Hello,

This is Aditya Singh Thakur from KPMG Data Analytics (Virtual Internship) team. We have reviewed the data sets which were provided by your company and during the data quality analysis, we have found some errors in the data sets.

The data quality analysis is the core phase and due to errors in the data set, we suggest the following mitigates to improve the data quality, which will eventually help us to drive better analytics and results for your company.

* We can take a mode year value for the missing records of the customer’s DOB.
* We can assign a uniform last name of customers, which values are missing.
* Replace gender ‘U’ concerning the customer name and make consistency.
* For tenure values, we can take a mean of the rest of the values and assign the mean value to the missing fields to maintain the consistency of data.
* Eliminate blank orders considering fake orders.

The following are the details of the errors encountered in the data set.

# Customer Demographic (Total records 4000)

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 01 record 1843  87 records Blanks |
| last\_name | 125 records Blanks |
| Gender | 88 records gender ‘U’  Values are not consistency M, Male, F, Female, female, U |
| job\_title | 506 records Blanks |
| job\_industry | 656 records mention ‘N/A’ |
| Default | 3317 records value ‘special characters’ includes null and Blanks |
| Tenure | 87 records Blanks |

**Transactions (Total records 20000 -past 3 months)**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Online\_order | 94 records Blanks |
| brand | 48 records Blanks |
| product\_line | 48 records Blanks |
| product\_class | 48 records Blanks |
| product\_size | 48 records Blanks |
| standard\_cost | 48 records Blanks |
| product\_first\_sold\_date | 48 records Blanks |

Regards,

KPMG (Data Analytics Team)