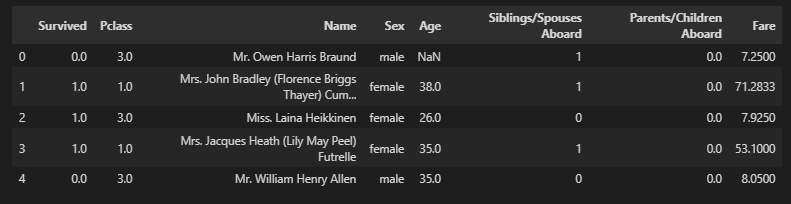
The data contains information about Titanic passengers.

Here is a screenshot of the first five rows of dataset:



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,  split column: title and name separately | 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 |  |  |  |  |  |