

The data contains information about Titanic passengers. Here is a screenshot of the first five rows of dataset:

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	0.0	3.0	Mr. Owen Harris Braund	male	NaN	1	0.0	7.2500
1	1.0	1.0	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0.0	71.2833
2	1.0	3.0	Miss. Laina Heikkinen	female	26.0	0	0.0	7.9250
3	1.0	1.0	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0.0	53.1000
4	0.0	3.0	Mr. William Henry Allen	male	35.0	0	0.0	8.0500

The validation criteria for data could be like:

Column name	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
Data type	float64	float64	object	object	float64	int64	float64	float64
Validation criteria	Numbers 0 or 1	Numbers between 0 and 3	Full name with title Mr., Ms. etc. First letters are capital	male/female in lower case	Age between 0-80	Values between 0 and 8	Values between 0 and 6	Values between 0 and 523.455
Optimal data type	Boolean, integer or categorical	Integer	object	object	integer	integer	integer	float
Numbers of data points	888	886	888	886	886	889	884	885
Numbers of missing data points	1	3	1	3	3	0	5	4
Errors found	Values 61, 225, 489 are incorrect	Values 77, 425, 443 and 660 are incorrect	Rows 45, 72 and 55,101 are duplicates. Value 346 have only title Miss.	Values 24(F), 173(Female), 316(fem), 346(40), 382(Male), 412(emale), 448(M), 749(boy) are incorrect	Values 78, 305, 346, 468, 643, 753, 801, 829 are incorrect. All values below 1	-	Values 217, 288, 346, 493 are incorrect	Values 256, 328, 378, 869, 870, 886 are incorrect