datacamp

Data Quality Dimensions
## Introduction to Data Quality

Learn more at www.DataCamp.com

What are Data Quality Dimensions?

Data Quality is a measurement of the degree to which data is fit for purpose. Good data quality generates trust in data. Data Quality Dimensions are a measurement of a specific attribute of a data's quality.

> Completeness

Completeness measures the degree to which all expected records in a dataset are present. At a data element level, completeness is the degree to which all records have data populated when expected.
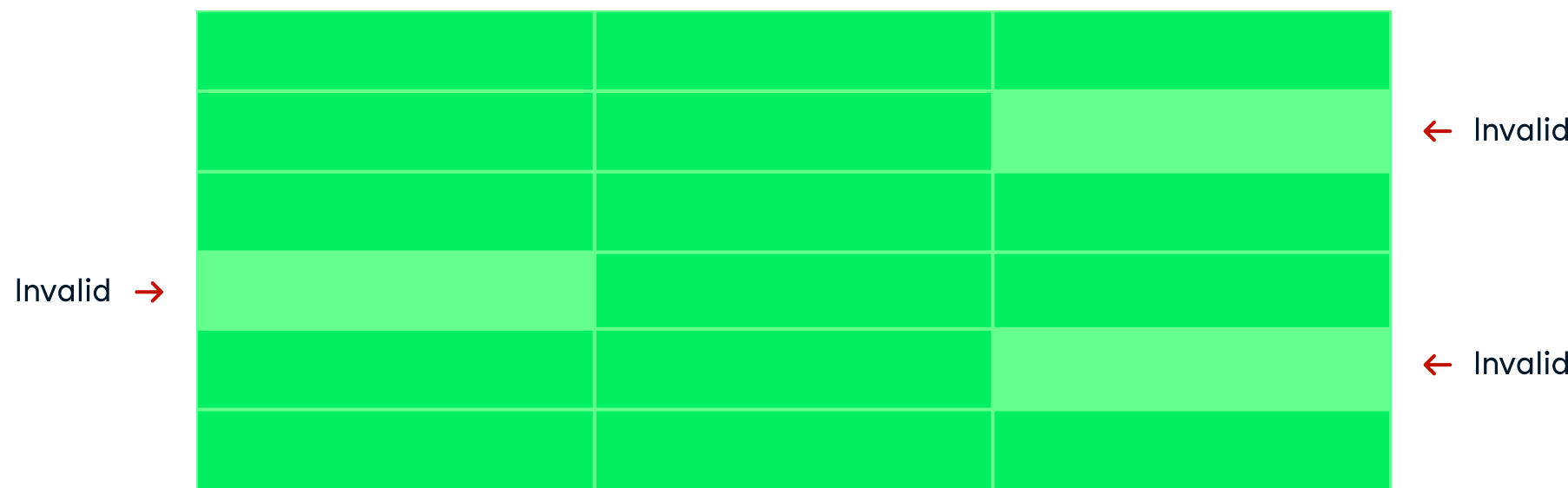
Missing ↓

↑ Missing
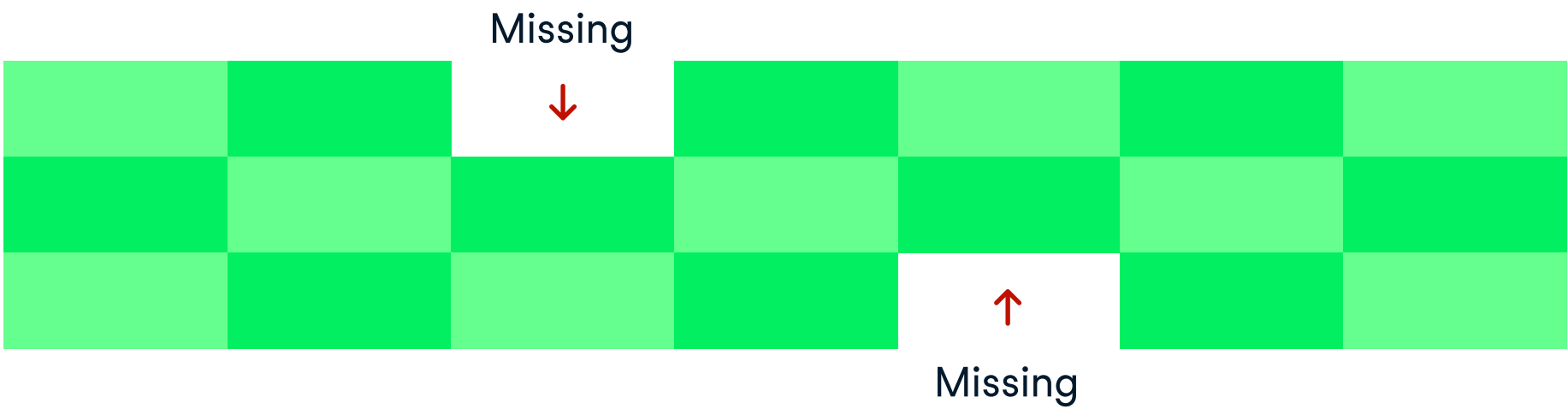
Completeness Example

All records must have a value populated in the CustomerName field.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

> Validity

Validity measures the degree to which the values in a data element are valid.

