_\_name.

## 6 Appendix

### 6.1 R-Package Version ‘DQAstats’

```
## [1] '0.1.7.9016'
```

### 6.2 R-Package Version ‘DIZutils’

```
## [1] '0.0.9.9001'
```

```
##
```

```
## ## SQL Statments
```

```
##
```

```
## ### Source Data System
```

```
##
```

```
## ### Target Data System
```