Hi Team,

I have reviewed the data sets which were provided by your company and during the data quality analysis, I have found few places where there is still scope for improvement in the data set.

Few Errors identified:

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 87 records Blanks |
| last\_name | 125 records Blanks |
| gender | 88 records gender ‘U’  Inconsistent values like M, Male, F, Female, Femal, U |
| job\_title | 506 records Blanks |
| job\_industry | 656 records with ‘N/A’ |
| Default | Inconsistent values including special characters |
| Tenure | 87 records Blanks |

|  |  |
| --- | --- |
| **FIELD NAME** | **Number of Blank Records** |
| Online\_order | 94 |
| brand | 48 |
| product\_line | 48 |
| product\_class | 48 |
| product\_size | 48 |
| standard\_cost | 48 |
| product\_first\_sold\_date | 48 |

Suggestion:

* Gender values ‘U’ can be replaced with reference to the customer name.
* Gender values ‘Femal’,’F’,’Female’ can all be replaced by ‘Female’ and values such as ‘Mal’,’M’ can be replaced by ‘Male’
* For tenure values, we can input mean of tenure .
* Blank orders can be eliminated.

Thanks

Rajatha