Invalid →

← Invalid

← Invalid

Validity Example

- CustomerBirthDate value must be a date in the past.
- CustomerAccountType value must be either Loan or Deposit.
- LatestAccountOpenDate value must be a date in the past.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

> Uniqueness

Uniqueness measures the degree to which the records in a dataset are not duplicated.

← Not Unique

← Not Unique

Uniqueness Example

All records must have a unique CustomerID and CustomerName.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

> Timeliness

Timeliness is the degree to which a dataset is available when expected and depends on service level agreements being set up between technical and business resources.

SLA	Table Load Time
08:00 am	07:59 am
10:00 am	09:59 am
11:00 am	11:01 am

← Missed the SLA

Timeliness Example

All records in the customer dataset must be loaded by the 9:00 am.

CustomerID	CustomerName	
100000192	01-01-2023	11:07 am
100000198	01-01-2023	11:07 am
100000120	01-01-2023	11:07 am

> Consistency

Consistency is a data quality dimension that measures the degree to which data is the same across all instances of the data. Consistency can be measured by setting a threshold for how much difference there can be between two datasets.
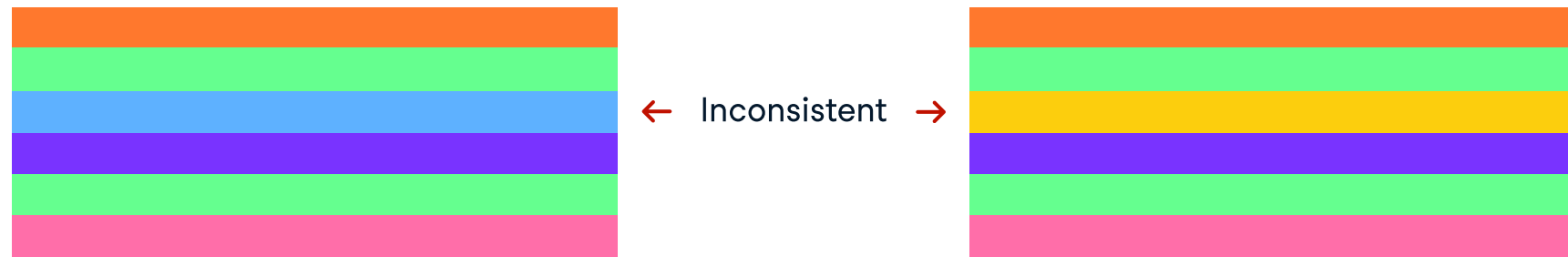
← Inconsistent →

Consistency Example

The count of records loaded today must be within +/- 5% of the count of records loaded yesterday.

| Count of records in TargetCustomerTable	Record count difference from previous day
10,000,000	4,909,797 ✕
5,090,203	75 ✓
5,090,128	1 ✓

The count of records loaded today must be within +/- 5% of the count of records loaded yesterday.

AccountTableCustomerID	CustomerTableCustomerID
108394858	108394858
192039482	192039482
203475849	NULL ✕
2930485953	NULL ✕
102832748	102832748

> Accuracy

Accuracy measures the degree to which data is correct and represents the truth.

Verified Source Document

Downstream Table
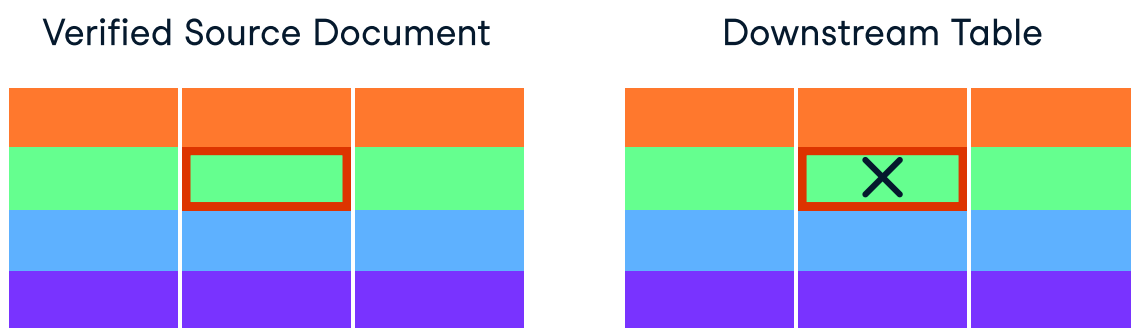
Accuracy Example

All records in the Customer Table must have accurate Customer Name, Customer Birthdate, and Customer Address fields when compared to the Tax Form.

Tax Form
Name: Ava Shiffer   Birthdate: 10/30/1990
Address: 910 Quality St
City: Washington   State: DC
Zip: 20008

CustomerName	CustomerBirthDate	CustomerAddress	CustomerCity	CustomerState	CustomerZip
Ava Shiffer	10/31/1990	910 Quality St	Washington	WA	20008