

Types of data validation

This reading describes the purpose, examples, and limitations of six types of data validation. The first five are validation types associated with the data (type, range, constraint, consistency, and structure) and the sixth type focuses on the validation of application code used to accept data from user input.

As a junior data analyst, you might not perform all of these validations. But you could ask if and how the data was validated before you begin working with a dataset. Data validation helps to ensure the integrity of data. It also gives you confidence that the data you are using is clean. The following list outlines six types of data validation and the purpose of each, and includes examples and limitations.



1) Data type

- Purpose:** Check that the data matches the data type defined for a field.
- Example:** Data values for school grades 1-12 must be a numeric data type.
- Limitations:** The data value 13 would pass the data type validation but would be an unacceptable value. For this case, data range validation is also needed.



2) Data range

- Purpose:** Check that the data falls within an acceptable range of values defined for the field.
- Example:** Data values for school grades should be values between 1 and 12.
- Limitations:** The data value 11.5 would be in the data range and would also pass as a numeric data type. But, it would be unacceptable because there aren't half grades. For this case, data constraint validation is also needed.

